



T.A. N.º 112/2023

CT. N.º 179/2020 (SEI 19.16.3687.0045763/2020-61)

CT. SIAD N.º 9263649

PRIMEIRO TERMO ADITIVO AO CONTRATO DE CONTRATO DE CONSULTORIA, QUE ENTRE SI CELEBRAM O MINISTÉRIO PÚBLICO DO ESTADO DE MINAS GERAIS, POR INTERMÉDIO DA PROCURADORIA-GERAL DE JUSTIÇA, E A FUNDAÇÃO DE DESENVOLVIMENTO DA PESQUISA – FUNDEP, NA FORMA AJUSTADA.

CONTRATANTE: Ministério Público do Estado de Minas Gerais, por intermédio da **Procuradoria-Geral de Justiça**, com sede na Av. Álvares Cabral, 1690, CEP 30.170-008, bairro Santo Agostinho, nesta Capital, inscrita no CNPJ sob o nº 20.971.057/0001-45, neste ato representado pelo Procurador-Geral de Justiça Adjunto Administrativo, **Márcio Gomes de Souza**.

CONTRATADA: Fundação de Desenvolvimento da Pesquisa - FUNDEP, pessoa jurídica de direito privado, sem fins lucrativos, inscrita no CNPJ sob o nº 18.720.938/0001-41, com sede na Avenida Antônio Carlos, nº 6627, 4º andar, bairro Pampulha, em Belo Horizonte/MG, CEP 31.270-901, neste ato representada por seu Presidente, **Jaime Arturo Ramírez**, CPF nº 554.155.556-68.

As partes acima qualificadas celebram o presente contrato de consultoria, nos termos da Lei Federal nº 8.666/93 e suas alterações, da Lei Federal 10.520/02, pelos Decretos nº. 5.450/05, 7.892/2013, pela legislação correlata, conforme Dispensa de Licitação nº. 076, de 02/12/2020, com fulcro no art. 24, inciso XIII, da Lei nº 8.666/93, mediante as cláusulas e condições seguintes:

CLÁUSULA PRIMEIRA – Do objeto

Constituem objetos do presente Termo Aditivo ao Contrato inicial, cujo objeto é a “contratação de consultoria técnica especializada para prestação de serviços de pesquisa e desenvolvimento na área da Inteligência Artificial Aplicada e soluções de *Big Data*, vinculada ao escopo do Programa de Capacidades Analíticas do MPMG- Fase 3, incluindo transferência de conhecimento, por meio de *mentoring* e operação assistida”, a prorrogação dos prazos de vigência e de execução e a alteração do Apenso II (Descritivo Técnico do Objeto) do Termo de Referência - Anexo Único do Contrato.

CLÁUSULA SEGUNDA – Da prorrogação

Prorrogam-se os prazos de vigência e de execução do contrato inicial por mais 12 (doze) meses, a partir de **29/06/2023 até 28/06/2024**, inclusive, com fulcro no art. 57, §1º, I da Lei nº 8.666/93.

CLÁUSULA TERCEIRA – Da alteração

Em virtude da alteração do escopo do contrato no que tange à adequação de entregáveis, mormente a inclusão de um grupo de atividades de apoio à implantação no ambiente do MPMG, suporte e manutenção dos softwares desenvolvidos (Plano de Manutenção e Suporte (doc. 5183626)), sem custo adicional, de modo a garantir a estabilidade das ferramentas após a fase de homologação, o Apenso II (Descritivo Técnico do Objeto) do Termo de Referência passa a vigorar conforme descrito no Anexo Único do presente instrumento.

CLÁUSULA QUARTA - Do valor

A presente prorrogação não ensejará qualquer despesa orçamentária.

CLÁUSULA QUINTA - Da vigência

O presente Termo Aditivo terá sua vigência a partir da data de sua assinatura.

CLÁUSULA SEXTA – Da publicação

Este Instrumento será publicado pela **Contratante** no *Diário Oficial Eletrônico do Ministério Público de Minas Gerais*, nos termos da Lei Federal 8.666/93.

CLÁUSULA SÉTIMA - Da continuidade contratual

Permanecem inalteradas e em pleno vigor todas as cláusulas e condições do Contrato inicial, naquilo em que não conflitarem com este instrumento.

ANEXO ÚNICO - TA N.º 112/2023**APENSO II DO TERMO DE REFERÊNCIA
DESCRITIVO TÉCNICO DO OBJETO**

Eixos de atividades:

O projeto consiste da execução paralela de 8 eixos de atividades, descritos na Tabela 1, a seguir:

Tabela 1 - Eixos de atividades.

ID	Tipo	Descrição
A	Planejamento Arquitetural	Planejamento e especificação dos componentes da plataforma do MPMG.
C	Coleta de Dados	Metodologias e módulos de software para coleta de dados.
D	Tratamento de Dados	Processamento, curadoria e identificação da qualidade dos dados.
E	Demandas Específicas	Atividades específicas no âmbito do MPMG.
F	Aplicações Finalísticas	Projeto e desenvolvimento de aplicações utilizando a plataforma e módulos de software para resolução de problemas específicos. (p.ex., trilhas de auditoria).
G	Gestão do Projeto	Administração do projeto, incluindo: coordenação, gestão, administração financeira (via FUNDEP), bem como garantia de qualidade e procedimentos de gerenciamento de risco.
M	Módulos de Software	Módulos de software implementam funcionalidades específicas e são a base das aplicações finalísticas.
P	Plataforma Tecnológica	A plataforma tecnológica provê o arcabouço metodológico e tecnológico para a integração de módulos e construção de aplicações.
I	Implantação, Suporte e Manutenção	Atividades para implantação no ambiente do MPMG das ferramentas e aplicações desenvolvidas, e suporte, manutenção e melhorias contínuas vinculadas.

2. GRUPOS DE TRABALHO

A seguir são descritos os Grupos de Trabalho pertencentes a cada Eixo de Atividades apresentado na Tabela 1.

A. Planejamento Arquitetural

Grupo de Trabalho	A01 - Arquitetura Tecnológica				
Grupos de Interação	A02, D01, D02, F06, G01, P01, P02, P03				
Coordenador	Dorgival Guedes				
Equipe	Nível	Quantidade	Nomes		
	Professor	1	Dorgival Guedes		
	Pós-Doc	0	TOTAL	3	
	Doutorando	1			
	Mestrando	0			
	Graduando	1			
	Técnico	0	--		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Sim	Sim	Sim	Sim	Sim
Descrição Geral					
<p>Avaliar as tecnologias de hardware e software mais indicadas para execução do trabalho dos demais grupos e aplicações do GSI, envolvendo a coleta, tratamento, armazenamento, pesquisa, análise e visualização de dados textuais estruturados e não estruturados, imagens, áudios e vídeos. Em particular, avaliar soluções para monitoração de métricas e coleta de logs da plataforma e seu processamento. As atividades de levantamento e avaliação de novas tecnologias serão desenvolvidas ao longo de toda a duração do projeto, como realizado no primeiro ano. Além dessas atividades, o trabalho é dividido em etapas ligadas ao desenvolvimento da solução para métricas e logs. Para esse objetivo, prevemos as etapas de levantamento, desenvolvimento e testes.</p>					
Objetivos					
<p>Continuar o processo de avaliação e integração de tecnologias de big-data aplicáveis ao ambiente do MPMG. Em particular, no primeiro ano o foco será em soluções para centralização e processamento de logs; no segundo ano, linhas de pesquisa possíveis incluem o uso das arquiteturas Lambda e/ou Kappa, a avaliação e possível aplicação do kubernetes^[1] como solução de provisionamento, adição de abstrações para automatização de fluxos (p.ex., Apache Airflow^[2]) em complementação ao Apache NiFi^[3].</p>					
Etapas					
E-A01.1		Início	M1	Fim	M3
	Título	Levantamento de informações sobre logs e métricas de interesse e tecnologias de processamento			

		Risco vinculado	R-A01.1		
			R-A01.2		
Tecnologias previstas	Tecnologias de monitoração e de coleta e processamento de logs a serem identificadas durante a etapa				
Tipos de dados necessários	Métricas e logs geradas pelas ferramentas e aplicações de interesse				
Descrição	Essa etapa envolve a identificação de quais ferramentas e aplicações deverão ser incluídas no processo de coleta e processamento de métricas e logs no sistema, bem como um levantamento das possíveis soluções tecnológicas para o processo de monitoração, coleta e armazenamento de logs.				
E-A01.2		Início	M4	Fim	M9
	Título	Implementação de um sistema de coleta e processamento de logs e métricas			
	Risco vinculado	R-A01.3	R-A01.5		
			R-A01.4		
Tecnologias previstas	Ferramentas disponíveis no ambiente de big data do GSI e tecnologias identificadas na etapa E-A01.1				
Tipos de dados necessários	Métricas e logs geradas pelas ferramentas e aplicações de interesse				
Descrição	Desenvolvimento do ambiente de coleta, armazenamento e processamento de métricas e logs da plataforma de big-data.				
E-A01.3		Início	M7	Fim	M12
	Título	Avaliação da solução de métricas e logs e integração de ferramentas de big data			
	Risco vinculado	R-A01.4	R-A01.6		

		R-A01.5	
Tecnologias previstas	Ferramentas disponíveis no ambiente de big data do GSI e tecnologias identificadas na etapa E-A01.1		
Tipos de dados necessários	Métricas e logs geradas pelas ferramentas e aplicações de interesse		
Descrição	Avaliação da solução de coleta, armazenamento e processamento de métricas e logs, levantamento da situação da arquitetura do MPMG ao final do projeto. Levantamento de opções disponíveis na comunidade de Software Livre que possam ser integradas ao ambiente.		
Entregáveis			
A01.1	E-A01.1	Relatório de levantamento de informações sobre logs e métricas de interesse e tecnologias de processamento	
	M3		
A01.2	E-A01.2	Protótipo do ambiente de métricas e logs operando com algumas fontes de dados selecionadas	
	M6		
A01.3	E-A01.3	Ambiente de coleta, armazenamento e processamento de métricas e logs	
	M9		
A01.4	E-A01.3	Relatório de avaliação do ambiente de métricas e logs e de outras tecnologias aplicadas ao ambiente de processamento do GSI	
	M12		
Milestones			
M-A01.1	Relatório mensal de acompanhamento das atualizações da infraestrutura		
M1			
M-A01.2	Relatório mensal de acompanhamento das atualizações da infraestrutura		
M2			
M-A01.3	Relatório mensal de acompanhamento das atualizações da infraestrutura		
M4			
M-A01.4	Relatório mensal de acompanhamento das atualizações da infraestrutura		
M5			
M-A01.5	Relatório mensal de acompanhamento das atualizações da infraestrutura		
M7			
M-A01.6	Relatório mensal de acompanhamento das atualizações da infraestrutura		
M8			

M-A01.7	Relatório mensal de acompanhamento das atualizações da infraestrutura
M10	
M-A01.8	Relatório mensal de acompanhamento das atualizações da infraestrutura
M11	

Riscos Críticos para a Implementação				
Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-A01.1	Dificuldade de encontrar graduando com experiência.	Médio	Mitigar	Aluno terá que ser treinado nas tecnologias envolvidas durante a evolução do projeto.
R-A01.2	Falta de clareza sobre as métricas e logs de interesse.	Baixo	Aceitar	Reforçar o processo de levantamento.
R-A01.3	Falta de informação clara sobre a origem dos dados a serem monitorados.	Baixo	Mitigar	Realizar reuniões com as equipes envolvidas no uso das ferramentas.
R-A01.4	Incompatibilidade entre ferramentas.	Médio	Mitigar	Reavaliar opções tecnológicas existentes e adotar outra combinação de ferramentas.
R-A01.5	Espaço de armazenamento insuficiente para as métricas e logs escolhidas.	Alto	Mitigar	Reduzir o conjunto de métricas e logs a serem tratados na versão entregue do sistema, planejar a expansão do armazenamento.
R-A01.6	Desempenho da solução resultante abaixo do esperado.	Médio	Aceitar	Limitar as consultas realizadas e planejar a expansão do parque computacional.

Cronograma das Entregas por Etapas

ETAPAS		ENTREGAS											
CÓD.	TÍTULO	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-A01.1	Levantamento de informações sobre logs e métricas de interesse e tecnologias de processamento			A01.1									
E-A01.2	Implementação de um sistema de coleta e processamento de logs e métricas						A01.2			A01.3			
E-A01.3	Avaliação da solução de métricas e logs e integração de ferramentas de big data												A01.4
MILESTONES		M-A01.1	M-A01.2		M-A01.3	M-A01.4		M-A01.5	M-A01.6		M-A01.7	M-A01.8	

Etapas - 2022					
E-A01.4		Início	M13	Fim	M15
	Título	Levantamento de informações sobre as tecnologias escolhidas			
	Descrição	De forma semelhante à etapa E-A01.1, essa etapa envolve a identificação de quais ferramentas e aplicações deverão ser incluídas na arquitetura para o segundo ano.			

E-A01.5		Início	M16	Fim	M21
	Título	Integração de novas tecnologias			
	Descrição	Extensão da arquitetura de big-data do MPMG com as tecnologias elencadas.			
E-A01.6		Início	M19	Fim	M24
	Título	Avaliação de desempenho da solução implementada			
	Descrição	Avaliação da solução implementada, levantamento da situação da arquitetura do MPMG ao final do projeto. Levantamento de opções disponíveis na comunidade de Software Livre que possam ser integradas ao ambiente no futuro.			
Entregáveis - 2022					
A01.5	E-A01.11	Relatório de levantamento de informações sobre as tecnologias de processamento que serão foco do segundo ano.			
	M15				
A01.6	E-A01.12	Protótipo com as novas tecnologias.			
	M18				
A01.7	E-A01.13	Ambiente de processamento integrado com as novas tecnologias.			
	M21				
A01.8	E-A01.14	Relatório de avaliação da arquitetura do MPMG ao final do projeto e de outras tecnologias que possam ser de interesse no futuro.			
	M24				

Grupo de Trabalho	A02 - Arquitetura Confiança Zero
Grupos de Interação	A01, D's, E's, F's, G01, M's, P's

Coordenador	José Marcos S. Nogueira				
Equipe	Nível	Quantidade	Nomes		
	Professor	3	José Marcos S. Nogueira, Dorgival Olavo Guedes Neto, Daniel Fernandes Macedo		
	Pós-Doc	0	TOTAL	6	
	Doutorando	0			
	Mestrando	0			
	Graduando	2			
	Técnico	1	Especialista em segurança de redes		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Sim	Sim	Sim	Não	Não
Descrição Geral					
<p>As atividades em segurança computacional previstas para 2021 e 2022 são decorrentes das recomendações constantes do Entregável A02.1 do projeto em 2020: Projeto de Segurança de Dados e Aplicações: DOCUMENTO DE ESPECIFICAÇÕES E RECOMENDAÇÕES DE SEGURANÇA PARA ACESSO ÀS APLICAÇÕES - SISTEMAS ATUAIS. Algumas recomendações foram tomadas como base, notadamente as R4, R5, R13 e R18.</p> <p>A atividade A02 - Segurança - pretende apoiar a implementação gradativa da abordagem de segurança zero no GSI, através de um projeto piloto. O protótipo de projeto piloto desenvolvido no primeiro ano do projeto será aperfeiçoado e complementado com novos módulos do modelo estrutural de confiança zero, empregando análise de dados de autenticação e autorização, o que trará segurança a mais ativos do GSI. A atividade objetiva dar apoio à implantação das funcionalidades autenticação de dois fatores e autenticação federada para os ativos.</p> <p>Este documento apresenta um escopo mínimo de atividades, que podem ser ajustadas desde que haja acordo prévio entre o GSI e a coordenação do A02.</p>					
Objetivos					
<p>O objetivo geral da atividade proposta é adicionar segurança aos ativos cibernéticos do GSI. A ideia é implementar gradualmente uma arquitetura de confiança zero para proteção dos ativos do GSI. Para isso alguns dos módulos do modelo estrutural de confiança zero serão construídos, integral ou parcialmente.</p> <p>Os objetivos específicos, que serão alcançados pela execução de tarefas, estão a seguir.</p>					

Objetivo 1: Aprimoramento da arquitetura de segurança do ambiente Nutanix do GSI pela adição de mecanismos de segurança às máquinas virtuais. A tarefa correspondente consiste na análise de viabilidade de autenticação das VMs e no controle dos fluxos entre VMs. Do ponto de vista do modelo estrutural de confiança zero, este mecanismo é parte do módulo "Sistema de diagnóstico e mitigação contínuos (CDM - Continuous Diagnostics and Mitigation)";

Objetivo 2: Promover métodos e técnicas para o gerenciamento de identidade e perfis de usuários, o gerenciamento de configuração de ativos, incluindo dispositivos e programas, com vista à definição e implantação de um algoritmo de confiança zero para estabelecer dinamicamente graus de confiança para autorização de uso de recursos. A tarefa correspondente consiste no desenvolvimento de mecanismos para estimar o grau de confiança associado a um pedido de autenticação pelo emprego de modelos de aprendizado de máquina (algoritmo de confiança). Do ponto de vista do modelo estrutural de confiança zero, esses mecanismos são parte do módulo "Ponto de Decisão de Políticas (PDP)";

Objetivo 3: Desenvolver uma solução para registro e análise de eventos relativos aos fluxos de dados dos processos de negócio. A solução inclui coleta e armazenamento de eventos, bem como a coleta de dados dos usuários no momento de autenticação, tendo em vista atributos como localização física, rede em que se encontra, horário de conexão e outros. A estrutura do data lake pode ser utilizada para implementar uma estrutura de análise de eventos registrados, que pode auxiliar na identificação de perfis de uso, situações de operação indesejada das ferramentas, análise de causas raiz (root cause) para falhas e identificação de ataques/invasões. A tarefa correspondente consiste no desenvolvimento de mecanismos para estimar se solicitações realizadas no ambiente de big data são potenciais ataques, pelo emprego de modelos de aprendizado de máquina. Do ponto de vista do modelo estrutural de confiança zero, estes mecanismos são parte do módulo "Sistema de informações de segurança e gerenciamento de eventos (SIEM - Security Information and Event Management)";

Objetivo 4: Dar continuidade ao apoio para implantação gradual da abordagem de confiança zero para os sistemas computacionais (ativos) do GSI, incluindo dar continuidade aos processos de implantação das funcionalidades de autenticação federada e autenticação dois fatores.

Objetivo 5: Evolução do sistema de diagnóstico e mitigação contínuos CDM. A construção inicial deste componente lógico está prevista no primeiro ano do projeto (2021), Objetivo 1. O objetivo aqui é dar continuidade aos trabalhos desenvolvidos. No caso, o processo de coleta de informações sobre o estado atual dos ativos do GSI e a aplicação de atualizações aos componentes de configuração e de software.

Objetivo 6: Evolução do Sistema de informações de segurança e gerenciamento de eventos (SIEM). Como continuidade de atividades relativas aos Objetivo 2/2021 e Objetivo 3/2021, este objetivo consiste na evolução da construção de um sistema corporativo que agrega logs de ativos, tráfego de rede, ações de acesso a recursos e outros eventos que fornecem informação em tempo real (ou em tempo quase real) sobre a postura de segurança dos sistemas de informações do GSI.

Objetivo 7: Aprimoramento do algoritmo de atribuição de confiança do mecanismo de decisão de política. As atividades deste objetivo para 2022 dá continuidade às atividades relativas ao Objetivo 2/2021 do projeto. Para cada sujeito que solicita algum tipo de acesso a um sistema, será auferido dinamicamente um grau de confiança para subsidiar a decisão de autorização. O algoritmo de confiança calcula um nível de confiança atribuído a um agente ou sujeito e o mecanismo de política (PE) usa políticas para ultimar a concessão ou negação do acesso a um recurso. O objetivo é considerar no algoritmo entradas de múltiplas fontes, como: base de dados de políticas com informações sobre usuários; atributos e papéis de usuários; históricos de padrão de comportamento de usuários; fontes de ameaças e outras fontes.

Objetivo 8: Evolução do Sistema de informações de segurança e gerenciamento de eventos (SIEM). A construção inicial deste componente lógico do modelo estrutural está prevista no primeiro ano do projeto, Objetivo 3/2021. O processo consiste de coleta de informações centradas na segurança para análise posterior. Os dados serão usados para refinar políticas e avisar sobre possíveis ataques contra ativos da organização.

Etapas

	Início	M1	Fim	M3
E-A02.1	Título Levantamento de dados - logs e gestão de VMs			
	Risco vinculado			R-A02.6
Tecnologias previstas		LDAP, WSO2, PostgreSQL		
Tipos de dados necessários		Logs de acesso das aplicações finalísticas do GSI, documentação de processos de gestão de VMs.		
Descrição	Nesta etapa, para o Objetivo 1, serão obtidos, junto ao GSI, logs de autenticação e autorização para as aplicações em uso. Em função dos dados coletados, serão definidas técnicas de aprendizado de máquina que serão empregadas em uma etapa posterior. Nesta etapa também será realizado um levantamento dos processos de gestão de VMs no GSI. Pretende-se identificar como as VMs são criadas, destruídas e mantidas. Será proposto um serviço para ser candidato a um piloto de gestão de VM baseado na confiança zero. Para o Objetivo 4, será dado			

	apoio à implantação de autenticação federada e de dois fatores, conforme demanda e disponibilidade.				
		Início	M4	Fim	M6
E-A02.2	Título	Segurança zero para VMs do GSI			
	Risco vinculado				
					R-A02.6
Tecnologias previstas	Ambientes de virtualização, WebAuthN, ferramentas de automação de gestão de redes (e.g. chef, ansible), mecanismos de assinatura de executáveis (e.g. tripwire)				
Tipos de dados necessários	N/A				
Descrição	Nesta etapa, para o Objetivo 1, será estudado e analisado, empregando o conceito de segurança zero, como prover segurança para as VMs do GSI. Serão analisados aspectos relacionados à autenticação de usuários nessas VMs, políticas de uso de softwares para scan de vulnerabilidades e saúde de softwares. Finalmente, em função da análise do resultados obtidos, serão investigadas técnicas para controlar as conexões de rede (entrando e potencialmente saindo também), autorizando-as ou negando-as em função das suas origens e destinos. Para o Objetivo 4, será dado apoio à implantação de autenticação federada e de dois fatores, conforme demanda e disponibilidade.				
		Início	M7	Fim	M12
E-A02.3	Título	Aprendizado de máquina para análise de segurança na autenticação			
	Risco vinculado	R-A02.1	R-A02.2		R-A02.3
		R-A02.4			R-A02.6
Tecnologias previstas	Python, toolkits de aprendizado de máquina, LDAP, WSO2, PostgreSQL				
Tipos de dados necessários	Logs de acesso das aplicações finalísticas do GSI				

Descrição	Nesta etapa, para o Objetivo 2, serão realizados estudos, empregando aprendizado de máquina, para caracterizar o risco de acesso às aplicações finalísticas do GSI. Serão propostos modelos prescritivos da população do GSI, que irão definir o comportamento padrão dos usuários em termos de autenticação e autorização. Estes modelos serão empregados para definir o grau de confiança dos usuários e das operações requisitadas. Caso os dados permitam, poderá ser proposto também modelos segmentados, de forma a analisar o comportamento de grupos de usuários ou o comportamento para sistemas específicos. A partir desta descrição, pretende-se avaliar se uma certa demanda de login, ou uma demanda de execução de uma operação, é uma demanda suspeita, ou uma demanda dentro dos padrões e que pode ser aceita/permitida. Para o Objetivo 4, será dado apoio à implantação de autenticação federada e de dois fatores, conforme demanda e disponibilidade.				
E-A02.4		Início	M7	Fim	M12
	Título	Aprendizado de máquina para análise de segurança no ambiente de Big Data			
	Risco vinculado	R-A02.1	R-A02.2	R-A02.3	
			R-A02.4		R-A02.6
Tecnologias previstas		Python, toolkits de aprendizagem de máquina, Ranger.			
Tipos de dados necessários		Logs de acesso dos registros e tabelas do ambiente de data lake			
Descrição	Nesta etapa, para o Objetivo 2, serão estudados os acessos aos registros e tabelas do ambiente de data lake, procurando caracterizar as operações ditas normais em cada um dos sistemas de armazenamento. Estes dados serão armazenados e analisados no ambiente de confiança zero no componente "Sistema de informações de segurança e gerenciamento de eventos (SIEM - Security Information and Event Management)". O objetivo do SIEM é marcar potenciais ataques para análise da equipe de redes do MPMG ou do GSI. Para o Objetivo 4, será dado apoio à implantação de autenticação federada e de dois fatores, conforme demanda e disponibilidade.				
Entregáveis					
A02.1	E-A02.1	Documento contendo:			

		<ul style="list-style-type: none"> - Uma descrição dos logs coletados de autenticação e autorização para as aplicações. - A identificação das técnicas de aprendizado de máquinas a serem empregadas posteriormente. - O resultado do levantamento dos processos de gestão de VMs. - Uma proposição inicial de um serviço candidato a piloto de gestão de VMs.
	M3	<ul style="list-style-type: none"> - Relato indicativo do progresso das atividades de apoio à implantação das autenticações nos sistemas do GSI.
A02.2	E-A02.2	<ul style="list-style-type: none"> - Resultado do estudo e análise sobre como prover segurança zero para as VMs do GSI.
	M6	<ul style="list-style-type: none"> - Resultado sobre a investigação de técnicas para controlar as conexões de rede. - Relato indicativo do progresso das atividades de apoio à implantação das autenticações nos sistemas do GSI.
A02.3	E-A02.3	<p>Documento contendo:</p> <ul style="list-style-type: none"> - Resultado dos estudos de caracterização do risco de acesso às aplicações finalísticas do GSI, empregando aprendizado de máquina. - Proposição de modelos prescritivos da população do GSI, definidores do comportamento padrão dos usuários em termos de autenticação e autorização.
	M9	<ul style="list-style-type: none"> - Relato sobre a possibilidade de avaliação de demandas de login ou de execução de operações quanto à suspeição ou confidência. - Relato indicativo do progresso das atividades de apoio à implantação das autenticações nos sistemas do GSI.
A02.4	E-A02.4	Documento contendo as observações e resultados finais das atividades do projeto relacionadas aos quatro primeiros objetivos esboçados.
	M12	
Milestones		
M-A02.1	Relatório mensal resumido indicativo do progresso das atividades.	
M1		
M-A02.2	Relatório mensal resumido indicativo do progresso das atividades.	

M2	
M-A02.3	Relatório mensal resumido indicativo do progresso das atividades.
M4	
M-A02.4	Relatório mensal resumido indicativo do progresso das atividades.
M5	
M-A02.5	Relatório mensal resumido indicativo do progresso das atividades.
M7	
M-A02.6	Relatório mensal resumido indicativo do progresso das atividades.
M8	
M-A02.7	Relatório mensal resumido indicativo do progresso das atividades.
M10	
M-A02.8	Relatório mensal resumido indicativo do progresso das atividades.
M11	

Riscos Críticos para a Implementação				
Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-A02.1	Falta de dados de acesso das aplicações finalísticas.	Médio	Mitigar	Gerar dados de alguns meses, durante o período do projeto, a partir de instrumentação de uma aplicação finalística.
R-A02.2	Falta de dados de acesso das tabelas e registros do data lake.	Médio	Mitigar	Gerar dados de alguns meses, durante o período do projeto, a partir de instrumentação de componentes do data lake (por exemplo, ranger).
R-A02.3	Pouco volume de dados coletados.	Alto	Aceitar	Precisão reduzida dos modelos de IA.
R-A02.4	Dados relativos a ataques rotulados como operação normal.	Médio	Aceitar	Sempre existe a possibilidade de um ataque não ser identificado, e por causa disso serem marcados uso legítimo. Assim, devemos aceitar que alguns ataques não serão capturados pelos modelos.
R-A02.5	Dificuldade em prover apoio às atividades de implantação de	Baixo	Aceitar	Renegociar atividades conforme disponibilidade da equipe.

	autenticações, devido a excesso de demanda.			
R-A02.6	Limitação de recursos pode não permitir a contratação da equipe prevista.	Baixo	Mitigar	Lista de objetivos pode precisar ser revista e ajustada à equipe disponível.

Cronograma das Entregas por Etapa													
Etapas		Entregas											
Cód.	Título	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-A02.1	Levantamento de dados - logs e gestão de VMs			A02.1									
E-A02.2	Segurança zero para VMs do GSI						A02.2						
E-A02.3	Aprendizado de máquina para análise de segurança na autenticação									A02.3			
E-A02.4	Aprendizado de máquina para análise de segurança no ambiente de big data												A02.4

MILESTONES	M-A02.1	M-A02.2		M-A02.3	M-A02.4		M-A02.5	M-A02.6		M-A02.7	M-A02.8	
-------------------	----------------	----------------	--	----------------	----------------	--	----------------	----------------	--	----------------	----------------	--

Etapas - 2022					
E-A02.5	Início	M13		Fim	M21
	Título	Sistema de diagnóstico e mitigação contínuos			
	Descrição	Desenvolvimento das atividades relativas à evolução do sistema de diagnóstico e mitigação contínuos CDM.			
E-A02.6	Início	M16		Fim	M24
	Título	Mecanismos de logs de atividades de rede e sistema			
	Descrição	Desenvolvimento das atividades relativas à evolução dos mecanismos de logs de atividades de rede e sistema.			
E-A02.7	Início	M13		Fim	M21
	Título	Algoritmo de atribuição de confiança do mecanismo de decisão de política			
	Descrição	Desenvolvimento das atividades relativas ao aprimoramento do algoritmo de atribuição de confiança.			
E-A02.8	Início	M16		Fim	M24
	Título	Sistema de informações de segurança e gerenciamento de eventos (SIEM).			

	Descrição	Desenvolvimento das atividades relativas à evolução do sistema de informações de segurança e gerenciamento de eventos.
Entregáveis - 2022		
A02.5	E-A02.11	Documento contendo a situação atual do desenvolvimento das etapas E-A02.5 e E-A02.7
	M15	
A02.6	E-A02.12	Documento contendo a situação atual do desenvolvimento das etapas E-A02.5, E-A02.6, E-A02.7 e E-A02.8
	M18	
A02.7	E-A02.13	Documento contendo a situação atual do desenvolvimento das etapas E-A02.5, E-A02.6, E-A02.7 e E-A02.8
	M21	
A02.8	E-A02.14	Documento contendo a situação final do desenvolvimento das etapas E-A02.6 e E-A02.8
	M24	

C. Coleta de Dados

Grupo de Trabalho	C01 - Coleta de Dados Públicos		
Grupos de Interação	F02, F03, F04, F05, G01, M01, M02, M03, M05, P01		
Coordenador	Rodrygo Santos		
Equipe	Nível	Quantidade	Nomes
	Professor	1	Rodrygo Santos

	Pós-Doc	0	TOTAL	11	
	Doutorando	2			
	Mestrando	4			
	Graduando	4			
	Técnico	1	Desenvolvedor web		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Não	Não	Sim	Não	Não
Descrição Geral					
O projeto visa o desenvolvimento de uma plataforma extensível para a coleta de dados de bases públicas, disponibilizadas na Web, de forma automatizada e recorrente, e a extração dos dados coletados para representação estruturada na plataforma de processamento e análise de dados do MPMG.					
Objetivos					
<ol style="list-style-type: none"> 1. Consolidar a plataforma de coleta desenvolvida na primeira fase do projeto em 2020. 2. Ampliar a cobertura de testes e generalidade da plataforma para a coleta de novas fontes de dados. 3. Aprimorar a eficiência e a robustez da plataforma, tanto para a configuração quanto para a execução de coletas. 4. Aprimorar a extração dos dados coletados. 					
Etapas					
E-C01.1		Início	M1	Fim	M3
	Título	Validação da plataforma atual			
	Risco vinculado	R-C01.1			
Tecnologias previstas		Django; Scrapy; Pyppeteer			
Tipos de dados necessários		Descritores de fontes diversas			
Descrição	Validação da plataforma desenvolvida para configuração e monitoramento de coletores e levantamento de requisitos para melhorias de usabilidade.				
E-C01.2		Início	M2	Fim	M4
	Título	Automação da configuração de coletores			

		Risco vinculado	R-C01.2		
Tecnologias previstas	Django; Scrapy; Pypeteer				
Tipos de dados necessários	Descritores de fontes diversas				
Descrição	Implementação de mecanismos para abstrair a complexidade do processo de configuração de novos coletores, incluindo a detecção automática de parâmetros obrigatórios e a injeção de valores para esses parâmetros.				
		Início	M4	Fim	M5
E-C01.3	Título	Execução em modo de sondagem			
	Risco vinculado	R-C01.3	R-C01.4	R-C01.5	
Tecnologias previstas	Django; Scrapy; Pypeteer				
Tipos de dados necessários	Descritores de fontes de interesse; descritores de regras por fonte				
Descrição	Ampliação dos mecanismos de sondagem atuais para configuração de regras arbitrárias (e.g., checklists) para monitoramento de fluxos de publicação em uma fonte.				
E-C01.4		Início	M5	Fim	M10
	Título	Melhorias na orquestração de coletas			

		Risco vinculado	R-C01.6		
Tecnologias previstas	Django; Scrapy-Cluster; Scrapy; Pypeteer				
Tipos de dados necessários	Descritores de fontes diversas				
Descrição	A plataforma atual contempla mecanismos ainda embrionários para orquestração de coletas em modo distribuído, bem como sua execução de modo incremental e contínuo. Há aqui várias oportunidades de melhoria da orquestração do sistema distribuído como um todo, bem como de seus mecanismos para monitoramento e escalonamento de coletas.				
		Início	M1	Fim	M12
E-C01.5	Título	Coleta de dados de municípios			
	Risco vinculado	R-C01.3	R-C01.4	R-C01.5	
Tecnologias previstas	Django; Scrapy-Cluster; Scrapy; Pypeteer				
Tipos de dados necessários	Descritores de fontes de interesse				
Descrição	Ampliação da cobertura de coletores para dados de municípios do Estado de Minas Gerais. O foco desta etapa é na coleta de dados dos portais de transparência dos municípios do Estado de Minas Gerais. Os dados coletados serão disponibilizados por demanda para os grupos de Aplicações Finalísticas.				

		Início	M1	Fim	M12
E-C01.6	Título	Levantamento de novos requisitos de coleta			
	Risco vinculado	R-C01.3			
Tecnologias previstas		Django; Scrapy-Cluster; Scrapy; Pypeteer			
Tipos de dados necessários		Descritores de fontes diversas			
Descrição	Dada a variabilidade de fontes de dados na Web, naturalmente serão identificados requisitos não contemplados pela plataforma atual. Esses requisitos deverão ser analisados e documentados a fim de orientar futuros desenvolvimentos.				
Entregáveis					
C01.1	E-C01.1	Relatório descrevendo requisitos de usabilidade levantados durante a validação da plataforma atual e planejamento de eventuais melhorias.			
	M3				
C01.2	E-C01.2	Versão da plataforma com componentes de automação de configuração integrados.			
	M4				
C01.3	E-C01.3	Versão da plataforma com suporte à configuração de regras de sondagem arbitrárias.			
	M5				
C01.4	E-C01.4	Versão da plataforma com componentes para coleta distribuída, incremental e contínua.			
	M10				
C01.5	E-C01.5	Coletores de dados de municípios mineiros configurados na plataforma.			
	M11				
C01.6	E-C01.6	Relatório descrevendo requisitos de coleta levantados durante a configuração de novos coletores e planejamento de eventuais melhorias.			
	M12				
Milestones					
M-C01.1	Relatório de progresso da etapa E-C01.1 e E-C01.5.				
M1					

M-C01.2	Relatório de progresso da etapa E-C01.1 e E-C01.5.
M2	
M-C01.3	Relatório de progresso das etapas E-C01.4 e E-C01.5.
M6	
M-C01.4	Relatório de progresso das etapas E-C01.4 e E-C01.5.
M7	
M-C01.5	Relatório de progresso das etapas E-C01.4 e E-C01.5.
M8	
M-C01.6	Relatório de progresso das etapas E-C01.4 e E-C01.5.
M9	

Riscos Críticos para a Implementação				
Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-C01.1	Não-adoção efetiva da plataforma por parte do MPMG.	Alto	Mitigar	Ações de treinamento; envolvimento do cliente.
R-C01.2	Explosão combinatória de parâmetros para coleta de uma única fonte.	Baixo	Mitigar	Amostragem de um subconjunto das combinações possíveis.
R-C01.3	Barreiras tecnológicas a coletas automáticas.	Médio	Mitigar	Incorporação de mecanismos de robustez, incluindo solucionadores de captchas e obfuscadores de identidade (e.g., rotação de user-agent e IP address); busca por fontes alternativas.
R-C01.4	Barreiras legais a coletas automáticas.	Médio	Mitigar	Busca por acordos de cooperação com o MPMG; busca por fontes alternativas.
R-C01.5	Explosão do número de fontes a coletar.	Médio	Mitigar	Amostragem de um subconjunto das fontes de interesse; aumento da equipe responsável pela configuração de

				novos coletores na plataforma; implantação de coletas distribuídas.
R-C01.6	Recursos de rede limitados.	Baixo	Mitigar	Negociação para disponibilização de múltiplos IPs públicos para coletas distribuídas.
R-C01.7	Restrições tecnológicas para integração.	Baixo	Mitigar	Avaliação junto à equipe do MPMG da melhor estratégia para integração da plataforma de coleta a outras plataformas; desacoplamento da plataforma de coleta como microsserviço.

Cronograma das Entregas por Etapas

Etapas		Entregas											
Cód.	Título	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-C01.1	Validação da plataforma atual			C01.1									
E-C01.2	Automação da configuração de coletores				C01.2								
E-C01.3	Execução em modo de sondagem					C01.3							
E-C01.4	Melhorias na orquestração de coletas										C01.4		
E-C01.5	Coleta de dados de municípios											C01.5	
E-C01.6	Levantamento de novos requisitos de coleta												C01.6

MILESTONES	M-C01.1	M-C01.2				M-C01.3	M-C01.4	M-C01.5	M-C01.6			
-------------------	----------------	----------------	--	--	--	----------------	----------------	----------------	----------------	--	--	--

Etapas - 2022												
E-C01.11				Início		M13		Fim			M15	
	Título	Coleta de portais de notícias										
	Descrição	Ampliação da cobertura de coletores para portais de notícias.										
E-C01.12				Início		M16		Fim			M18	
	Título	Coleta de bases de conhecimento										
	Descrição	Ampliação da cobertura de coletores para bases de conhecimento, como a Wikipedia.										
E-C01.13				Início		M13		Fim			M21	
	Título	Coleta de bases de dados de outros estados										
	Descrição	Ampliação da cobertura de coletores para bases de dados de interesse providas por entidades com base fora de Minas Gerais.										
E-C01.14				Início		M19		Fim			M24	
	Título	Integração a outras plataformas										
	Descrição	Integração da plataforma de coleta para operação conjunta com outras plataformas existentes no MPMG, como Nifi e Lemonade.										
Entregáveis - 2022												

C01.11	E-C01.11	Novos coletores configurados na plataforma.
	M15	
C01.12	E-C01.12	Novos coletores configurados na plataforma.
	M18	
C01.13	E-C01.13	Novos coletores configurados na plataforma.
	M21	
C01.14	E-C01.14	Versão da plataforma de coleta integrada a plataformas existentes.
	M24	

D. Tratamento de Dados

Grupo de Trabalho	D01 - Integração de Repositórios de Dados				
Grupos de Interação	A01, D02, F02, F03, G01, M02, M03, M04				
Coordenador	Marcos André Gonçalves Sub-coordenador: Alberto Laender				
Equipe	Nível	Quantidade	Nomes		
	Professor	2	Marcos André Gonçalves, Alberto Laender		
	Pós-Doc	0	TOTAL	7	
	Doutorando	2			
	Mestrando	1			
	Graduando	2			
	Técnico	1	Engenheiro de Dados		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Sim	Sim	Sim	Sim	Não
Descrição Geral					
- Generalizar as soluções de integração de dados testando em várias bases com foco na construção de MDMs.					
Objetivos					

- Testes em outras bases de Dados
- Implementação e Avaliação com Pareamento Probabilístico
- Testes e Experimentos com Aprendizado Ativo
- Avaliação Substancial dos Métodos
- Avaliação do Pareamento Incremental

Etapas

E-D01.1	Início	M1	Fim	M2
	Título	Pareamento Probabilístico		
	Risco vinculado		R-D01.1	R-D01.2
Tecnologias previstas		Hive, Apache, Linguagens Python e Java		
Tipos de dados necessários		Dados de multiplas bases		
Descrição	Pareamento Probabilístico (Pessoa) - Testes e Avaliação.			
E-D01.2	Início	M1	Fim	M12
	Título	Aprendizado Ativo		
	Risco vinculado		R-D01.3	
Tecnologias previstas		Hive, Apahe, Linguagens Python e Java		
Tipos de dados necessários		Sob demanda		
Descrição	Avaliação do Aprendizado Ativo para o pareamento de dados (Continuo).			
E-D01.3	Início	M1	Fim	M4
	Título	Pareamento Incremental		
	Risco vinculado		R-D01.1	R-D01.2

Tecnologias previstas	Hive, Apahe, Linguagens Python e Java				
Tipos de dados necessários					
Descrição	Implementação, Testes e Avaliação do Pareamento Incremental (Pessoa)				
E-D01.4		Início	M3	Fim	M4
	Título	Caracterização (Nova Base)			
		Risco vinculado		R-D01.1	R-D01.2
Tecnologias previstas	Hive, Apache, Linguagens Python e Javac				
Tipos de dados necessários	Licitações (ou outra MDM)				
Descrição	Integração de documentos de licitações (ou outra MDM) - Caracterização. Possibilidades de MDMs: telefone, endereço, produtos/serviços (para grupo de compras públicas).				
E-D01.5		Início	M5	Fim	M8
	Título	Pareamento (Novas Base) - Primeiros Teste			
		Risco vinculado		R-D01.1	R-D01.2
				R-D01.3	
Tecnologias previstas	Hive, Apache, Linguagens Python e Java				
Tipos de dados necessários	Licitações ou outra MDM				
Descrição	Integração de documentos de licitações (ou outra MDM) - Primeiras Soluções e testes - Casamento Determinístico e Probabilístico.				

Possibilidades de MDMs: telefone, endereço, produtos/serviços (para grupo de compras públicas).					
E-D01.6		Início	M9	Fim	M12
	Título	Pareamento (Nova Base) - Aprimoramentos			
	Risco vinculado			R-D01.1	R-D01.2
				R-D01.3	
Tecnologias previstas		Hive, Apache, Linguagens Python e Java			
Tipos de dados necessários		Licitações ou outra MDM			
Descrição	- Integração de documentos de licitações (ou outra MDM) - Aprimoramento de Soluções, Aprendizado Ativo Possibilidades de MDMs: telefone, endereço, produtos/serviços (para grupo de compras públicas).				
Entregáveis					
D01.1	E-D01.1	Avaliação do Pareamento Probabilístico			
	M2				
D01.2	E-D01.3	- Pareamento Incremental			
	M4	- Caracterização do novo Pareamento			
D01.3	E-D01.4	- Resultados Experimentais do Novo Pareamento			
	M4	- Aprendizado Ativo			
D01.4	E-D01.5	- Relatório Parcial			
	M8	- Código Integrado na Infra-Estrutura do GDI/MPMG			
D01.5	E-D01.6	- Relatório Final			
	M12	- Código Integrado na Infra-Estrutura do GDI/MPMG			
Milestones					

M-D01.1	Parecer sobre quais dados serão utilizados nas Etapa D01.4, Etapa D01.5 e Etapa D01.6 e como eles serão obtidos
M1	
M-D01.2	Descrição sucinta dos resultados preliminares do pareamento Incremental e da Caracterização novo Pareamento
M3	
M-D01.3	Testes iniciais do Novo Pareamento; Avaliação do Novo Pareamento
M5	
M-D01.4	Melhorias do Novo Pareamento; Aprendizado Ativo
M6	
M-D01.5	Melhorias do Novo Pareamento; Aprendizado Ativo
M7	
M-D01.6	Melhorias do Novo Pareamento; Aprendizado Ativo; Início da Integração das Soluções na Arquitetura Geral
M9	
M-D01.7	Melhorias do Novo Pareamento; Aprendizado Ativo; Integração das Soluções na Arquitetura Geral
M10	
M-D01.8	Melhorias do Novo Pareamento; Aprendizado Ativo; Início da Integração das Soluções na Arquitetura Geral; Preparação pro Relatório Final
M11	

Riscos Críticos para a Implementação				
Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-D01.1	Baixa precisão / revocação nos procedimentos de integração. A integração é muito dependente das características dos dados.	Baixo	Mitigar	Testes e refinamentos sucessivos dos procedimentos implementados.
R-D01.2	Dados ruidosos/com baixa qualidade.	Baixo	Mitigar	Testes e refinamentos sucessivos da ferramenta a ser implementada, inclusive para limpeza dos dados.

R-D01.3	Robustez da avaliação dos métodos de aprendizado ativo. Soluções baseadas em aprendizado ativo precisam de uma avaliação mais robusta de longo prazo baseado no uso real das ferramentas e do feedback dos usuários com possíveis reformulações constantes.	Baixo	Mitigar	Testes e refinamentos sucessivos da ferramenta a ser implementada baseados no uso real da ferramenta.
----------------	---	-------	---------	---

Cronograma das Entregas por Etapas													
Etapas		Entregas											
Cód.	Título	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-D01.1	Pareamento Probabilístico		D01.1										
E-D01.2	Aprendizado Ativo												
E-D01.3	Pareamento Incremental				D01.2								
E-D01.4	Caracterização (Nova Base)				D01.3								
E-D01.5	Pareamento (Novas Base) - Primeiros Teste								D01.4				
E-D01.6	Pareamento (Nova Base) - Aprimoramentos												D01.5
MILESTONES		M-D01.1		M-D01.2		M-D01.3	M-D01.4	M-D01.5		M-D01.6	M-D01.7	M-D01.8	

Etapas - 2022					
E-D01.7		Início	M13	Fim	M14
	Título	Caracterização de Dados - Novas MDMs			

	Descrição	Integração de outras MDMs - Caracterização - Possibilidades de MDMs: telefone, endereço, produtos/serviços (para grupo de compras públicas).			
Entregáveis - 2022					
E-D01.8		Início	M15	Fim	M19
	Título	Pareamento de novas MDMs - Soluções Inicias e Teses			
	Descrição	Integração de outras MDMs - Primeiras Soluções, testes e avaliação para o Casamento Determinístico e Probabilístico. Possibilidades de MDMs: telefone, endereço, produtos/serviços (para grupo de compras públicas).			
E-D01.9		Início	M20	Fim	M24
	Título	Pareamento (Nova MDS) - Aprimoramentos			
	Descrição	Integração de outra MDMs - Aprimoramento de Soluções, Aprendizado Ativo Possibilidades de MDMs: telefone, endereço, produtos/serviços (para grupo de compras publicas)			
Entregáveis - 2022					
D01.6	E-D01.11	Relatório com a Caracterização das Novas Bases			
	M14				
D01.7	E-D01.12	- Relatório Parcial com a descrição da Implementação dos testes com o Pareamentos Determinísticos e Probabilísticos nas Novas Bases - Código Integrado na Infra-Estrutura do GDI/MPMG			
	M19				
D01.8	E-D01.13	- Relatório Final com a descrição da Implementação dos testes com o Pareamentos Determinísticos e Probabilísticos nas Novas Bases - Código Integrado na Infra-Estrutura do GDI/MPMG			
	M24				

Grupo de Trabalho	D02 - Curadoria de dados				
Grupos de Interação	C01, D01, F02, F03, G01, P01, P02				
Coordenador	Anisio Mendes Lacerda				
Equipe	Nível	Quantidade	Nomes		
	Professor	3	Anísio Mendes Lacerda, Gisele L Pappa, Wagner Meira Jr.		
	Pós-Doc	0	TOTAL	6	
	Doutorando	0			
	Mestrando	2			
	Graduando	1			
	Técnico	0	--		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Sim	Sim	Não	Sim	Não
Descrição Geral					
<p>Ao trabalhar com os dados de licitações presentes nos bancos de dados do MP no projeto de 2020, percebemos que os dados apresentam várias anomalias e dados faltantes. Por exemplo, existem licitações cujo valor licitado é de 1 trilhão de reais, existem cidades para as quais nenhuma licitação se encontra na base, entre outros problemas. Dada a visão do MP em usar inteligência para resolver muitos de seus problemas, hoje tratados manualmente, uma etapa de curadoria e identificação da qualidade dos dados já presentes e futuramente carregados no banco de dados da instituição é crucial.</p> <p>Assim, propomos que para o ano de 2021 o foco do trabalho seja os dados estruturados e não estruturados relacionados a licitações. Já para o ano de 2022 serão abordados os dados de obras e compras públicas.</p> <p>Por fim, a implementação de indicadores de qualidade para dados não estruturados depende fortemente da definição da tarefa, do seu escopo e da sua análise de viabilidade. É importante frisar que esta tarefa não tem como objetivos o processamento de dados para a sua eventual estruturação, correção ou classificação de contexto e conteúdo.</p>					
Objetivos					
<p>O principal objetivo deste grupo de trabalho é gerar indicadores de qualidade dos dados relacionados a licitação, obras públicas e compras públicas, desde que estes já estejam disponíveis nos bancos de dados do MP.</p> <p>Esses indicadores serão usados para gerar alertas quando a qualidade do dado no banco for inferior a um limiar mínimo desejado.</p>					
Etapas					
E-D02.1		Início	M1	Fim	M7

	Título	Levantamento de fontes de dados estruturadas e não estruturadas utilizadas em aplicações finalísticas e caracterizações iniciais			
	Risco vinculado	R-D02.1	R-D02.2		R-D02.3
		R-D02.4			
Tecnologias previstas		Hive, Druid, SQL, Spark, Python, PySpark, Apache Deequ			
Tipos de dados necessários		Acesso aos bancos de dados de aplicações finalísticas do MPMG e metadados correspondentes. Acesso a especialistas de conteúdo.			
Descrição	Realizar um mapeamento das fontes de dados e um estudo de viabilidade acerca das métricas a serem implementadas para as referidas fontes. Alguns exemplos de métricas são: i) O percentual de preenchimento de campos; ii) O percentual de valores válidos para campos onde é possível realizar esta validação; e iii) A presença de <i>outliers</i> em campos numéricos.				
E-D02.2		Início	M4	Fim	M12
	Título	Implementação dos mecanismos de geração de alerta de informações discrepantes ou faltantes.			
	Risco vinculado	R-D02.2	R-D02.3		R-D02.4
Tecnologias previstas		Hive, Druid, SQL, Spark, Python, PySpark, Apache Deequ			
Tipos de dados necessários		Acesso aos bancos de dados de aplicações finalísticas do MPMG e metadados correspondentes. Acesso a especialistas de conteúdo.			
Descrição	Realizar a implementação das métricas levantadas na etapa anterior, além da implementação de um relatório gerencial consolidando as informações de qualidade para as bases de dados acordadas.				
Entregáveis					

D02.1	E-D02.1	Relatório contendo o mapeamento das bases de dados estruturadas e fontes não estruturadas utilizadas em cada uma das aplicações finalísticas (licitações, obras públicas e compras públicas), bem como um levantamento dos fluxos de ingestão de dados que geraram as bases em questão e seus possíveis pontos de falha.
	M3	
D02.2	E-D02.1	Relatório contendo levantamento inicial e análise de viabilidade das métricas e indicadores de qualidade a serem implementadas, bem como as suas respectivas caracterizações iniciais. Definição de cronograma de implementação considerando as bases de dados, métricas e indicadores acordados.
	M7	
D02.3	E-D02.2	Relatório contendo a descrição das métricas implementadas até o momento, segundo o cronograma acordado no relatório D02.2, bem como os resultados das mesmas nas bases de dados mapeadas no relatório D02.1, com foco em indicadores de licitações.
	M9	
D02.4	E-D02.2	Códigos e relatório contendo a descrição das métricas acordadas no relatório D02.2, bem como os resultados das mesmas nas bases de dados mapeadas no relatório D02.1. Os códigos incluem as métricas, indicadores e a implementação de relatório gerencial para as bases de dados acordadas.
	M12	
Milestones		
M-D02.1	Relatório contendo solicitação de acessos a bases de dados e sistemas que ainda estejam pendentes, levantamento de metadados ausentes e mapeamento inicial das bases e fluxos de ingestão de dados que as geraram.	
M1		
M-D02.2	Status do mapeamento das bases de dados e fluxos de ingestão de dados.	
M2		
M-D02.3	Status preliminar das métricas levantadas para cada base de dados com os status de discussão junto aos parceiros do MPMG (métrica aprovada, em discussão, em análise de viabilidade, etc).	
M4		
M-D02.4	Status de desenvolvimento das métricas de acordo com o cronograma em em D02.2	
M5		
M-D02.5	Proposta de formato de relatório de relatório gerencial. Avaliação da necessidade de readequação de cronograma.	
M6		
M-D02.6	Status de desenvolvimento das métricas de acordo com o cronograma acordado em D02.2	
M8		
M-D02.7	Status de desenvolvimento das métricas de acordo com o cronograma acordado em D02.2	

	Status de desenvolvimento do relatório gerencial a partir das métricas já desenvolvidas Primeira versão do relatório gerencial para sugestões e ajustes
M10	Atualização do levantamento das fontes de dados relacionadas a obras públicas e compras públicas com foco no desenvolvimento da tarefa em 2022
M-D02.8	Status de desenvolvimento das métricas de acordo com o cronograma em em D02.2
M11	Status de desenvolvimento do relatório gerencial a partir das métricas já desenvolvidas Segunda versão do relatório gerencial para sugestões e ajustes

Riscos Críticos para a Implementação				
Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-D02.1	Falta de consenso na definição das métricas a serem implementadas e no escopo de aplicação da tarefa.	Médio	Travar	Subir a negociação para as instâncias imediatamente superiores. Replanejar o cronograma de atividades.
R-D02.2	Alteração na estrutura do banco de dados de licitações (EDW).	Alto	Travar	Interação com as equipes do MP para entendimento do planejamento das alterações ou alterações do sistema de banco de dados do MP após o projeto de 2020. Só será possível iniciar qualquer implementação desse grupo de trabalho após este alinhamento.
R-D02.3	Saída de membros da equipe responsável.	Alto	Mitigar	Contratação de substitutos e/ou reatribuição de responsabilidades dentro da equipe.
R-D02.4	Mudança das tecnologias sendo utilizadas no desenvolvimento do projeto por parte do MP.	Alto	Mitigar	Readequar o projeto e treinar o pessoal nas tecnologias não dominadas. Pode atrasar o cronograma.

Cronograma das Entregas por Etapas													
ETAPAS		ENTREGAS											
CÓD.	TÍTULO	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-D02.1	Levantamento de fontes de dados estruturadas utilizadas em aplicações finalísticas e caracterizações iniciais			D02.1				D02.2					
E-D02.2	Implementação dos mecanismos de geração de alerta de informações discrepantes ou faltantes.								D02.3				D02.4
MILESTONES		M-D02.1	M-D02.2		M-D02.3	M-D02.4	M-D02.5		M-D02.6		M-D02.7	M-D02.8	

Etapas - 2022					
		Início	M13	Fim	M15
E-D02.11	Título	Levantamento dos dados estruturados e não estruturados, utilizados pelas aplicações finalísticas e módulos de softwares relacionados a obras públicas e compras públicas.			
	Descrição	Será feito o levantamento dos dados estruturados e não estruturados, que são utilizados pelas aplicações finalísticas e módulos de softwares relacionados a obras públicas e compras públicas.			

		Início	M15	Fim	M17
E-D02.12	Título	Análise de viabilidade de aplicação das métricas implementadas em 2021 nos dados de obras públicas e compras públicas			
	Descrição	Será feita uma análise de viabilidade de aplicação das métricas implementadas em 2021, referente aos dados de obras públicas e compras públicas			
		Início	M17	Fim	M24
E-D02.13	Título	Implementação das métricas de qualidade aos contextos de obras públicas e compras públicas			
	Descrição	Serão implementadas métricas de qualidade aos contextos de obras públicas e compras públicas.			
Entregáveis - 2022					
D02.11	E-D02.11	Relatório contendo o levantamento atualizado das fontes de dados a serem analisadas bem como proposta das métricas a serem implementadas nos contextos de obras públicas e compras públicas.			
	M15				
D02.12	E-D02.12	Relatório contendo análise de viabilidade e cronograma de implementação para as métricas levantadas em D02.11.			
	M17				
D02.13	E-D02.13	Relatório contendo status parcial de implementação das métricas conforme cronograma proposta em D02.12.			
	M20				

D02.14	E-D02.14	Códigos e relatório contendo a descrição das métricas acordadas no relatório D02.12, bem como os resultados das mesmas nas bases de dados mapeadas no relatório D02.11. Os códigos incluem as métricas, indicadores e a atualização de relatório gerencial para as bases de dados acordadas.
	M24	

E. Demandas Específicas

Grupo de Trabalho	E01 - Construção de modelos de linguagem específicos para transcrição automática de áudios				
Grupos de Interação	F04, G01, M01				
Coordenador	Adriano Alonso Veloso				
Equipe	Nível	Quantidade	Nomes		
	Professor	1	Adriano Alonso Veloso		
	Pós-Doc	0	TOTAL	5	
	Doutorando	0			
	Mestrando	1			
	Graduando	2			
	Técnico	1	Desenvolvedor Web		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Não	Não	Não	Não	Não
Descrição Geral					
Ferramenta com funcionalidade de transcrição (degravação) de arquivos de áudio no idioma português do Brasil. Trata-se da evolução natural do projeto de transcrição de áudios, com um foco em modelos que consigam transcrever áudios oriundos de cenários específicos.					
Objetivos					
O objetivo deste projeto é construir modelos de linguagens específicos para cada cenário de aplicação de transcrição de áudio. O sistema de transcrição já implementado faz uso de um modelo de linguagem geral, porém resultados melhores podem ser obtidos caso a caso, através do uso de modelos de linguagem específicos. Dessa forma, pretendemos utilizar o sistema de transcrição existente como um modelo pré-treinado, o qual pode ser refinado criando-se modelos de linguagem específicos.					

Está previsto, de acordo com a demanda e necessidade do MPMG, a evolução da interface Web de anotação desenvolvida em 2020.

Etapas

		Início	M1	Fim	M2
E-E01.1	Título	Definição dos cenários de aplicação			
	Risco vinculado	R-E01.1			
Tecnologias previstas		Nenhuma			
Tipos de dados necessários		Áudio e texto pareado			
Descrição	As tarefas se resumirão a reuniões para definição de dois cenários de aplicação para o modelo de transcrição.				
		Início	M2	Fim	M8
E-E01.2	Título	Construção, avaliação e aprimoramento do modelo de transcrição específico para o primeiro cenário de aplicação.			
	Risco vinculado	R-E01.1			
Tecnologias previstas		Modelo de transcrição automática			
Tipos de dados necessários		Áudio e texto pareado			
Descrição	As tarefas se resumirão ao treinamento de rede neural profunda para a realização da transcrição automática. O treinamento será um aprimoramento a partir do modelo de transcrição geral, entregue no projeto de 2020.				
E-E01.3		Início	M8	Fim	M12

	Título	Construção, avaliação e aprimoramento do modelo de transcrição específico para o segundo cenário de aplicação.		
	Risco vinculado	R-E01.1		
Tecnologias previstas		Modelo de transcrição automática		
Tipos de dados necessários		Áudio e texto pareado		
Descrição	As tarefas se resumirão ao treinamento de rede neural profunda para a realização da transcrição automática. O treinamento será um aprimoramento a partir do modelo de transcrição geral, entregue no projeto de 2020.			
Entregáveis				
E01.1	E-E01.1	Relatório com a análise de viabilidade de possíveis cenários de aplicação. Definição de dois cenários de aplicação mais promissores.		
	M2			
E01.2	E-E01.2	Implementação de uma rede neural profunda capaz de transcrever áudios automaticamente, considerando o primeiro cenário de aplicação (que será definido).		
	M8			
E01.3	E-E01.3	Implementação de uma rede neural profunda capaz de transcrever áudios automaticamente, considerando o segundo cenário de aplicação (que será definido).		
	M12			
Milestones				
M-E01.1	Início do projeto. Levantamento da equipe.			
M1				
M-E01.2	Levantamento e aquisição dos dados necessários para o treinamento da rede neural profunda.			
M3				
M-E01.3	Construção de modelo de linguagem específico para o primeiro cenário de aplicação.			
M4				
M-E01.4	Elaboração de técnicas de transferência de aprendizado, de forma a aprimorar a rede neural profunda geral produzida no projeto de 2020.			
M5				
M-E01.5	Implementação e avaliação das técnicas de transferência de aprendizado, de forma a aprimorar a rede neural profunda geral produzida no projeto de 2020. Foco no primeiro cenário de aplicação.			
M6				
M-E01.6	Aprimoramento dos algoritmos desenvolvidos.			
M7				

M-E01.7	Construção de modelo de linguagem específico para o segundo cenário de aplicação.
M9	
M-E01.8	Implementação e avaliação das técnicas de transferência de aprendizado, de forma a aprimorar a rede neural profunda geral produzida no projeto de 2020. Foco no segundo cenário de aplicação.
M10	
M-E01.9	Aprimoramento dos algoritmos desenvolvidos.
M11	

Riscos Críticos para a Implementação				
Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-E01.1	Falta de dados necessários para o treinamento do modelo de linguagem	Alto	Travar	Não há

Cronograma das Entregas por Etapas													
ETAPAS		ENTREGAS											
CÓD.	TÍTULO	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-E01.1	Definição dos cenários de aplicação		E01.1										
E-E01.2	Construção, avaliação e aprimoramento do modelo de transcrição específico para o primeiro cenário de aplicação.								E01.2				
E-E01.3	Construção, avaliação e												E01.3

	aprimoramento do modelo de transcrição específico para o segundo cenário de aplicação.											
MILESTONES	M-E01.1		M-E01.2	M-E01.3	M-E01.4	M-E01.5	M-E01.6		M-E01.7	M-E01.8	M-E01.9	

Etapas - 2022					
E-E01.4		Início	M13	Fim	M14
	Título	Definição dos cenários de aplicação			
	Descrição	As tarefas se resumirão a reuniões para definição de dois cenários de aplicação para o modelo de transcrição em 2022.			
E-E01.5		Início	M14	Fim	M20
	Título	Construção, avaliação e aprimoramento do modelo de transcrição específico para o primeiro novo cenário de aplicação.			
	Descrição	As tarefas se resumirão ao treinamento de rede neural profunda para a realização da transcrição automática. O treinamento será um aprimoramento a partir do modelo de transcrição geral, entregue no projeto de 2021.			
E-E01.6		Início	M20	Fim	M24
	Título	Construção, avaliação e aprimoramento do modelo de transcrição específico para o segundo novo cenário de aplicação.			

	Descrição	As tarefas se resumirão ao treinamento de rede neural profunda para a realização da transcrição automática. O treinamento será um aprimoramento a partir do modelo de transcrição geral, entregue no projeto de 2021.
Entregáveis - 2022		
E01.4	E-E01.11	Relatório com a análise de viabilidade de possíveis cenários de aplicação. Definição de dois cenários de aplicação mais promissores.
	M14	
E01.5	E-E01.12	Implementação de uma rede neural profunda capaz de transcrever áudios automaticamente, considerando o primeiro novo cenário de aplicação (que será definido para 2022).
	M20	
E01.6	E-E01.13	Implementação de uma rede neural profunda capaz de transcrever áudios automaticamente, considerando o segundo novo cenário de aplicação (que será definido para 2022).
	M24	

Grupo de Trabalho	E02 - Arcabouço computacional para detecção de notícias falsas				
Grupos de Interação	C01, F04, G01, M01				
Coordenador	Fabrício Benevenuto				
Equipe	Nível	Quantidade	Nomes		
	Professor	3	Fabrício Benevenuto, Jussara M. Almeida, Ana Paula Couto		
	Pós-Doc	1	TOTAL	9	
	Doutorando	2			
	Mestrando	1			
	Graduando	2			
	Técnico	0			
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Não	Não	Sim	Não	Não
Descrição Geral					
Este módulo de software visa desenvolver um arcabouço computacional para identificação de notícias falsas disseminadas em mídias sociais, com base em dados oriundos de sites de checagem de fatos e algoritmos de aprendizado de máquina supervisionado.					
Objetivos					

O principal objetivo a ser alcançado neste projeto é a construção de um arcabouço computacional capaz identificar, com acurácia, informações verdadeiras e falsas. A seguir desmembramos nosso projeto em três objetivos que constituem passos necessários para atingirmos nosso objetivo final.

1) Construção de uma base de dados rotulada para o problema de detecção de informações falsas. Primeiramente, será necessário desenvolver uma série de coletores de dados de sites de checagem de fatos brasileiros, como Agência Lupa, Aos Fatos, Fato ou Fake, Projeto Comprova, etc. O foco será nas agências reconhecidas por entidades do jornalismo brasileiro ou internacionais, como a IFCN (International Fact-Checking Network). Depois disso, uma tarefa complexa consiste em implementar uma abordagem capaz de correlacionar adequadamente uma notícia ou informação falsa a uma checagem. Isso é necessário pois normalmente os sites de checagem apenas expõem fatos que são falsos ou verdadeiros, não exibindo a fonte da desinformação. Pretendemos nessa etapa utilizar dados de redes sociais a partir de ferramentas já desenvolvidas e aplicativos de mensagens, como WhatsApp e Telegram, para a identificação da informação falsa a que se refere a checagem. A ideia geral é buscar por informações compartilhadas em especial no WhatsApp contendo mensagens (ex. imagens) com informação que possa ser rotulada como falsa ou verdadeira, de acordo com checagens disponíveis nos mais diversos sites de checagem de fatos nacionais. Cabe ressaltar que a tarefa de detecção de notícias falsas é dependente do contexto, ou seja, um modelo treinado com uso de dados do contexto político pode não apresentar um desempenho adequado quando aplicado em um contexto diferente, como no caso da saúde. A criação de um modelo capaz de identificar qualquer tipo de notícia falsa ainda está distante e trata-se de um problema em aberto em que vários pesquisadores do mundo estão explorando. Nossa abordagem aqui consiste na definição de um escopo factível e que corrobora com esse esforço global em busca de um mecanismo de detecção automática de notícias falsas. Nossa abordagem consiste na exploração do contexto político, o que pode resultar em uma ferramenta útil para o MPMG nas eleições seguintes ao projeto. Por fim, durante esta etapa também será realizado um levantamento dos datasets rotulados disponíveis e comumente utilizados para a tarefa de detecção de notícias falsas. A ideia é que esses dados podem ser utilizados em conjunto com a base de dados construída para melhoria da etapa de treinamento dos algoritmos de aprendizado de máquina.

2) Levantamento e implementação de features (ex. padrões típicos) para detecção de notícias falsas. Existem trabalhos recentes que propõem features que podem ser úteis para a detecção de notícias falsas. Esta etapa do projeto inclui um levantamento destes trabalhos para identificação das principais features utilizadas para detecção de notícias falsas. Depois disso, nosso esforço será direcionado a investigar a viabilidade, e implementar as principais features considerando a base de dados criadas na etapa 1).

3) Desenvolvimento de um arcabouço computacional para identificação de notícias falsas. Nesta etapa, iremos explorar o conjunto de dados construído na etapa 1) e as features implementadas na etapa 2) para a construção de um modelo baseado em aprendizado de máquina supervisionado que, dada uma informação (ex. uma imagem disseminada no WhatsApp), fornecerá como resultado a probabilidade dela ser falsa.

Etapas

E-E02.1	Início	M1	Fim	M4
Título	Base de dados de Desinformação			

		Risco vinculado	R-E02.1		
			R-E02.2		
Tecnologias previstas		Python			
Tipos de dados necessários					
Descrição	Construção de uma base de dados rotulada para o problema de detecção de conteúdo falsas.				
E-E02.2		Início	M5	Fim	M7
	Título	Levantamento e Implementação de Features com poder discriminatório			
		Risco vinculado	R-E02.1		
			R-E02.2		
Tecnologias previstas		Python			
Tipos de dados necessários					
Descrição	Levantamento, análise de viabilidade e implementação de features com poder discriminatório para detecção de desinformação.				
E-E02.3		Início	M8	Fim	M12
	Título	Detecção de desinformação através de Aprendizado de Máquina			
		Risco vinculado	R-E02.1	R-E02.4	
			R-E02.3		
Tecnologias previstas		Python (Sklearn)			
Tipos de dados necessários					
Descrição	Desenvolvimento de um arcabouço computacional para identificação de desinformação.				
Entregáveis					
E02.1	E-E02.1	Relatório contendo informações gerais sobre a base de dados construída, como detalhes do processo, volume de dados, etc e o código dos coletores dos sites de			
	M4				

E02.2	E-E02.2	cheecagem de fatos. Documento contendo a descrição das features implementadas bem como uma análise de viabilidade e códigos utilizados para implementação das mesmas.
	M7	
E02.3	E-E02.3	Documentação e código do arcabouço computacional para identificação de notícias falsas. Como arcabouço computacional pretendemos entregar um API em python contendo uma série de funções que permitem o uso do modelo de detecção de notícias falsas. Um exemplo de função recebe uma imagem compartilhada no WhatsApp e retorna a probabilidade da informação contida nesta mensagem ser falsa.
	M12	
Milestones		
M-E02.1	Definição da equipe.	
M1		
M-E02.2	Definição dos sites de checagem a serem coletados.	
M2		
M-E02.3	Implementação dos coletores dos sites de checagem.	
M3		
M-E02.4	Levantamento das features para detecção de notícias falsas que serão implementadas.	
M5		
M-E02.5	Implementação das features para detecção de notícias falsas.	
M6		
M-E02.6	Definição das abordagens de aprendizado de máquina que serão testadas.	
M8		
M-E02.7	Testes com abordagens de aprendizado de máquina supervisionado definidas anteriormente.	
M9		
M-E02.8	Definição do algoritmo que será utilizado para proposição do modelo final.	
M10		
M-E02.9	Versão inicial (para testes locais) da API em python que irá conter algumas funções que permitem o uso do modelo de detecção de notícias falsas.	
M11		

Riscos Críticos para a Implementação

Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-E02.1	Mudança de equipe.	Médio	Aceitar	Realizar novo processo para contratação de recurso.
R-E02.2	A coleta de dados de checagens de fatos será feita com base na estrutura HTML das páginas. Assim, um risco é a mudança dessa estrutura HTML dos sites de checagem coletados depois do desenvolvimento dos coletores.	Médio	Aceitar	Uma vez identificada uma mudança na estrutura dos sites ainda na fase da criação da base de dados, pretendemos adequar os coletores implementados para considerar a novo layout do site coletado.
R-E02.3	O desempenho do modelo a ser construído está relacionada à qualidade e volume dos dados utilizados durante a etapa de treino.	Médio	Mitigar	Ampliar o conjunto de dados de treino com o uso de base para detecção de notícias falsas disponíveis na literatura.
R-E02.4	Ao final, o modelo desenvolvido pode não atingir um alto nível de precisão.	Baixo	Aceitar	<p>Nesse caso, pretendemos apresentar uma série de esforços e direcionamentos com resultados empíricos de tentativas de melhorar a capacidade de predição do arcabouço construído. Cabe ressaltar, que nosso grupo tem grande experiência com algoritmos de aprendizado de máquina e sabe explorar com metodologia científica diversas possibilidades capazes de melhorar o resultado final para essa tarefa. Além disso, trabalhos anteriores do nosso grupo [1] realizado no contexto político norte americano resultou em modelos de aprendizado promissores, o que nos dá confiança que poderemos realizar a detecção de notícias falsas no contexto brasileiro com boa acurácia ao final do projeto. Sendo assim, acreditamos que esse risco é pequeno, mas somente a execução do projeto poderá confirmar nossas expectativas. [1] Supervised Learning for Fake News Detection.</p> <p>Julio Reis, André Correia, Fabrício Murai, Adriano Veloso, Fabrício Benevenuto.</p>

In IEEE Intelligent Systems. Volume 34, Number 2. March 2019.

Cronograma das Entregas por Etapas													
Etapas		Entregas											
Cód.	Título	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E- E02.1	Base de dados de Desinformação				E02.1								
E- E02.2	Levantamento e Implementação de Features com poder discriminatório							E02.2					
E- E02.3	Deteção de desinformação através de Aprendizado de Máquina												E02.3
MILESTONES		M- E02.1	M- E02.2	M- E02.3		M- E02.4	M- E02.5		M- E02.6	M- E02.7	M- E02.8	M- E02.9	

Etapas - 2022			
E- E02.4	Início	M13	
	Fim	M16	
	Título	Construção de bases de dados rotuladas para o problema de deteção de conteúdo falsas em múltiplos contextos.	
Descrição	Conforme mencionado anteriormente a tarefa de deteção de notícias falsas é dependente do contexto. Assim, para o projeto de 2022, pretendemos explorar o problema considerando múltiplos contextos. Particularmente nosso foco principal será saúde, com o uso de dados relacionados à pandemia de COVID-19 em diferentes países. Neste cenário será necessário construirmos bases de dados rotuladas que nos permitam explorar adequadamente o problema de deteção de notícias falsas em múltiplos contextos.		

		Início	M17	Fim	M19
E-E02.5	Título	Enriquecimento das bases de dados construídas e implementação, e análise de features.			
	Descrição	Uma etapa metodológica fundamental está relacionada ao enriquecimento dos dados coletados na etapa anterior (e que serão utilizados para treinamento dos modelos) com informações de propagação das notícias em plataformas sociais, e outras features que possam ser úteis para detecção acurada de notícias falsas em múltiplos contextos. Essa investigação é importante pois as features implementadas durante o projeto de 2021, cujo foco é o contexto político, podem não ser suficientemente informativas para a detecção de notícias falsas em outros contextos. Nós esperamos que nesta etapa do projeto sejamos capazes de fornecer uma perspectiva sobre features que se mantêm ou não úteis para a detecção de notícias falsas considerando diferentes contextos.			
		Início	M20	Fim	M24
E-E02.6	Título	Explorar métodos de detecção de notícias falsas em múltiplos contextos.			
	Descrição	Nesta etapa pretendemos explorar diferentes estratégias de aprendizado de máquina para detectar notícias falsas disseminadas em múltiplos contextos usando dados e features explorados durante as etapas iniciais. Particularmente, planejamos investigar o potencial preditivo de outras categorias de métodos de aprendizagem de máquina, como por exemplo, abordagens ativas para atualização constante do modelo e o uso de abordagens baseadas em aprendizado profundo.			
Entregáveis - 2022					
E02.4	E-E02.11	Relatório contendo informações gerais acerca das bases de dados (rotulados) oriundos de múltiplos contextos.			
	M16				
E02.5	E-E02.12	Relatório contendo uma descrição do processo de enriquecimento das bases de dados, bem como análise de viabilidade e features implementadas para detecção de notícias falsas em múltiplos contextos.			
	M19				
E02.6	E-E02.13	Documento e métodos para detecção de notícias falsas em múltiplos contextos incorporados ao arcabouço computacional desenvolvido durante o ano de 2021.			
	M24				

Grupo de Trabalho	E03 - Análise preditiva através de aprendizado explicativo				
Grupos de Interação	E05, G01, M01				
Coordenador	Adriano Alonso Veloso				
Equipe	Nível	Quantidade	Nomes		
	Professor	1	Adriano Alonso Veloso		
	Pós-Doc	0	TOTAL	4	
	Doutorando	1			
	Mestrando	0			
	Graduando	1			
	Técnico	1	--		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Não	Não	Não	Não	Não
Descrição Geral					
Solução para predição, antecipação ou identificação de desfechos de dois temas do interesse do MPMG; Solução de auxílio no entendimento de fenômenos e processos complexos através do uso de conceitos de explicabilidade de predições.					
Objetivos					
Com base nos locais de ocorrência de eventos, que podem ser, por exemplo, crimes, determinar onde as forças e/ou medidas de combate podem ser posicionadas. Prover explicações acerca das predições, de forma a permitir contramedidas de prevenção/ação.					
Etapas					

E-E03.1		Início	M1	Fim	M3
	Título	Definição do tipo de evento a ser trabalhado e aquisição de dados necessários para o treinamento do modelos preditivo.			
		Risco vinculado	R-E03.1		
Tecnologias previstas		nenhuma			
Tipos de dados necessários					
Descrição	As tarefas se resumirão a reuniões para definição sobre dados disponíveis, bem como os mecanismos necessários para aquisição dos dados.				
E-E03.2		Início	M3	Fim	M7
	Título	Modelo preditivo para antecipação de ocorrência de eventos.			
		Risco vinculado	R-E03.1		
Tecnologias previstas		Modelo preditivo capaz de antecipar ocorrências de diferentes tipos de eventos.			
Tipos de dados necessários					
Descrição	Tarefas relacionadas à predição de futuros eventos. Construção de modelos preditivos. Melhoria da representação dos dados. Aprimoramento dos modelos preditivos.				
E-E03.3		Início	M7	Fim	M12
	Título	Modelo explicativo capaz de fornecer intuições sobre a efetividade das medidas preventivas e/ou de combate..			

		Risco vinculado	R-E03.1		
Tecnologias previstas		Modelo explicativo capaz de fornecer intuições acerca das predições realizadas. Tais explicações são utilizadas para a formação do entendimento acerca da ocorrência de eventos.			
Tipos de dados necessários					
Descrição	Tarefas relacionadas à construção de modelo explicativos. Remoção de características redundantes. Remoção de sinais confundidores. Implementação de metáforas visuais que permitam o entendimento das predições.				
Entregáveis					
E03.1	E-E03.1	Relatório com a análise de viabilidade de possíveis cenários de aplicação. Definição de dois cenários de aplicação mais promissores.			
	M3				
E03.2	E-E03.2	Modelo preditivo implementado, testado e validado.			
	M7				
E03.3	E-E03.3	Modelo explicativo implementado, testado e validado com os especialistas.			
	M12				
Milestones					
M-E03.1	Início do projeto. Levantamento da equipe e definição dos eventos a serem trabalhados.				
M1					
M-E03.2	Implementação de mecanismos para obtenção dos dados necessários para o treinamento do modelo preditivo.				
M2					
M-E03.3	Seleção das variáveis mais relevantes para a modelagem de eventos.				

M4	
M-E03.4	Criação de variáveis mais discriminativas a partir das variáveis fornecidas. Construção do modelo preditivo.
M5	
M-E03.5	Avaliação da efetividade das predições. Aprimoramento final do modelo preditivo.
M6	
M-E03.6	Implementação do modelo explicativo.
M8	
M-E03.7	Avaliação do modelo explicativo.
M9	
M-E03.8	Implementação de metáforas visuais que permitam a formação de entendimento acerca do policiamento.
M10	
M-E03.9	Validação das explicações com especialistas. Explicações consideradas inválidas (i.e., discriminação, injustiça etc.) propiciarão a retirada de variáveis do modelo preditivo.
M11	

Riscos Críticos para a Implementação

Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-E03.1	Falta de dados apropriados para treinamento do modelo preditivo	Alto	Travar	Nenhum

Cronograma das Entregas por Etapas

ETAPAS	ENTREGAS

CÓD.	TÍTULO	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E- E03.1	Definição do tipo de evento a ser trabalhado e aquisição de dados necessários para o treinamento do modelos preditivo			E03.1									
E- E03.2	Modelo preditivo para antecipação de ocorrência de eventos							E03.2					
E- E03.3	Modelo explicativo capaz de fornecer intuições sobre a efetividade das medidas preventivas e/ou de combate												E03.3
MILESTONES		M- E03.1	M- E03.2		M- E03.3	M- E03.4	M- E03.5		M- E03.6	M- E03.7	M- E03.8	M- E03.9	

Etapas - 2022

E-E03.4		Início	M13	Fim	M15
	Título	Definição do tipo de evento a ser trabalhado e aquisição de dados necessários para o treinamento do modelos preditivo.			
	Descrição	As tarefas se resumirão a reuniões para definição sobre dados disponíveis, bem como os mecanismos necessários para aquisição dos dados.			
E-E03.5		Início	M15	Fim	M19
	Título	Modelo preditivo para antecipação de ocorrência de eventos.			
	Descrição	Tarefas relacionadas à predição de futuros eventos. Construção de modelos preditivos. Melhoria da representação dos dados. Aprimoramento dos modelos preditivos.			
E-E03.6		Início	M19	Fim	M24
	Título	Modelo explicativo capaz de fornecer intuições sobre a efetividade das medidas preventivas e/ou de combate..			
	Descrição	Tarefas relacionadas à construção de modelo explicativos. Remoção de características redundantes. Remoção de sinais confundidores. Implementação de metáforas visuais que permitam o entendimento das predições.			
Entregáveis - 2022					
E03.4	E-E03.11	Relatório com a análise de viabilidade de possíveis cenários de aplicação. Definição de dois cenários de aplicação mais promissores.			
	M15				
E03.5	E-E03.12	Modelo preditivo implementado, testado e validado.			
	M19				
E03.6	E-E03.13	Modelo explicativo implementado, testado e validado com os especialistas.			

M24

Grupo de Trabalho	E04 - Reconhecimento de Pessoas e Detecção de Objetos em Imagens e Vídeos				
Grupos de Interação	C01, E05, F05, G01				
Coordenador	Jefersson Alex dos Santos				
Equipe	Nível	Quantidade	Nomes		
	Professor	1	Jefersson Alex dos Santos		
	Pós-Doc	0	TOTAL	8	
	Doutorando	1			
	Mestrando	1			
	Graduando	3			
	Técnico	2	Desenvolvedor Web		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Não	Não	Não	Não	Não
Descrição Geral					
Ferramentas e funcionalidades para reconhecimento de pessoas e objetos-chave em imagens e vídeos.					
Objetivos					
Desenvolver ferramentas robustas para reconhecimento de pessoas e detecção de objetos-chave em imagens e vídeos.					
Etapas					

E-E04.1	Início	M1	Fim	M2	
	Título	Levantamento de requisitos.			
	Risco vinculado				
Tecnologias previstas		--			
Tipos de dados necessários		Será necessário definir com a equipe do MPMG os objetos-chave e as fontes de imagens para reconhecimento de pessoas.			
Descrição	Levantamento de requisitos para elaboração de bases de dados para reconhecimento de pessoas e objetos-chave. Essa etapa contará com a interação com a equipe do MPMG.				
E-E04.2	Início	M3	Fim	M4	
	Título	Criação de conjuntos de dados para aprendizado supervisionado.			
	Risco vinculado	R-E04.1			
Tecnologias previstas		--			
Tipos de dados necessários		O conjunto de dados para treinamento e avaliação dos modelos será composto por imagens coletadas de datasets públicos e/ou indicações do MP-MG. Estamos considerando que será necessária anotação manual de imagens e vídeos nesse processo.			
Descrição	Montagem e coleta de dados para treinamento dos modelos. Serão feitas tanto a coleta de dados de repositórios públicos quanto eventuais anotações de imagens fornecidas pelo MP-MG.				
E-E04.3	Início	M5	Fim	M7	

	Título	Levantamento de abordagens do estado da arte			
	Risco vinculado				
Tecnologias previstas		--			
Tipos de dados necessários		--			
Descrição	<p>Estudo completo e implementação de modelos do estado da arte para:</p> <p>(1) retornar pessoas parecidas baseadas em similaridades faciais;</p> <p>(2) detecção de objetos-chave em imagens e vídeos.</p>				
		Início	M8	Fim	M10
E-E04.4	Título	Consolidação e desenvolvimento das ferramentas			
	Risco vinculado		R-E04.2		
Tecnologias previstas		--			
Tipos de dados necessários		--			
Descrição	Consolidação dos modelos e desenvolvimento das ferramentas para duas aplicações (reconhecimento facial e detecção de objetos).				
		Início	M11	Fim	M12
E-E04.5	Título	Integração e ajuste das ferramentas desenvolvidas			
	Risco vinculado				

			R-E04.2		
Tecnologias previstas					
Tipos de dados necessários					
Descrição	Integração e ajustes dos módulos e ferramentas desenvolvidas. Serão realizados testes simulando situações reais de uso das ferramentas desenvolvidas.				
Entregáveis					
E04.1	E-E04.1	Documento de requisitos.			
	M2				
E04.2	E-E04.2	Conjuntos de dados rotulados.			
	M4				
E04.3	E-E04.3	Relatório e códigos dos principais métodos levantados/implementados.			
	M7				
E04.4	E-E04.4	Códigos das ferramentas consolidadas e testadas.			
	M10				
E04.5	E-E04.5	Relatório de simulações realizadas com dados reais após integração dos módulos.			
	M12				
Milestones					
M-E04.1	Versão preliminar do documento de requisitos.				
M1					
M-E04.2	Versão preliminar dos dados públicos coletados para treinamento supervisionado dos modelos.				
M3					

M-E04.3	Listagem dos principais métodos existentes na literatura para cada aplicação.
M5	
M-E04.4	Relatório com resultados preliminares dos principais métodos existentes na literatura para cada aplicação.
M6	
M-E04.5	Versão preliminar dos códigos implementados.
M8	
M-E04.6	Códigos e relatório com resultados preliminares.
M9	
M-E04.7	Relatório preliminar de simulações realizadas com dados reais após integração dos módulos.
M11	

Riscos Críticos para a Implementação

Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-E04.1	Dificuldades em encontrar repositórios acessíveis para realização de simulação de situações reais de uso das ferramentas.	Baixo	Mitigar	Deveremos utilizar e adaptar conjuntos de dados públicos para simular os efeitos do uso em situações reais.
R-E04.2	Dificuldades para montagem da equipe com a experiência exigida em visão computacional e aprendizado de máquina.	Médio	Mitigar	Treinamento rápido de alunos e/ou técnicos nos dois primeiros meses do projeto.

Cronograma das Entregas por Etapas

Etapas		Entregas											
Cód.	Título	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12

E- E04.1	Levantamento de requisitos.		E04.1										
E- E04.2	Criação de conjuntos de dados para aprendizado supervisionado.				E04.2								
E- E04.3	Levantamento de abordagens do estado da arte.							E04.3					
E- E04.4	Consolidação e desenvolvimento das ferramentas.										E04.4		
E- E04.5	Integração e ajuste das ferramentas desenvolvidas.												E04.5
MILESTONES		M- E04.1		M- E04.2		M- E04.3	M- E04.4		M- E04.5	M- E04.6		M- E04.7	

Etapas - 2022

	Início	M13	Fim	M14
E- E04.11	Título	Levantamento de requisitos.		
	Descrição	Levantamento de requisitos para elaboração de bases de dados para reconhecimento de placas de automóveis. Essa etapa contará com a interação com a equipe do MPMG.		

E- E04.12		Início	M14	Fim	M15
	Título	Criação de conjuntos de dados para reconhecimento de placas.			
	Descrição	O conjunto de dados para treinamento e avaliação dos modelos será composto por imagens coletadas de datasets públicos e/ou indicações do MP-MG.			
E- E04.13		Início	M16	Fim	M20
	Título	Desenvolvimento dos métodos			
	Descrição	Levantamento e desenvolvimento de abordagens do estado da arte para reconhecimento de placas de automóveis.			
E- E04.14		Início	M21	Fim	M24
	Título	Consolidação e integração com outros módulos			
	Descrição	Consolidação, integração e realização de testes simulando situações reais de uso.			
Entregáveis - 2022					
E04.11	E-E04.11	Documento de requisitos sobre reconhecimento de placas de automóveis.			
	M14				
E04.12	E-E04.12	Conjunto de dados anotados para aprendizado supervisionado de padrões de placas de automóveis.			
	M15				
E04.13	E-E04.13	Códigos dos métodos implementados e testados.			
	M20				
E04.14	E-E04.14	Relatório de testes realizados simulando situações reais de uso das ferramentas.			

M24

Grupo de Trabalho	E05 - Serviços de Geocodificação				
Grupos de Interação	A01, F05, G01				
Coordenador	Clodoveu Davis				
Equipe	Nível	Quantidade	Nomes		
	Professor	2	Clodoveu Davis, Jefersson A. dos Santos		
	Pós-Doc	0	TOTAL	6	
	Doutorando	1			
	Mestrando	1			
	Graduando	2			
	Técnico	0	--		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Não	Não	Não	Não	Não
Descrição Geral					
<p>O termo geocodificação define a capacidade de determinar posições ou identificar lugares com base em descrições textuais. A fonte mais comum de descrições é o endereçamento postal urbano. Outras fontes incluem nomes de lugares e hierarquias de subdivisão territorial cujos componentes são nomeados. De forma mais ampla, o conceito de geocodificação pode ser estendido para admitir como descrição de entrada uma imagem, que seria comparada a imagens localizadas contidas em uma base de referência, para que se obtenha uma lista de lugares candidatos por ordem de similaridade. Denomina-se geocodificação reversa a capacidade de produzir descrições a partir de coordenadas ou delimitações espaciais dadas. A geocodificação reversa é particularmente interessante quando se busca caracterizar lugares cujas coordenadas são conhecidas (por exemplo a partir do rastreamento de smartphones usando GPS ou outras técnicas) para uso de investigadores.</p>					

Objetivos

- (1) Prover serviços capazes de realizar a geocodificação de endereços dentro do estado de Minas Gerais, combinando fontes de endereçamento abertas e serviços comerciais;
- (2) Prover serviços capazes de identificar lugares associados a um nome fornecido, com base em um dicionário toponímico, construído a partir de múltiplas fontes de dados abertas;
- (3) Prover serviços de geração da descrição de lugares a partir de coordenadas geográficas fornecidas (geocodificação reversa);
- (4) Prover serviços de indicação de prováveis lugares associados a imagens fornecidas como entrada, por meio de técnicas de recuperação de imagens.

Etapas

		Início	M1	Fim	M2
E-E05.1	Título	Levantamento de métodos para geotagging			
	Risco vinculado	R-E05.1			
Tecnologias previstas					
Tipos de dados necessários					
Descrição	Revisão de literatura e testes preliminares com métodos para geotagging.				
		Início	M1	Fim	M3
E-E05.2	Título	Construção do dicionário toponímico de Minas Gerais			
	Risco vinculado	R-E05.1			
Tecnologias previstas		--			
Tipos de dados necessários		Fontes de dados abertos com endereços do Estado de Minas Gerais.			

Descrição	Criação de um dicionário toponímico construído a partir de múltiplas fontes de dados abertas. Toponímia é o estudo dos nomes de lugares/regiões.				
E-E05.3		Início	M2	Fim	M4
	Título	Construção do banco de dados de endereçamento			
	Risco vinculado	R-E05.1			
Tecnologias previstas	--				
Tipos de dados necessários	Fontes de dados abertos de endereçamento urbano				
Descrição	Criação de uma base de referência, para que se obtenha uma lista de lugares candidatos para realização de geotagging.				
E-E05.4		Início	M1	Fim	M6
	Título	Ampliação da base de imagens de referência.			
	Risco vinculado	R-E05.1			
Tecnologias previstas	--				
Tipos de dados necessários	--				
Descrição	Ampliação e eventuais ajustes na base de imagens coletadas no projeto com o MPMG em 2021. A ideia é obter um conjunto de imagens com diversidade e robustez suficiente para permitir a realização de geotagging.				

E-E05.5	Início	M3	Fim	M8
	Título	Desenvolvimento da API: geocodificação de endereços		
	Risco vinculado	R-E05.1		
Tecnologias previstas		--		
Tipos de dados necessários		--		
Descrição	Desenvolvimento da API de geocodificação, ou seja, ferramentas para determinar posições ou identificar lugares com base em descrições textuais.			
E-E05.6	Início	M2	Fim	M7
	Título	Desenvolvimento da API: identificação de lugares nomeados		
	Risco vinculado	R-E05.1		
Tecnologias previstas		--		
Tipos de dados necessários		--		
Descrição	Desenvolvimento de API para identificação e recomendação de imagens a partir de descrições textuais.			

E-E05.7	Início	M6	Fim	M9
	Título	Desenvolvimento da API: geocodificação reversa		
	Risco vinculado	R-E05.1		
Tecnologias previstas		--		
Tipos de dados necessários		--		
Descrição	Desenvolvimento de API para agregar descrições textuais a uma dada coordenada geográfica.			
E-E05.8	Início	M5	Fim	M11
	Título	Desenvolvimento da API: localização de imagens		
	Risco vinculado			
Tecnologias previstas		--		
Tipos de dados necessários		--		
Descrição	Desenvolvimento de API para recomendação de imagens a uma dada coordenada geográfica.			
E-E05.9	Início	M3	Fim	M12
	Título	Testes e atividades de integração com outros projetos		

		Risco vinculado			
Tecnologias previstas		--			
Tipos de dados necessários		--			
Descrição	Integração dos módulos do projeto e treinamento dos modelos com base de dados atualizada. Deverão ser realizados testes simulando situações reais de uso das ferramentas.				
Entregáveis					
E05.1	E-E05.1	Relatório dos principais métodos levantados.			
	M2				
E05.2	E-E05.2	Dicionário toponímico de Minas Gerais			
	M3				
E05.3	E-E05.3	Banco de dados de endereçamento urbano aberto			
	M4				
E05.4	E-E05.4	Banco de imagens de referência			
	M6				
E05.5	E-E05.6	API de identificação de lugares nomeados			
	M7				
E05.6	E-E05.5	API de geocodificação de endereços			
	M8				
E05.7	E-E05.7	API de geocodificação reversa			

	M9	
E05.8	E-E05.8	API de identificação da localização associada a imagens
	M11	
E05.9	E-E05.9	Códigos e relatórios dos testes realizados.
	M12	
Milestones		
M-E05.1	Relatório parcial dos métodos levantados.	
M1		
M-E05.2	Relatório parcial com descrição dos conjuntos de dados de endereçamento e imagens de referência.	
M5		
M-E05.3	Relatório parcial com testes preliminares das APIs em desenvolvimento.	
M6		
M-E05.4	Relatório parcial da API de recomendação de imagens de endereços.	
M10		

Riscos Críticos para a Implementação

Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-E05.1	Alocação de estudantes.	Alto	Mitigar	Antecipar as buscas por candidatos.

Cronograma das Entregas por Etapas

Etapas		Entregas											
Cód.	Título	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E- E05.1	Levantamento de métodos para geotagging		E05.1										
E- E05.2	Construção do dicionário toponímico de Minas Gerais			E05.2									
E- E05.3	Construção do banco de dados de endereçamento				E05.3								
E- E05.4	Ampliação da base de imagens de referência.						E05.4						
E- E05.5	Desenvolvimento da API: geocodificação de endereços								E05.6				
E- E05.6	Desenvolvimento da API: identificação de lugares nomeados							E05.5					
E- E05.7	Desenvolvimento da API: geocodificação reversa									E05.7			
E- E05.8	Desenvolvimento da API:											E05.8	

	localização de imagens												
E-E05.9	Testes e atividades de integração com outros projetos												E05.9
MILESTONES		M-E05.1				M-E05.2	M-E05.3					M-E05.4	

Etapas - 2022					
E-E05.11		Início	M13	Fim	M18
	Título	Expansão das fontes de referência para o Brasil			
	Descrição				
E-E05.12		Início	M19	Fim	M24
	Título	Expansão da funcionalidade de geocodificação e geocodificação reversa			
	Descrição				
Entregáveis - 2022					
E05.11	E-E05.11	Gazetteer brasileiro e base de endereçamento nacional.			
	M18				
E05.12	E-E05.12	API de geocodificação expandida.			
	M22				

Cronograma das Entregas por Etapas													
ETAPAS		ENTREGAS											
CÓD.	TÍTULO	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-E06.1	Levantamento Bibliográfico				E06.3								
E-E06.2	Definição de Padrão de Dados			E06.1									
E-E06.3	Definição de Protocolos			E06.2									
E-E06.4	Implementação de Piloto						E06.4						
E-E06.5	Melhorias constantes da solução/expansão dos casos de testes									E06.5			E06.6
MILESTONES		M-E06.1	M-E06.2			M-E06.3		M-E06.4	M-E06.5		M-E06.6	M-E06.7	

Etapas 2022					
E-E06.11		Início	M13	Fim	M18
	Título	Definição de Novo Padrão de Dados			
	Descrição	Será definido o padrão de dados de acordo com a demanda do MPMG.			
E-E06.12		Início	M13	Fim	M18

	Título	Definição de Protocolos			
	Descrição	Serão definidos os protocolos de comunicação.			
		Início	M13	Fim	M18
E-E06.13	Título	Implementação do Piloto			
	Descrição	Será desenvolvido e disponibilizado um módulo para testes.			
		Início	M19	Fim	M24
E-E06.14	Título	Melhorias constantes da solução/expansão dos casos de testes			
	Descrição	Ajustes e melhorias da solução proposta e testada na etapa E-E06.13.			
Entregáveis - 2022					
E06.11	E-E06.11	Relatório Parcial dos Padrões de Dados			
	M15				
E06.12	E-E06.12	Relatório Parcial dos Protocolos			
	M15				
E06.13	E-E06.13	Descrição do Projeto Piloto			
	M18				
E06.14	E-E06.14	Relatório Parcial das melhorias do projeto pilot/expansão da solução			
	M21				
E06.15	E-E06.14	Relatório Final das melhorias do projeto pilot/expansão da solução			

M24

Grupo de Trabalho		E07 - Plataforma Inteligente para a Investigação de Movimentações Bancárias e Vínculos Financeiros			
Grupos de Interação	A01, A02, D01, E06, G01, M04, P04				
Coordenador	Wagner Meira Jr.				
Equipe	Nível	Quantidade	Nomes		
	Professor	6	Alberto Laender, Daniel Macedo, Dorgival Guedes, José Marcos Nogueira, Marcos André Gonçalves, Wagner Meira Jr.		
	Pós-Doc	2	TOTAL	19	
	Doutorando	2			
	Mestrando	2			
	Graduando	2			
	Técnico	5	3 Desenvolvedores Java, 1 Especialista em segurança de redes, 1 Desenvolvedor Web		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Sim	Sim	Sim	Sim	Não
Descrição Geral					
A Plataforma Inteligente para a Investigação de Movimentações Bancárias e Vínculos Financeiros é formada por um conjunto de processos, módulos e normas para tráfego (seguro) de dados bancários entre instituições financeiras e órgãos governamentais.					
Objetivos					
Disponibilizar uma plataforma completa, que permita ao Órgãos Governamentais construir e conduzir investigações utilizando inteligência computacional para a análise de movimentações bancárias e vínculos financeiros.					

Para atingir esse objetivo, será construída uma Plataforma descentralizada e inteligente, que permita a investigação de movimentações bancárias, a quantificação e a sumarização das transações e a análise de vínculos financeiros. A Plataforma será formada por 6 (seis) macro-módulos:

1. **Validador:** Utilizado pelas instituições financeiras para validar os dados de movimentações bancárias, garantindo que estão no formato definido pelo Bacen (Carta Circular 3454/2010), e em caso positivo, efetuar a transmissão segura dos dados.
 1. Permitir o cadastro de Casos, Processos e Inquéritos;
 2. Realizar a validação de formatos e conteúdos dos arquivos de sigilo;
 3. Criar os arquivos para serem transmitidos;

2. **Transmissor:** Utilizado pelas instituições financeiras para criar um par de chaves (uma pública e outra privada) para transmissão dos dados gerados e validados. A chave pública é enviada por meio da Internet aos computadores do respectivo órgão governamental. Quando o cadastro da instituição financeira é aprovado, a chave pública é assinada pelo órgão governamental e enviada automaticamente ao módulo que gerou a chave (instituição financeira). A partir deste ponto, a instituição financeira que possui a chave assinada pode utilizá-la para transmitir dados bancários ao respectivo órgão governamental, sempre que for requisitado.
 1. Transmitir arquivos pela internet para os servidores dos órgãos;
 2. Garantir a integridade dos dados (HASH);
 3. Garantir a Confidencialidade (Criptografia e SSL);
 4. Garantir o não repúdio (Autenticidade);
 5. Gerar chaves;
 6. Gerar comprovantes de entrega.

3. **Quarentena:**
 1. Processar os arquivos recebidos;
 2. Consultar arquivos recebidos;
 3. Analisar o conteúdo;
 4. Confirmar/aceitar ou recusar os dados recebidos;

4. IA (Inteligência Artificial):

1. Realizar a análise de vínculos por meio de grafos;
2. Disponibilizar os dados em formato para integração com a Plataforma Áduna;
3. Disponibilizar os dados para integração com o módulo de visualização (Painel de BI).

5. Visualização:

1. Disponibilizar metáforas visuais que reflitam os vínculos e movimentações financeiras;
2. Disponibilizar metáforas visuais que quantifiquem os montantes envolvidos nas transações e vínculos;
3. Disponibilizar todas as visualizações em um único Painel de *Business Intelligence* (BI).

6. Relatórios:

1. Gerar relatórios detalhados que sumarizem os vínculos identificados e quantifiquem os valores financeiros movimentados;
2. Disponibilizar os relatórios de análises.

Os módulos 1 e 2 (Validador e Transmissor) são executados na infraestrutura das respectivas instituições financeiras. Os demais módulos são executados na infraestrutura dos respectivos órgãos governamentais.

Etapas

	Início	M1	Fim	M3
E-E07.1	Título	Padrões e protocolos de dados		
	Risco vinculado			
Tecnologias previstas				
Tipos de dados necessários		Documentos descritivos dos padrões/protocolos utilizados pelo SIMBA.		

Descrição	Entendimento dos padrões e protocolos utilizados pelo SIMBA para a transmissão dos dados. Proposta de ajustes/modificações nos padrões e protocolos atuais.				
E-E07.2	Início	M1	Fim	M3	
	Título	Modelagem: Validador e Transmissor			
	Risco vinculado				
Tecnologias previstas					
Tipos de dados necessários					
Descrição	Modelagem e especificação dos requisitos funcionais e não-funcionais dos módulos Validador e Transmissor.				
E-E07.3	Início	M2	Fim	M3	
	Título	Modelagem: Quarentena			
	Risco vinculado				
Tecnologias previstas					
Tipos de dados necessários					
Descrição	Modelagem e especificação dos requisitos funcionais e não-funcionais do módulo Quarentena.				
E-E07.4	Início	M2	Fim	M4	

	Título	Modelagem e especificação dos bancos de dados necessários para o armazenamento dos dados e integração da plataforma			
		Risco vinculado			
Tecnologias previstas					
Tipos de dados necessários					
Descrição	Definição das tecnologias a serem utilizadas, do esquema de dados e do modelo entidade-relacionamento, levando em consideração os dados transmitidos e os módulos Validador, Transmissor e Quarentena.				
E-E07.5		Início	M1	Fim	M12
	Título	Segurança da comunicação entre os módulos			
		Risco vinculado			
Tecnologias previstas					
Tipos de dados necessários					
Descrição	Especificação, modelagem e implementação da segurança na comunicação entre os diversos módulos da plataforma.				
E-E07.6		Início	M1	Fim	M6
	Título	Desenvolvimento: Validador e Transmissor			
		Risco vinculado			

Tecnologias previstas					
Tipos de dados necessários					
Descrição	Implementação dos módulos Validador e Transmissor.				
		Início	M2	Fim	M6
E-E07.7	Título	Desenvolvimento: Quarentena			
	Risco vinculado				
Tecnologias previstas					
Tipos de dados necessários					
Descrição	Implementação do módulo Quarentena.				
		Início	M6	Fim	M8
E-E07.8	Título	Integração: Validador, Transmissor e Quarentena			
	Risco vinculado				
Tecnologias previstas					
Tipos de dados necessários					
Descrição	Integração, teste e validação entre os módulos Validador, Transmissor e Quarentena.				

E-E07.9	Início	M4	Fim	M6
	Título	Implementação dos bancos de dados		
	Risco vinculado			
Tecnologias previstas		PostgreSQL		
Tipos de dados necessários				
Descrição	Criação dos bancos de dados necessários para dar suporte ao armazenamento dos dados transmitidos, funcionamento e integração dos módulos da plataforma.			
E-E07.10	Início	M8	Fim	M10
	Título	Testes de integridade e segurança		
	Risco vinculado			
Tecnologias previstas				
Tipos de dados necessários				
Descrição	Testes de integridade e segurança dos dados e comunicação entre os módulos da plataforma.			
E-E07.11	Início	M1	Fim	M12
	Título	Plataforma Inteligente de Investigação de Movimentações Bancárias e Vínculos Financeiros		

		Risco vinculado			
Tecnologias previstas					
Tipos de dados necessários					
Descrição	Disponibilização de versões preliminares e final da Plataforma.				
Entregáveis					
E07.1	E-E07.5	Documento de especificação dos requisitos de segurança dos módulos e plataforma e tecnologias a serem utilizadas.			
	M2				
E07.2	E-E07.1	Documento com a definição dos padrões de dados e protocolos de transmissão a serem utilizados pela plataforma.			
	M3				
E07.3	E-E07.2	Documento com a modelagem e especificação de requisitos dos módulos Validador e Transmissor.			
	M3				
E07.4	E-E07.3	Documento com a modelagem e especificação de requisitos do módulo Quarentena.			
	M3				
E07.5	E-E07.5	Documento com a definição das tecnologias e protocolos utilizados para a comunicação e transmissão de dados pela Plataforma.			
	M3				
E07.6	E-E07.4	Documento com a modelagem e especificação dos bancos de dados necessários para o armazenamento dos dados e integração da plataforma.			
	M4				
E07.7	E-E07.6	Implantação da primeira versão dos módulos Validador e Transmissor na infraestrutura do MPMG.			
	M6				

E07.8	E-E07.7	Implantação da primeira versão do módulo Quarentena na infraestrutura do MPMG.
	M6	
E07.9	E-E07.9	Criação e instanciação dos bancos de dados na infraestrutura do MPMG.
	M6	Documento de detalhamento do esquema de dados e modelo entidade-relacionamento.
E07.10	E-E07.11	Implantação da 1ª versão da plataforma na infraestrutura do MPMG.
	M8	Documento detalhado descrevendo o funcionamento e a integração entre os módulos Validador, Transmissor e Quarentena.
E07.11	E-E07.10	Documento detalhado de testes de integridade e segurança.
	M10	
E07.12	E-E07.11	Implantação da 2ª versão da plataforma na infraestrutura do MPMG.
	M12	

Milestones

M-E07.1	Detalhamento das atividades desenvolvidas nas etapas 1, 2, 3, 4 e 5.
M1	
M-E07.2	Detalhamento das atividades desenvolvidas nas etapas 1, 2, 3 e 4.
M2	
M-E07.3	Detalhamento das atividades desenvolvidas na etapa 4.
M3	
M-E07.4	Detalhamento das atividades de integração da etapa E-E07.5 (segurança) com o desenvolvimento dos módulos Validador, Transmissor e Quarentena.
M4	
M-E07.5	Detalhamento das atividades desenvolvidas na etapa E-E07.9.

M4	
M-E07.6	Detalhamento das atividades de integração da etapa E-E07.5 (segurança) com o desenvolvimento dos módulos Validador, Transmissor e Quarentena.
M5	
M-E07.7	Detalhamento das atividades desenvolvidas na etapa E-E07.8.
M7	
M-E07.8	Detalhamento das atividades desenvolvidas na integração das etapas E-E07.5 + E-E07.8 + E-E07.10.
M9	
M-E07.9	Detalhamento das atividades desenvolvidas.
M11	

Riscos Críticos para a Implementação

Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-E07.1	Impossibilidade de formação de equipe técnica qualificada.	Alto	Mitigar	Avaliar junto ao MPMG a alteração no cronograma de entregas para possibilitar a contratação de pessoal.
R-E07.2	Saída de algum dos desenvolvedores das atividades do grupo de trabalho.	Alto	Mitigar	Disparar novo processo seletivo imediato para a contratação de novo desenvolvedor.
R-E07.3	Saída de algum bolsista do grupo de trabalho: graduando, mestrando, doutorando ou pós- doutorando.	Médio	Mitigar	O coordenador deve iniciar processo seletivo imediatamente para a contratação de novo bolsista.
R-E07.4	Falta de especialistas (consultores) em sistemas de movimentações bancárias.	Alto	Travar	Cabe à equipe do GSI/MPMG garantir, por meio de parcerias institucionais, a disponibilidade de servidores com experiência no desenvolvimento e,

				também, na utilização de sistemas de movimentações bancárias, como o SIMBA.
R-E07.5	Dificuldade de contato com especialistas em sistemas de movimentações bancárias.	Médio	Transferir	Solicitar à equipe técnica do GSI/MPMG providências para uma maior disponibilidade dos especialistas em sistemas de movimentações bancárias.
R-E07.6	Complexidade de desenvolvimento "muito" além do identificado pela equipe DCC/UFGM durante a fase de projeto.	Baixo	Mitigar	Acordar com a equipe técnica do GSI/MPMG ajustes no cronograma de entregas e complexidade dos módulos a serem desenvolvidos, primando por atingir os objetivos principais do projeto.
R-E07.7	Ausência de contato e acesso a instituições financeiras parceiras para a realização de testes de carga, confiabilidade e segurança.	Alto	Transferir	É de responsabilidade do GSI/MPMG a impossibilidade de testes reais, com pelo menos uma instituição financeira. Necessário firmar parcerias, se possível. As avaliações locais (na infraestrutura do GSI/MPMG) serão feitas pela equipe.
R-E07.8	Dificuldade de acesso à infraestrutura de rede de dados, certificados, chaves e VMs para o desenvolvimento da solução, testes diversos e implantação do sistema.	Alto	Transferir	Cabe à equipe do GSI/MPMG disponibilizar a infraestrutura necessária para o bom andamento do cronograma de atividades.

Cronograma das Entregas por Etapas

ETAPAS		ENTREGAS												
CÓD.	TÍTULO	M0	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-E07.1	Padrões e de protocolos de dados				E07.2									

E- E07.2	Modelagem: Validador e Transmissor				E07.3									
E- E07.3	Modelagem: Quarentena				E07.4									
E- E07.4	Modelagem e especificação dos bancos de dados necessários para o armazenamento dos dados e integração da plataforma					E07.6								
E- E07.5	Segurança da comunicação entre os módulos			E07.1	E07.5									
E- E07.6	Desenvolvimento: Validador e Transmissor							E07.7						
E- E07.7	Desenvolvimento: Quarentena							E07.8						
E- E07.8	Integração: Validador, Transmissor e Quarentena													
E- E07.9	Implementação dos bancos de dados							E07.9						
E- E07.10	Testes de integridade e segurança											E07.11		

E- E07.11	Plataforma Inteligente de Investigações Bancárias e Vínculos Financeiros									E07.10				E07.12
MILESTONES		M- E07.1	M- E07.2	M- E07.3	M- E07.4 M- E07.5	M- E07.6		M- E07.7		M- E07.8			M- E07.9	

Etapas - 2022					
E-E07.12	Início	M13	Fim	M20	
	Título	Módulo de inteligência: Análise de vínculos financeiros			
	Descrição	Realiza a análise de vínculos por meio de grafos; Disponibiliza os dados em formato para integração com a Plataforma Áduna; Disponibiliza dados para integração com o módulo de visualização.			
E-E07.13	Início	M15	Fim	M24	
	Título	Módulo de visualização			
	Descrição	Disponibiliza metáforas visuais que reflitam os vínculos e movimentações financeiras; Disponibiliza metáforas visuais que quantifiquem os montantes envolvidos nas transações e vínculos; Disponibiliza todas as visualizações em um único Painel de Business Intelligence (BI).			

		Início	M19	Fim	M24
E-E07.14	Título	Módulo de Relatórios de inteligência			
	Descrição	Gera relatórios detalhados que sumarizem os vínculos identificados e quantifiquem os valores financeiros movimentados; Disponibiliza os relatórios de análises.			
Entregáveis - 2022					
E07.11	E-E07.12	Documento de detalhamento dos dados e formato para integração com o módulo de visualização.			
	M17				
E07.12	E-E07.12	Documento de detalhamento dos dados e formato para integração com a Plataforma Áduna.			
	M18				
E07.13	E-E07.12	Implantação do módulo de inteligência e análise de vínculos na infraestrutura do MPMG.			
	M19				
E07.14	E-E07.13	Painel de BI com visualização para análise de vínculo.			
	M20				
E07.15	E-E07.13	Painel de BI com visualização do montante das movimentações financeiras.			
	M21				
E07.16	E-E07.14	1ª versão da plataforma integrada ao módulo de relatórios de inteligência.			
	M22				
E07.17	E-E07.14	Versão final da plataforma.			
	M24				

F. Aplicações Finalísticas

Grupo de Trabalho	F01 - Processamento dos dados coletados dos portais de transparência dos municípios do Estado de Minas Gerais				
Grupos de Interação	C01, G01, P04				
Coordenador	Rodrygo Santos				
Equipe	Nível	Quantidade	Nomes		
	Professor	2	Rodrygo Santos, Renato Vimieiro		
	Pós-Doc	0	TOTAL	8	
	Doutorando	0			
	Mestrando	2			
	Graduando	2			
	Técnico	2	Cientista de dados, Engenheiro de dados		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Não	Não	Não	Não	Não
Descrição Geral					
Desenvolvimento das etapas de preparação dos dados oriundos de portais da transparência dos municípios do Estado de Minas Gerais.					
Objetivos					
Construir etapas de pré processamento, enriquecimento e estruturação dos dados de portais da transparência dos municípios do Estado de Minas Gerais, que foram coletados pelo grupo C01. O objetivo principal do grupo é disponibilizar, ao grupo P04, os dados de forma estruturada para serem consumidos pelo Painel de BI.					
Etapas					
E-F01.1		Início	M1	Fim	M12

	Título	Pré processamento dos dados de portais da transparência dos municípios do Estado de Minas Gerais			
	Risco vinculado	R-F01.1	R-F01.3		
		R-F01.2			
Tecnologias previstas		Python			
Tipos de dados necessários		Dados de portais da transparência coletados pelo grupo C01.			
Descrição	Etapa de limpeza, transformação e estruturação dos dados coletados pelo grupo C01 de portais da transparência dos municípios do Estado de Minas Gerais.				
		Início	M1	Fim	M3
E-F01.2	Título	Caracterização dos dados de portais da transparência dos municípios do Estado de Minas Gerais			
	Risco vinculado	R-F01.1	R-F01.3		
		R-F01.2			
Tecnologias previstas		Python			
Tipos de dados necessários		Dados de portais da transparência processados na etapa E-F01.1.			
Descrição	Será feita a caracterização e detalhamento dos dados resultantes de resultados parciais da etapa E-F01.1 de forma que o grupo de Plataforma Tecnológica (P04) possa entender melhor as características dos dados dos portais da transparência dos municípios do Estado de Minas Gerais.				
		Início	M3	Fim	M12
E-F01.3	Título	Enriquecimento dos dados de portais da transparência dos municípios do Estado de Minas Gerais			

		Risco vinculado	R-F01.1	R-F01.3	
			R-F01.2		
Tecnologias previstas		Python			
Tipos de dados necessários		Dados de portais da transparência processados na etapa E-F01.1.			
Descrição	Enriquecimento dos dados através de interação com o grupo M01 (Extração de Entidades).				
E-F01.4		Início	M2	Fim	M12
	Título	Checagem automatizada do cumprimento dos normativos vigentes para os portais da transparência dos municípios do Estado de Minas Gerais			
		Risco vinculado	R-F01.1	R-F01.3	
			R-F01.2		
Tecnologias previstas		Python			
Tipos de dados necessários		Dados de portais da transparência processados na etapa E-F01.1.			
Descrição	Checagem automatizada do cumprimento dos normativos vigentes para os portais da transparência dos municípios do Estado de Minas Gerais. Nesta etapa será necessária a interação com especialistas de conteúdo do MPMG.				
Entregáveis					
F01.1	E-F01.2	Relatório de caracterização dos dados coletados dos portais de transparência dos municípios do Estado de Minas Gerais.			
	M3				
F01.2	E-F01.4	Relatório descritivo das informações disponibilizadas nos portais da transparência dos municípios até então processados.			
	M4				

F01.3	E-F01.4	Relatório descritivo das informações disponibilizadas nos portais da transparência dos municípios até então processados.
	M6	
F01.4	E-F01.4	Relatório descritivo das informações disponibilizadas nos portais da transparência dos municípios até então processados.
	M8	
F01.5	E-F01.4	Relatório descritivo das informações disponibilizadas nos portais da transparência dos municípios até então processados.
	M10	
F01.6	E-F01.4	Relatório descritivo das informações disponibilizadas nos portais da transparência dos municípios até então processados.
	M12	
Milestones		
M-F01.1	Relatório parcial sobre as atividades desenvolvidas.	M1
M1		
M-F01.2	Relatório parcial sobre as atividades desenvolvidas.	M2
M2		
M-F01.3	Relatório parcial contendo o indicativo de quais municípios serão abordados na entrega prevista para M4.	M3
M3		
M-F01.4	Relatório parcial contendo o indicativo de quais municípios serão abordados na entrega prevista para M6.	M5
M5		
M-F01.5	Relatório parcial contendo o indicativo de quais municípios serão abordados na entrega prevista para M8.	M7
M7		
M-F01.6	Relatório parcial contendo o indicativo de quais municípios serão abordados na entrega prevista para M10.	

M9	Relatório parcial contendo o indicativo de quais municípios serão abordados na entrega prevista para M12.
M-F01.7	
M11	

Riscos Críticos para a Implementação				
Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-F01.1	Indisponibilidade dos dados de portais da transparência dos municípios do Estado de Minas Gerais.	Alto	Travar	Interação com a equipe do grupo C01 e a equipe do MPMG com o objetivo de buscar uma solução para a ausência de dados.
R-F01.2	Ausência de profissionais de ciência de dados e engenharia de dados para o desenvolvimento das etapas do grupo.	Alto	Mitigar	Iniciar processo de contratação imediato.
R-F01.3	Vacância do cargo de doutorando ou mestrando ou graduando.	Médio	Mitigar	Iniciar processo de contratação imediato.

Cronograma das Entregas por Etapas													
ETAPAS		ENTREGAS											
CÓD.	TÍTULO	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-F01.1	Pré processamento dos dados de portais da transparência dos municípios do Estado de Minas Gerais												

E-F01.2	Caracterização dos dados de portais da transparência dos municípios do Estado de Minas Gerais			F01.1								
E-F01.3	Enriquecimento dos dados de portais da transparência dos municípios do Estado de Minas Gerais											
E-F01.4	Checagem automatizada do cumprimento dos normativos vigentes para os portais da transparência dos municípios do Estado de Minas Gerais				F01.2	F01.3	F01.4	F01.5	F01.6			
MILESTONES		M-F01.1	M-F01.2	M-F01.3		M-F01.4		M-F01.5		M-F01.6		M-F01.7

Etapas - 2022

		Início	M13	Fim	M24	
E-F01.5	Título	Pré processamento dos dados de portais da transparência dos municípios do Estado de Minas Gerais				
	Descrição	Etapa de limpeza, transformação e estruturação dos dados coletados pelo grupo C01 de portais da transparência dos municípios do Estado de Minas Gerais.				

		Início	M14	Fim	M24
E-F01.6	Título	Enriquecimento dos dados de portais da transparência dos municípios do Estado de Minas Gerais			
	Descrição				
		Início	M14	Fim	M24
E-F01.7	Título	Checagem automatizada do cumprimento dos normativos vigentes para os portais da transparência dos municípios do Estado de Minas Gerais			
	Descrição	Checagem automatizada do cumprimento dos normativos vigentes para os portais da transparência dos municípios do Estado de Minas Gerais. Nesta etapa será necessária a interação com especialistas de conteúdo do MPMG.			
Entregáveis - 2022					
F01.1	E-F01.7	Relatório de caracterização dos dados coletados dos portais de transparência dos municípios do Estado de Minas Gerais.			
	M15				
F01.2	E-F01.7	Relatório descritivo das informações disponibilizadas nos portais da transparência dos municípios até então processados.			
	M16				
F01.3	E-F01.7	Relatório descritivo das informações disponibilizadas nos portais da transparência dos municípios até então processados.			
	M18				
F01.4	E-F01.7	Relatório descritivo das informações disponibilizadas nos portais da transparência dos municípios até então processados.			
	M20				
F01.5	E-F01.7	Relatório descritivo das informações disponibilizadas nos portais da transparência dos municípios até então processados.			

	M22	
F01.6	E-F01.7	Relatório descritivo das informações disponibilizadas nos portais da transparência dos municípios até então processados.
	M24	

Grupo de Trabalho	F02 - Análise de Fraudes em Licitações				
Grupos de Interação	C01, D01, D02, G01, M01, M03, M04, P02				
Coordenador	Gisele Lobo Pappa				
Equipe	Nível	Quantidade	Nomes		
	Professor	3	Gisele Lobo Pappa, Anísio Mendes Lacerda, Wagner Meira Jr.		
	Pós-Doc	0	TOTAL	11	
	Doutorando	1			
	Mestrando	3			
	Graduando	3			
	Técnico	1	Desenvolvedor Web para alterações necessárias no Painel de BI		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Sim	Não	Sim	Sim	Não
Descrição Geral					
Este grupo de trabalho está relacionado a construção de trilhas de auditoria para detecção de indícios de fraudes em licitações. Além da construção das próprias trilhas, ele também engloba mecanismos de agregação dos resultados dessas trilhas e o construção de meta-trilhas. Estas últimas serão geradas seguindo padrões encontrados no processo de criação manual de trilhas no LEMONADE.					

Objetivos

São contemplados três objetivos principais:

1. Construção de até 6 trilhas de auditagem (TA1, TA2, TA3, TA4, TA5, TA6), considerando metade das trilhas no ano de 2021 e a outra metade em 2022, utilizando técnicas de aprendizado de máquina. Dado o conhecimento que adquirimos durante o primeiro ano de projeto, sugerimos alguns exemplos de trilhas, que podem ser substituídas por quaisquer outras demandas de trilha com as mesmas características e que sejam de interesse do MP. São elas:

1. Trilha para a identificação de perdedores frequentes e os vencedores associados;
2. Identificação de Fracionamento/ Parcelamento de compras para fraudar os limites de contratação;
3. Detecção de rodízio entre licitantes;
4. Direcionamento de licitação para ME ou EPP onde o vencedor faz parte de um grupo econômico maior;
5. Trilha para compra de itens em quantidades absurdas (relacionado ao F03);
6. Identificação de itens estranhos (fora do padrão) ao loteamento da licitação (relacionado ao F03).

2. Uso das saídas das trilhas implementadas como entrada para um modelo de classificação, que retornaria a probabilidade de uma licitação ser fraudulenta baseado nesses conjuntos de alertas. Essa tarefa estaria condicionada a viabilidade de termos um conjunto de licitações rotuladas como fraude.

3. Construção de meta-trilhas a partir de padrões encontrados na estrutura das trilhas já desenvolvidas. Essa tarefa depende do uso do sistema manual de construção de trilhas para obtermos dados necessários para a construção automática das meta-trilhas.

Ainda durante o ano de 2021 deverão ser levantadas as trilhas de auditagem a serem implementadas em 2022, bem como a necessidade de estender as tarefas dos objetivos 2 e 3 para o referido ano.

Etapas

		Início	M1	Fim	M12
E-F02.1	Título	Análises de viabilidade			
	Risco vinculado		R-F02.1	R-F02.4	R-F02.6

Tecnologias previstas		--			
Tipos de dados necessários		Acesso aos bancos de dados do MPMG e metadados correspondentes Acesso a especialistas de conteúdo			
Descrição	Estudo de viabilidade e definição das 3 trilhas (TA1, TA2 e TA3) que serão implementadas no projeto e do modelo de predição de fraude.				
E-F02.2		Início	M3	Fim	M12
	Título	Desenvolvimento de trilhas			
	Risco vinculado	R-F02.1	R-F02.2	R-F02.3	
		R-F02.4	R-F02.5		
Tecnologias previstas		SQL, Hive, python, pyspark			
Tipos de dados necessários		Bancos de dados de despesas públicas Metadados correspondentes			
Descrição	Projeto, implementação e análise de resultados das 3 trilhas selecionadas.				
E-F02.3		Início	M3	Fim	M12
	Título	Implementação de meta-trilhas			
	Risco vinculado	R-F02.2	R-F02.3	R-F02.4	
		R-F02.6	R-F02.7		
Tecnologias previstas		SQL, Hive, python, pyspark, Lemonade			
Tipos de dados necessários		Bancos de dados de despesas públicas			

Metadados correspondentes				
Descrição	Projeto, implementação e análise de resultados de meta-trilhas parametrizáveis utilizando o Lemonade.			
E-F02.4	Início	M3	Fim	M10
	Título	Implementação de modelo de predição de fraudes em licitações a partir de alertas gerados pelas trilhas		
	Risco vinculado	R-F02.2	R-F02.3	R-F02.4
		R-F02.6	R-F02.7	
Tecnologias previstas	SQL, Hive, python, pyspark			
Tipos de dados necessários	Bancos de dados de despesas públicas Metadados correspondentes			
Descrição	Projeto, implementação e análise de um modelo de predição de fraudes através dos alertas gerados pelas trilhas já implementadas. Os métodos de aprendizado a serem utilizados para o desenvolvimento dessa tarefa sofrerão forte influência da disponibilidade de dados rotulados.			
Entregáveis				
F02.1	E-F02.1	Relatório com a descrição detalhada dos objetivos das trilhas que serão implementadas ao longo do projeto, seus riscos e dependências.		
	M2			
F02.2	E-F02.2	Relatório contendo a descrição da trilha de alertas de fraude de acordo com a TA1, a ser definida na Etapa E-F02.1		
	M7			
F02.3	E-F02.2	Relatório contendo a descrição da trilha de alertas de fraude de acordo com a TA2, a ser definida na Etapa E-F02.1		
	M10			
F02.4	E-F02.4	Modelo de classificação capaz de retornar a probabilidade de uma licitação ser fraudulenta com base nos alertas gerados.		

	M10	
F02.5	E-F02.2	Relatório contendo a descrição da trilha de alertas de fraude de acordo com a TA3, a ser definida na Etapa E-F02.1
	M12	
F02.6	E-F02.3	Conjunto de meta-trilhas que podem ser pré-carregadas no LEMONADE e parametrizadas pelo usuário.
	M12	
Milestones		
M-F02.1	Status do processo de definição das trilhas.	
M1		
M-F02.2	Status do levantamento dos dados e modelagem da TA1	
M3	Status da construção de um repositório de dados de trilhas conhecidas como fraudulentas.	
M-F02.3	Status do levantamento dos dados e modelagem da TA1.	
M4	Status da construção de um repositório de dados de trilhas conhecidas como fraudulentas.	
	Status do levantamento das trilhas já criadas por especialistas no LEMONADE.	
M-F02.4	Status da implementação da TA1.	
M5	Status da modelagem do problema de classificação de trilhas fraudulentas a partir dos alertas das trilhas.	
	Status da caracterização das trilhas já criadas por especialistas no LEMONADE.	
M-F02.5	Status da implementação e testes da TA1.	

	Status do levantamento dos dados e modelagem da TA2.
M6	Status da modelagem e implementação do problema de classificação de trilhas fraudulentas a partir dos alertas das trilhas. Primeira proposta de meta-trilhas.
M-F02.6	Status da modelagem e implementação da TA2. Status do levantamento dos dados e modelagem da TA3.
M8	Status da modelagem, implementação e testes do problema de classificação de trilhas fraudulentas a partir dos alertas das trilhas. Status da implementação das primeiras meta-trilhas fora do ambiente LEMONADE.
M-F02.7	Levantamento inicial das trilhas, meta trilhas e refinamentos para o modelo de predição de fraudes a serem desenvolvidos em 2022. Status da implementação e testes da TA2. Status da implementação da TA3.
M9	Status da modelagem, implementação e testes do problema de classificação de trilhas fraudulentas a partir dos alertas das trilhas. Status da primeira implementação das meta-trilhas na ferramenta LEMONADE.
M-F02.8	Status das negociações junto ao MPMG sobre as trilhas, meta trilhas e refinamentos para o modelo de predição de fraudes a serem desenvolvidos em 2022.
M11	Status da implementação e testes da TA3. Status das meta-trilhas finais a serem integradas ao LEMONADE.

Riscos Críticos para a Implementação

Código	Descrição	Impacto	Tratamento	Plano de Contingência
--------	-----------	---------	------------	-----------------------

R-F02.1	Atraso na definição das trilhas a serem desenvolvidas ou ausência de especialistas de conteúdo capazes de auxiliar no detalhamento das mesmas.	Médio	Travar	Subir a negociação de trilhas para as instâncias imediatamente superiores. Replanejar o cronograma de atividades Realizar estudos próprios para a definição da trilha (pode gerar prejuízos ao entendimento do problema).
R-F02.2	Alteração na estrutura do banco de dados de licitações (EDW).	Alto	Travar	Interação com as equipes do MP para entendimento do planejamento das alterações ou alterações do sistema de banco de dados do MP após o projeto de 2020. Só será possível iniciar qualquer implementação desse grupo de trabalho após este alinhamento.
R-F02.3	Qualidade dos dados nos bancos de dados.	Alto	Travar	Correção dos problemas por parte do MP.
R-F02.4	Saída de membros da equipe responsável.	Alto	Mitigar	Contratação de substitutos e/ou

				reatribuição de responsabilidades dentro da equipe.
R-F02.5	Mudança das tecnologias sendo utilizadas no desenvolvimento do projeto por parte do MP.	Alto	Mitigar	Readequar o projeto e treinar o pessoal nas tecnologias não dominadas. Pode atrasar o cronograma.
R-F02.6	Necessidade de refatoração de trilhas (incluindo as trilhas implementadas em 2020) por alteração na estrutura do banco de dados.	Médio	Mitigar	Analisar impacto das alterações, definir prioridades e ajustar o cronograma de entregas.
R-F02.7	Ausência de dados rotulados para a trilha de predição de fraudes.	Médio	Transferir	Definição das alterações necessárias para o desenvolvimento do modelo. Podem incluir alteração para tarefas não supervisionadas, construção de rótulos a partir de variáveis pré-existentes ou rotulação manual.

Cronograma das Entregas por Etapas

Etapas		Entregas											
Cód.	Título	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12

E-F02.1	Análises de viabilidade		F02.1										
E-F02.2	Desenvolvimento de trilhas							F02.2			F02.3		F02.5
E-F02.3	Implementação de meta-trilhas												F02.6
E-F02.4	Implementação de modelo de predição de fraudes em licitações a partir de alertas gerados pelas trilhas									F02.4			
MILESTONES		M-F02.1		M-F02.2	M-F02.3	M-F02.4	M-F02.5		M-F02.6	M-F02.7		M-F02.8	

Etapas - 2022

E-F02.5	Início	M13	Fim	M15
	Título	Análises de viabilidade das tarefas definidas para o ano de 2022		
	Descrição	Estudos de viabilidade para implementação das trilhas de auditoria e meta-trilhas, definidas para o ano de 2022, e dos refinamentos propostos para o modelo de predição de fraudes.		
E-F02.6	Início	M17	Fim	M24
	Título	Implementação das trilhas de auditoria definidas para o ano de 2022		
	Descrição	Esta etapa diz respeito à modelagem, implementação, documentação e integração das trilhas TA4, TA5 e TA6.		
E-F02.7	Início	M17	Fim	M24
	Título	Implementação das meta trilhas definidas para o ano de 2022		

	Descrição	Esta etapa diz respeito à modelagem, implementação, documentação e integração das meta trilhas definidas para o ano de 2022.		
Entregáveis - 2022				
E-F02.8	Início	M17	Fim	M20
	Título	Implementação dos refinamentos propostos para o modelo de predição de fraudes em licitação		
	Descrição	Esta etapa diz respeito à modelagem, implementação e documentação dos refinamentos propostos para o modelo de predição de fraudes em licitação.		
F02.7	E-F02.5	Relatório de análise de viabilidade para a implementação das trilhas de auditoria, meta trilhas e refinamentos propostos para o modelo de predição de fraudes em licitação.		
	M15			
F02.8	E-F02.6	Relatório contendo códigos e descrições das trilhas segundo o cronograma acordado em E-F02.11.		
	M24			
F02.9	E-F02.7	Relatório contendo códigos e descrições das meta trilhas segundo o cronograma acordado em E-F02.11.		
	M24			
F02.10	E-F02.8	Relatório contendo códigos e descrições dos refinamentos propostos para o modelo de predição de fraudes segundo o cronograma acordado em E-F02.11.		
	M20			

Grupo de Trabalho	F03 - Análise de Fraudes em Compras Públicas
Grupos de Interação	C01, D02, F01, G01, M01, M03, P02
Coordenador	Anísio Mendes Lacerda

Equipe	Nível	Quantidade	Nomes		
	Professor	3	Anísio Mendes Lacerda, Gisele L. Pappa, Wagner Meira Jr.		
	Pós-Doc	0	TOTAL	7	
	Doutorando	1			
	Mestrando	1			
	Graduando	1			
	Técnico	1	Desenvolvedor Web		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Sim	Não	Não	Sim	Não
Descrição Geral					
A tarefa F03 é uma das mais desafiadoras tecnicamente, dentre as tarefas que compreendem as aplicações finalísticas no projeto de 2020. Até o fim do projeto em vigência, entregaremos uma versão funcional de um agrupamento de objetos de licitação semelhantes junto a uma estimativa de preço para cada item. Porém, não teremos tempo para nenhuma avaliação manual neste ano. A ideia da continuação deste grupo de trabalho é refinar as estratégias desenvolvidas até o momento, após uma avaliação manual dos agrupamentos e precificação dos itens, seguida da organização dos resultados em um banco de preços, disponível online.					
Objetivos					
<ol style="list-style-type: none"> 1. Avaliar qualitativamente os resultados obtidos pelo projeto do ano de 2020 e realizar refinamentos quando necessários. 2. Propor e implementar refinamentos para as técnicas já desenvolvidas de acordo com o feedback obtido em 1. 3. Criar um "Banco de Preços", onde pode-se consultar o valor praticado atualmente para o objeto sendo licitado. 					
Etapas					
E-F03.1		Início	M1	Fim	M4
	Título	Análise qualitativa dos resultados obtidos no projeto de 2020.			
		Risco vinculado	R-F03.2	R-F03.3	R-F03.4

Tecnologias previstas		--			
Tipos de dados necessários		Grupos de itens e valores de preços dos itens determinados em outras fases da tarefa.			
Descrição		A análise qualitativa nessa etapa deverá ser realizada tanto pela equipe do F03 quanto por especialistas de conteúdo do MPMG. O objetivo dessa tarefa é, a partir de uma amostra de grupos gerados a partir dos algoritmos desenvolvidos em 2020, avaliar o quão coesos esses grupos são. Ou seja, avaliar se os itens dos grupos apresentados dizem respeito ao mesmo objeto ou não. Por parte dos integrantes do F03 essa análise será realizada através de caracterizações das bases. Já as equipes do MPMG receberão uma amostra de itens e deverão responder questionários a respeito da percepção dos mesmos em relação à qualidade dos grupos.			
E-F03.2	Início	M1		Fim	M10
	Título	Refinamentos da técnica proposta no projeto de 2020.			
	Risco vinculado	R-F03.1	R-F03.2	R-F03.3	
		R-F03.4			
Tecnologias previstas		python, pyspark, Hive, SQL, biblioteca gerais de ML			
Tipos de dados necessários		Análise qualitativa do sistema entregue em 2020, dados estruturados de itens			
Descrição		Os refinamentos propostos nesta etapa dizem respeito tanto a sugestões referentes às análises qualitativas realizadas pelos especialistas de conteúdo quanto a testes com diferentes formas de representar as descrições dos itens. Por exemplo, teste com diferentes algoritmos para geração de embeddings de palavras, alterações no método de estruturação das sentenças proposto na etapa corrente do projeto, etc. Esses refinamentos podem incluir a adição de fontes de dados não utilizadas até então, desde que as mesmas se encontrem estruturadas e disponíveis nos bancos de dados do MPMG. Além disso, a inclusão das mesmas também depende de análise de viabilidade.			
E-F03.3	Início	M1		Fim	M12
	Título	Banco de preços em versão Web			

		Risco vinculado	R-F03.3	R-F03.4	R-F03.5
Tecnologias previstas		React, Node.Js, Hive, Apache Druid, python, pyspark, SQL			
Tipos de dados necessários		Grupos de itens e valores de preços dos itens determinados em outras fases da tarefa.			
Descrição	Projeto, e implementação de um portal WEB que funcione como um banco de preços de preços de itens. Através deste portal o usuário pode inserir a descrição de um item específico e obter como resposta o conjunto de itens que mais se aproxima da descrição inserida bem o estatísticas de preço desses itens. A construção do grupo de itens e das estatísticas de preço são de responsabilidade da Etapa 2.				
Entregáveis					
F03.1	E-F03.1	Relatório com análise qualitativa do sistema a ser finalizado no projeto corrente e possíveis refinamentos.			
	M2				
F03.2	E-F03.1	Amostra de grupos de itens para avaliação qualitativa por especialistas de conteúdo. Idealmente o tempo de avaliação pelos especialistas deve ser de 1 a 2 meses (implica entrega até o final de M6).			
	M4				
F03.3	E-F03.3	Especificação de requisitos + Modelagem conceitual do Banco de Dados do Banco de Preços e de API do para consumo externo dos dados do banco de preço.			
	M5				
F03.4	E-F03.2	Relatório contendo os resultados da análise qualitativa, a descrição dos refinamentos incorporados e análise comparativa com o modelo antes do refinamento.			
	M10				
F03.5	E-F03.3	Banco de preços em versão Web.			
	M12	API para consumo do banco de preços.			
Milestones					
M-F03.1	Levantamento do perfil dos analistas de conteúdo para análise qualitativa das descrições dos objetos pertencentes a diferentes grupos.				
M1					

M-F03.2	1. Levantamento das análises que serão encaminhadas para os especialistas de conteúdo.
M3	2. Checkpoint parcial dos requisitos do Banco de Preços.
M-F03.3	1. Status da implementação do banco de dados do Banco de Preços.
M6	2. Descrição das estratégias de reagrupamento (pós processamento em geral) implementadas como refinamento da técnica de agrupamento atual.
M-F03.4	1. Descrição da estrutura do BD implementado.
M7	2. Status dos desenhos das telas e da implementação da primeira iteração de desenvolvimento do back-end do Banco de Preços. 3. Propostas de alteração no método de agrupamento de itens atual de acordo com a análise qualitativa.
M-F03.5	1. Status de implementação da primeira iteração de desenvolvimento do front-end e da segunda iteração de desenvolvimento do back-end do Banco de Preços.
M8	2. Status das propostas de alteração no método de agrupamento de itens atual e resultados quantitativos preliminares.
M-F03.6	1. Status de implementação da segunda iteração de desenvolvimento do front-end e da terceira iteração de desenvolvimento do back-end do Banco de Preços.
M9	2. Status das propostas de alteração do método de agrupamento de itens e resultados quantitativos preliminares.
M-F03.7	1. Incorporação das alterações realizadas no método de agrupamento de itens ao banco de dados.
M11	2. Última iteração de desenvolvimento do front-end e back-end do Banco de Preços.

Riscos Críticos para a Implementação

Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-F03.1	Atraso da realização da análise qualitativa pelos especialistas de conteúdo.	Médio	Mitigar	Realizar parte das análises internamente. Trará impactos no tempo de entrega.
R-F03.2	Alteração na estrutura do banco de dados de licitações (EDW).	Alto	Travar	Interação com as equipes do MP para entendimento do planejamento das alterações ou alterações do sistema de

				banco de dados do MP após o projeto de 2020. Só será possível iniciar qualquer implementação desse grupo de trabalho após este alinhamento.
R-F03.3	Qualidade dos dados nos bancos de dados de licitações.	Alto	Travar	Correção dos problemas por parte do MP. Exemplos de problemas críticos encontrados atualmente: os itens que não foram licitados por lote não estão disponíveis nas tabelas do banco de dados do LINS às quais temos acesso.
R-F03.4	Saída de membros da equipe responsável.	Alto	Mitigar	Contratação de substitutos e/ou reatribuição de responsabilidades dentro da equipe.
R-F03.5	Mudança das tecnologias sendo utilizadas no desenvolvimento do projeto por parte do MP.	Alto	Mitigar	Pode atrasar o cronograma.

Cronograma das Entregas por Etapas

Etapas		Entregas											
Cód.	Título	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-F03.1	Análise qualitativa dos resultados obtidos no projeto de 2020.		F03.1		F03.2								
E-F03.2	Refinamentos da técnica proposta no projeto de 2020.										F03.4		

E-F03.3	Banco de preços em versão Web.					F03.3							F03.5
MILESTONES		M-F03.1		M-F03.2			M-F03.3	M-F03.4	M-F03.5	M-F03.6			M-F03.7

Etapas - 2022					
E-F03.4		Início	M13	Fim	M16
	Título	Levantamento de gastos públicos a serem monitorados em 2022			
	Descrição	Levantamento de gastos específicos para monitoramento temporal. Alguns exemplos são: Gastos em saneamento básico, gastos em diárias, gastos em produtos eletrônicos, gastos em combustível, etc.			
E-F03.5		Início	M16	Fim	M24
	Título	Modelagem para detecção de gastos públicos fora do padrão			
	Descrição	Através do agrupamento dos municípios pelas suas características demográficas é possível realizar a comparação dos gastos levantados nas etapas anteriores e posterior disparo de alertas para gastos fora do padrão.			
Entregáveis - 2022					
F03.6	E-F03.11	Relatório contendo o levantamento de requisitos e cronograma de desenvolvimento para o monitoramento temporal de gastos públicos para áreas específicas.			
	M16				
F03.7	E-F03.12	Relatório contendo descrição da modelagem proposta e primeiros resultados dos monitoramentos de acordo com o cronograma definido em E-F03.11.			
	M19				
F03.8	E-F03.13	Relatório e códigos dos modelos de monitoramento propostos em E-F03.11, bem como a sua inclusão em seção específica do Banco de Preços WEB.			
	M24				

Grupo de Trabalho	F04 - Identificação e caracterização de padrões associados a fraudes eleitorais em redes sociais online, WhatsApp e Telegram				
Grupos de Interação	C01, E02, E04, E05, F01, G01				
Coordenador	Jussara M Almeida				
Equipe	Nível	Quantidade	Nomes		
	Professor	4	Jussara M. Almeida, Ana Paula Couto, Virgílio Almeida, Fabrício Benevenuto		
	Pós-Doc	1	TOTAL	10	
	Doutorando	1			
	Mestrando	2			
	Graduando	2			
	Técnico	0	--		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Não	Não	Sim	Não	Não
Descrição Geral					
Este projeto visa desenvolver métodos para identificação e caracterização de padrões associados a fraudes eleitorais em dados de redes sociais, WhatsApp e Telegram. Em especial, este projeto tem como foco um tipo de padrão de fraude eleitoral específico que é a disseminação de propagandas políticas fora de época em redes sociais online, WhatsApp e Telegram. Busca-se desenvolver métodos para identificação com base no uso de palavras-chave pré-selecionadas bem como viabilizar análises de padrões associados a este tipo de conteúdo em um conjunto de dados públicos pré-coletados.					
Objetivos					
O principal objetivo a ser alcançado é o desenvolvimento de um método para identificar a presença de propaganda política fora de época em dados públicos previamente coletados de redes sociais, WhatsApp e Telegram bem como demonstrar a aplicabilidade do mesmo na					

análise de padrões associados em uma amostra de dados públicos previamente coletada. Para se atingir este objetivo, o projeto pode ser desdobrado em dois grandes passos principais.

1) Desenvolvimento de ferramenta para identificação de propaganda política fora de época pelo uso de palavras chaves pré-definidas em conteúdo textual (i.e., mensagens textuais, textos extraídos de imagens e áudios transcritos) advindos de dados públicos previamente coletados de redes sociais (a saber Twitter, YouTube, Instagram e possivelmente Facebook), WhatsApp e Telegram, utilizando os coletores desenvolvidos em 2020. Nesta etapa buscaremos desenvolver uma estratégia que dada uma mensagem textual (seja um texto original, áudio transcrito ou elementos textuais extraídos de uma imagem) atribua uma pontuação/score (p.ex: um número entre 0 e 1) que esteja relacionada com a chance do conteúdo conter propaganda política fora de época. A estratégia a ser desenvolvida explorará a presença de palavras chave pré-definidas no texto de entrada.

2) Realização de um estudo de pequeno porte para analisar a presença de propaganda política fora de época em uma amostra de dados previamente coletados de até duas aplicações pré-selecionadas (dentre aquelas para as quais coletores foram desenvolvidos em 2020), visando demonstrar a aplicabilidade do método desenvolvido na etapa 1). Este estudo deve contemplar um conjunto de análises previamente acordado junto ao MPMG considerando os seguintes aspectos. Para redes sociais (a saber, Twitter, YouTube, Instagram, ou, se coletor estiver disponível, Facebook), o estudo deve contemplar análises de grão fino visando identificar padrões relacionados a conteúdo específicos bem como usuários a eles associados. Para o WhatsApp ou Telegram, o estudo deve contemplar apenas uma caracterização agregada restrita ao volume de propagandas identificadas por grupo monitorado.

Etapas

	Início	M1	Fim	M4
E-F04.1	Título	Versão inicial da ferramenta de identificação de propaganda política fora de época em mensagens textuais por uso de palavras-chave		
	Risco vinculado	R-F04.1	R-F04.3	
R-F04.2				
Tecnologias previstas	Python (Sklearn)			
Tipos de dados necessários	Dados de redes sociais, WhatsApp, Telegram previamente coletados utilizando ferramentas desenvolvidas em 2020			
Descrição	Desenvolvimento de uma primeira versão de ferramenta para identificar a presença de propaganda política fora de época em conteúdo textual (sejam mensagens textuais, áudios transcritos ou atributos textuais extraídos de imagens) previamente coletado. Esta ferramenta deverá explorar o uso de palavras-chave extraídas de um vocabulário controlado,			

<p>previamente definido. Dado um texto de entrada e, potencialmente, atributos relacionados (p.ex: instante em que mensagem foi compartilhada), a ferramenta deverá retornar uma pontuação/score que está relacionado à probabilidade de que o texto de entrada contenha propaganda política fora de época. Realização de experimentos de testes para avaliar a eficácia da abordagem desenvolvida.</p>					
E-F04.2		Início	M5	Fim	M8
	Título	Consolidação da ferramenta de identificação de propaganda política fora de época em mensagens textuais por uso de palavras-chave			
	Risco vinculado	R-F04.1	R-F04.3		
		R-F04.2			
Tecnologias previstas	Python (Sklearn)				
Tipos de dados necessários	Dados de redes sociais, WhatsApp, Telegram previamente coletados utilizando ferramentas desenvolvidas em 2020				
Descrição	Considerando os resultados obtidos com a primeira versão da ferramenta, estratégias alternativas que possam levar a uma eficácia maior da ferramenta serão investigadas. Por exemplo, a expansão do vocabulário controlado e o uso de algoritmos de aprendizado de máquina supervisionado alternativos serão considerados. Busca-se nesta etapa consolidar a ferramenta para que a mesma possa ser usada como entrada para as ferramentas de análise que serão desenvolvidas na próxima etapa.				
E-F04.3		Início	M9	Fim	M12
	Título	Estudo sobre a presença de propagandas políticas fora de época em dados de redes sociais e grupos de WhatsApp e Telegram			
	Risco vinculado	R-F04.1			
		R-F04.3			

Tecnologias previstas		Python (Sklearn)
Tipos de dados necessários		Dados de redes sociais, WhatsApp, Telegram previamente coletados utilizando ferramentas desenvolvidas em 2020
Descrição	<p>Será realizado um estudo de pequeno porte sobre a presença de propaganda política fora de época, identificada utilizando o método desenvolvido, em uma amostra de dados previamente coletados de redes sociais (Twitter, YouTube, Instagram e, caso coletor disponível, Facebook) e/ou de grupos públicos do WhatsApp e Telegram, utilizando os coletores desenvolvidos em 2020. Considerando as características muito distintas destas aplicações, no que tange os padrões de interação do usuário, este estudo poderá contemplar até duas aplicações de interesse, selecionadas dentre aquelas para as quais os coletores estão disponíveis. Este estudo visa demonstrar a aplicabilidade do método desenvolvido mais amplamente e poderá ser repetido em outras aplicações posteriormente pela equipe do MPMG. Para realização deste estudo, será necessário inicialmente um alinhamento entre as equipes da UFMG e do MPMG sobre quais análises deverão ser feitas e, conseqüentemente, os aspectos a serem caracterizados para cada aplicação. Como exemplo, um análise possível corresponde a: qual o número de mensagens com propaganda política fora de época compartilhadas em um determinado intervalo de tempo dado um conjunto de dados coletados do Twitter? As seguintes restrições deverão ser consideradas na especificação das análises a serem realizadas. Para dados de redes sociais, poderão ser feitas análises focadas tanto em conteúdos específicos (p.ex: quantas mensagens com propaganda fora de época foram compartilhadas em determinado período?) quanto em usuários (p.ex: verificar a presença de propaganda fora de época dentre os conteúdos postados por um usuário específico), restrita aos dados públicos previamente coletados. Para dados de WhatsApp e Telegram, serão consideradas apenas análises agregadas que permitam quantificar o volume de propagandas fora de época disseminadas em grupos específicos. Análises por conteúdo específico e/ou por usuário não fazem parte do escopo do projeto.</p>	
Entregáveis		
F04.1	E-F04.1	Código de uma versão preliminar da ferramenta de identificação de propaganda política fora de época em conteúdo textual e relatório descrevendo o seu funcionamento, formatos de entrada e saída e resultados de avaliação preliminar.
	M4	
F04.2	E-F04.2	Código da versão consolidada da ferramenta de identificação de propaganda política fora de época em conteúdo textual e relatório descrevendo o seu funcionamento bem como formatos de entrada e saída e resultados de avaliação preliminar.
	M8	
F04.3	E-F04.3	Relatório descrevendo o estudo de caracterização da presença de propaganda política fora de época em uma amostra de dados. Código script em python que implementa as análises realizadas juntamente com descrição de seu funcionamento.
	M12	
Milestones		
M-F04.1	Definição da equipe.	

M1	
M-F04.2	Estudo de métodos de identificação de propaganda política disponíveis na literatura. Definição de vocabulário controlado para detecção de propaganda política fora de época.
M2	
M-F04.3	Implementação de uma versão preliminar do método de identificação de propaganda política fora de época em redes sociais, WhatsApp e Telegram
M3	
M-F04.4	Priorização de refinamentos a serem investigados considerando os resultados obtidos na avaliação preliminar.
M5	
M-F04.5	Resultados da implementação de refinamento.
M6	
M-F04.6	Resultados da implementação de refinamento.
M7	
M-F04.7	Definição das análises que irão compor o estudo de caracterização da presença de propaganda política fora de época bem como das duas aplicações que serão alvo do estudo
M9	
M-F04.8	Resultados da caracterização para a primeira aplicação.
M10	
M-F04.9	Resultados da caracterização para a segunda aplicação.
M11	

Riscos Críticos para a Implementação

Código	Descrição	Impacto	Tratamento	Plano de Contingência
--------	-----------	---------	------------	-----------------------

R-F04.1	Mudança de equipe.	Médio	Aceitar	Realizar novo processo para contratação de recurso.
R-F04.2	A estratégia de identificação de propagandas fora de época pode não atingir as expectativas em termos de qualidade.	Baixo	Mitigar	Buscar estender vocabulário controlado e realizar refinamentos na estratégia. Investigar também métodos alternativos na literatura.
R-F04.3	Baixa disponibilidade e/ou qualidade de dados para testes da ferramenta.	Baixo	Mitigar	A disponibilidade de dados de redes sociais, WhatsApp e Telegram, previamente coletados, é essencial para podermos avaliar o método de identificação de propaganda política e demonstrar sua aplicabilidade. Especificamente, os resultados do método desenvolvido bem como das análises realizadas são fortemente dependentes dos dados disponíveis. Caso os dados disponibilizados não tenham qualidade e/ou quantidade suficientes, será necessário realizar coleta adicional ou buscar coleções de dados alternativas que permitam demonstrar a aplicabilidade do método desenvolvido.

Cronograma das Entregas por Etapas

Etapas		Entregas											
Cód.	Título	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-F04.1	Versão inicial da ferramenta de identificação de propaganda política fora de época em mensagens textuais por uso de palavras-chave				F04.1								
E-F04.2	Consolidação da ferramenta de identificação de								F04.2				

	propaganda política fora de época em mensagens textuais por uso de palavras-chave											
E-F04.3	Estudo sobre a presença de propagandas políticas fora de época em dados de redes sociais e grupos de WhatsApp e Telegram											F04.3
MILESTONES		M-F04.1	M-F04.2	M-F04.3		M-F04.4	M-F04.5	M-F04.6		M-F04.7	M-F04.8	M-F04.9

Etapas - 2022					
		Início	M13	Fim	M16
E-F04.4	Título	Detecção de disparos em massa de propagandas políticas em uma aplicação de rede social, WhatsApp ou Telegram			
	Descrição	Dando continuidade ao trabalho desenvolvido nas etapas anteriores, pretende-se explorar um outro tipo de padrão associado a propaganda política que é o disparo em massa de mensagens com propaganda política Nesta etapa busca-se desenvolver um método para detecção deste padrão em dados de uma aplicação alvo, selecionada entre redes sociais (Twitter, YouTube, Facebook e Instagram), WhatsApp e Telegram. Como estes disparos podem ser executados via ações coordenadas, em múltiplas contas de usuários, durante esta etapa, buscaremos desenvolver um método que identifique padrões associados a atividades coordenadas.			
		Início	M17	Fim	M18
E-F04.5	Título	Refinamentos no método de detecção de disparos em massa			
	Descrição	A partir da realização de experimentos iniciais com o método desenvolvido, buscaremos refinamentos que levem a melhorias na detecção.			

		Início	M19	Fim	M24
E-F04.6	Título	Detecção de disparos em massa entre aplicações			
	Descrição	Nesta etapa pretende-se estender o método desenvolvido, até então focado no fluxo de mensagens em uma única aplicação, para detectar disparos em massa entre plataformas de redes sociais, por exemplo entre o Instagram e o WhatsApp. Serão selecionadas 2 aplicações alvo e o método desenvolvido será estendido para contemplar as mensagens enviadas nas duas aplicações bem como potenciais diferenças de abordagens adotadas nas plataformas.			
Entregáveis - 2022					
F04.6	E-F04.11	Código da primeira versão do método de detecção de disparos em massa e relatório descrevendo o seu funcionamento, formatos de entrada e saída e resultados de avaliação preliminar.			
	M16				
F04.7	E-F04.12	Código da versão refinada do método de detecção de disparos em massa e relatório descrevendo o seu funcionamento, formatos de entrada e saída e resultados de avaliação.			
	M18				
F04.8	E-F04.13	Código do método de detecção de disparos em massa entre aplicações e relatório descrevendo o seu funcionamento, formatos de entrada e saída e resultados de avaliação.			
	M24				

Grupo de Trabalho	F05 - Acompanhamento da execução de obras públicas			
Grupos de Interação	C01, E05, M02			
Coordenador	Jefersson Alex dos Santos			
Equipe	Nível	Quantidade	Nomes	
	Professor	2	Jefersson Alex dos Santos, George Luiz Medeiros Teodoro	
	Pós-Doc	0	TOTAL	6

		Doutorando	1		
		Mestrando	1		
		Graduando	1		
		Técnico	1	Desenvolvedor Mobile	
Plataforma vínculo	de	Lins	Áduna	Lemonade	Trilhas
		Não	Não	Não	Não
Descrição Geral					
Ferramentas para o acompanhamento da execução de obras públicas no Estado.					
Objetivos					
Desenvolver e consolidar abordagens computacionais para monitoramento de obras públicas nas áreas de educação, saúde, segurança pública e/ou infraestrutura a partir do reconhecimento de padrões em imagens.					
Etapas					
E-F05.1		Início	M1	Fim	M2
	Título	Definição do escopo das atividades			
	Risco vinculado				
Tecnologias previstas		--			
Tipos de dados necessários		Será necessário definir com a equipe do MPMG o escopo das tarefas que serão monitoradas pelas imagens.			
Descrição	Levantamento de requisitos para elaboração da base de dados inicial e casos de uso. Essa etapa contará com a interação com a equipe do MPMG.				

E-F05.2		Início	M1	Fim	M6
	Título	Extração de cronogramas a partir de documentos oficiais			
	Risco vinculado	R-F05.1			
Tecnologias previstas		--			
Tipos de dados necessários		Documentos oficiais que descrevem planejamento e projeto das obras monitoradas. Será avaliado interação com a tarefa M02 para detecção de entidades.			
Descrição	Desenvolvimento de algoritmos e ferramentas para extração automática dos cronogramas a partir de documentos oficiais				
E-F05.3		Início	M3	Fim	M5
	Título	Criação de conjuntos de dados para treinamento de modelos			
	Risco vinculado				
Tecnologias previstas		--			
Tipos de dados necessários		Imagens coletadas de repositórios públicos e por meio de eventuais campanhas pelo app Trena.			
Descrição	Montagem e coleta de dados para treinamento dos modelos para novos tipos de obras.				
E-F05.4		Início	M6	Fim	M8
	Título	Implementação e testes com abordagens do estado da arte.			

		Risco vinculado	R-F05.3		
Tecnologias previstas		--			
Tipos de dados necessários		Conjuntos de imagens rotuladas.			
Descrição	Implementação e testes com abordagens do estado da arte. Será feito um levantamento completo da literatura visando analisar as principais técnicas que poderão ser utilizadas na resolução das tarefas descritas nessa linha.				
E-F05.5		Início	M4	Fim	M7
	Título	Desenvolvimento do módulo interativo			
	Risco vinculado				
Tecnologias previstas		--			
Tipos de dados necessários		Conjuntos de imagens do F05 de 2020 e de 2021.			
Descrição	Desenvolvimento de um módulo de aprendizado interativo com feedback do usuário para o app Trena. O app deverá recomendar estágio das obras assim que o usuário tirar uma foto. A recomendação poderá ser revisada pelo usuário do app.				
E-F05.6		Início	M7	Fim	M10
	Título	Consolidação das ferramentas.			
	Risco vinculado				

Tecnologias previstas		--			
Tipos de dados necessários		--			
Descrição	Desenvolvimento, teste e consolidação das ferramentas e modelos.				
Entregáveis					
E-F05.7		Início	M11	Fim	M12
	Título	Integração dos módulos.			
	Risco vinculado				
Tecnologias previstas		--			
Tipos de dados necessários		--			
Descrição	Integração dos módulos do projeto e treinamento dos modelos com base de dados atualizada. Deverão ser realizados testes simulando situações reais de uso das ferramentas.				
Entregáveis					
F05.1	E-F05.1	Documento de requisitos e descrição do escopo.			
	M2				
F05.2	E-F05.3	Conjuntos de dados rotulados.			
	M5				
F05.3	E-F05.2	Scripts para extração de cronogramas a partir de documentos textuais.			
	M6				
F05.4	E-F05.5	Códigos do módulo interativo do App TRENA.			
	M7				

F05.5	E-F05.4	Códigos das abordagens implementadas e testadas.
	M8	
F05.6	E-F05.6	Códigos das ferramentas consolidadas e testadas.
	M10	
F05.7	E-F05.7	Relatório de simulações realizadas com dados reais após integração dos módulos.
	M12	
Milestones		
M-F05.1	Versão preliminar do documento de requisitos.	
M1		
M-F05.2	Relação de repositórios de documentos oficiais que serão utilizados para desenvolvimento do módulo de extração de cronogramas.	
M3		
M-F05.3	Listagem da coleta preliminar de documentos e detalhamento do módulo interativo do App TRENA.	
M4		
M-F05.4	Versão preliminar dos códigos das ferramentas consolidadas.	
M9		
M-F05.5	Relatório com resultados preliminares das simulações realizadas.	
M11		

Riscos Críticos para a Implementação

Código	Descrição	Impacto	Tratamento	Plano de Contingência
---------------	------------------	----------------	-------------------	------------------------------

R-F05.1	Acesso a documentos com planejamento de obras atuais com imagens disponíveis.	Alto	Mitigar	Simular/adaptar documentos que possam servir para desenvolvimento das ferramentas.
----------------	---	------	---------	--

Cronograma das Entregas por Etapas

Etapas		Entregas											
Cód.	Título	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-F05.1	Definição do escopo das atividades		F05.1										
E-F05.2	Extração de cronogramas a partir de documentos oficiais						F05.3						
E-F05.3	Criação de conjuntos de dados para treinamento de modelos					F05.2							
E-F05.4	Implementação e testes com abordagens do estado da arte.								F05.5				
E-F05.5	Desenvolvimento do módulo interativo							F05.4					
E-F05.6	Consolidação das ferramentas.										F05.6		

E-F05.7	Integração dos módulos.												F05.7
MILESTONES		M-F05.1		M-F05.2	M-F05.3					M-F05.4		M-F05.5	

Etapas - 2022					
E-F05.11		Início	M13	Fim	M14
	Título	Definição do escopo das atividades para 2021			
	Descrição	Levantamento de requisitos para monitoramento de novas obras. Essa etapa contará com a interação com a equipe do MPMG.			
E-F05.12		Início	M15	Fim	M24
	Título	Desenvolvimento de abordagens e ferramentas para monitoramento de obras ainda não contempladas			
	Descrição	Criação de datasets, levantamento do estado da arte e implementação das novas abordagens.			
Entregáveis - 2022					
F05.11	E-F05.11	Documento de requisitos e descrição do escopo.			
	M14				
F05.12	E-F05.12	Códigos das ferramentas desenvolvidas.			
	M24				

Grupo de Trabalho	F06 - Chatbot Adaptativo e Geração de Relatórios Analíticos				
Grupos de Interação	A01, A02, G01, P03				
Coordenador	Wagner Meira Jr.				
Equipe	Nível	Quantidade	Nomes		
	Professor	4	Wagner Meira Jr., Regina Helena Silva, Raquel Prates, Adriana Pagano		
	Pós-Doc	4	TOTAL	16	
	Doutorando	2			
	Mestrando	2			
	Graduando	4			
	Técnico	0	--		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Não	Não	Não	Não	Não
Descrição Geral					
Este projeto visa desenvolver um chatbot adaptativo para atuar na ouvidoria do MPMG e a geração de relatórios analíticos a partir dos resultados das estatísticas, análises provenientes da utilização do chatbot e das informações relevantes extraídas da base de conhecimento continuamente enriquecida.					
Objetivos					
Os principais objetivos deste projeto são instanciar o chatbot implementado pela plataforma adaptativa inteligente para coleta de dados e construção de bases de conhecimento para o contexto da ouvidoria do MPMG e a geração de relatórios estatísticos do uso do chatbot e das informações extraídas a partir das bases de conhecimentos. Para atingir este objetivo, o projeto pode ser dividido em três passos principais: 1) Levantamento dos requisitos para a instanciação do chatbot para utilização na Ouvidoria do MPMG e das informações de interesse a serem extraídas das bases de conhecimento; 2) Desenvolvimento de um chatbot adaptativo refletindo o contexto da ouvidoria do MPMG e; 3) Geração de relatórios analíticos com os resultados das estatísticas e análises a partir da utilização do chatbot e informações relevantes extraídas a partir da base de conhecimento continuamente enriquecida.					

Etapas					
E-F06.1		Início	M1	Fim	M3
	Título	Detalhamento do Chatbot e dos Relatórios Analíticos			
		Risco vinculado	R-F06.1		
	R-F06.2				
Tecnologias previstas		--			
Tipos de dados necessários		--			
Descrição	Levantamento de requisitos para a instanciação do chatbot no contexto da ouvidoria do MPMG e definição das informações a serem incluídas nos relatórios analíticos				
E-F06.2		Início	M4	Fim	M8
	Título	Chatbot 1.0			
		Risco vinculado	R-F06.1	R-F06.3	
	R-F06.2				
Tecnologias previstas		--			
Tipos de dados necessários		--			
Descrição	Nesta etapa pretende-se especificar cenários de utilização do chatbot e os principais componentes para a implementação do chatbot (questões, textos de interação), além da implementação dos mecanismos de adaptação do modelo de linguagem, a partir de análise linguística de campo, sintonia e modo.				
E-F06.3		Início	M9	Fim	M12
	Título	Chatbot 2.0			

		Risco vinculado	R-F06.1	R-F06.3	
			R-F06.2		
Tecnologias previstas		--			
Tipos de dados necessários		--			
Descrição	A partir do feedback da utilização da versão anterior do chatbot, o objetivo desta etapa será a readequação de textos de interação e da implementação dos mecanismos de adaptação do modelo de linguagem, a partir de análise linguística de campo, sintonia e modo.				
		Início	M4	Fim	M6
E-F06.4	Título	Geração de Relatórios 1.0			
	Risco vinculado	R-F06.1			
Tecnologias previstas		--			
Tipos de dados necessários		--			
Descrição	Versão 1.0 dos Relatórios Analíticos.				
		Início	M7	Fim	M9
E-F06.5	Título	Geração de Relatórios 2.0			
	Risco vinculado	R-F06.1			

Tecnologias previstas		--			
Tipos de dados necessários		--			
Descrição	Versão 2.0 dos Relatórios Analíticos.				
Entregáveis					
E-F06.6	Início	M10	Fim	M12	
	Título	Geração de Relatórios 3.0			
	Risco vinculado	R-F06.1			
Tecnologias previstas		--			
Tipos de dados necessários		--			
Descrição	Versão 3.0 dos Relatórios Analíticos.				
Entregáveis					
F06.1	E-F06.1	Relatório com a descrição dos requerimentos funcionais e não-funcionais do chatbot e especificação das análises analíticas			
	M3				
F06.2	E-F06.4	Dashboard com as análises analíticas			
	M6				
F06.3	E-F06.2	Aplicativo chatbot versão 1.0			
	M8				

F06.4	E-F06.5	Dashboard com as análises analíticas
	M9	
F06.5	E-F06.3	Aplicativo chatbot versão 2.0
	M12	
F06.6	E-F06.6	Dashboard com as análises analíticas
	M12	
Milestones		
M-F06.1	Definição da equipe	
M1		
M-F06.2	Acompanhamento das atividades desenvolvidas.	
M2		
M-F06.3	Acompanhamento das atividades desenvolvidas.	
M4		
M-F06.4	Acompanhamento das atividades desenvolvidas.	
M5		
M-F06.5	Acompanhamento das atividades desenvolvidas.	
M7		

M-F06.6	Acompanhamento das atividades desenvolvidas.
M10	
M-F06.7	Acompanhamento das atividades desenvolvidas.
M11	

Riscos Críticos para a Implementação				
Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-F06.1	Mudança de equipe	Médio	Aceitar	Realizar novo processo para contratação de recurso.
R-F06.2	Desenvolvimento de tecnologia que depende de pesquisa na fronteira do conhecimento	Médio	Mitigar	O projeto será conduzido por equipe de pesquisadores experientes nas áreas envolvidas, que trabalham na fronteira do conhecimento de suas áreas. Serão realizadas entregas parciais com melhorias incrementais progressivas.
R-F06.3	Desempenho esperado desconhecido	Alto	Mitigar	Por se tratar de tecnologia nova aliada à pesquisa, os resultados da aplicação final são de difícil previsibilidade em situações práticas. Para mitigar, cada desenvolvimento incremental será baseado também no feedback da utilização prática da versão anterior. Usar metodologia de "startup enxuta" para validar conceitos e ideias na prática antes da implementação?

Cronograma das Entregas por Etapas													
ETAPAS		ENTREGAS											
CÓD.	TÍTULO	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12

E-F06.1	Detalhamento do Chatbot e dos Relatórios Analíticos			F06.1									
E-F06.2	Chatbot 1.0								F06.3				
E-F06.3	Chatbot 2.0												F06.5
E-F06.4	Geração de Relatórios 1.0						F06.2						
E-F06.5	Geração de Relatórios 2.0									F06.4			
E-F06.6	Geração de Relatórios 3.0												F06.6
MILESTONES		M-F06.1	M-F06.2		M-F06.3	M-F06.4		M-F06.5			M-F06.6	M-F06.7	

Etapas - 2022					
E-F06.11		Início	M13	Fim	M16
	Título	Chatbot 3.0			
	Descrição	Versão 3.0 do Chatbot Adaptativo para utilização na Ouvidoria do MPMG			
E-F06.12		Início	M17	Fim	M20
	Título	Chatbot 4.0			
	Descrição	Versão 4.0 do Chatbot Adaptativo para utilização na Ouvidoria do MPMG			

E-F06.13		Início	M21	Fim	M24
	Título	Chatbot 5.0			
	Descrição	Versão 5.0 do Chatbot Adaptativo para utilização na Ouvidoria do MPMG			
Entregáveis - 2022					
F06.11	E-F06.11	Geração de Relatórios 4.0			
	M16				
F06.12	E-F06.12	Geração de Relatórios 5.0			
	M20				
F06.13	E-F06.13	Geração de Relatórios 6.0			
	M24				

G. Gestão do Projeto

Grupo de Trabalho	G01 - Gestão do Projeto				
Grupos de Interação	Todos				
Coordenador	Wagner Meira Jr.				
Equipe	Nível	Quantidade	Nomes		
	Professor	1	Wagner Meira Jr.		
	Pós-Doc	0	TOTAL	7	

	Doutorando	0				
	Mestrando	0				
	Graduando	0				
	Técnico	6	Gerente do projeto			
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social	
	Sim	Sim	Sim	Sim	Sim	
Descrição Geral						
Grupo de trabalho responsável por toda a administração do projeto, incluindo: coordenação, gestão, administração financeira (via FUNDEP), bem como garantia de qualidade e procedimentos de gerenciamento de risco.						
Objetivos						
<p>O principal objetivo é o cumprimento bem-sucedido dos objetivos gerais do projeto, e de cada grupo de atividades, de forma eficiente e eficaz, em tempo hábil, enquadrando-se no orçamento e na qualidade esperada pela parceria MPMG-DCC/UFMG. Para isso, será realizado o monitoramento das atividades, por meio do contato direto com os coordenadores de cada grupos de trabalho (e/ou seus líderes delegados), incluindo medidas a serem implementadas para garantir a sinergia e orientação das equipes.</p> <p>Objetivos específicos:</p> <ul style="list-style-type: none"> - Definição e implementação dos procedimentos de comunicação a serem seguidos dentro do projeto. - Garantir que as questões de pesquisa, desenvolvimento, administrativas e financeiras estejam dentro do cronograma. - Acompanhamento da evolução dos grupos de trabalho, reuniões internas do projeto (físicas e virtuais). - Plano de gestão com procedimentos para resolver quaisquer gargalos e atrasos que possam surgir por meio de medidas corretivas e estratégias de contingência. 						

- Coordenação e acompanhamento da equipe de engenharia de dados (analistas) lotados no MPMG, contratados via FUNDEP.
- Conduzir e gerenciar o cronograma de reuniões mensais de coordenação (acompanhamento das atividades de cada grupo de trabalho).
- Conduzir e gerenciar o cronograma de reuniões das equipes com a participação de integrantes da equipe técnica do MPMG.
- Organizar, em conjunto com a equipe de gestão de projetos/contratos do MPMG, as reuniões gerenciais do projeto.
- Criar e disponibilizar os modelos para os documentos entregáveis, milestones (marcos) e apresentações do projeto.

Etapas

		Início	M0	Fim	M0		
E-G01.0	Título	Detalhamento das etapas e entregas do projeto a serem executadas nos primeiros 12 meses					
	Risco vinculado	R-G01.7					
		R-G01.8					
Descrição	<p>Coordenar o detalhamento das atividades de todos os grupos de trabalho, de modo a disponibilizar, de forma mais detalhada, as atividades a serem desenvolvidas nos primeiros 12 meses do projeto.</p> <p>Todos os coordenadores de grupo de atividades têm obrigação nesta etapa.</p>						
E-G01.1		Início	M2	Fim	M12		
	Título	Acompanhar as integrações dos grupos de trabalho					
	Risco vinculado	R-G01.1	R-G01.3				
		R-G01.2	R-G01.4				

Descrição	Garantir, através de reuniões com as equipes, as necessidade de integrações entre os diferente módulos do projeto.				
		Início	M1	Fim	M12
E-G01.2	Título	Gerir o escopo do projeto			
	Risco vinculado	R-G01.2	R-G01.4	R-G01.6	
		R-G01.3	R-G01.5	R-G01.7	
Descrição	Garantir que o projeto irá incluir todos os passos necessários para ser concluído com sucesso.				
		Início	M1	Fim	M12
E-G01.3	Título	Gestão de pessoas			
	Risco vinculado	R-G01.1	R-G01.3		
		R-G01.2	R-G01.4		
Descrição	Encaminhar a contratação de pessoas, coletar assinaturas do termo de sigilo, determinar (quando necessário) as pessoas responsáveis por cada etapa do projeto, quais habilidades necessárias para cada cargo, desenvolvimento das competências e o acompanhamento dos colaboradores durante a realização do projeto. Definir quem precisará contratar para garantir a conclusão do projeto.				
		Início	M1	Fim	M12
E-G01.4	Título	Gestão da comunicação do projeto			

		Risco vinculado	R-G01.6			
Descrição	Gerir a comunicação do projeto, incluindo o planejamento, a coleta de informações, a distribuição e armazenamento. Além de fazer a interface de comunicação entre as equipes do MPMG e as equipes do DCC/UFMG.					
		Início	M1	Fim	M12	
E-G01.5	Título	Gestão de riscos				
	Risco vinculado	Todos os riscos definidos no projeto.				
Descrição	Identificar, planejar, avaliar e controlar os riscos que estão envolvidos no projeto. O objetivo dessa etapa é evitar ou pelo menos diminuir a chance de riscos negativos afetarem o desenvolvimento do projeto. Visa, também, aumentar a chance de acontecerem riscos de oportunidade (riscos positivos).					
		Início	M1	Fim	M12	
E-G01.6	Título	Coordenação da equipe de engenharia de dados				
	Risco vinculado	R-G01.3				
Descrição	Coordenação das atividades da equipe de engenharia de dados que ficará alocada no MPMG. Garantir a interação dos engenheiros de dados com as equipes do projeto que necessitarem de tal profissional.					

	Conduzir, em conjunto com a equipe técnica do MPMG, as atividades da equipe de engenharia de dados, de forma que as soluções entregues no projeto sejam incorporados ao ambiente do MPMG e disponibilizados para acesso e requisição por outros grupos de atividades.				
E-G01.7	Início	M1	Fim	M12	
	Título	Condução e disponibilização dos entregáveis do projeto			
	Risco vinculado	R-G01.5	R-G01.7		
		R-G01.6			
Descrição	Disponibilizar à equipe técnica do MPMG, por e-mail, os documentos entregáveis de cada um dos grupos de trabalho, sempre no último dia útil de cada mês.				
E-G01.8	Início	M1	Fim	M1	
	Título	Criação dos modelos dos documentos entregáveis do projeto			
	Risco vinculado				
Descrição	Criação e disponibilização dos modelos para os documentos entregáveis, milestones e apresentações do projeto.				
E-G01.9	Início	M6	Fim	M10	
	Título	Condução do detalhamento do projeto 2022			
	Risco vinculado	R-G01.6			

Descrição	Organizar as equipe/coordenadores para a escrita/detalhamento do projeto 2022.				
Entregáveis					
G01.0	E-G01.0	Projeto detalhado para os primeiros 12 meses de atividades de todos os grupos de trabalho.			
	M0				
G01.1	E-G01.8	Modelos para os documentos entregáveis, milestones e apresentações do projeto.			
	M1				
G01.2	E-G01.6	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.			
	M1				
G01.3	E-G01.2	Cronograma das reuniões de coordenação, reuniões gerenciais e reuniões gerenciais de detalhamento do projeto 2022.			
	M1				
G01.4	E-G01.6	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.			
	M2				
G01.5	E-G01.6	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.			
	M3				
G01.6	E-G01.6	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.			
	M4				
G01.7	E-G01.6	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.			
	M5				

G01.8	E-G01.6	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.	
	M6		
G01.9	E-G01.9	Plano de condução do detalhamento do projeto 2022.	
	M6		
G01.10	E-G01.6	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.	
	M7		
G01.11	E-G01.9	Segunda versão da proposta de etapas e entregáveis para 2022.	
	M7	*A primeira versão está disponível no projeto 2021.	
G01.12	E-G01.6	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.	
	M8		
G01.13	E-G01.9	Primeira versão detalhada do Projeto 2022.	
	M8		
G01.14	E-G01.6	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.	
	M9		
G01.15	E-G01.6	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.	
	M10		
G01.16	E-G01.9	Versão final do Projeto 2022.	
	M10		
G01.17	E-G01.6	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.	

	M11		
G01.18	E-G01.6	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.	
	M12		

Riscos Críticos				
Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-G01.1	Saída de algum bolsista dos grupos de trabalho: graduando, mestrando, doutorando ou pós-doutorando.	Médio	Transferir	Solicitar ao coordenador a seleção de novo bolsista.
R-G01.2	Saída de bolsista líder de algum grupo de trabalho.	Alto	Transferir	Solicitar ao coordenador a seleção (urgente) de novo bolsista líder.
R-G01.3	Saída de algum dos engenheiros de dados alocados no MPMG.	Alto	Mitigar	Iniciar seleção imediata para contratação de outro engenheiro de dados.
R-G01.4	Vacância de qualquer cargo técnico do projeto.	Alto	Mitigar	Iniciar seleção imediata para contratação de outro profissional para o cargo vago.
R-G01.5	Atraso em mais de 7 (sete) dias úteis por parte da equipe técnica do MPMG em dar parecer técnico sobre as entregas de cada mês.	Alto	Transferir	Solicitar à equipe de Gestão de Contratos do MPMG a apreciação imediata dos entregáveis do mês anterior.
R-G01.6	Atraso na marcação de reuniões com especialistas de conteúdo do MPMG.	Médio	Transferir	Informar ao MPMG que a não ocorrência da reuniões pode impactar as atividades do respectivo grupo de trabalho.
R-G01.7	Atraso no pagamento de fatura do projeto por conta do Risco R-G01.5.	Alto	Mitigar	Reunião Gerencial de urgência para solução do problema.
R-G01.8	Período de férias de algum coordenador de grupo de atividades.	Baixo	Mitigar	Solicitar ao coordenador titular que designe um coordenador interino.

Cronograma das Entregas por Etapas														
ETAPAS		ENTREGAS												
CÓD.	TÍTULO	M0	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-G01.0	Detalhamento das etapas e entregas do projeto a serem executadas nos primeiros 12 meses	G01.0												
E-G01.1	Gerir as integrações dos grupos de trabalho													
E-G01.2	Gerir o escopo do projeto		G01.3											
E-G01.3	Gestão de pessoas													
E-G01.4	Gestão da comunicação do projeto													
E-G01.5	Gestão de riscos													
E-G01.6	Coordenação da equipe de engenharia de dados		G01.2	G01.4	G01.5	G01.6	G01.7	G01.8	G01.10	G01.12	G01.14	G01.15	G01.17	G01.18
E-G01.7	Condução e disponibilização													

	dos entregáveis do projeto													
E-G01.8	Criação dos modelos dos documentos entregáveis do projeto		G01.1											
E-G01.9	Condução do detalhamento do projeto 2022						G01.9	G01.11	G01.13			G01.16		
MILESTONES		--	--	--	--	--	--	--	--	--	--	--	--	--

Etapas - 2022					
E-G01.10	Início	M14		Fim	M24
	Título	Acompanhar as integrações dos grupos de trabalho			
Descrição	Garantir, através de reuniões com as equipes, as necessidade de integrações entre os diferente módulos do projeto.				
E-G01.11	Início	M13		Fim	M24
	Título	Gerir o escopo do projeto			
Descrição	Garantir que o projeto irá incluir todos os passos necessários para ser concluído com sucesso.				
E-G01.12	Início	M13		Fim	M24
	Título	Gestão de pessoas			

Descrição	Encaminhar a contratação de pessoas, coletar assinaturas do termo de sigilo, determinar (quando necessário) as pessoas responsáveis por cada etapa do projeto, quais habilidades necessárias para cada cargo, desenvolvimento das competências e o acompanhamento dos colaboradores durante a realização do projeto. Definir quem precisará contratar para concluir o projeto.			
E-G01.13	Início	M13	Fim	M24
	Título	Gestão da comunicação do projeto		
Descrição	Gerir completamente a comunicação do projeto, incluindo o planejamento, a coleta de informações, a distribuição e armazenamento. Além de fazer a interface de comunicação entre as equipes do MPMG e as equipes do DCC/UFMG.			
E-G01.14	Início	M13	Fim	M24
	Título	Gestão de riscos		
Descrição	Identificar, planejar, avaliar e controlar os riscos que estão envolvidos no projeto. O objetivo dessa etapa é evitar ou pelo menos diminuir a chance de riscos negativos afetarem o desenvolvimento do projeto. Visa, também, aumentar a chance de acontecerem riscos de oportunidade, também conhecidos como riscos positivos.			
E-G01.15	Início	M13	Fim	M24
	Título	Coordenação da equipe de engenharia de dados		
Descrição	<p>Coordenação das atividades da equipe de engenharia de dados que ficará alocada no MPMG.</p> <p>Garantir a interação dos engenheiros de dados com as equipes do projeto que necessitarem de tal profissional.</p> <p>Conduzir, em conjunto com a equipe técnica do MPMG, as atividades da equipe de engenharia de dados, de forma que as soluções entregues no projeto sejam incorporados ao ambiente do MPMG e disponibilizados para acesso e requisição por outros grupos de atividades.</p>			
E-G01.16	Início	M13	Fim	M24

	Título	Condução e disponibilização dos entregáveis do projeto
Descrição	Disponibilizar à equipe técnica do MPMG, por e-mail, os documentos entregáveis de cada um dos grupos de trabalho.	
Entregáveis - 2022		
G01.2	E-G01.15	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.
	M13	
G01.3	E-G01.15	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.
	M14	
G01.4	E-G01.15	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.
	M15	
G01.5	E-G01.15	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.
	M16	
G01.6	E-G01.15	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.
	M17	
G01.7	E-G01.15	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.
	M18	
G01.8	E-G01.15	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.
	M19	
G01.9	E-G01.15	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.
	M20	
G01.10	E-G01.15	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.

	M21	
G01.11	E-G01.15	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.
	M22	
G01.12	E-G01.15	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.
	M23	
G01.13	E-G01.15	Relatórios das atividades desenvolvidas pela equipe de engenharia de dados.
	M24	

M. Módulos de software

Grupo de Trabalho	M01 - Processamento de linguagem natural para extração e classificação de entidades e suas relações em textos não-estruturados		
Grupos de Interação	D01, E02, F02, F03, F04, G01, M02, M05, P01		
Coordenador	Marcos André Gonçalves		
Equipe	Nível	Quantidade	Nomes
	Professor	1	Marcos André Gonçalves
	Pós-Doc	1	TOTAL 5
	Doutorando	0	
	Mestrando	0	
	Graduando	2	
	Técnico	1	Desenvolvedor <i>front-end</i> responsável pelo Software de Anotação.

Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
		Sim	Sim	Não	Sim
Descrição Geral					
Implementação, extensão e aplicação de técnicas avançadas de aprendizado ativo e de processamento de linguagem natural para extração e classificação automática de entidades e seus relacionamentos a partir de textos não estruturados. Suporte a outras aplicações e à rotulação de dados de outros processos de aprendizado de máquina.					
Objetivos					
Extração de relacionamentos; Ampliação de Aplicações; Expansão do uso das técnicas semi-supervisionadas para rotulação de outros processos de aprendizado de máquina.					
Etapas					
E-M01.1		Início	M1	Fim	M2
	Título	Caracterização de Dados			
	Risco vinculado	R-M01.1			
Tecnologias previstas		Linguagens Python e Java			
Tipos de dados necessários		Licitações/diários oficiais (C04), dados de redes sociais (C01, C02).			
Descrição	Caracterização de relacionamentos entre entidades.				
E-M01.2		Início	M3	Fim	M7
	Título	Reconhecimento de Relacionamentos entre Entidades e Aplicações			
	Risco vinculado	R-M01.2	R-M01.3		

Tecnologias previstas		Linguagens Python e Java.			
Tipos de dados necessários		Amostra dos dados utilizados na etapa anterior (ou dados similares) rotuladas com entidades e seus relacionamentos.			
Descrição	Extração e classificação de relacionamentos entre entidades; Aplicação do reconhecedor de entidades e seus relacionamentos para novos tipos de dados (ex.: redes sociais).				
Entregáveis					
E-M01.3	Início	M8	Fim	M12	
	Título	Refinamentos e Extensões dos Métodos Propostos			
	Risco vinculado	R-M01.3	R-M01.4		
Tecnologias previstas		Linguagens Python e Java			
Tipos de dados necessários		Amostra de dados rotulados da etapa anterior (expandida)			
Descrição	Refinamentos do módulo de reconhecimento de entidades e seus relacionamentos; Extensão das técnicas de aprendizado ativo para auxiliar outros processos de rotulação de aprendizado de máquina				
Entregáveis					
M01.1	E-M01.1	Relatório contendo a caracterização das bases de dados quanto aos relacionamentos entre entidades			
	M2				
M01.2	E-M01.2	Módulo de extração e classificação de relacionamentos entre entidades; Relatório contendo avaliação experimental			
	M7				
M01.3	E-M01.3	Módulo de aprendizado ativo; Relatório contendo avaliação experimental			
	M12				
Milestones					

M-M01.1	Parecer sobre quais dados serão utilizados na Etapa M01.1 e como eles serão obtidos
M1	
M-M01.2	Estudo do estado-da-arte em reconhecimento de relacionamentos entre entidades. Descrição sucinta dos principais métodos
M3	
M-M01.3	Implementação dos primeiros métodos de referência em reconhecimento de relacionamentos entre entidades
M4	
M-M01.4	Amostra dos dados rotulada com o auxílio de método de aprendizado ativo e, caso já esteja disponível, da interface Web de rotulação
M5	
M-M01.5	Resultados da aplicação do reconhecedor de entidades e seus relacionamentos em outros tipos de dados
M6	
M-M01.6	Resumo de todas as interações com outros Grupos de Atividades e respectivos resultados obtidos até o momento
M8	
M-M01.7	Modelos de reconhecimento de entidades específicos para diferentes tipos de dados (ex.: licitações, notícias, dados de redes sociais)
M9	
M-M01.8	Método de aprendizado ativo geral para tarefas diversas de aprendizado de máquina
M10	
M-M01.9	Modelos de reconhecimento de entidades e seus relacionamentos melhorados através da seleção de novos dados rotulados
M11	

Riscos Críticos para a Implementação

Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-M01.1	Não obtenção dos dados desejados de forma completa	Baixo	Mitigar	Buscar dados públicos com características similares às dos dados desejados
R-M01.2	Baixa precisão / revocação no reconhecimento de relações. A mineração de relações entre entidades é um problema ainda mais complexo que o reconhecimento das entidades, podendo ter um baixo desempenho	Baixo	Mitigar	Testes e refinamentos sucessivos da ferramenta a ser implementada
R-M01.3	Dados ruidosos/com baixa qualidade (ex.: dados de redes sociais, que não são bem comportados como os de editais e outros textos legais)	Baixo	Mitigar	Testes e refinamentos sucessivos da ferramenta a ser implementada, inclusive para limpeza dos dados
R-M01.4	Robustez da avaliação dos métodos de aprendizado ativo. Soluções baseadas em aprendizado ativo precisam de uma avaliação mais robusta de longo prazo baseado no uso real das ferramentas e do feedback dos usuários com possíveis reformulações constantes.	Baixo	Mitigar	Testes e refinamentos sucessivos da ferramenta a ser implementada baseados no uso real da ferramenta

Cronograma das Entregas por Etapas

Etapas		Entregas											
Cód.	Título	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-M01.1	Caracterização de Dados		M01.1										
E-M01.2	Reconhecimento de Relacionamentos entre Entidades e Aplicações							M01.2					

E-M01.3	Refinamentos e Extensões dos Métodos Propostos												M01.3
MILESTONES		M-M01.1		M-M01.2	M-M01.3	M-M01.4	M-M01.5		M-M01.6	M-M01.7	M-M01.8	M-M01.9	

Etapas - 2022					
E-M01.4		Início	M13	Fim	M16
	Título	Expansão de Aplicações			
	Descrição	Avaliação mais robusta, de longo prazo, baseada no uso real das ferramentas por diferentes Grupos de Atividade; Suporte a esses grupos; Interação com o grupo de transcrição de áudio; Identificação das entidades nos textos de áudios transcritos.			
E-M01.5		Início	M17	Fim	M20
	Título	Aprendizado Ativo			
	Descrição	Obtenção do feedback de outros Grupos de Atividade e avaliação mais robusta dos métodos de Aprendizado Ativo; Suporte, por meio das técnicas de Aprendizado Ativo implementadas, a outros Grupos que necessitem rotulagem manual de dados			
E-M01.6		Início	M21	Fim	M24
	Título	Refinamentos no Reconhecedor de Entidades e Relacionamentos			
	Descrição	Melhorias no reconhecedor de entidades e seus relacionamentos baseadas principalmente em coleções robustas de dados rotulados gerados a partir do uso da ferramenta por diferentes Grupos de Atividade			

Entregáveis - 2022		
M01.4	E-M01.11	Relatório descrevendo as interações com outros Grupos de Atividade e as aplicações que foram produto dessas interações
	M16	
M01.5	E-M01.12	Relatório contendo a avaliação experimental da ferramenta geral de aprendizado ativo e como ela auxiliou outros Grupos de Atividade; Código atualizado
	M20	
M01.6	E-M01.13	Relatório contendo a avaliação experimental da ferramenta de reconhecimento de entidades/relacionamentos no âmbito de cada aplicação; Código atualizado
	M24	

Grupo de Trabalho	M02 - Classificação Semântica		
Grupos de Interação	D01, D02, F02, F03, G01, M01		
Coordenador	Marcos André Gonçalves		
Equipe	Nível	Quantidade	Nomes
	Professor	1	Marcos André Gonçalves
	Pós-Doc	0	TOTAL 5
	Doutorando	2	
	Mestrando	0	
	Graduando	2	
	Técnico	0	--

Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Sim	Sim	Sim	Sim	Não
Descrição Geral					
Esse projeto visa consolidar os resultados da classificação semântica de atributos de tabelas estruturadas e iniciar a classificação das Seções (segmentos) de Diários Oficiais. Identificar os tópicos do segmentos e o que cada seção corresponde (semântica) .					
Objetivos					
I- Melhoria da Classificação Semântica de Atributos II- Classificação Semântica de Segmentos de Diários Oficiais					
Etapas					
E-M02.1		Início	M1	Fim	M2
	Título	Caracterização das Bases			
		Risco vinculado	R-M02.1		R-M02.3
Tecnologias previstas		Linguagens Python e Java			
Tipos de dados necessários		Dados de tabelas; Dados de Diários oficiais, incluindo licitações com rotulação			
Descrição	- Consolidação dos resultados anteriores em Tabelas estruturadas - Caracterização dos dados dos Diários Oficiais				
E-M02.2		Início	M3	Fim	M7
	Título	Segmentação de Textos			
		Risco vinculado	R-M02.1	R-M02.2	R-M02.3

Tecnologias previstas		Linguagens Python e Java			
Tipos de dados necessários		Dados de Diários oficiais com rotulação			
Descrição		<ul style="list-style-type: none"> - Segmentação dos textos dos diários; - Testes iniciais com os algoritmos de classificação; - Aferição de qualidade do processo de segmentação e classificação 			
Entregáveis					
E-M02.3		Início	M8	Fim	M12
	Título	Segmentação e Classificação - Aprimoramento das Soluções			
	Risco vinculado	R-M02.1	R-M02.2	R-M02.3	
Tecnologias previstas		Linguagens Python e Java			
Tipos de dados necessários		Dados de Diários oficiais com rotulação			
Descrição		<ul style="list-style-type: none"> - Refinamentos dos módulo de segmentação e classificação; - Utilização de técnicas de aprendizado ativo para auxiliar os aprendizado de máquina; - Integração das soluções na Infra-estrutura do MPMG; - Integração com outras aplicações finalísticas. 			
M02.1	E-M02.1	Relatório contendo a caracterização das bases de dados quanto aos diários oficiais			
	M2				
M02.2	E-M02.2	Módulo de segmentação e classificação; Relatório contendo avaliação experimental			

	M7	
M02.3	E-M02.3	Módulos Avançados de segmentação e classificação; Impacto do Aprendizado ativo no processo; Relatório contendo avaliação experimental
	M12	
Milestones		
M-M02.1	Parecer sobre quais dados serão utilizados na Etapa M02.1 e como eles serão obtidos	
M1		
M-M02.2	Estudo do estado-da-arte em segmentação de textos. Descrição sucinta dos principais métodos.	
M3		
M-M02.3	Implementação dos primeiros métodos de segmentação e classificação	
M4		
M-M02.4	Amostra dos dados rotulada com o auxílio de método de aprendizado ativo e, caso já esteja disponível, da interface Web de rotulação	
M5		
M-M02.5	Resultados da aplicação do módulos de de segmentação e classificação	
M6		
M-M02.6	Resumo de todas as interações com outros Grupos de Atividades e respectivos resultados obtidos até o momento	
M8		
M-M02.7	Modelos de segmentação e classificação para diferentes tipos de eventos como leis, decretos, editais de concursos, nomeação de servidores públicos	
M9		
M-M02.8	Método de aprendizado ativo geral para tarefas diversas de aprendizado de máquina	

M10	Modelos de segmentação e classificação melhorados através da seleção de novos dados rotulados
M-M02.9	
M11	

Riscos Críticos para a Implementação

Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-M02.1	Não obtenção dos dados desejados de forma completa.	Baixo	Mitigar	Buscar dados públicos com características similares às dos dados desejados.
R-M02.2	Baixa precisão / revocação na classificação.	Baixo	Mitigar	Testes e refinamentos sucessivos dos métodos implementados.
R-M02.3	Dados ruidosos/com baixa qualidade.	Baixo	Mitigar	Testes e refinamentos sucessivos dos métodos implementados. Limpeza de Dados.

Cronograma das Entregas por Etapas

Etapas		Entregas											
Cód.	Título	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-M02.1	Caracterização das Bases		M02.1										
E-M02.2	Segmentação de Textos							M02.2					
E-M02.3	Segmentação e Classificação - Aprimoramento das Soluções												M02.3

MILESTONES	M- M02.1		M- M02.2	M- M02.3	M- M02.4	M- M02.5		M- M02.6	M- M02.7	M- M02.8	M- M02.9	
-------------------	---------------------	--	---------------------	---------------------	---------------------	---------------------	--	---------------------	---------------------	---------------------	---------------------	--

Etapas - 2022					
E-M02.4		Início	M13	Fim	M14
	Título	- Classificação Semântica de outra Aplicação e/ou Classificação Hierárquica de Dados (Hierarquias de Classificação) - Caracterização de Bases			
	Descrição	- Caracterização de Bases em aplicações novas ou Hierarquia de Classificação			
E-M02.5		Início	M15	Fim	M19
	Título	- Classificação Semântica de outra Aplicação e/ou Classificação Hierárquica de Dados (Hierarquias de Classificação) - Implementação e Testes das Primeiras Soluções			
	Descrição	- Testes iniciais com os algoritmos de classificação hierárquica ou classificação de nova aplicação - Aferição de qualidade do processo de classificação			
E-M02.6		Início	M20	Fim	M24
	Título	- Classificação Semântica de outra Aplicação e/ou Classificação Hierárquica de Dados (Hierarquias de Classificação) - Aprimoramento de Soluções			
	Descrição	- Refinamentos dos módulo de classificação hierárquica ou nova aplicação; - Utilização de técnicas de aprendizado ativo para auxiliar os aprendizado de máquina; - Integração das soluções na Infra-estrutura do MPMG;			

	- Integração com outras aplicações finalísticas.
--	--

Entregáveis - 2022	
--------------------	--

M02.4	E-M02.11	Relatório contendo a caracterização das novas bases de dados ou hierarquias.
	M14	
M02.5	E-M02.12	Módulo de classificação (hierárquica); Relatório contendo avaliação experimental.
	M19	
M02.6	E-M02.13	Módulos Avançados de classificação (hierárquica); Impacto do Aprendizado ativo no processo; Relatório contendo avaliação experimental.
	M24	

Grupo de Trabalho	M03 - Classificação e Enriquecimento de Dados de Despesas Públicas		
Grupos de Interação	C01, D02, F02, F03, G01, M01, M02		
Coordenador	Gisele L Pappa		
Equipe	Nível	Quantidade	Nomes
	Professor	3	Gisele L Pappa, Anisio M Lacerda, Wagner Meira Jr
	Pós-Doc	0	TOTAL 10
	Doutorando	1	
	Mestrando	3	
	Graduando	3	
	Técnico	0	--

Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Sim	Não	Não	Sim	Não
Descrição Geral					
<p>Após a coleta de dados referentes a licitações, identificar que arquivo corresponde a qual documento da licitação é de extrema importância (ex: edital, ata, anexo, errata). Existe uma situação, em especial, que pode ser crítica. Licitações onde são publicadas erratas possuem dados divergentes nos bancos de dados do MP, que consideram apenas o texto original. Para o caso de geração de alertas de fraude, conhecer o texto final do edital pode ser de suma importância. Considerando um segundo exemplo, conhecer o valor dos lances dados por diferentes empresas pode ser crucial para algumas trilhas de auditoria. Esse mesmo problema pode ser encontrado em coletas de outras despesas públicas, onde tratamentos semelhantes podem ser necessários.</p>					
Objetivos					
<p>Os objetivos desse grupo de trabalho são:</p> <ol style="list-style-type: none"> 1. Dadas as coletas já realizadas, identificar o tipo de documento da licitação presente em um determinado arquivo, incluindo arquivo do edital de licitação, anexos, errata, ata, projeto, memorial descritivo, entre outros. Serão considerados os documentos coletados dos portais de transparência dos municípios (desde que estejam em arquivos separados) e dados do Lins (Tribunal de contas do Estado). 2. Identificar seções do site Fiscalizando o TCE de onde podemos coletar dados de lances oferecidos por empresa em licitações, para que sejam coletadas pelo C04. 3. Atualizar os dados dos arquivos de erratas das licitações nos bancos do MP. 4. Extrair dos dados coletados pelo C04 (e listados no objetivo 2 desse grupo de trabalho) as informações de valores de lances dados para cada um dos itens da licitação. 5. Interação com os grupos M02 e M05 para integração dos dados. Após coletados pelo C04, cada arquivo será rotulado pelo M03 e entregue junto com o seu respectivo rótulo para o M02. <p>Este, por sua vez, realizará a extração de entidades do arquivo de acordo com o rótulo do mesmo e encaminhará para o M05, responsável pela integração dos dados.</p> <ol style="list-style-type: none"> 6. Aplicação das etapas descritas anteriormente em dados de coleta de outros tipos de despesa pública diferentes de licitação. 					
Etapas					
E-M03.1		Início	M1	Fim	M4

	Título	Levantamento de requisitos para as tarefas de classificação de documentos e processamento de lances de licitações.			
	Risco vinculado	R-M03.1	R-M03.2	R-M03.3	
		R-M03.5	R-M03.7		
Tecnologias previstas		NIFI, Hive, HDFS, HBase, SQL, Python, Pyspark			
Tipos de dados necessários		Arquivos de coleta ligados a licitações (pdf, csv, html, doc, etc) Resultados de 2020 do grupo M02 Acesso ao banco de dados de licitações do MPMG Metadados dos respectivos bancos			
Descrição	Essa fase tem por objetivo identificar os tipos de documento da licitação de interesse para o MP, e uma primeira análise dos documentos coletados para avaliar se esses documentos estão realmente presentes na coleta. Serão considerados os documentos coletados dos portais de transparência dos municípios (desde que estejam em arquivos separados) e dados do Lins (Tribunal de contas do Estado). Nela, também analisaremos de onde podem ser coletados os lances no site do TCE, tarefa que será executada pelo C04.				
E-M03.2	Início	M1	Fim	M12	
	Título	Projeto e Implementação de Classificador de Documentos			
	Risco vinculado	R-M03.3	R-M03.5	R-M03.6	
Tecnologias previstas		NIFI, Hive, HDFS, HBase, SQL, Python, Pyspark			
Tipos de dados necessários		Arquivos de coleta ligados a licitações (pdf, csv, html, doc, etc) Amostra de arquivos rotulados Resultados do grupo M02			

Descrição	Essa fase tem por objetivo modelar e implementar um classificador para identificar automaticamente a qual documento da licitação ele se refere.				
E-M03.3	Início	M6	Fim	M12	
	Título	Atualização dos dados de licitações de acordo com erratas			
	Risco vinculado	R-M03.2	R-M03.3	R-M03.4	
		R-M03.6			
Tecnologias previstas	NIFI, Hive, HDFS, HBase, SQL, Python, Pyspark				
Tipos de dados necessários	Acesso ao banco de dados de licitações do MPMG Metadados dos respectivos bancos				
Descrição	Essa etapa focará na atualização dos dados extraídos dos arquivos de erratas das licitações nos banco do MP. Para realização dessa tarefa, precisamos do extrator de entidades criado pelo M02. A cobertura da atualização dos dados estará diretamente ligada a qualidade do extrator.				
E-M03.4	Início	M5	Fim	M12	
	Título	Processamento da coleta de lances em licitações			
	Risco vinculado	R-M03.2	R-M03.3	R-M03.4	
		R-M03.6	R-M03.7		
Tecnologias previstas	NIFI, Hive, HDFS, HBase, SQL, Python, Pyspark				
Tipos de dados necessários	Arquivos de coleta ligados a licitações (pdf, csv, html, doc, etc) Resultados do grupo M02 Acesso ao banco de dados de licitações do MPMG				

Metadados dos respectivos bancos		
Descrição	Extração das informações de valores de lances dados para cada um dos itens da licitação a partir dos dados coletados pelo C04 do portal do Fiscalizando o TCE.	
Entregáveis		
M03.1	E-M03.1	Relatório contendo análise de viabilidade, e levantamento das fontes de coleta para a tarefa de processamento de lances de licitação.
	M3	
M03.2	E-M03.1	Relatório contendo o levantamento de requisitos para a tarefa de classificação de documentos de licitação e caracterização inicial dos documentos coletados.
	M4	
M03.3	E-M03.2	Relatório contendo os resultados iniciais da tarefa de classificação de documentos.
	M7	
M03.4	E-M03.4	Relatório e código dos mecanismos de processamento da coleta de lances de licitação e inserção dos lances nas bases do MPMG.
	M10	
M03.5	E-M03.2	Códigos, documentação e relatório contendo os resultados da tarefa de classificação de documentos.
	M12	
M03.6	E-M03.3	Códigos, documentação e relatório sobre o mecanismo de correção/atualização de dados de licitações.
	M12	
Milestones		
M-M03.1	Alinhamento de responsabilidades com as equipes C04, M02 e M05.	
M1	Mapeamento inicial das fontes de coleta para a coleta de lances de licitação	
M-M03.2	Mapeamento inicial dos arquivos coletados no âmbito de licitações. Inclui as diferentes fontes dos arquivos, extensões, volumes, etc. Análise de necessidade de rotulação de documentos	
M2		

M-M03.3	Status das coletas dos lance de licitações
M5	Alinhamento com a equipe M05 para especificação e projeto das bases de dados que serão utilizadas para armazenar os lances de licitações Status das rotulações de documentos
M-M03.4	Status das coletas dos lance de licitações Status das implementações dos scripts de processamento e limpeza das coletas de lances de licitação Status das implementações dos atributos necessários às tarefas de classificação de documentos
M6	Status das implementações dos modelos de classificação de documentos Proposta da implementação do mecanismo de correção/atualização de dados de licitações. Necessidade de alinhamento com M02 para extração de entidades dos documentos e com o MPMG para processo de atualização das bases
M-M03.5	Status e resultados iniciais do processamento e limpeza das coletas de lances de licitação
M8	Status da implementação do mecanismo de atualização de dados
M-M03.6	Resultados preliminares e alinhamento dos ajustes necessários para finalização da tarefa de armazenamento dos lances de licitações
M9	Status da implementação do mecanismo de atualização de dados Status das implementações dos modelos de classificação de documentos Status das implementações dos atributos necessários às tarefas de classificação de documentos
M-M03.7	Status da implementação do mecanismo de atualização de dados
M11	Status das implementações dos modelos de classificação de documentos Status das implementações dos atributos necessários às tarefas de classificação de documentos

Riscos Críticos para a Implementação				
Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-M03.1	Dificuldade na atribuição de tarefas e responsabilidades com os grupos de interação.	Alto	Mitigar	Prolongar as discussões e realizar ajustes no cronograma e, se for o caso, no projeto.

R-M03.2	Alteração na estrutura do banco de dados de licitações (EDW).	Alto	Travar	Interação com as equipes do MP para entendimento do planejamento das alterações ou alterações do sistema de banco de dados do MP após o projeto de 2020. Só será possível iniciar qualquer implementação desse grupo de trabalho após este alinhamento.
R-M03.3	Saída de membros da equipe responsável.	Alto	Mitigar	Contratação de substitutos e/ou reatribuição de responsabilidades dentro da equipe.
R-M03.4	Mudança das tecnologias sendo utilizadas no desenvolvimento do projeto por parte do MP.	Alto	Mitigar	Readequar o projeto e treinar o pessoal nas tecnologias não dominadas. Pode atrasar o cronograma.
R-M03.5	Falta de consenso junto ao MPMG a respeito do contexto e dos arquivos a serem tratados na tarefa de classificação de documentos.	Alto	Travar	Encaminhar as decisões para as instâncias imediatamente superiores.
R-M03.6	Atraso, incompatibilidade e/ou falhas nos resultados provenientes de outros grupos de iteração (tanto do MPMG quanto da UFMG).	Médio	Transferir	Ajustes de cronograma, reuniões para discussão das incompatibilidades e/ou falhas, planejamento de correção dos problemas junto aos grupos responsáveis.
R-M03.7	Ausência e/ou Insuficiência de fontes de coleta ou das informações presentes nas fontes.	Alto	Travar	Levantamento de novas fontes de coleta junto aos parceiros do MPMG Ajustes nos objetivos da tarefa Ajustes de cronograma.

Cronograma das Entregas por Etapas

ETAPAS		ENTREGAS											
CÓD.	TÍTULO	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12

E-M03.1	Levantamento de requisitos para as tarefas de classificação de documentos e processamento de lances de licitações			M03.1	M03.2								
E-M03.2	Projeto e Implementação de Classificador de Documentos							M03.3					M03.5
E-M03.3	Atualização dos dados de licitações de acordo com erratas												M03.6
E-M03.4	Processamento da coleta de lances em licitações										M03.4		
MILESTONES		M-M03.1	M-M03.2			M-M03.3	M-M03.4		M-M03.5	M-M03.6		M-M03.7	

Etapas - 2022			
		Início	Fim
E-M03.11	Título	M13	M18
	Descrição	Refinamento das etapas de integração de dados de erratas e lances em licitações processados no ano de 2021	
		Dada a complexidade da tarefa de integração de dados é possível que exista a necessidade de revisão e refinamentos dos processos implementados no ano anterior.	

E-M03.12		Início	M13	Fim	M16
	Título	Levantamento de outras informações relevantes, para o contexto de despesas públicas presentes nas coletas realizadas.			
	Descrição	Informações relacionadas a despesas públicas podem ser encontradas em diferentes fontes como portais da transparência e Diários Oficiais. Essa etapa tem como objetivo realizar o levantamento das informações relevantes para o contexto de despesas públicas presentes nessas fontes e a viabilidade de extração das mesmas. Um exemplo são informações de aditamentos e ou suspensões de contratos presentes em Diários Oficiais (assumindo que estes sejam segmentados em 2021)			
Entregáveis - 2022					
E-M03.13		Início	M17	Fim	M24
	Título	Extração das informações levantadas na etapa E-M03.12			
	Descrição	Modelagem e implementação dos extratores para as informações levantadas, bem como interação com as equipes responsáveis pela integração dos dados extraídos.			
M03.11	E-M03.11	Relatório contendo mapeamento de informações relevantes para o contexto de despesas públicas presentes nas coletas realizadas, análise da viabilidade de extração das informações encontradas e mapeamento das atividades com os Grupos de Interação.			
	M16				
M03.12	E-M03.12	Relatório descrevendo os refinamentos executados para as etapas de integração de dados iniciadas em 2021			
	M18				
M03.13	E-M03.13	Relatório e código dos extratores de informação e dos fluxos necessários para a integração de dados			
	M24				

Grupo de Trabalho	M04 - Mineração de grafos como suporte para a construção de trilhas
--------------------------	--

Grupos de Interação	F01, F02, F04, G01 e outros sob demanda.				
Coordenador	Wagner Meira Jr.				
Equipe	Nível	Quantidade	Nomes		
	Professor	4	Wagner Meira, Renato Vimieiro, Loïc Cerf, Adriano Veloso		
	Pós-Doc	1	TOTAL	7	
	Doutorando	0			
	Mestrando	2			
	Graduando	0			
	Técnico	0	--		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Não	Sim	Sim	Sim	Não
Descrição Geral					
Esta etapa consiste na formulação de metodologias baseadas em mineração de padrões complexos em grafos para a construção de trilhas de auditoria. Ela terá forte interação com a etapa F02.					
Objetivos					
Projetar, implementar e validar algoritmos de mineração de grafos, assim como aplicá-los no suporte à construção de trilhas de auditoria.					
Etapas					
E-M04.1		Início	M1	Fim	M4
	Título	Levantamento de demandas por algoritmos de mineração de grafos			
	Risco vinculado	R-M04.1			
		R-M04.6			

Tecnologias previstas		Fractal, GraphFrames			
Tipos de dados necessários					
Descrição	Identificação de melhorias no sistema atual e levantamento de novas possibilidades.				
E-M04.2	Início	M5	Fim	M10	
	Título	Projeto e implementação dos algoritmos demandados			
	Risco vinculado	R-M04.2	R-M04.5		
		R-M04.4			
Tecnologias previstas		Fractal, GraphFrames			
Tipos de dados necessários					
Descrição	Implementação das propostas levantadas no E-M04.1.				
E-M04.3	Início	M11	Fim	M12	
	Título	Especificação e implantação de estudos de caso no âmbito de trilhas de auditoria			
	Risco vinculado	R-M04.3	R-M04.5		
		R-M04.4	R-M04.6		
Tecnologias previstas		Fractal, GraphFrames			
Tipos de dados necessários		Acesso aos dados de licitações e licitantes.			
Descrição	Utilização das propostas implementadas em E-M04.2 na construção de trilhas de auditoria. Estudo de caso com trilhas do F02.				

Entregáveis		
M04.1	E-M04.1	Relatório com descrição do processo/fluxo para implementação de um novo segmento de mineração de grafos: (1) modelagem do grafo → (2) algoritmo → (3) métricas de avaliação dos padrões. De acordo com este processo, serão mostrados resultados obtidos com o sistema corrente para avaliação qualitativa por especialistas de conteúdo.
	M2	
M04.2	E-M04.1	Relatório com o mapeamento de possíveis melhorias com relação ao sistema corrente, incluindo ajustes, novos algoritmos, modelagens de grafos e métricas de avaliação bem como a viabilidade dos mesmos.
	M4	
M04.3	E-M04.2	Relatório contendo a descrição detalhada dos refinamentos, novos algoritmos, modelagens e métricas selecionados como prioridade.
	M8	
M04.4	E-M04.2	Relatório contendo a descrição técnica do sistema implementado, incluindo instalação, especificação de requisitos, parametrização e execução dos novos algoritmos.
	M10	
M04.5	E-M04.3	Relatório com estudo de casos da aplicação dos novos algoritmos e métricas em trilhas determinadas pelos especialistas.
	M12	
Milestones		
M-M04.1		Descrição do processo/fluxo para implementação de um novo segmento de mineração de grafos: (1) modelagem do grafo → (2) algoritmo → (3) métricas de avaliação dos padrões. A ser submetido à equipe do MP para apreciação. Referência E-M04.1.
	M1	
M-M04.2		Levantamento de novas métricas de avaliação dos resultados obtidos pelo sistema corrente. Referência E-M04.1.
	M3	
M-M04.3		Identificação de ajustes necessários ao sistema corrente para adequação às métricas de avaliação. Referência E-M04.1.
	M3	
M-M04.4		Levantamento de dados, possíveis atores e relacionamentos a serem considerados na modelagem de trilhas de auditoria usando grafos. Referência E-M04.2.
	M5	

M-M04.5	Identificação de algoritmos para mineração de grafos, considerando a modelagem proposta anteriormente. Referência E-M04.1.
M6	
M-M04.6	Revisão teórica dos algoritmos identificados na etapa anterior. Referência E-M04.2.
M7	
M-M04.7	Levantamento dos requisitos de hardware e software dos algoritmos. Referência E-M04.2.
M7	
M-M04.8	Levantamento e análise de bibliotecas e/ou pacotes de software para implementação dos algoritmos. Verificação da adequação dos recursos disponíveis para implementar as funcionalidades dos algoritmos. Referência E-M04.2.
M8	
M-M04.9	Desenho do projeto. Detalhamento da implementação dos algoritmos em linguagem apropriada para garantir manutenibilidade e explicabilidade da implementação. Referência E-M04.2.
M8	
M-M04.10	Ajuste do projeto conforme sugestões da equipe do MP. Referência E-M04.2.
M9	
M-M04.11	Implementação dos algoritmos conforme projeto. Referência E-M04.2.
M9	
M-M04.12	Teste e validação das implementações. Referência E-M04.2.
M9	
M-M04.13	Instalação no ambiente do MP. Referência E-M04.2.
M10	
M-M04.14	Levantamento dos requisitos para o estudo de caso. Referência E-M04.3.
M11	

M-M04.15	Obtenção dos dados e modelagem usando grafos para utilização dos algoritmos implementados. Referência E-M04.3.
M11	
M-M04.16	Aplicação dos algoritmos e geração de resultados preliminares. Compilar métricas e montar relatório parcial para discussão interna com membros da equipe F02. Referência E-M04.3.
M11	

Riscos Críticos para a Implementação

Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-M04.1	Atraso da realização da análise qualitativa pelos especialistas de conteúdo.	Médio	Mitigar	Realizar parte das análises internamente. Trará impactos no tempo de entrega.
R-M04.2	Alteração na estrutura dos algoritmos de mineração de padrões para outros (ex: aprendizado de máquina).	Alto	Travar	Interação com as equipes do MP para entendimento do planejamento das alterações ou alterações do sistema corrente após o projeto de 2020. Só será possível iniciar qualquer implementação desse grupo de trabalho após este alinhamento.
R-M04.3	Qualidade dos dados nos bancos de dados de licitações.	Alto	Travar	Correção dos problemas por parte do MP.
R-M04.4	Saída de membros da equipe responsável.	Alto	Mitigar	Contratação de substitutos e/ou reatribuição de responsabilidades dentro da equipe. Treinamento dos novos contratados.
R-M04.5	Mudança das tecnologias sendo utilizadas no desenvolvimento do projeto por parte do MP.	Alto	Mitigar	Readequação do cronograma de entregas.
R-M04.6	Grafos fora da capacidade de processamento dos algoritmos.	Médio	Mitigar	Interação com as equipes do MP para tomada de decisões em relação à realização de parte das análises com grafos menores.

Cronograma das Entregas por Etapas

ETAPAS		ENTREGAS											
CÓD.	TÍTULO	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-M04.1	Levantamento de demandas por algoritmos de mineração de grafoS		M04.1		M04.2								
E-M04.2	Projeto e implementação dos algoritmos demandados.								M04.3		M04.4		
E-M04.3	Especificação e implantação de estudos de caso no âmbito de trilhas de auditoria.												M04.5
MILESTONES		M-M04.1		M-M04.2 M-M04.3		M-M04.4	M-M04.5	M-M04.6 M-M04.7	M-M04.8 M-M04.9	M-M04.10 M-M04.11 M-M04.12	M-M04.13	M-M04.14 M-M04.15 M-M04.16	

Etapas - 2022					
E-M04.11		Início	M13	Fim	M14
	Título	Avaliação dos resultados do estudo de caso			
	Descrição	Levantar deficiências e qualidades da metodologia empregada para construção de trilhas de auditoria baseadas em grafos.			
E-M04.12		Início	M14	Fim	M20
	Título	Especificação e implantação de novos algoritmos para mineração de grafos			
	Descrição				

	Descrição	Levantamento, projeto e implementação de novos algoritmos de mineração de grafos para construção de trilhas de auditoria baseadas em grafos. Avaliação de métricas de qualidade.			
E- M04.13		Início	M21	Fim	M22
	Título	Estudo de caso para construção de trilhas de auditoria através de mineração de grafos			
	Descrição	Aplicação dos métodos desenvolvidos na etapa anterior para construção de trilhas de auditoria (estudo de caso).			
Entregáveis - 2022					
M04.11	E- M04.11	Relatório descrevendo novas possibilidades de modelagem, métricas e algoritmos para construção de trilhas de auditoria baseadas em grafos.			
	M14				
M04.12	E- M04.12	Relatório detalhando o projeto e algoritmos implementados. Implementação e instalação do sistema no ambiente do MP.			
	M20				
M04.13	E- M04.13	Relatório com apresentação do estudo de caso, enfatizando a metodologia utilizada.			
	M22				

Grupo de Trabalho	M05 - Busca em Coleções Não-Estruturadas			
Grupos de Interação	C01, D02, E02, E06, F04, G01, M01, M03			
Coordenador	Rodrygo Santos			
Equipe	Nível	Quantidade	Nomes	
	Professor	1	Rodrygo Santos	
	Pós-Doc	0	TOTAL	7
	Doutorando	2		

	Mestrando	2			
	Graduando	2			
	Técnico	0	--		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Não	Sim	Não	Não	Não
Descrição Geral					
O projeto visa estender a plataforma de busca atualmente em desenvolvimento para possibilitar a exploração da estrutura subjacente ao esquema de cada fonte de dados bem como do feedback fornecido pelos usuários da busca.					
Objetivos					
<ol style="list-style-type: none"> 1. Aprimorar a estruturação dos dados coletados. 2. Aprimorar a eficiência e robustez do motor de busca, tanto para indexação quanto para recuperação de informação. 3. Aprimorar os mecanismos de avaliação da qualidade da busca, por meio da coleta de feedback implícito dos usuários-alvo. 4. Aprimorar os mecanismos de ranking implementados, por meio da exploração do feedback coletado. 5. Estender o motor de busca para novas modalidades de interação, como recomendação personalizada de novos conteúdos e busca conversacional para exploração das coleções indexadas. 					
Etapas					
E-M05.1		Início	M1	Fim	M3
	Título	Integração da plataforma atual			
	Risco vinculado	R-M05.1			
Tecnologias previstas		Áduna; Elasticsearch			
Tipos de dados necessários		Documentação da plataforma Áduna			
Descrição	Integração da plataforma de busca à plataforma Áduna.				

E-M05.2		Início	M4	Fim	M6
	Título	Suporte a filtragem semântica			
		Risco vinculado	R-M05.2		
Tecnologias previstas		Elasticsearch			
Tipos de dados necessários		Coleções indexadas; anotações semânticas			
Descrição	Exploração de estratégias para filtragem de documentos a partir da imposição de anotações semânticas (e.g., entidades nomeadas) ao conteúdo não-estruturado dos documentos.				
E-M05.3		Início	M7	Fim	M9
	Título	Suporte a modelos de ranking semântico			
		Risco vinculado	R-M05.2		
Tecnologias previstas		Elasticsearch			
Tipos de dados necessários		Coleções indexadas; anotações semânticas			
Descrição	Exploração de estratégias para representação semântica de documentos e consultas, seja de forma explícita, a partir da exploração das anotações semânticas disponíveis, ou implícita, a partir do aprendizado não supervisionado de representações de consultas e documentos.				

E-M05.4		Início	M10	Fim	M12
	Título	Suporte a filtragem personalizada			
	Risco vinculado	R-M05.1	R-M05.2		
Tecnologias previstas		Elasticsearch; extensões para user profiling			
Tipos de dados necessários		Coleções indexadas; anotações semânticas; logs de cliques			
Descrição	Exploração de modelos de filtragem, capazes de aprender o perfil de interesse de cada usuário (e.g., a partir de suas consultas históricas), a fim de prover um mecanismo de notificação automática sempre que novas informações relevantes estiverem disponíveis no sistema.				
E-M05.5		Início	M4	Fim	M12
	Título	Indexação de novas coleções			
	Risco vinculado	R-M05.3			
Tecnologias previstas		Elasticsearch			
Tipos de dados necessários		Coleções a serem indexadas			
Descrição	Indexação de novas coleções de documentos não-estruturados à medida que forem disponibilizadas.				
Entregáveis					
M05.1	E-M05.1	Versão da plataforma de busca operando sob a plataforma Áduna.			
	M3				
M05.2	E-M05.2	Versão da plataforma de busca com suporte à filtragem por entidades de interesse.			

	M6	
M05.3	E-M05.3	Versão da plataforma de busca com suporte a ranking semântico.
	M9	
M05.4	E-M05.4	Versão da plataforma de busca com suporte a filtragem personalizada.
	M12	
M05.5	E-M05.5	Índices atualizados.
	M12	
Milestones		
M-M05.1	Relatório de progresso da etapa E-M05.1.	
M1		
M-M05.2	Relatório de progresso da etapa E-M05.1.	
M2		
M-M05.3	Relatório de progresso da etapa E-M05.2.	
M4		
M-M05.4	Relatório de progresso da etapa E-M05.2.	
M5		
M-M05.5	Relatório de progresso da etapa E-M05.3.	
M7		
M-M05.6	Relatório de progresso da etapa E-M05.3.	
M8		

M-M05.7	Relatório de progresso da etapa E-M05.4.
M10	
M-M05.8	Relatório de progresso da etapa E-M05.4.
M11	

Riscos Críticos para a Implementação

Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-M05.1	Não-adoção efetiva da plataforma por parte do MPMG.	Alto	Mitigar	Ações de treinamento; envolvimento do cliente.
R-M05.2	Ineficácia ou ineficiência dos módulos de anotação semântica.	Médio	Mitigar	Alinhamento com os grupos responsáveis pelos módulos de anotação semântica (e.g., reconhecimento de entidades); implementação de mecanismos robustos para ponderação adequada da evidência semântica disponível como critério de ranking; implementação de mecanismos para representação de semântica latente.
R-M05.3	Indisponibilidade de novas coleções.	Baixo	Travar	
R-M05.4	Ineficácia ou ineficiência do módulo de interação conversacional.	Baixo	Mitigar	Investigação do estado-da-arte para o problema de busca conversacional; utilização de interfaces tradicionais como contingência.

Cronograma das Entregas por Etapas

Etapas		Entregas											
Cód.	Título	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-M05.1	Integração da plataforma			M05.1									

	atual												
E-M05.2	Suporte a filtragem semântica						M05.2						
E-M05.3	Suporte a modelos de ranking semântico									M05.3			
E-M05.4	Suporte a filtragem personalizada												M05.4
E-M05.5	Indexação de novas coleções												M05.5
MILESTONES		M-M05.1	M-M05.2		M-M05.3	M-M05.4		M-M05.5	M-M05.6		M-M05.7	M-M05.8	

Etapas - 2022					
E-M05.11		Início	M13	Fim	M18
	Título	Suporte a modelos de ranking supervisionados			
	Descrição	Exploração do feedback dos usuários, na forma de cliques, para o desenvolvimento de modelos de aprendizado de máquina supervisionado ou mesmo de aprendizado por reforço para ranking.			
E-M05.12		Início	M19	Fim	M24
	Título	Suporte a busca conversacional			

	Descrição	Pesquisa e o desenvolvimento de interfaces interativas para a busca (e.g., chatbots), capazes de reter uma memória da interação do usuário com o sistema na sessão de busca atual e em sessões anteriores, e elicitare a necessidade de informação do usuário na sessão atual por meio de mecanismos conversacionais.		
E-M05.13	Início	M13	Fim	M24
	Título	Indexação de novas coleções		
	Descrição	Indexação de novas coleções de documentos não-estruturados à medida que forem disponibilizadas.		
Entregáveis - 2022				
M05.11	E-M05.11	Versão da plataforma de busca com suporte a ranking supervisionado.		
	M18			
M05.12	E-M05.12	Versão da plataforma de busca com suporte a interação conversacional.		
	M24			

P. Plataforma Tecnológica

Grupo de Trabalho	P01 - Evolução da Plataforma Lemonade			
Grupos de Interação	A01, A02, D01, D02, F02, F03, G01, M01, M02, P02			
Coordenador	Wagner Meira Junior			
Equipe	Nível	Quantidade	Nomes	
	Professor	4	Wagner Meira Junior, Renato Ferreira, George Teodoro, Dorgival Guedes	
	Pós-Doc	0	TOTAL	8

	Doutorando	0			
	Mestrando	0			
	Graduando	2			
	Técnico	2	--		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Não	Não	Sim	Sim	Não
Descrição Geral					
A Plataforma Lemonade é utilizada para a criação de fluxos de trabalho e trilhas de auditoria usando abstrações que simplificam essa atividade. Além de facilitar a criação de fluxos e trilhas sem a necessidade de conhecimento sobre programação, a plataforma Lemonade também abstrai diversas integrações implementadas no projeto em 2020, incluindo integração com a parte de segurança, dados e infraestrutura. Para este projeto, o foco principal será na disseminação do uso da plataforma, com efetiva adoção no MPMG. Também são consideradas atividades de manutenção e evolução.					
Objetivos					
<ol style="list-style-type: none"> 1. Facilitar e acelerar a criação de experimentos, fluxos e trilhas de auditoria por parte do MPMG; 2. Permitir a utilização dos recursos de infraestrutura, dados e outras tecnologias de forma integrada e transparente para o usuário final, simplificando a manutenção por parte da equipe de TI e da GSI; 3. Servir como ferramenta de apoio para o aprendizado dos conceitos associados à ciência dos dados; 4. Servir como plataforma de ensaio para experimentos sobre os dados, realizados por diferentes perfis de usuários no MPMG, expandindo a possibilidade de descoberta de informação a partir dos dados do <i>data lake</i>. 					
Etapas					
E-P01.1		Início	M2	Fim	M12
	Título	Capacitação para adoção e uso da Plataforma Lemonade			
	Risco vinculado	R-P01.1	R-P01.4	R-P01.5	
Tecnologias previstas		Plataforma Lemonade instalada em ambiente de treinamento; Ambiente de EaD para treinamentos; Sistema de Help-Desk (a definir) para suporte de 2o. nível.			

Tipos de dados necessários		Amostra dos dados mais relevantes de acordo com o MPMG para a capacitação dos usuários.			
Descrição		Iniciativas para a adoção da Plataforma Lemonade (visão do Especialista). A UFMG já possui material desenvolvido para treinamento em ciência dos dados utilizando o Lemonade como plataforma de apoio. A ideia é que o MPMG designe pessoas que serão capacitadas a utilizar e manter a ferramenta e depois possam difundi-la internamente.			
E-P01.2		Início	M1	Fim	M12
	Título	Integração do Lemonade com tecnologias e infraestrutura.			
	Risco vinculado	R-P01.2	R-P01.6		
Tecnologias previstas		PosgreSQL; Ambari; Apache Atlas; Apache Druid			
Tipos de dados necessários		Detalhes sobre os parâmetros para uso das tecnologias no ambiente do MPMG.			
Descrição		Atividades de integração com a infraestrutura e tecnologias existentes e/ou que serão usadas pelo MPMG. Entre as possíveis tecnologias candidatas, estão o PostgreSQL (para migração da base de dados interna do Lemonade), Apache Druid (para dados OLAP), Apache Ambari (criação de clusters Spark/HDFS) e Apache Atlas (linhagem de dados integrada ao Lemonade).			
E-P01.3		Início	M1	Fim	M9
	Título	Adaptação do Lemonade para manutenção de informações sobre linhagem de artefatos e avaliação de qualidade de dados			
	Risco vinculado	R-P01.3	R-P01.4	R-P01.6	
Tecnologias previstas		Apache Atlas, Amazon Deequ			
Tipos de dados necessários		Dados do <i>data lake</i> , dados pós-processados.			

Descrição	Expansão da Plataforma para manutenção de linhagem e avaliação da qualidade dos dados, linhagem de modelos, fluxos, trilhas e relatórios. Criação de APIs para exportar as informações e produção de relatórios.				
E-P01.4		Início	M6	Fim	M11
	Título	Inclusão de algoritmos para análise sobre interpretabilidade e justiça para algoritmos			
	Risco vinculado	R-P01.5			
Tecnologias previstas	Apache Spark, Scikit-Learn, SHAP, TensorBoard, Google What-If-Tool				
Tipos de dados necessários	Dados da execução das trilhas, questionários de usuários.				
Descrição	Expansão da Plataforma para incluir análises sobre interpretabilidade e justiça para algoritmos. Estes 2 tópicos são relevantes para o MPMG por trazerem outras métricas para o resultado dos algoritmos que não simplesmente a acurácia. No caso de transparência, a questão é poder explicar para as pessoas as decisões que foram tomadas pelos algoritmos. Já para justiça, evitar que dados viesados causem algum tipo de discriminação de indivíduos por conta de algum atributo sensível (raça, cor, idade, religião, sexo, etc).				
E-P01.5		Início	M8	Fim	M12
	Título	Suporte à implantação de web services a partir de fluxos			
	Risco vinculado	R-P01.7			
Tecnologias previstas	Lemonade, Kubernetes, Ambari				
Tipos de dados necessários	---				
Descrição	Um dos possíveis objetivos da construção de modelos de aprendizado de máquina é a disponibilização de modelos para serem utilizados por outras aplicações e sistemas. O objetivo desta etapa é ter, no Lemonade, a possibilidade de implantar um fluxo de				

trabalho em um *web service*, que poderá ser consumido através de uma API.

Entregáveis

P01.1	E-P01.1	Plataforma Lemonade instalada para treinamento em infraestrutura fornecida pelo MPMG; material de estudo sobre a Plataforma, configurado em ambiente de EaD.
	M3	
P01.2	E-P01.2	Versão do Lemonade com suporte ao banco de dados PostgreSQL.
	M4	
P01.3	E-P01.1	Relatório sobre uso da Plataforma por parte do MPMG, detalhando métricas relacionadas aos usuários, trilhas, execuções e suporte.
	M5	
P01.4	E-P01.1	Versão do Lemonade com melhorias na usabilidade a partir das sugestões e resultados obtidos na capacitação
	M6	
P01.5	E-P01.2	Versão do Lemonade com suporte ao banco de dados Druid.
	M6	
P01.6	E-P01.2	Versão do Lemonade integrada com outras tecnologias a serem elencadas durante os primeiros meses do projeto.
	M9	
P01.7	E-P01.3	Versão do Lemonade com suporte à análise da qualidade dos dados (Lemonade Data Quality)
	M9	
P01.8	E-P01.4	Versão do Lemonade com suporte às ferramentas de transparência e justiça para algoritmos.
	M11	
P01.9	E-P01.1	Relatório final sobre o uso do Lemonade.
	M12	

P01.10	E-P01.2	Versão 2.0 do Lemonade.
	M12	
Milestones		
M-P01.1	Avaliação do impacto da migração da base de dados do Lemonade para o banco PostgreSQL.	
M1		
M-P01.2	Definição sobre as atividades de capacitação no MPMG; Apoio dos gestores do MPMG à iniciativa.	
M2		
M-P01.3	Recursos para a capacitação disponibilizados; envolvidos elencados.	
M3		
M-P01.4	Implantação de versão.	
M4		
M-P01.5	Relatório inicial sobre uso do Lemonade.	
M5		
M-P01.6	Implantação de versão.	
M6		
M-P01.7	Definição de técnicas, ferramentas e algoritmos para interpretabilidade e justiça	
M7		
M-P01.8	Fechamento das tecnologias a serem integradas ao Lemonade.	
M8		
M-P01.9	Implantação de versão.	

M9	
M-P01.10	Fechamento de escopo da versão 2.0
M10	
M-P01.11	Fechamento de bugs críticos.
M11	
M-P01.12	Implantação de versão; relatório de uso. Ter alguma trilha implantada como serviço web.
M12	

Riscos Críticos para a Implementação

Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-P01.1	Não-adoção efetiva da Plataforma por parte do MPMG.	Alto	Mitigar	Ações de treinamento; acompanhamento dos usuários por meio de suporte de 2o. nível; envolvimento do cliente.
R-P01.2	Alteração de tecnologias usadas na infraestrutura.	Médio	Mitigar	Realização de provas de conceito; apoio de empresa parceira do MPMG.
R-P01.3	Dados usados nos fluxos e trilhas são de baixa qualidade.	Baixo	Mitigar	Definir plano de qualidade de dados; estender o Lemonade para facilitar a análise qualitativa.
R-P01.4	Recursos da infraestrutura incapazes de atender à demanda para o processamento.	Alto	Transferir	Apresentar dados quantitativos sobre uso da infraestrutura; avaliar a expansão dos recursos.
R-P01.5	Dificuldade na análise dos resultados da plataforma.	Médio	Mitigar	Capacitação; expansão para interpretabilidade.
R-P01.6	Dependência da adoção do Apache Atlas pelo MPMG.	Médio	Travar	Acompanhar a evolução da adoção do Atlas; rever necessidade de integração caso haja mudança de direção.
R-P01.7	Indefinição sobre solução para criação de web services sob demanda (Kubernetes).	Médio	Mitigar	Propor o uso de Kubernetes ou avaliar o uso de ferramentas como o Ambari.

Cronograma das Entregas por Etapas

Etapas		Entregas											
Cód.	Título	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E- P01.1	Capacitação para adoção e uso da Plataforma Lemonade			P01.1		P01.3	P01.4						P01.9
E- P01.2	Integração do Lemonade com tecnologias e infraestrutura.				P01.2		P01.5			P01.6			P01.10
E- P01.3	Adaptação do Lemonade para manutenção de informações sobre linhagem de artefatos e avaliação de qualidade de dados.									P01.7			
E- P01.4	Inclusão de algoritmos para análise sobre interpretabilidade e justiça para algoritmos.											P01.8	
E- P01.5	Suporte à implantação de web services a partir de fluxos.												
MILESTONES		M-P01.1	M-P01.2	M-P01.3	M-P01.4	M-P01.5	M-P01.6	M-P01.7	M-P01.8	M-P01.9	M-P01.10	M-P01.11	M-P01.12

Etapas - 2022

E- P01.6	Início	M13	Fim	M24
	Título	Integração do Lemonade com tecnologias e infraestrutura		
	Descrição	Atividades de integração com a infraestrutura e tecnologias existentes e/ou que serão usadas pelo MPMG. Entre as possíveis tecnologias candidatas, estão o ElasticSearch, Keras + TensorFlow, Apache Kafka, entre outros.		
E- P01.7	Início	M13	Fim	M16
	Título	Identificação de requisitos para melhoria da Plataforma Lemonade		
	Descrição	Identificar novos requisitos para o processamento de dados através do Lemonade.		
E- P01.8	Início	M15	Fim	M22
	Título	Expansão da Plataforma Lemonade para processamento de dados geoespaciais.		
	Descrição	Inclusão de operações e algoritmos de geoprocessamento. Possíveis candidatos incluem operações para normalização e conversão de coordenadas (pontos) e polígonos, operação para localização de ponto em polígono, interseção e sobreposição de polígonos, junção (join) baseado em coordenadas geográficas, agrupamento espacial com ST-DBScan. Suporte a formatos como GeoJSON, OSM-XML e outros. Criação de índices geoespaciais (ZOrderCurves). Funções para computar distâncias, centroids, reduzir precisão e outras. Tecnologia de referência: Apache Sedona (em incubação).		
E- P01.9	Início	M16	Fim	M24
	Título	Expansão da Plataforma Lemonade para processamento de grafos.		
	Descrição	Definição de principais operações relacionadas a grafos (redes), tais como conversão de dados, execução de algoritmos (PageRank, componentes conectados, label propagation, SVD++, componentes fortemente conectados e contagem de triângulos). A tecnologia de referência é a biblioteca Spark GraphX. Requer a colaboração com os grupos de visualização e de processamento de grafos para propostas de novas visualizações e algoritmos.		

Entregáveis - 2022		
P01.11	E-P01.11	Versão do Lemonade com novas tecnologias integradas.
	M18	
P01.12	E-P01.12	Especificação de requisitos para as melhorias no Lemonade
	M16	
P01.13	E-P01.13	Versão do Lemonade com implementação das melhorias identificadas na especificação de requisitos.
	M20	
P01.14	E-P01.14	Versão do Lemonade com suporte a processamento de dados geoespaciais.
	M22	
P01.15	E-P01.15	Versão do Lemonade com suporte a processamento de grafos.
	M24	

Grupo de Trabalho	P02 - Evolução da Plataforma de Processamento de Trilhas		
Grupos de Interação	A01, A02, D01, D02, F02, F03, G01, M01, M02, P01		
Coordenador	Wagner Meira Junior		
Equipe	Nível	Quantidade	Nomes
	Professor	3	Renato Ferreira, George Teodoro, Dorgival Guedes
	Pós-Doc	0	TOTAL 5
	Doutorando	0	
	Mestrando	1	

	Graduando	1			
	Técnico	0	--		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Não	Não	Sim	Sim	Não
Descrição Geral					
<p>A Plataforma de Processamento de Trilhas permite que diferentes tipos de usuários interajam com fluxos de trabalho definidos no Lemonade. Neste caso, os fluxos são chamados de trilhas. Usuários <i>Especialistas</i> podem construir, experimentar, testar e disponibilizar trilhas parametrizáveis, usando as diferentes fontes de dados disponíveis no MPMG. Usuários finais usam as trilhas, especificando parâmetros para execução, como por exemplo, filtros e ordenação, além de analisar os resultados gerados pelo processamento das trilhas. Tais resultados podem ser desde novas tabelas sendo criadas, até visualizações mais elaboradas. Neste projeto, a Plataforma de Processamento de Trilhas terá avanços principalmente na parte de construção facilitada do fluxo e na facilidade na interpretação, análise e julgamento dos resultados por parte dos usuários. Atividades para a disseminação do uso e adoção pelo MPMG, bem como capacitação de usuários, são consideradas entre os objetivos.</p>					
Objetivos					
<p>1. Fornecer uma solução para a construção e utilização de trilhas de auditorias que seja simples do ponto de vista de aprendizado, mas que seja capaz de processar grandes volumes de dados através do uso de fluxos de trabalho. 2. Atender a diferentes perfis de uso: procuradores de justiça que usam as trilhas para obter informações em processos e especialistas técnicos que irão construir as trilhas usando recursos de aprendizado de máquina e a infraestrutura de processamento de dados. 3. Fornecer funções para análise dos dados sob o ponto de vista de novos conceitos quanto ao uso de algoritmos de aprendizado de máquina, tais como justiça algorítmica (evitar vieses de natureza discriminatória) e interpretabilidade dos resultados (entendimento dos passos que levaram à tomada de decisão pelo algoritmo).</p>					
Etapas					
E-P02.1		Início	M4	Fim	M11
		Título	Capacitação para adoção e uso da Plataforma de Execução de Trilhas		
		Risco vinculado	R-P02.1		
Tecnologias previstas		Plataforma Lemonade instalada em ambiente de treinamento; Ambiente de EaD para treinamentos; Sistema de Help-Desk (a definir) para suporte de 2o. nível.			
Tipos de dados necessários		Amostra dos dados mais relevantes de acordo com o MPMG para a capacitação dos usuários.			

Descrição	Realização de ações para capacitação de usuários e adoção da Plataforma de Processamento de Trilhas (visão do usuário final). Nesta etapa, o objetivo é capacitar o usuário final e também avaliar a Plataforma de Execução de Trilhas. Diferente da etapa de capacitação existente em P01, esta etapa é menos abrangente conceitualmente, focando principalmente na facilidade de uso/usabilidade, utilização e interpretação dos resultados.				
		Início	M2	Fim	M9
E-P02.2	Título	Avaliação de ferramentas de visualização e exploração de dados			
	Risco vinculado	R-P02.2			
Tecnologias previstas	Apache Superset, Metabase				
Tipos de dados necessários	Amostra dos dados, de diferentes tipos, usados no MPMG.				
Descrição	Avaliação de tecnologias para visualização e exploração de dados com Apache Superset e Metabase, com possíveis adaptações para autenticação e autorização, integração com fontes de dados do Lemonade e tradução da interface (partes ausentes). O resultado desta etapa será usado para determinar se compensa a adoção da ferramenta pelo MPMG, se é possível sua integração com o Lemonade e a parte de processamento de trilhas. Esta etapa é complementar à apresentada em E-F01.5 (avaliativa).				
		Início	M2	Fim	M10
E-P02.3	Título	Interface para exploração de dados no Lemonade			
	Risco vinculado	R-P02.3	R-P02.1		
Tecnologias previstas	Lemonade, Dataiku (inspiração)				
Tipos de dados necessários	Amostra dos dados, de diferentes tipos, usados no MPMG.				

Descrição	Esta nova interface para análise e exploração sobre qualidade dos dados e geração de trilhas, chamada de Lemonade Data Explorer, tem como objetivo simplificar a construção de trilhas. A ideia é que seja construída uma interface de usuário onde ele comece a partir das tabelas que quer analisar. A partir daí, o usuário poderá transformar os dados de uma maneira que remete, em partes, ao que ele faria em uma planilha de dados. Em seguida ele poderia criar novas bases e/ou visualizações para analisar os resultados. Todos os passos seriam registrados pela interface e ao fim, permitiria que o usuário salvasse-os na forma de um fluxo de trabalho ou mesmo uma trilha. Consideramos que a ferramenta Dataiku (comercial) servirá como inspiração para o desenvolvimento.				
E-P02.4	Início	M6	Fim	M10	
	Título	Nova interface para análise e exploração dos dados (Lemonade Notebook)			
	Risco vinculado	R-P02.1	R-P02.3		
Tecnologias previstas	Jupyter, Zeppelin (serão usados como inspiração)				
Tipos de dados necessários					
Descrição	Notebooks são ferramentas para análise, exploração e visualização de dados. Entre os principais produtos, encontram-se o Apache Zeppelin e o Jupyter Notebook. O Lemonade Notebook se propõe a ser diferente das ferramentas existentes por não requerer que o usuário programe, mas sim que parametrize as várias células que compõem o <i>notebook</i> .				
E-P02.5	Início	M6	Fim	M7	
	Título	Avaliação do uso de AutoML para a criação de trilhas			
	Risco vinculado	R-P02.4			
Tecnologias previstas	Databrick automl-toolkit				
Tipos de dados necessários	Diversos				

Descrição	O aprendizado de máquina automatizado, ou AutoML é o processo de automatizar o processo de aprendizado de máquina a partir das entradas e das saídas esperadas. Em tese, AutoML permitiria que o cientista de dados construísse modelos de aprendizado de máquina mais rapidamente. Tarefas como o tratamento dos dados (<i>feature engineering</i>), escolha de parâmetros e algoritmos e validações cruzadas seriam automatizados.				
E-P02.6	Início	M8	Fim	M12	
	Título	Aplicação de AutoML para a criação de trilhas			
	Risco vinculado	R-P02.4			
Tecnologias previstas	Databrick automl-toolkit				
Tipos de dados necessários	Diversos				
Descrição	Esta etapa depende do resultado de E-P02.5. Seu objetivo é prover uma interface no Lemonade para que um usuário especialista possa criar rapidamente trilhas usando AutoML.				
E-P02.7	Início	M1	Fim	M12	
	Título	Incorporação de resultados à Plataforma de Processamento de Trilhas			
	Risco vinculado	R-P02.5	R-P02.6		
Tecnologias previstas	Várias				
Tipos de dados necessários	Diversos				
Descrição	Nesta etapa, o objetivo é incorporação dos novos resultados das atividades dos demais grupos (contínuo) à Plataforma de Processamento de Trilhas. Para que essa incorporação seja feita da maneira mais simples e eficiente, poderão ser feitas reuniões entre os grupos para alinhar as estratégias, algoritmos e tecnologias a serem utilizadas.				

Entregáveis		
P02.1	E-P02.2	Especificação de requisitos para adaptações do Apache Superset
	M2	
P02.2	E-P02.3	Especificação de casos de uso do Lemonade Data Explorer
	M2	
P02.3	E-P02.2	Relatório sobre a avaliação das ferramentas para visualização e exploração de dados.
	M3	
P02.4	E-P02.3	Primeira versão da interface do Lemonade Data Explorer
	M4	
P02.5	E-P02.1	Plataforma Lemonade instalada para treinamento em infraestrutura fornecida pelo MPMG. Módulo I de capacitação de usuários definido.
	M6	
P02.6	E-P02.5	Relatório de avaliação sobre o uso de AutoML no Lemonade para uso no MPMG.
	M7	
P02.7	E-P02.2	Adaptações da parte de segurança e integração com o Lemonade para o uso da ferramenta de visualização e exploração de dados no MPMG.
	M8	
P02.8	E-P02.3	Especificação de casos de uso do Lemonade Notebook, revisão da Especificação do Lemonade Data Explorer
	M8	
P02.9	E-P02.2	Versão final do Apache Superset com as adaptações para o MPMG
	M9	
P02.10	E-P02.3	Versão final da interface do Lemonade Data Explorer e Lemonade Notebook

	M10	
P02.11	E-P02.1	Plataforma Lemonade instalada para treinamento em infraestrutura fornecida pelo MPMG. Módulo II de capacitação de usuários definido, incorporando as novas ferramentas voltadas para esse público (possivelmente, Apache Superset e Lemonade Data Explorer).
	M11	
Milestones		
M-P02.1	Implementação de trilhas do projeto anterior instaladas.	
M1		
M-P02.2	Definição dos objetivos do uso do Apache Superset, Lemonade Data Explorer, Lemonade Notebook	
M2		
M-P02.3	Apache Superset instalado e configurado.	
M3		
M-P02.4	Versão instalada do Lemonade + Lemonade Data Explorer	
M4		
M-P02.5	Revisão de especificações e requisitos	
M5		
M-P02.6	Apache Superset usando autenticação integrada por LDAP e comunicando-se com o Lemonade.	
M6		
M-P02.7	Avaliação e decisão sobre uso de AutoML.	
M7		
M-P02.8	Revisão de especificações e requisitos	
M8		

M-P02.9	Reunião de definição de integração com os demais grupos
M9	
M-P02.10	Versão instalada do Lemonade + Lemonade Data Explorer e Lemonade Notebook (versão final)
M10	
M-P02.11	Produtos instalados
M11	
M-P02.12	Reunião de fechamento
M12	

Riscos Críticos para a Implementação

Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-P02.1	Não-adoção efetiva da Plataforma por parte do MPMG.	Alto	Mitigar	Ações de treinamento; acompanhamento dos usuários por meio de suporte de 2o. nível; envolvimento do cliente.
R-P02.2	Incompatibilidade entre as tecnologias e o Apache Superset.	Alto	Aceitar	Avaliar alternativas para a integração, definir claramente os objetivos da integração e critérios para aceitação.
R-P02.3	Complexidade da proposta de novas interfaces para exploração de dados.	Alto	Mitigar	Limitar o escopo das funcionalidades, incorporando prioritariamente aquelas que trazem maior facilidade de uso.
R-P02.4	Tecnologia não se mostrar útil ou com resultados ruins.	Médio	Aceitar	Rever a adoção. Caso negativo, realocar esforços.
R-P02.5	Atrasos na execução das atividades dos outros	Médio	Mitigar	Tentar gerenciar através de reuniões de acompanhamento, estabelecendo responsabilidades e prazos.

	grupos.			
R-P02.6	Incompatibilidade das tecnologias desenvolvidas pelos outros grupos e a Plataforma.	Médio	Mitigar	Avaliar possibilidade de adaptação; coordenar esforços para alinhar as tecnologias com os demais grupos.

Cronograma das Entregas por Etapas

Etapas		Entregas											
Cód.	Título	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-P02.1	Capacitação para adoção e uso da Plataforma de Execução de Trilhas						P02.5					P02.11	
E-P02.2	Avaliação de ferramentas de visualização e exploração de dados		P02.1	P02.3					P02.7	P02.9			
E-P02.3	Interface para exploração de dados no Lemonade		P02.2		P02.4				P02.8		P02.10		
E-P02.4	Nova interface para análise e exploração dos dados (Lemonade Notebook)												
E-P02.5	Avaliação do uso de AutoML para a criação de trilhas							P02.6					
E-P02.6	Aplicação de AutoML para a criação de trilhas												

E-P02.7	Incorporação de resultados à Plataforma de Processamento de Trilhas												
MILESTONES		M-P02.1	M-P02.2	M-P02.3	M-P02.4	M-P02.5	M-P02.6	M-P02.7	M-P02.8	M-P02.9	M-P02.10	M-P02.11	M-P02.12

Etapas - 2022				
E-P02.8	Início	M13	Fim	M24
	Título	Incorporação de resultados à Plataforma de Processamento de Trilhas		
	Descrição	Nesta etapa, o objetivo é incorporação dos novos resultados das atividades dos demais grupos (contínuo) à Plataforma de Processamento de Trilhas. Para que essa incorporação seja feita da maneira mais simples e eficiente, poderão ser feitas reuniões entre os grupos para alinhar as estratégias, algoritmos e tecnologias a serem utilizadas.		
E-P02.9	Início	M13	Fim	M20
	Título	Melhorias nas interfaces de usuário propostas para o ano de 2021 (Lemonade Data Explorer, Lemonade Notebook e Apache Superset).		
	Descrição	Dada a variedade de interfaces que estão sendo propostas para 2021, espera-se que correções e melhorias sejam necessárias. A partir do retorno dos usuários em relação ao uso da Plataforma de Trilhas, pretende-se corrigir as falhas identificadas e melhorar a experiência de usuário. Também serão feitas ações para adoção da ferramenta por parte dos usuário.		
E-P02.10	Início	M13	Fim	M24
	Título	Adaptação da Plataforma de Trilhas para que usuário envie suas próprias bases de dados.		

	Descrição	Esta etapa requer a colaboração com diversos grupos de trabalho para determinar como o usuário poderá enviar e gerenciar suas próprias bases de dados. Aspectos como onde esses dados serão armazenados (landing zone), como a segurança será garantida, como os dados serão armazenados e a linhagem mantida são alguns dos pontos que deverão ser discutidos nesta tarefa. A expectativa é que com a construção de uma interface mais intuitiva (Lemonade Data Explorer), que, guardadas as proporções, se assimilaria, em termos de uso a uma planilha, permita que os próprios usuários executassem várias ações para transformação dos dados e pudessem contar com uma infraestrutura robusta para tal fim.
Entregáveis - 2022		
P02.12	E-P02.8	Versão do Lemonade com suporte às implementações provenientes das atividades dos demais grupos (aquelas que estejam prontas).
	M18	
P02.13	E-P02.8	Segunda versão do Lemonade com suporte às implementações provenientes das atividades dos demais grupos (aquelas que estejam prontas).
	M24	
P02.14	E-P02.9	Versão do Lemonade com as melhorias na Plataforma de Trilhas. Análise de adoção pelo MPMG.
	M22	
P02.15	E-P02.10	Versão da Plataforma com suporte ao envio de dados pelos usuários.
	M18	

Grupo de Trabalho	P03 - Plataforma Adaptativa Inteligente para Coleta de Dados e Construção de Bases de Conhecimento		
Grupos de Interação	A01, A02, G01, F06		
Coordenador	Wagner Meira Jr.		
Equipe	Nível	Quantidade	Nomes
	Professor	5	Wagner Meira Jr., Ana Paula Couto da Silva, Marcos Gonçalves, Adriana Pagano, Rodrygo Santos
	Pós-Doc	4	TOTAL 17

	Doutorando	2			
	Mestrando	2			
	Graduando	4			
	Técnico	0	--		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Não	Não	Não	Não	Não
Descrição Geral					
Esta projeto visa desenvolver uma plataforma computacional de coleta e integração de dados de variadas fontes em bases de conhecimento, possibilitando sua utilização para implementar um chatbot dinâmico.					
Objetivos					
<p>O principal objetivo deste projeto é desenvolver uma plataforma de coleta e integração de dados de variadas fontes em bases de conhecimento enriquecidas de forma contínua, visando implementar chatbots dinâmicos. A seguir desmembramos nosso projeto em três objetivos que constituem passos necessários para atingirmos nosso objetivo final.</p> <p>1) Desenvolvimento de uma plataforma adaptativa de coleta, enriquecimento e engenharia de dados que considerem dados de diferentes tipos (sociais, econômicos, demográficos, geográficos, conhecimentos em linguística,) e provenientes de fontes heterogêneas de uma plataforma de coleta de dados de diferentes tipos (sociais, econômicos, demográficos, geográficos) e diferentes fontes (web, redes sociais online, IBGE, SUS, SMEs).</p> <p>2) Construção de uma base de conhecimento que integra, de forma efetiva e eficiente, várias fontes de dados heterogêneas, multimodais, históricas, garantindo a qualidade e a proveniência dos dados integrados de forma a tornar o processo de integração transparente, replicável, verificável, auditável e explicável.</p> <p>3) Desenvolvimento de um chatbot adaptativo, que personaliza experiência das pessoas a partir da determinação do seu perfil e segmentação da base de usuários, ao mesmo tempo que refletem alterações contextuais.</p>					
Etapas					
E-P03.1		Início	M1	Fim	M3
	Título	Detalhamento da Plataforma Adaptativa			

		Risco vinculado	R-P03.1		
Tecnologias previstas		--			
Tipos de dados necessários		--			
Descrição	Levantamento de requisitos para o desenvolvimento da plataforma				
		Início	M4	Fim	M8
E-P03.2	Título	Plataforma Adaptativa 1.0			
	Risco vinculado	R-P03.1			
		R-P03.2			
Tecnologias previstas		--			
Tipos de dados necessários		--			
Descrição	Versão 1.0 da plataforma adaptativa				
		Início	M9	Fim	M12
E-P03.3	Título	Plataforma Adaptativa 2.0			
	Risco vinculado	R-P03.1			
		R-P03.2			
Tecnologias previstas		--			
Tipos de dados necessários		--			

Descrição	Versão 2.0 da plataforma adaptativa	
Entregáveis		
P03.1	E-P03.1	Relatório com a descrição dos requerimentos funcionais e não-funcionais da plataforma adaptativa (fonte de dados, análises, especificação do chatbot)
	M3	
P03.2	E-P03.2	Plataforma Adaptativa (códigos, bases de conhecimentos)
	M8	
P03.3	E-P03.3	Plataforma Adaptativa (códigos, bases de conhecimentos)
	M12	
Milestones		
M-P03.1	Definição da equipe	
M1		
M-P03.2	Acompanhamento das atividades desenvolvidas.	
M2		
M-P03.3	Acompanhamento das atividades desenvolvidas.	
M4		
M-P03.4	Acompanhamento das atividades desenvolvidas.	
M5		
M-P03.5	Acompanhamento das atividades desenvolvidas.	
M6		
M-P03.6	Acompanhamento das atividades desenvolvidas.	

M7	
M-P03.7	Acompanhamento das atividades desenvolvidas.
M9	
M-P03.8	Acompanhamento das atividades desenvolvidas.
M10	
M-P03.9	Acompanhamento das atividades desenvolvidas.
M11	

Riscos Críticos para a Implementação

Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-P03.1	Mudança de equipe.	Médio	Aceitar	Realizar novo processo para contratação de recurso.
R-P03.2	Baixa qualidade dos dados coletados a partir das fontes especificadas.	Baixo	Mitigar	Procurar fontes de dados alternativas para os dados de interesse.

Cronograma das Entregas por Etapas

ETAPAS		ENTREGAS											
CÓD.	TÍTULO	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-P03.1	Detalhamento da Plataforma Adaptativa			P03.1									
E-P03.2	Plataforma Adaptativa								P03.2				

	1.0												
E-P03.3	Plataforma Adaptativa 2.0												P03.3
MILESTONES		M-P03.1	M-P03.2		M-P03.3	M-P03.4	M-P03.5	M-P03.6		M-P03.7	M-P03.8	M-P03.9	

Etapas - 2022					
E-P03.11		Início	M13	Fim	M16
	Título	Plataforma Adaptativa 3.0			
	Descrição	Versão 3.0 da plataforma adaptativa			
E-P03.12		Início	M17	Fim	M20
	Título	Plataforma Adaptativa 4.0			
	Descrição	Versão 4.0 da plataforma adaptativa.			
E-P03.13		Início	M21	Fim	M24
	Título	Plataforma Adaptativa 5.0			
	Descrição	Versão 5.0 da plataforma adaptativa.			
Entregáveis - 2022					
P03.11	E-P03.11	Plataforma Adaptativa (códigos, bases de conhecimentos)			
	M16				

P03.12	E-P03.12	Plataforma Adaptativa (códigos, bases de conhecimentos)
	M20	
P03.13	E-P03.13	Plataforma Adaptativa (códigos, bases de conhecimentos)
	M24	

Grupo de Trabalho	P04 - Proposta e implementação de visualizações interativas de dados (Painel de BI)				
Título					
Grupos de Interação	C01, F01, F02, F03, G01, M01, P01, P02				
Coordenador	Raquel Cardoso de Melo Minardi				
Equipe	Nível	Quantidade	Nomes		
	Professor	1	Raquel Cardoso de Melo Minardi		
	Pós-Doc	2	TOTAL	3	
	Doutorando	0			
	Mestrando	0			
	Graduando	0			
	Técnico	0	--		
Plataforma de vínculo	Lins	Áduna	Lemonade	Trilhas	Mapa social
	Não	Não	Não	Não	Não
Descrição Geral					

Projeto e implementação de representações visuais interativas capazes de evidenciar tendências, padrões, anomalias e outros padrões em conjuntos de dados de interesse.

Objetivos

Construir módulos de visualização que possam ser integrados ao Painel de BI construído em 2020. Esses módulos serão construídos usando a biblioteca D3.js e DataTables.

Será continuado o desenvolvimento do Painel de BI implementado no primeiro ano do projeto (2020), pelo então grupo F01.

Serão adicionadas visualizações relativas aos dados oriundos de portais de transparência do municípios do Estado de Minas Gerais, que foram previamente processados, enriquecidos e estruturados pelo grupo de trabalho F01.

Etapas

		Início	M1	Fim	M2
E-P04.1	Título	Planejamento dos casos de uso a serem trabalhados na primeira etapa.			
	Risco vinculado	R-P04.3			
Tecnologias previstas	Nenhuma				
Tipos de dados necessários	Descrições dos dados a serem trabalhados para os casos de uso da primeira etapa.				
Descrição	<p>Esta etapa visa o planejamento dos casos de uso a serem desenvolvidos na primeira etapa.</p> <p>Será feita uma avaliação dos dados que são disponibilizados nos portais da transparência dos municípios do Estado de Minas Gerais, de forma que seja possível verificar quais visualizações podem ser disponibilizadas no Painel de BI.</p> <ul style="list-style-type: none"> - Três casos de uso, se a equipe for composta por 1 bolsista - Seis casos de uso, se a equipe for composta por 2 bolsistas 				
E-P04.2		Início	M2	Fim	M6
	Título	Projeto e implementação das visualizações dos casos de uso definidos na E-P04.1.			

		Risco vinculado	R-P04.1	R-P04.2	R-P04.3
Tecnologias previstas		Módulos em D3/DataTables e documentação das implementações			
Tipos de dados necessários		Arquivos em formato json			
Descrição	Projeto e implementação de três casos de uso definidos na etapa E-P04.1				
E-P04.3		Início	M7	Fim	M8
	Título	Planejamento dos três casos de uso a serem trabalhados na segunda etapa.			
		Risco vinculado	R-P04.3		
Tecnologias previstas		Nenhuma			
Tipos de dados necessários		Descrições dos dados a serem trabalhados para os casos de uso da segunda etapa.			
Descrição	<p>Esta etapa visa o planejamento dos casos de uso que serão desenvolvidos na segunda etapa.</p> <p>Será feita uma avaliação dos dados que são disponibilizados nos portais da transparência dos municípios do Estado de Minas Gerais, de forma que seja possível verificar quais outras visualizações podem ser disponibilizadas no Painel de BI.</p>				
E-P04.4		Início	M9	Fim	M12
	Título	Projeto e implementação das visualizações dos casos de uso definidos na E-P04.3.			
		Risco vinculado	R-P04.1	R-P04.2	R-P04.3

Tecnologias previstas		Módulos em d3/DataTables e documentação das implementações			
Tipos de dados necessários		Arquivos em formato json			
Descrição		Projeto e implementação dos três casos de uso na etapa E-P04.3.			
Entregáveis					
E-P04.5	Início	M1	Fim	M2	
	Título	Avaliação do uso de uma ferramenta de visualização/BI a ser definida pelo MPMG para exploração de dados pela equipe do MPMG			
	Risco vinculado				
Tecnologias previstas		Relatório			
Tipos de dados necessários		Conjuntos de dados estruturados (em formato a definir)			
Descrição		Parecer sobre vantagens e limitações da ferramenta para análise visual e exploratória de dados para a equipe de TI do MPMG.			
P04.1	E-P04.1	Relatório contendo as descrições dos três primeiros casos de uso a serem desenvolvidos			
	M2				
P04.2	E-P04.5	Relatório de avaliação da ferramenta de visualização/BI.			
	M2				
P04.3	E-P04.2	Módulo de visualização de dados definido no E-P04.1			
	M4				
P04.4	E-P04.2	Módulo de visualização de dados definido no E-P04.1			
	M5				

P04.5	E-P04.2	Módulo de visualização de dados definido no E-P04.1
	M6	
P04.6	E-P04.3	Relatório contendo as descrições dos três próximos casos de uso a serem desenvolvidos
	M8	
P04.7	E-P04.4	Módulo de visualização de dados definido no E-P04.6
	M9	
P04.8	E-P04.4	Módulo de visualização de dados definido no E-P04.6
	M10	
P04.9	E-P04.4	Módulo de visualização de dados definido no E-P04.6
	M11	
Milestones		
M-P04.1	Relatório parcial sobre as reuniões realizadas e os caso de uso já definidos para a etapa E-P04.2.	
M1		
M-P04.2	Relatório parcial sobre as reuniões realizadas e os caso de uso já definidos para a etapa E-P04.	
M3		
M-P04.3	Relatório parcial sobre as reuniões realizadas e os caso de uso já definidos para a etapa E-P04.4.	
M7		
M-P04.4	Fechamento da atividade com entrega de todos as documentações.	
M12		

Riscos Críticos para a Implementação

Código	Descrição	Impacto	Tratamento	Plano de Contingência
R-P04.1	Volume grande de dados que impossibilite a representação visual. Exemplo: uma rede com milhões de nós não pode ser visualizada de uma vez. Solução: agregar os nós, visualização sub-redes, modularizar, etc.	Alto	Transferir	Propor agregações e filtros que serão implementados pela equipe de TI do MPMG ou de outras equipes do DCC de forma a prover dados compatíveis com a visualização.
R-P04.2	Dados em formato inadequado.	Alto	Transferir	Como a equipe do A05 tratará apenas das visualizações no front-end, todos os dados devem chegar nos formatos adequados, evitando atrasos nas entregas.
R-P04.3	Atraso na definição dos casos de uso.	Médio	Mitigar	Renegociar os prazos de entrega.

Cronograma das Entregas por Etapas

ETAPAS		ENTREGAS											
CÓD.	TÍTULO	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
E-P04.1	Planejamento dos casos de uso a serem trabalhados na primeira etapa.		P04.1										
E-P04.2	Projeto e implementação das visualizações dos casos de uso definidos na E-P04.1.				P04.3	P04.4	P04.5						
E-P04.3	Planejamento dos 3 casos de uso a serem trabalhados na segunda etapa.								P04.6				
E-P04.4	Projeto e implementação das visualizações dos casos de uso definidos na E-P04.3.									P04.7	P04.8	P04.9	

E-P04.5	Avaliação do uso de uma ferramenta de visualização/BI a ser definida pelo MPMG para exploração de dados pela equipe do MPMG		P04.2										
MILESTONES		M-P04.1	--	M-P04.2	--	--	--	M-P04.2	--	--	--	--	M-P04.3

Etapas - 2022				
E-P04.6	Início	M13	Fim	M14
	Título	Planejamento dos casos de uso a serem trabalhados na terceira etapa.		
	Descrição	Esta etapa visa o planejamento dos casos de uso a serem desenvolvidos na primeira etapa. - Três casos de uso, se a equipe for composta por 1 bolsista - Seis casos de uso, se a equipe for composta por 2 bolsistas		
E-P04.7	Início	M14	Fim	M18
	Título	Projeto e implementação das visualizações dos casos de uso definidos na E-P04.6.		
	Descrição	Projeto e implementação dos 3 casos de uso na etapa E-P04.6		
E-P04.8	Início	M19	Fim	M20
	Título	Planejamento dos 3 casos de uso a serem trabalhados na quarta etapa.		
	Descrição	Esta etapa visa o planejamento dos casos 3 de uso que serão desenvolvidos na quarta etapa.		

		Início	M21	Fim	M24
E-P04.9	Título	Projeto e implementação das visualizações dos casos de uso definidos na E-P04.8.			
	Descrição	Projeto e implementação dos 3 casos de uso na etapa E-P04.8			
Entregáveis - 2022					
P04.10	E-P04.6	Relatório contendo as descrições dos 3 próximos casos de uso a serem desenvolvidos			
	M14				
P04.11	E-P04.7	3 módulos de visualização de dados definido no P04.10			
	M18				
P04.12	E-P04.8	Relatório contendo as descrições dos 3 próximos casos de uso a serem desenvolvidos			
	M20				
P04.13	E-P04.9	3 módulos de visualização de dados definido no P04.12			
	M24				

I. Implantação, Manutenção e Suporte

GRUPO	EIXO / OBJETO	META	ATIVIDADES	1° TRI	2° TRI	3° TRI
ISM-01	Implantação no ambiente do MPMG	A solução de software deve estar implementada e funcional no ambiente de homologação do MPMG	Salvar o código fonte no GitLab do GSI			
			Disponibilizar aplicações em docker			
			Disponibilizar cada uma das aplicações separada em			

GRUPO	EIXO / OBJETO	META	ATIVIDADES	1° TRI	2° TRI	3° TRI	
			módulos: front-end, back-end, banco de dados etc				
			API's com documentação swagger				
			API's integradas com WSO2 API Manager				
			Aplicação integrada com autenticação do WSO2IS				
			Pipeline no Jenkins para deploy nos diversos ambientes				
	Suporte, manutenção e melhoria contínua	A solução de software deve disponibilizar um manual de utilização com exemplos para os usuários finais	Elaborar e disponibilizar manuais e exemplos de uso				
			A solução de software deve possuir video de apresentação geral	Criar vídeo de apresentação do propósito geral de cada ferramenta			
			Interações com os usuários finais (conteudistas) para apresentação do software	Realizar reunião de apresentação da ferramenta			
				Realizar treinamentos síncronos online			
			Corrigir falhas, otimizar processos, garantir a disponibilidade dos softwares e implementar novas funcionalidades	Corrigir falhas identificadas na ferramenta			
Otimizar funcionalidades da ferramenta							

GRUPO	EIXO / OBJETO	META	ATIVIDADES	1º TRI	2º TRI	3º TRI
		que não representem modificações estruturais	Implementar melhorias não estruturais na ferramenta			
		Engenharia de Dados	Continuidade das atividades desempenhadas pela equipe de Engenharia de Dados alocada junto ao MPMG			

** 1º Trim: maio, junho e julho/23

2º Trim: agosto, setembro e outubro/23

3º Trim: novembro, dezembro e janeiro/24

[1] <https://kubernetes.io/pt/>

[2] <https://airflow.apache.org/>

[3] <https://nifi.apache.org/>

Assim ajustadas, as partes assinam o presente Instrumento, para um só efeito de direito, por meio de assinatura/senha eletrônica, na presença de duas testemunhas.

Contratante:

Márcio Gomes de Souza

Procurador-Geral de Justiça Adjunto Administrativo

Contratada:

Jaime Arturo Ramírez
Presidente da FUNDEP

Testemunhas:

- 1)
- 2)



Documento assinado eletronicamente por **Jaime Arturo Ramírez, Usuário Externo**, em 26/06/2023, às 11:56, conforme art. 22, da Resolução PGJ n. 27, de 17 de dezembro de 2018.



Documento assinado eletronicamente por **MARCIO GOMES DE SOUZA, PROCURADOR-GERAL DE JUSTICA ADJUNTO ADMINISTRATIVO**, em 26/06/2023, às 14:23, conforme art. 22, da Resolução PGJ n. 27, de 17 de dezembro de 2018.



Documento assinado eletronicamente por **LETICIA SANTANA SANTOS ROCHA, OFICIAL DO MINIST. PUBLICO - QP**, em 26/06/2023, às 14:56, conforme art. 22, da Resolução PGJ n. 27, de 17 de dezembro de 2018.



Documento assinado eletronicamente por **PATRICIA FERREIRA FIALHO, OFICIAL DO MINIST. PUBLICO - QP**, em 26/06/2023, às 15:17, conforme art. 22, da Resolução PGJ n. 27, de 17 de dezembro de 2018.



A autenticidade do documento pode ser conferida no site <http://www.mpmg.mp.br/sei/processos/verifica>, informando o código verificador **5430858** e o código CRC **31BF7F18**.

