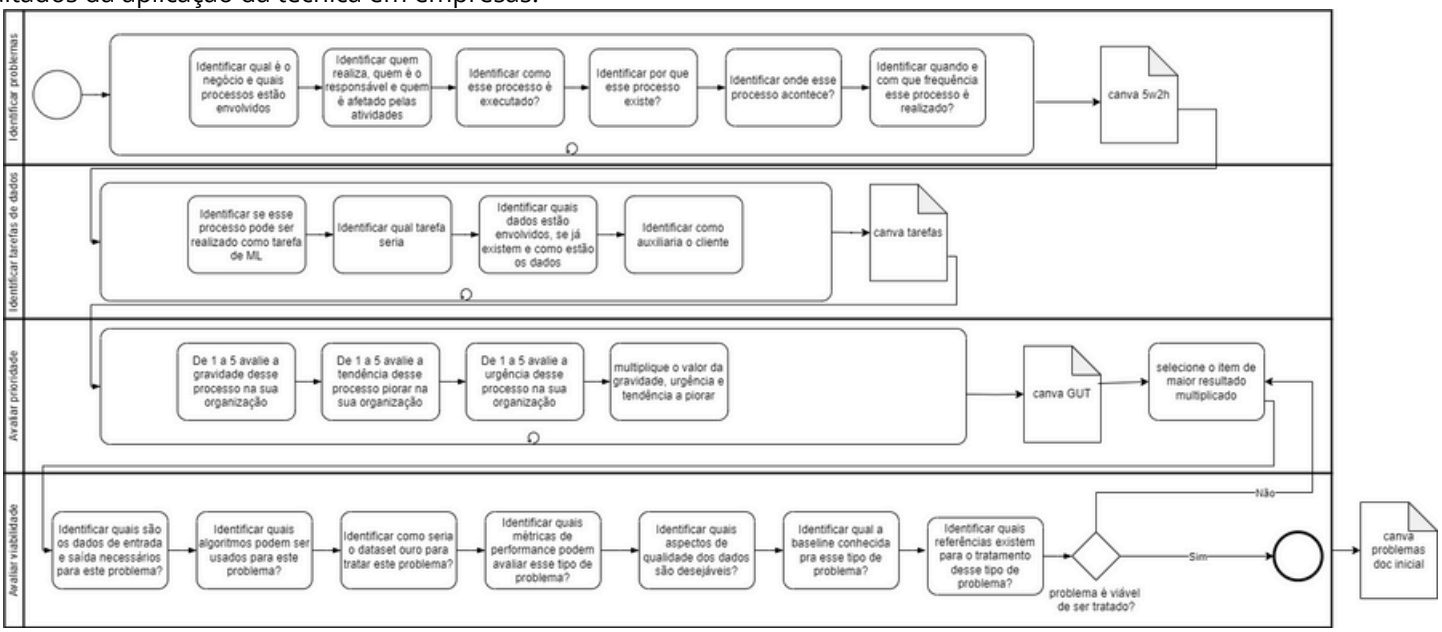
Identification Problems for AI Projects Innovations

Ao participar de projetos de inovação no CEIA-UFG e entrevistar coordenadores de projetos percebemos que a maior dificuldade apresentada foi a identificação do problema. Etapa inicial no desenvolvimento de soluções ML e que tem intersecção com a engenharia de requisitos. Propusemos um processo de RE baseado na ISO/EC 12207:2017. No entanto notamos que adicionava bastante complexidade, indesejável em projetos de inovação. Por isso, adotamos o Design Thinking para auxiliar na identificação de problemas de forma ágil.

A ideia é aplicar o processo identificando inicialmente o negócio e os processos existentes envolvidos no negócio. Também buscamos identificar (caso exista um possível problema a ser tratado com solução ML). Então aplicamos a matriz GUT que avalia a gravidade, urgência e tendência a piorar de cada um dos processos ou problemas apresentados. Multiplicamos o valor a fim de identificar qual processo é mais prioritário a ser tratado no projeto. A seguir Aplicamoss o 5W2H para identificacr tarefas existentes no processo, envolvidos, como: responsável, quem reaiza a atividade e quemé afetado por ela. Também identificamos onde a atividade acontece, como ela é reaizada e registrada (ou seja, como os dados são gerados e armazenados), por quê determinada atividade acontece e sua importância e com que frequência essa atividade é realizada. A seguir contamos com o cientista de dados para identificar a tarefa de dados associada aos problemas identificados, quais dados são necessários para tratar cada problema e como pderíamos obtê-los e também listamos como essa solução ajuda o cliente. Por fim, aplicamos a identificação do problema a ser tratado no projeto, quais são os dados de entrada e saíde, quais são as características de qualidade dos dados necessárias e qual seria o dataset ouro. como seriam as métricas de performance, qual baseline existente para esse tipos e problema e as referências.

A abordagem de processo foi criada durante o doutorado de Mariana Martins no Instituto de Informática da Universidade de Goiás, supervisionada pelo Prof. Renato Bulcão-Neto com o apoio da Profa. Taciana Novo Kudo. A técnica foi aplicada primeiramente em projetos de aprendizado de máquina do CEIA - UFG e evoluiu ao longo de uma série de estudos científicos. Hoje ela está sendo usada por diversas empresas. Abaixo está uma lista de alguns dos estudos científicos, que inclusive relatam resultados da aplicação da técnica em empresas.



MARTINS, M.; KUDO, T.; BULCÃO-NETO, R. A qualitative study on requirements engineering practices in an artificial intelligence unit of the brazilian industrial research and innovation company. In: Anais do XXVII Congresso Ibero-Americano em Engenharia de Software, p. 46–60, Porto Alegre, RS, Brasil, 2024. SBC.

MARTINS, M.; KUDO, T.; BULCÃO-NETO, R. A requirements engineering process for machine learning innovation projects. In: Anais do XXVII Workshop de Engenharia de Requisitos, Buenos Aires, Argentina, 2024. Even3.

MARTINS, M.; MANCINE, L.; RODRIGUES, J.L.; KUDO, T.; BULCÃO-NETO, R. Research Synthesis in Requirements Engineering for Machine Learning-Based AI Systems: A Tertiary Study. (SUBMETIDO).

Marcos Vinicius Satil Medeiros é um cientista de dados aluno do curso de Pós graduação da UFG.

Wagner Hélio da Silva é um engenheiro de dados aluno do curso de Pós graduação da UFG.

Pedro Koziel Diniz é um cientista de dados aluno do curso de Pós graduação da UFG.

Identificação de Objetivos do Negócio

Propósito: Identificar o contexto da organização, processos, problemas e oportunidades que podem ser resolvidas com soluções de inteligência artificial (IA) em projetos de pesquisa, desenvolvimento e inovação (PDI)

Os stakeholders envolvidos no preenchimento deste artefato e identificação de problemas para projetos de PDI são: Business Owner (BO), Especialista de Domínio (ED) e Cientista de Dados (CD)

Objetivo do negócio 1

- Responder LAIs de maneira sistêmica utilizando Inteligência Artificial e LLMs para busca em base dados e entrega de resultados para o cliente.

BO

Justificativas 2

- A Secretaria de Segurança Pública lida com grandes volumes de dados provenientes de investigações, laudos periciais e registros criminais.
- A análise manual desses dados é lenta e sujeita a erros, prejudicando a velocidade e a precisão dos resultados.

BO

Benefícios 3

- Redução de erros humanos.
- Eficiência: Necessidade de busca manual apenas em casos mais específicos.
- Facilitação de cruzamento de dados entre inquéritos e ocorrências.
- Transparência: Histórico de consultas e resultados acessíveis para auditoria e controle.

BO

KPIs 4

- Segurança e auditoria das ocorrências e inquéritos.
- Tempo médio de resposta às consultas (reduzir em X%)
- Taxa de erro nas buscas de informações (reduzir em X%)
- Redução de retrabalho em consultas repetidas (reduzir em X%)

BO

Processos para atingir o objetivo 5

- Definir diagrama de processos.
- Configuração e customização do ambiente.
- Correlacionamento de informações entre múltiplas bases de dados.
- Interface simplificada ou api para os atendentes acessarem o agente de IA.

BO

Problemas necessários de serem tratados 6

- Lentidão na busca manual de dados, por conta de autorização de burocráticas de órgãos dentetores da informação.
- Falta de integração entre diferentes bases de dados (banco de nao relacionais e NoSQL e bancos transacionais relacionais).
- Erros humanos ao localizar ou interpretar informações críticas.
- Sobrecarga de trabalho nos atendentes, que perdem tempo em buscas manuais repetidas.
- Falta de padronização na criação de relatórios e laudos.

BO

Resultados esperados: Lista de possíveis problemas a serem tratados com solução de AM em projetos PDI

Identificação de Problemas

Propósito: Identificar um ou mais problemas a serem tratados no projeto como uma solução IA e as respectivas prioridades de problemas ou oportunidades

Tarefas de cada processo 7

1 - O atendente recebe a solicitação da LAI e Verifica a idoniedade do solicitante.
2 - Analisa a Viabilidade de Resposta se os dados são cabiveis.
3 - Acessa Diferentes DB e Dashboars correlacionando as informações para responder a LAI

BO ED

Por quê 8

- Agiliza o fornecimento de informações críticas para investigações e laudos.

BO ED

Como é feito 9

- Consulta direta aos sistemas RAI (Registro de atendimento Integrado) e SPP (Sistema de Procedimentos Policiais) e análise manual das respostas por um operador.

BO ED

Envolvidos 10

- Time de desenvolvimento;
- Equipe do observatório e SSP-GO;
- Time de imprensa de Goiás;

BO ED

Onde 11

- Secretaria de Segurança Pública (SSP), do Estado de Goiás e entorno de Brasília.

BO ED

Com que frequência 12

- Diariamente.
- Tempo real.

BO ED

Gravidade 13

- 5

BO

Urgência 14

- 3

BO

Tendência 15

- 4

BO

Multiplicação 16

- 5 × 3 × 4 = 60

Tarefa de dados 13

- Automatização da busca;
- Extração de informações relevantes;
- Classificação de registros;

CD

Como ajuda o cliente 14

- Valor agregado: Reduz o tempo gasto em consultas e melhora a precisão na análise.
- Métricas: Tempo médio de busca, taxa de erros nas consultas.

CD

Dados necessários 15

- RAIs e Inqueritos;
- Logs de consultas;

CD

Resultados esperados: Detalhamento dos problemas baseado em 5W2H, lista de problemas priorizados conforme a matriz GUT e identificação de possíveis soluções de AM para os problemas analisados

Análise de Viabilidade

Propósito: Avaliar se o tratamento do(s) problema(s) é viável de ser tratado com solução de IA neste projeto analisando aspectos de dados, soluções e riscos

Problema prioritário 17

- Converter bases de dados relacionais e não relacionais em informação útil para os LLMs.
- Busca de informações relevantes nas bases de dados;

BO CD

Dados entrada 18

- Query do usuário, LAIs.
- Estrutura dos banco de dados, Ocorrências (Banco SQL e NoSQL)

BO ED CD

Qualidade dos dados 19

- Completude: o banco de dados deve conter todas as informações relevantes.
- Consistência: Padronização nos formatos dos campos.
- Atualização: Dados devem ser atualizados periodicamente.

BO CD

Dados saída 20

- Resposta à query feita, em linguagem natural.
- Logs de acesso
- Dashboard interativo.

CD

Métricas de desempenho 21

- Taxa de precisão das respostas.
- Redução no tempo de geração das LAIs
- Quantidade de consultas repetidas reduzidas;

CD

Riscos 22

- Dados incompletos ou inconsistentes; Resistência dos usuários ao novo sistema; Falha de integração com sistemas legados;

Descrição - Impacto - Probabilidade - Contingência

- identifique riscos - julgue o impacto de cada risco (ex. atrasos no projeto) - informe a probabilidade dos risco acontecer como baixa, média e alta - descreva as ações a serem tomadas caso o risco aconteça

BO CD

Algoritmos 23

- Modelos de linguagem para a tarefa de Text2SQL:
  - Modelos open source.
- Modelos de embeddings e semelhança - RAG.

CD

Baseline 24

- liste a *baseline* conhecida para esse tipo de problema ou problemas semelhantes

CD

Referências 25

- liste artigos e referências usadas para conhecer e tratar aspectos do problema

CD

Resultados esperados: Problemas viáveis de serem tratados com soluções AM em projetos de PDI