



Challenge Sprint 2 – 2024

AI for Robotic Process Automation

Prof Samuel Perroni da Rocha

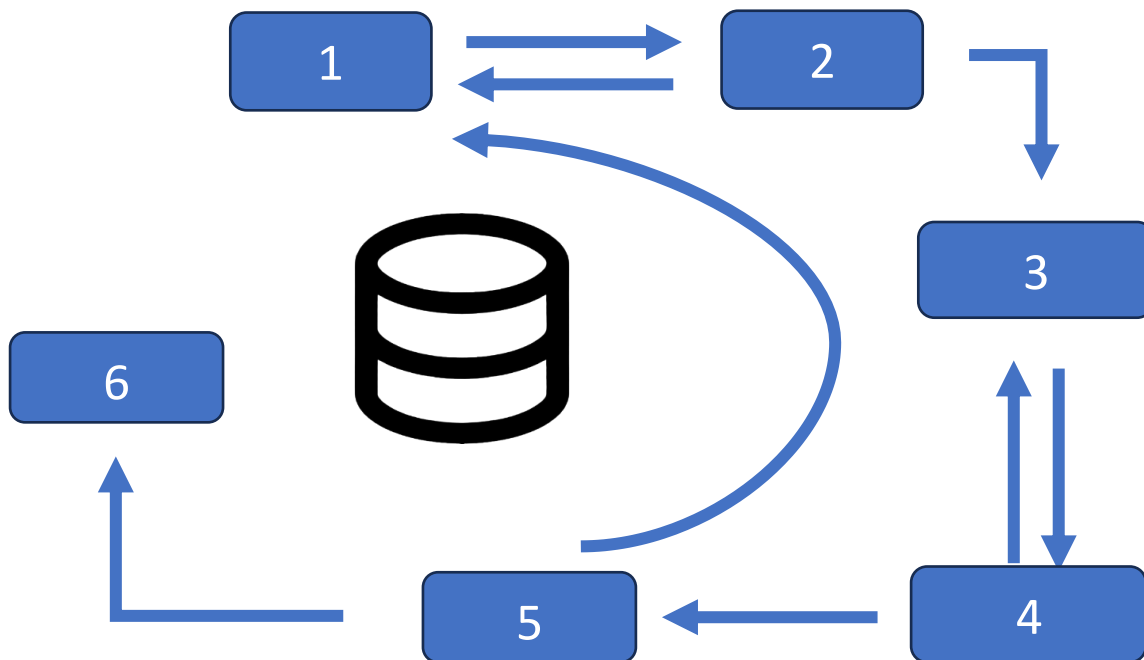
Camilly Alves RM 550210

João Vitor Martins RM 98744

Murilo Krauss RM 98262

Rafael Lima RM 88444

Seguindo as ideias prepostas no checkpoint 1, de um modelo de Random Forrest, para prever aqueles funcionários que tendem a deixar a empresa, e com isso aplicar algum programa de desenvolvimento, montamos o seguinte cronograma, baseado na metodologia CRISP-DM:



1. Entendimento do negócio (20/05): Definição do objetivo/ideia do projeto;
2. Entendimento dos dados (20/05): Entendimento de quais são os dados necessários para o projeto e a coleta do mesmo;
3. Preparação dos dados (14/06): Momento da limpeza do dataset, análises prévias (descritiva e exploratória dos dados), tratamento dos dados nulos, ausentes, converter em binários dentre outras alterações;
4. Criação e treinamento do Modelo (24/06): Etapa de criação do modelo, definição de camadas e outros atributos;
5. Avaliação do modelo (13/08): Avaliar se o resultado corresponde à expectativa do projeto. Caso os resultados não sejam satisfatórios, devemos reavaliar os entendimentos do negocio ou apenas aprimorar o modelo;
6. Deployment (19/08): Ultima etapa do projeto, momento que o modelo é colocado em produção.

Seguindo com o notebook, a base publica para dados de Attrition, foi a Employee Attrition (<https://www.kaggle.com/datasets/colearninglounge/employee-attrition>), encontrada no Kaggle. Ela é composta tanto por dados numéricos quanto categóricos, possuindo

Link do colab:

https://colab.research.google.com/drive/1O93q19ww4u_9vhQL3CJXGq0UBw8bNdrA?usp=sharing

