

Etapa DASHBOARD – Grupo 03.

Integrantes:

- Artur Carneiro de Lélis
- Carlos Eduardo da Silva Cerqueira
- Pablo Marcos Barbosa Souza
- Pedro Afonso Vieira Fernandes Braga
- Carla Vitoria Sales Muniz dos Santos

Relatório 2: Algoritmo de Aprendizado de Máquina

No presente relatório, é disposto as etapas envolvidas na aplicação dos algoritmos de aprendizado de máquina após a limpeza e análise da base de dados. Inicialmente, aplicamos modelos de algoritmos que pudessem alimentar os dados gerando as informações detalhadas sobre o desempenho e comportamento das aplicações. Para isso, utilizamos diferentes algoritmos, como Random Forest e Decision Tree, para o treinamento dos dados. Em seguida, após o treinamento, realizamos a validação dos modelos utilizando conjuntos de dados de validação e teste para avaliar a precisão e eficácia dos algoritmos.

Depois do pré-processamento inicial realizado pelo programador sênior (Carlos), a rodagem dos dados foi realizada para garantir que os algoritmos de aprendizado de máquina pudessem ser alimentados corretamente. Então, a preparação e rodagem dos dados permitiram melhores resultados que serviram para uma reestruturação mais detalha da base, na qual, houve uma remoção de valores nulos, duplicados e tratamento de dados inconsistentes.

A partir do treinamento dos modelos em cima da base, realizamos a plotagem dos resultados em gráficos para obter as informações requeridas e permitindo a visualização do desempenho dos modelos, além de identificar padrões ou anomalias nos dados. Dessa forma, as visualizações ajudaram na geração dos resultados de forma clara e eficaz, permitindo uma melhor compreensão do impacto das diferentes variáveis no modelo, como por exemplo: na matriz de

confusão para visualizar a precisão da classificação dos modelos, e nos gráficos de importância de variáveis para identificar quais características têm maior impacto nos modelos.

Por fim, realizamos uma análise aprofundada dos hiperparâmetros utilizando o método do RandomizedSearchCV para identificar os melhores parâmetros para os nossos modelos e produção do dashboard para visualização dos dados. Então, esse processo incluiu a definição de um espaço de busca, onde selecionamos um conjunto de possíveis valores de hiperparâmetros a serem testados, com uma execução da busca aleatória para realizar pesquisas dentro do espaço definido, maximizando então a eficiência na identificação dos melhores hiperparâmetros. Portanto, analisamos os resultados obtidos para então escolher a combinação de hiperparâmetros que ofereceu o melhor desempenho para a aplicação ao dashboard.