

Tópicos Contemporâneos 3

Projeto da Disciplina - AV1

Antonio Neto, Davi Cesar, João Ricardo

Ciência da Computação –CESAR School
Avenida, Cais do Apolo, 77, Recife-PE - 50030-22

Introdução

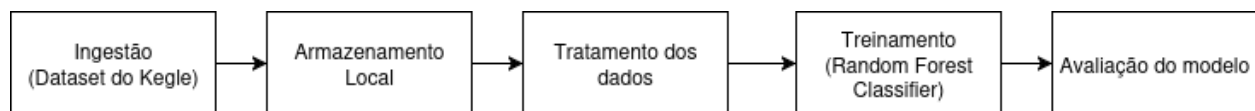
Esse documento apresenta as especificações do projeto da cadeira de Tópicos Contemporâneos 3, onde o grupo realizou o armazenamento, transformação e análise do dataset: [\(LoL\) League of Legends Ranked Games](#), com o objetivo final de criar um modelo de predição.

Para essa primeira entrega foram utilizados as seguintes ferramentas:

- Goggle Colab
- Python
- Pandas
- Sklearn (Random Forest Classifier)

Diagrama

Abaixo encontra-se o diagrama seguido para o desenvolvimento dessa atividade:



Arquitetura Parcial

A arquitetura do projeto foi desenvolvida com o objetivo de estruturar todas as etapas do processo de análise e modelagem preditiva sobre o dataset *League of Legends Ranked Games*. O fluxo de trabalho foi dividido em três grandes etapas: armazenamento, transformação e análise preditiva.

Na primeira etapa, de armazenamento, os dados foram obtidos a partir do dataset público disponível no Kaggle e importados para o ambiente Google Colab. O conjunto de dados, em formato CSV, foi carregado utilizando a biblioteca Pandas, sendo manipulado em memória por meio de DataFrames. Essa estrutura permitiu realizar consultas, filtragens e transformações de forma eficiente durante todo o processo.

Em seguida, foi executada a etapa de limpeza e transformação dos dados, fundamental para preparar o dataset para a modelagem. Inicialmente, foram identificados registros duplicados com

base na coluna *gameId*, representando partidas repetidas. Para garantir a integridade da análise, manteve-se apenas uma instância de cada partida. Posteriormente, foram realizadas filtragens de colunas, mantendo apenas aquelas relevantes para o objetivo do projeto — especificamente, as colunas referentes aos campeões utilizados nas partidas e a coluna que indicava o time vencedor, que representa a variável alvo da predição.

Ainda nesta etapa, foi conduzida uma análise exploratória para identificar os 10 campeões mais utilizados nas partidas. Essa análise permitiu compreender padrões de escolha e possíveis relações entre determinados campeões e as vitórias, além de auxiliar na interpretação dos resultados do modelo.

A etapa final consistiu na análise e modelagem preditiva. O modelo escolhido foi o Random Forest Classifier, disponível na biblioteca Scikit-learn (Sklearn). Esse algoritmo foi selecionado por sua robustez e capacidade de lidar com dados categóricos e complexos, como os relacionados à composição de campeões. O modelo foi treinado utilizando os dados tratados e posteriormente testado para avaliar seu desempenho na edição do time vencedor. Durante essa fase, também foram aplicadas métricas de avaliação que permitiram mensurar a precisão e a confiabilidade das predições.

Em resumo, o fluxo do projeto pode ser descrito da seguinte forma:

1. Coleta e armazenamento do dataset no ambiente Google Colab.
2. Remoção de duplicatas e seleção das colunas relevantes.
3. Análise dos campeões mais utilizados.
4. Treinamento do modelo Random Forest Classifier.
5. Predição e avaliação dos resultados obtidos.

Divisão de Tarefas:

- **Antonio Neto:** Ingestão dos Dados e Tratamento;
- **Davi Cesar:** Desenvolvimento do algoritmo preditivo;
- **João Ricardo:** Documentação;