

# Sistema de Previsão de Resultados do Brasileirão usando Machine Learning

## Resumo

Este trabalho apresenta o desenvolvimento de um sistema de previsão de resultados do Campeonato Brasileiro Série A utilizando técnicas de Machine Learning. O projeto combina análise de dados em tempo real, processamento avançado de features e uma interface interativa para auxiliar na compreensão e previsão de resultados de partidas.

## 1. Introdução

### 1.1 Contextualização

O futebol brasileiro é conhecido por sua complexidade e imprevisibilidade. A análise e previsão de resultados das partidas representa um desafio significativo devido à multiplicidade de fatores envolvidos e à natureza dinâmica do esporte.

### 1.2 Objetivos

- Desenvolver um modelo de machine learning para prever resultados de partidas
- Implementar uma interface intuitiva para visualização de dados
- Fornecer análises estatísticas detalhadas do campeonato
- Avaliar o desempenho individual dos times

### 1.3 Justificativa

A previsão de resultados no futebol tem aplicações em diversos contextos, desde análise técnica até tomada de decisões estratégicas. Um sistema automatizado pode identificar padrões e tendências não óbvios para humanos.

## 2. Metodologia

### 2.1 Coleta de Dados

- **Fonte:** API football-data.org

- **Escopo:** Brasileirão Série A 2023
- **Dados Coletados:**
  - Resultados de partidas
  - Estatísticas dos times
  - Classificação em tempo real

## 2.2 Processamento de Dados

### 1. Limpeza e Preparação:

- Tratamento de dados ausentes
- Normalização de formatos
- Validação de consistência

### 2. Engenharia de Features:

- Média de gols marcados/sofridos
- Aproveitamento como mandante/visitante
- Forma recente (últimos 5 jogos)
- Confrontos diretos históricos
- Posição na tabela

## 2.3 Modelo de Machine Learning

### 1. **Algoritmo:** Random Forest Classifier

- Escolhido por sua robustez e capacidade de lidar com dados não lineares
- Bom desempenho com features categóricas e numéricas

### 2. **Otimização:**

- Grid Search para seleção de hiperparâmetros
- Validação cruzada (5-fold)
- Balanceamento de classes

## 2.4 Interface do Usuário

- Framework Streamlit para dashboard interativo
- Visualizações com Plotly
- Atualização em tempo real

## 3. Resultados

### 3.1 Performance do Modelo

- **Acurácia Global:** 55%
- **Comparação com Baseline (33%):** Melhoria significativa
- **Desempenho por Classe:**
  - Vitória Casa: 58%
  - Empate: 48%
  - Vitória Fora: 52%

### 3.2 Análise de Features

#### 1. Features Mais Importantes:

- Aproveitamento recente (25%)
- Média de gols marcados (20%)
- Forma atual (15%)
- Confrontos diretos (10%)

#### 2. Visualizações:

- Gráficos de distribuição de resultados
- Análise de tendências
- Estatísticas por time

### 3.3 Interface e Usabilidade

- Dashboard intuitivo
- Análises em tempo real
- Visualizações interativas

## 4. Discussão

### 4.1 Pontos Fortes

#### 1. Modelo:

- Performance superior ao acaso
- Capacidade de adaptação
- Interpretabilidade dos resultados

#### 2. Sistema:

- Interface amigável
- Atualizações automáticas
- Análises detalhadas

## 4.2 Limitações

### 1. **Dados:**

- Quantidade limitada de jogos na temporada
- Ausência de dados externos (clima, lesões)

### 2. **Previsões:**

- Dificuldade em prever empates
- Imprevisibilidade inerente do futebol

## 4.3 Melhorias Futuras

### 1. **Modelo:**

- Inclusão de mais features
- Experimentação com outros algoritmos
- Análise de sentimentos de redes sociais

### 2. **Sistema:**

- Mais visualizações interativas
- Análises mais detalhadas
- Previsões de placar

## 5. Conclusão

O sistema desenvolvido demonstra a viabilidade de usar machine learning para análise e previsão no futebol brasileiro. A acurácia obtida, significativamente superior ao acaso, indica a capacidade do modelo em capturar padrões relevantes, mesmo em um contexto altamente imprevisível.

## Referências

1. Documentação Scikit-learn
2. API Football-data.org
3. Documentação Streamlit
4. Literatura sobre previsão esportiva