

# Universidade Federal da Paraíba

Centro de Ciências Sociais Aplicadas  
Curso: Ciências de Dados para Negócios  
Disciplina: Aprendizagem Supervisionada  
Professor: Aléssio Tony

## Modelo de previsão de pontuação de clubes da série A com base no valor de mercado do elenco

Aluno: Leonardo Feitosa Barroso

### 1. Introdução

O Campeonato Brasileiro da Série A é a principal competição de clubes do Brasil, movimentando cifras bilionárias anualmente. Entre os diversos indicadores utilizados para avaliar os clubes, destaca-se o *valor de mercado*, divulgado em plataformas especializadas como o Transfermarkt. Este projeto tem como objetivo prever a pontuação final de clubes com base em dados históricos de valor de mercado entre os anos de 2020 a 2024.

A contribuição do estudo está na aplicação de técnicas de aprendizado supervisionado para analisar a relação entre desempenho esportivo e valor financeiro dos elencos, utilizando métricas estatísticas para avaliar a robustez das previsões.

### 2. Revisão da Literatura

Estudos como Frick (2007) e Szymanski (2010) indicam uma correlação significativa entre investimento financeiro no futebol e desempenho esportivo. Técnicas como regressão linear e modelos baseados em árvores são amplamente empregadas em análises preditivas no esporte, por sua capacidade de lidar com variáveis explicativas numéricas e categóricas.

### 3. Dados

Os dados foram obtidos via **web scraping** do site Transfermarkt, extraindo-se o HTML de cada temporada da Série A entre 2020 e 2024. A base foi consolidada em um arquivo **CSV único**, com a adição manual dos pontos finais de cada clube por temporada, utilizando a função PROCV no Excel. O conjunto de dados final está contido na aba **"todos\_clubes\_serieA"** da planilha `valor_de_clube_por_temporada_com_pontos.xlsx`.

#### Variáveis

- **Clube:** nome do clube
- **Plantel:** número de jogadores no elenco

- $\emptyset$  Idade: idade média dos jogadores
- Estrangeiros: número de jogadores estrangeiros
- $\emptyset$  Valor de Mercado: valor médio do clube por jogador
- Valor de Mercado Total: valor total de mercado do clube
- Ano: temporada correspondente
- Pontos: pontuação final do clube

A aba **crud** contém os dados organizados e padronizados para facilitar a montagem final da aba principal.

## Pré-processamento

Foram realizados os seguintes procedimentos:

- Junção dos CSVs de cada ano em um único arquivo
- Padronização dos nomes dos clubes
- Remoção de valores ausentes
- Adição dos pontos por temporada via Excel

## 4. Estratégia Empírica

O problema foi tratado como uma tarefa de **regressão supervisionada**, com a variável-alvo sendo a pontuação final dos clubes.

### Modelo Utilizado

- **Regressão Linear**: modelo base para avaliar tendência geral

### Métricas de Avaliação

- Correlação de Pearson
- $R^2$  (Coeficiente de Determinação)
- MAE e RMSE (não apresentados aqui por brevidade, mas disponíveis no código)

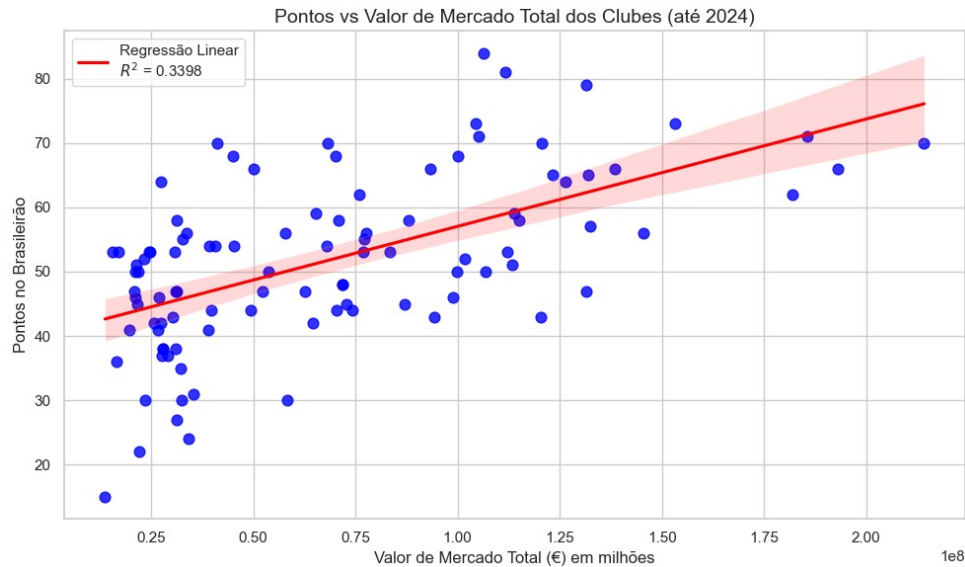


Figura 1: Relação entre Pontuação e Valor de Mercado Total

## 5. Resultados e Discussão

Com base na regressão linear, os principais resultados foram:

- **Correlação entre Valor de Mercado Total e Pontos:** 0,5829
- **R<sup>2</sup> do modelo:** 0,3398

Isso indica que cerca de 34% da variação da pontuação pode ser explicada pelo valor de mercado dos clubes, evidenciando uma relação significativa, mas não determinística.

Isso indica que cerca de 34% da variação da pontuação pode ser explicada pelo valor de mercado dos clubes, evidenciando uma relação significativa, mas não determinística.

### Exemplo do resultado de Previsão

Clube	Valor de Mercado Total	Pontos Previstos
SE Palmeiras	R\$ 238.750.000	80,23
CR Flamengo	R\$ 219.150.000	76,96

Outros modelos como Random Forest e XGBoost estão em fase de ajuste e validação cruzada para posterior comparação de desempenho.

## 6. Considerações Finais

Este estudo demonstrou que é possível prever, com razoável precisão, a pontuação de clubes com base no valor de mercado total de seus elencos. A regressão linear apresentou resultados interpretáveis e eficazes, sendo o modelo base adotado.

## Limitações

- Fatores como desempenho técnico, lesões, mudanças de técnico e mando de campo não foram considerados.
- Dados limitados às cinco últimas temporadas.

## Perspectivas Futuras

- Incorporação de variáveis táticas e extracampo
- Avaliação de modelos mais complexos com tuning de hiperparâmetros
- Desenvolvimento de interface web para visualização das previsões

## Anexos

O projeto completo, incluindo dados, scripts e notebook com análise, está disponível em: [https://github.com/leonardofeitos4/Previsao\\_DePontosBrasileirao – SerieA](https://github.com/leonardofeitos4/Previsao_DePontosBrasileirao-SerieA)