

FGA0083 - Aprendizado de Máquina	Semestre: 2025.1
Docente: Sérgio Antônio Andrade	Turma: 01
Grupo 05	
Integrantes: Harryson Campos Martins Pedro Henrique Muniz de Oliveira Flávio Gustavo Araújo de Melo Leandro de Almeida Oliveira Jefferson Sena Oliveira José André Rabelo Rocha	Matrícula: 211039466 200059947 211030602 211030827 200020323 211062016

Mini Trabalho 5:

Seleção de modelos com potencial

Análise Comparativa dos Modelos

1. Objetivos

O objetivo deste trabalho é desenvolver modelos de aprendizado de máquina capazes de prever os resultados de partidas do Campeonato Brasileiro de Futebol. O problema é tratado como uma **classificação multiclasse**, onde as possíveis saídas são: **vitória**, **empate** ou **derrota** do time mandante.

2. Modelos Implementados

2.1 Regressão Logística

Características:

- Modelo linear simples e interpretável
- Rápido para treinar e fazer previsões
- Boa base para comparação com modelos mais complexos

2.2 Random Forest

Características:

- Ensemble de árvores de decisão
- Robusto a overfitting
- Capaz de capturar relações não-lineares
- Lida bem com features numéricas e categóricas

2.3 Support Vector Machine (SVM)

Características:

- Kernel RBF para capturar relações não-lineares
- `class_weight='balanced'` para lidar com desbalanceamento
- Normalização dos dados para melhor performance
- Boa capacidade de generalização

3. Métricas de Avaliação

3.1 Accuracy (Precisão Geral)

- Mede a proporção total de previsões corretas
- Útil para uma visão geral do desempenho

3.2 F1-Score (Macro Average)

- Equilibra precisão e recall
- Especialmente importante devido ao possível desbalanceamento das classes
- Considera igualmente todas as classes

3.3 Matriz de Confusão

- Visualiza os erros do modelo
- Permite identificar padrões de erro específicos
- Ajuda a entender como o modelo confunde as classes

4. Análise Crítica

4.1 Vantagens e Desvantagens

Regressão Logística:

- **Vantagens:**
 - Simples e interpretável
 - Rápido para treinar
- **Desvantagens:**
 - Pode não capturar relações complexas
 - Performance limitada em problemas não-lineares

Random Forest:

- **Vantagens:**
 - Alta capacidade de capturar padrões complexos
 - Robusto a overfitting
 - Lida bem com features diversas
- **Desvantagens:**
 - Pode ser mais lento para treinar
 - Menos interpretável

SVM:

- **Vantagens:**

- Boa capacidade de generalização
- Efetivo em espaços de alta dimensionalidade
- Tratamento adequado de classes desbalanceadas

- **Desvantagens:**

- Sensível à escala dos dados
- Pode ser computacionalmente intensivo

4.2 Considerações sobre os Dados

- Necessidade de normalização para o SVM
- Importância do balanceamento de classes
- Relevância das features selecionadas

5. Recomendações

5.1 Para Próximos Passos

- Experimentar com outros kernels no SVM
- Ajustar hiperparâmetros dos modelos
- Considerar técnicas de *feature engineering*
- Avaliar ensemble methods combinando os modelos

5.2 Melhorias Potenciais

- Adicionar mais features relevantes
- Implementar validação cruzada
- Explorar técnicas de redução de dimensionalidade
- Testar outros algoritmos como **XGBoost** ou **LightGBM**

6. Conclusão

Os três modelos implementados oferecem diferentes abordagens para o problema de previsão de resultados de futebol. Cada um tem suas particularidades e pode ser mais adequado dependendo dos objetivos específicos do projeto.

A análise comparativa sugere que modelos mais complexos como Random Forest e SVM podem oferecer melhor performance, mas a escolha final deve considerar também aspectos como interpretabilidade, velocidade de treinamento e requisitos computacionais.

O próximo passo seria **refinar os modelos mais promissores**, otimizando seus hiperparâmetros e possivelmente combinando-os em ensembles para melhorar ainda mais a performance preditiva.