

Tópicos Especiais em Engenharia Elétrica A (75H)

# Utilização de Algoritmos *Random Forest* e *XGBoost* para a Predição de Resultados Esportivos.

Aluno Lucas Peres de Medeiros  
Professor Lucian Soares Schiavon

Instituto Federal de Educação, Ciência e  
tecnologia Sul-rio-grandense  
Campus Pelotas

Pelotas, dezembro de 2019



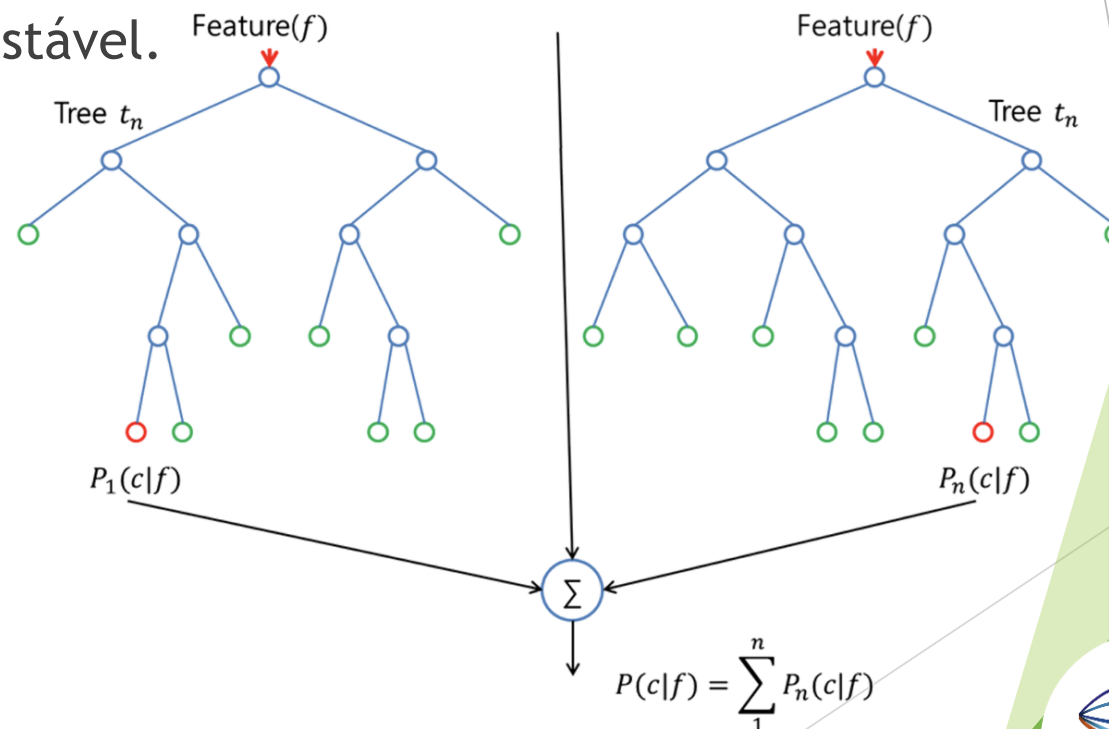
# Motivação

- ▶ Evolução do universo de investimentos com o acesso universal à internet.
- ▶ Mais de **R\$ 7.000.000.000,00** movimentados em apostas esportivas **APENAS EM 2018.** \*
- ▶ Possibilidade de utilização da “Inteligência Computacional” como ferramentas de redução de riscos.



# Random Forest

- ▶ Algoritmo Flexível
  - ▶ Combinação de Árvores de Decisão
    - ▶ Maior acurácia e mais estável.

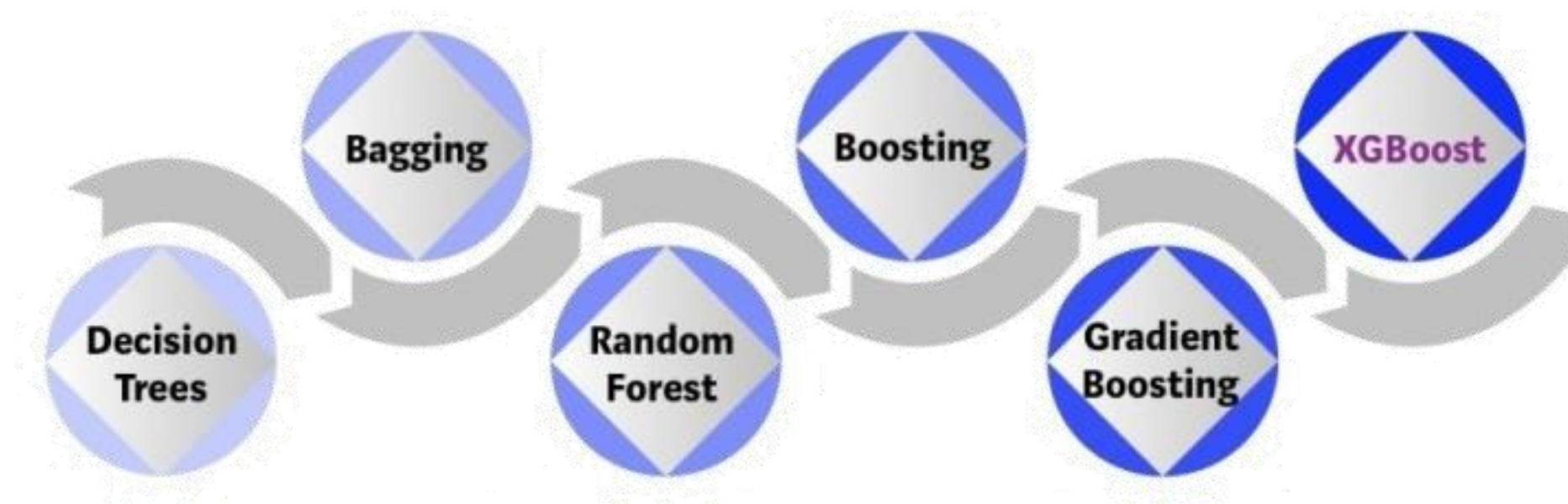


# XGBoost

- ▶ Algoritmo de aprendizado de máquina baseado em árvore de decisão.
  - ▶ Utiliza uma estrutura aprimorada de *Gradient boosting*.
- ▶ Numa média geral o algoritmo tem sido o mais vitorioso no *Kaggle*.
- ▶ Suporte para as principais linguagens de programação
  - ▶ Python, C++, R, Julia, etc., independente de S.O.



# *XGBoost*



# Dataset Base

- ▶ “Historical Football Results and Betting Odds Data” da Premier League 2018/2019.
  - ▶ Estatísticas por jogo.
  - ▶ Dados de casas de apostas.
  - ▶ Etc.
- ▶ **PROBLEMA:**
  - ▶ As estatísticas só são conhecidas após a partida, impedindo a utilização das mesmas como elementos de entrada para a predição.



# Dataset Desenvolvido

- ▶ Média até o momento da partida:
  - ▶ Pontos conquistados por confronto
  - ▶ Gols marcados por confronto
  - ▶ Finalizações por confronto
  - ▶ Finalizações no gol por confronto
  - ▶ Mais de um gol e meio por confronto
  - ▶ Mais de dois gols e meio por confronto



# Dataset Desenvolvido

- ▶ Possíveis saídas
  - ▶ Equipe vencedora (Mandante, empate, visitante)
  - ▶ Equipe mandante ganhou? (Sim / Não)
  - ▶ Equipe visitante ganhou? (Sim / Não)
  - ▶ Houve mais de um gol e meio? (Sim / Não)
  - ▶ Houve mais de dois gols e meio? (Sim / Não)





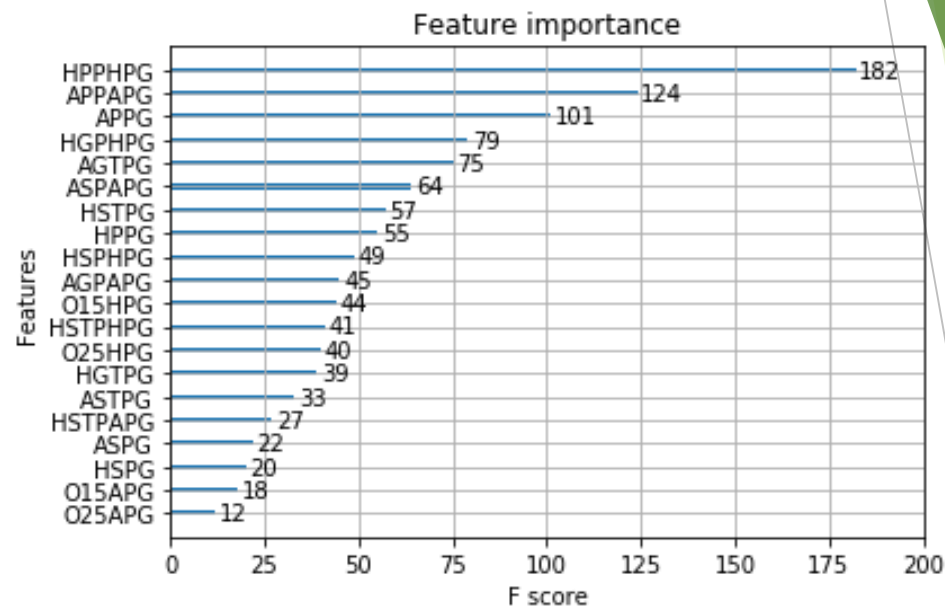
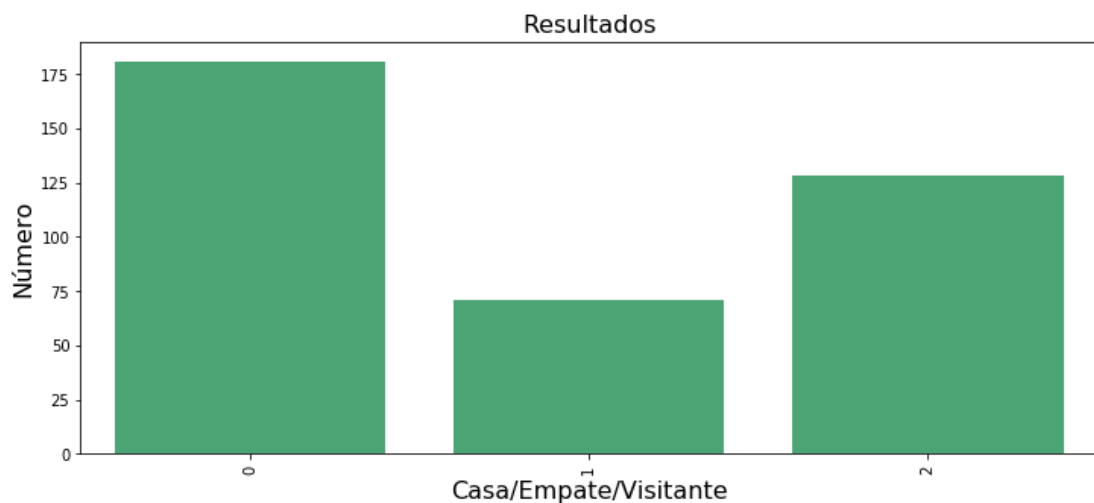
# Aplicação

- ▶ Dados lidos nos algoritmos
  - ▶ *Random Forest*: Desenvolvido pelo autor
  - ▶ *XGBoost*: Aplicação para Python
- ▶ *Predição*:
  - ▶ Validação cruzada: *K-fold* ( $k = 10$ )



# RESULTADOS

## Equipe Vencedora



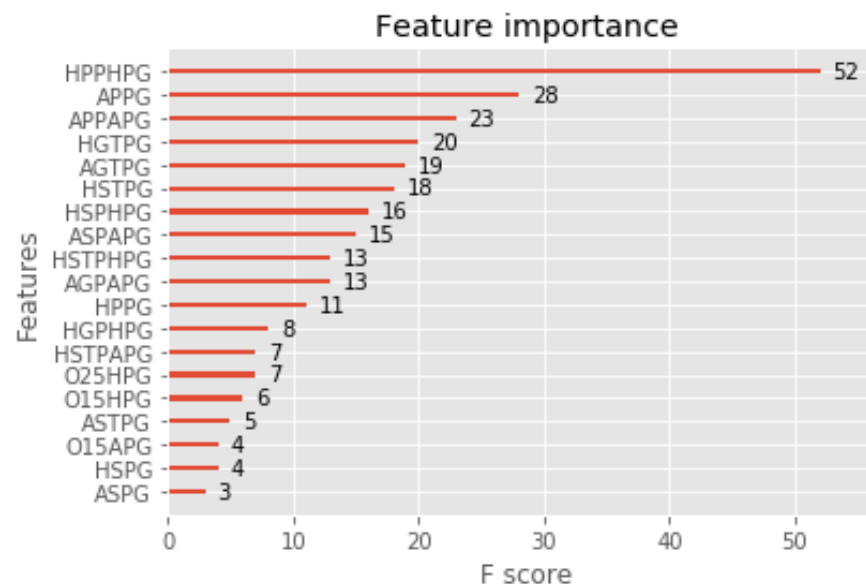
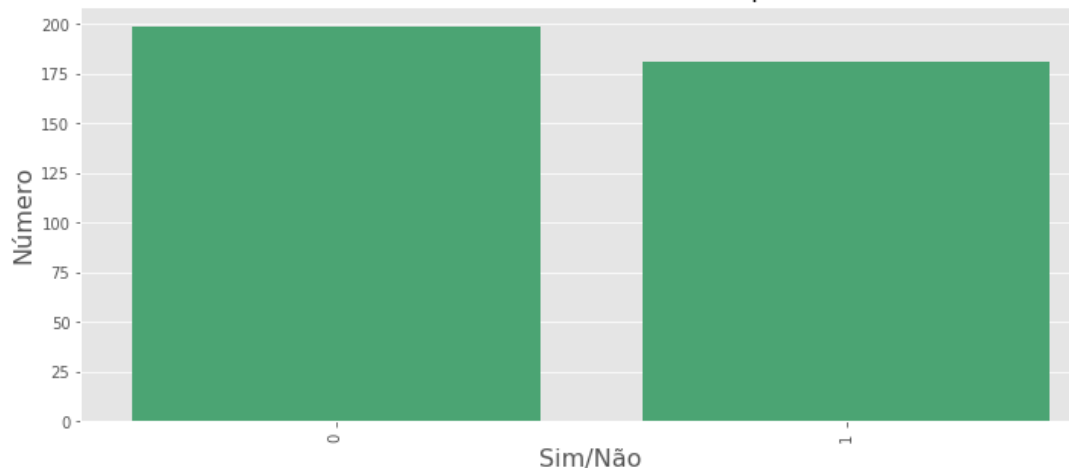
*Random Forest: 63,1 %*

*XGBoost: 64,74 %*

# RESULTADOS

## Mandante ganha?

Vitórias do Time da Casa na Temporada



*Random Forest: 74,74 %*

*XGBoost: 75,26 %*



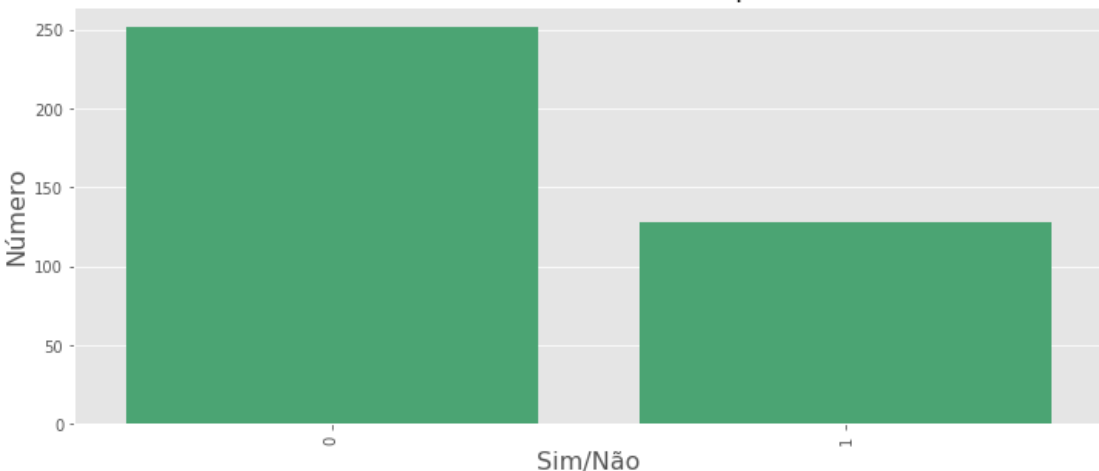
ENGENHARIA  
ELÉTRICA



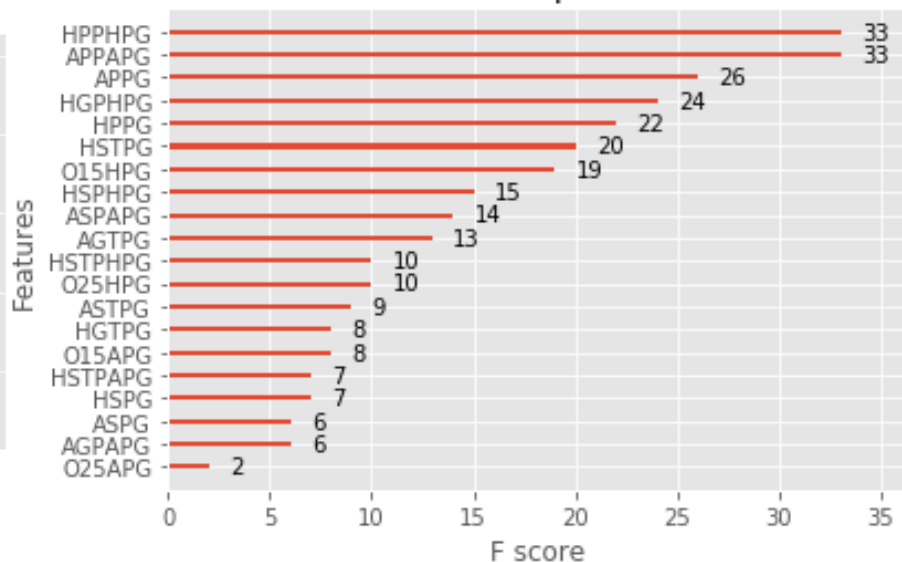
# RESULTADOS

## Visitante ganha?

Vitórias do Time Visitante Temporada



Feature importance



*Random Forest: 78,42 %*

*XGBoost: 80,26 %*

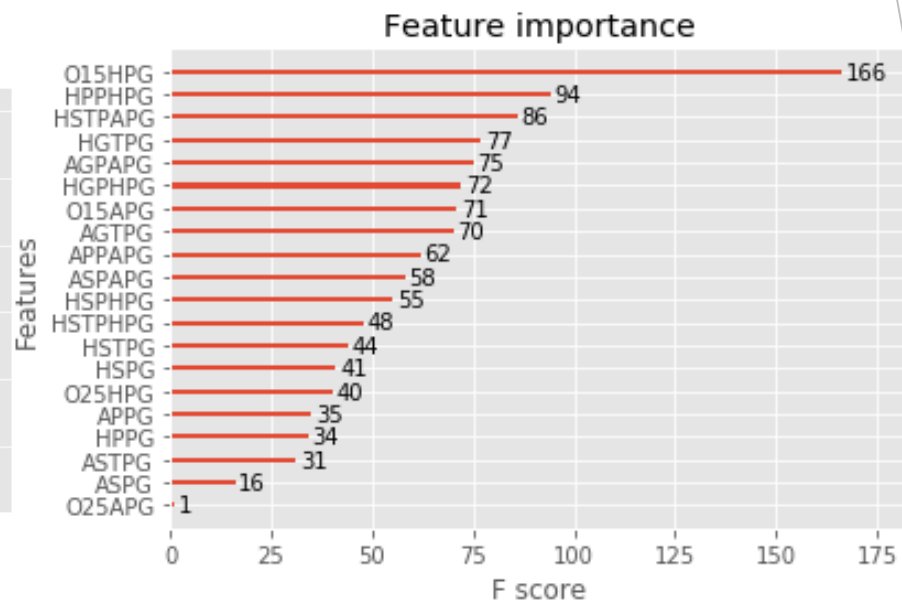
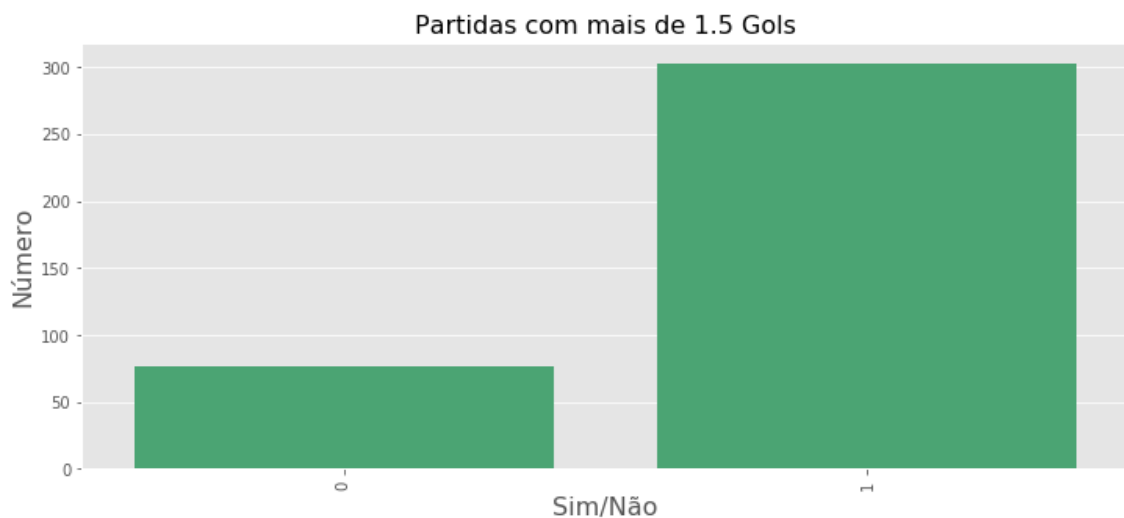


ENGENHARIA  
ELÉTRICA



# RESULTADOS

## Over 1,5 gols

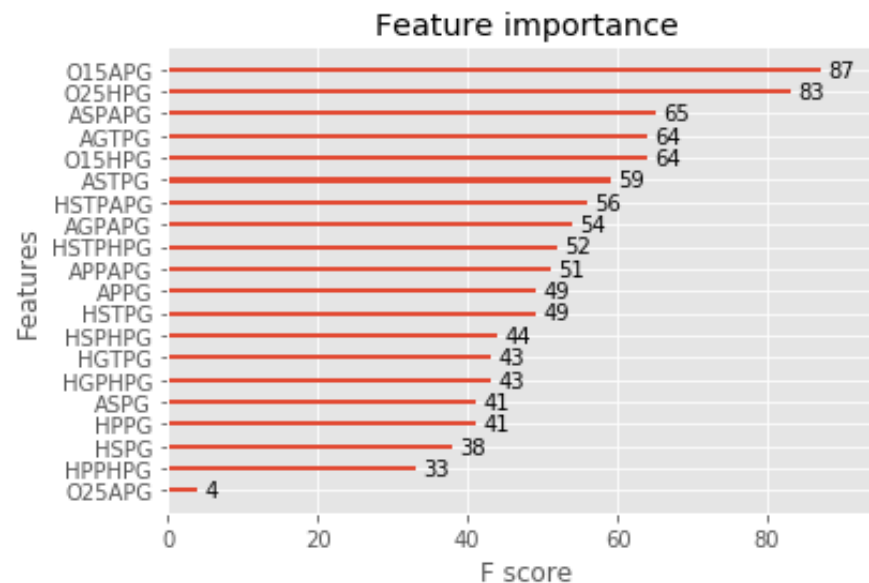
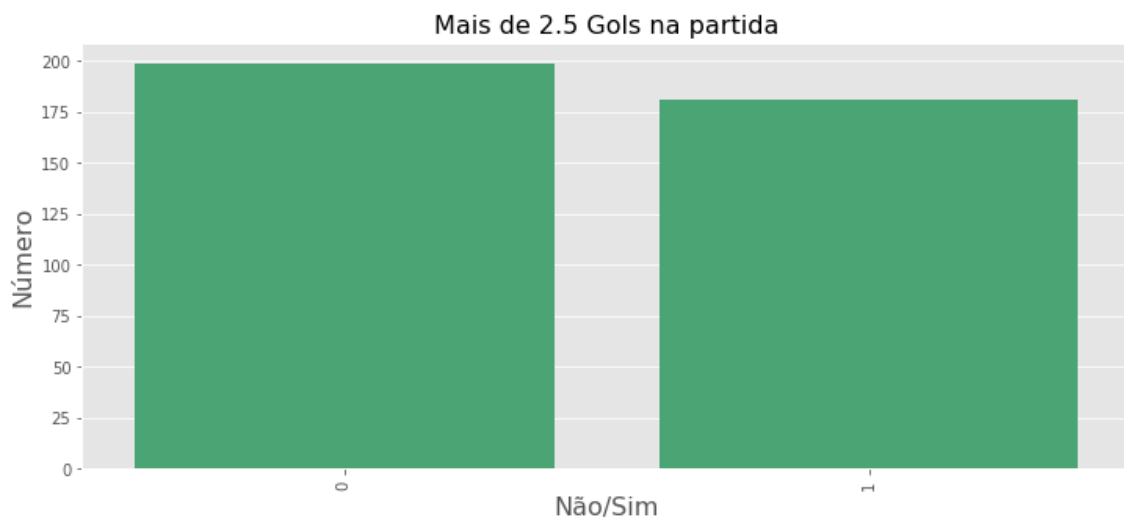


*Random Forest: 79,02 %*

*XGBoost: 80,52 %*

# RESULTADOS

## Over 2,5 gols



*Random Forest: 54,41 %*

*XGBoost: 73,42 %*

# Quadro Comparativo

	Quesito				
Algoritmo	Vencedor	V. Mandante	V. Visitante	Over 1.5	Over 2.5
Random Forest	63,10%	74,74%	78,42%	79,20%	54,21%
XGBoost	64,70%	75,03%	80,26%	80,52%	73,42%
Diferença	1,60%	0,29%	1,84%	1,32%	19,21%

# Considerações Finais e próximos passos

- ▶ Resultados
  - ▶ Ligeira vantagem do XGBoost
  - ▶ Bons resultados em ambos os algoritmos
- ▶ Aprimoramento do *dataset*
  - ▶ Mais informações e maior validade
- ▶ Estudo sobre a auto calibração do XGBoost
  - ▶ Melhor resultados





# Referências

ANICETO, Maria. Classificadores Ensemble, tipos Bagging e Boosting. 2017. Disponível em: <https://lamfo-unb.github.io/2017/09/27/BaggingVsBoosting/>. Acesso em: 15 dez. 2019.

Pedregosa et al. Scikit-learn: Machine Learning in Python. JMLR 12, pp. 2825-2830, 2011.

J. D. Hunter. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, 2007.

SILVA, Josenildo Costa da. Aprendendo em uma floresta aleatória. Machina Sapiens. 2018. Disponível em: <https://medium.com/machina-sapiens/o-algoritmo-da-floresta-aleat%C3%B3ria-3545f6babdf8>. Acesso em: 14 dez. 2019.



Tópicos Especiais em Engenharia Elétrica A (75H)

# Utilização de Algoritmos *Random Forest* e *XGBoost* para a Predição de Resultados Esportivos.

## Obrigado!

Pelotas, dezembro de 2019

