

Projeto Ames

A ser entregue dia 22/10/2023, final do dia.

Em duplas.

Descrição

Você vai construir um projeto de regressão sobre o dataset “Ames”

(<https://www.openintro.org/book/statdata/?data=ames>) de imóveis na cidade de Ames, Iowa, Estados Unidos.

O objetivo do projeto é demonstrar suas habilidades na construção de um projeto de regressão, para além de leitura de dados e preparação simples de dados.

No repositório <https://github.com/FabioAyresInsper/ames> você encontra três *notebooks* Jupyter que fazem a leitura dos dados, a análise exploratória e um pouco de preparação de dados, e uma regressão linear simples. Você pode usar livremente este material em seu projeto.

Você deverá, portanto:

- Ler, entender e executar os *notebooks* do repositório acima.
- Realizar *feature engineering* conforme necessário OU JUSTIFICAR que a *feature engineering* já feita é suficiente (ou seja, explicar porque).
- Construir mais de um modelo de regressão (que já pode incluir o modelo simples dado no repositório do professor) e fazer o ajuste de hiperparâmetros destes.
- Comparar os desempenhos de modelos de modo adequado.
- Escolhido o melhor modelo, treiná-lo no conjunto de treinamento e medir o desempenho no conjunto de teste, para estimar o desempenho de generalização do modelo.
- Escrever suas conclusões à respeito do que foi aprendido acerca do modelo, por exemplo:
 - Quais as consequências do desempenho do modelo final para a aplicação de negócios?
 - Quais features são mais importantes na determinação do preço do imóvel? Note que essa pergunta pode ou não ter resposta, dependendo das capacidades dos modelos de regressão que você escolher.
- (Bonus 1pt): Fazer o “*deploy*” do modelo em uma API REST simples. Basta um *endpoint*, que recebe um JSON com os atributos de um imóvel e retorna o preço predito. Você deverá retreinar o modelo com os dados completos para esse “*deploy*”

Rubrica

Os itens a serem avaliados no projeto são:

FEAT – Fazer *feature engineering* ou justificar adequadamente sua ausência (ou seja, está usando a preparação de dados fornecida pelo professor)

HYPER – Para cada modelo, fez ajuste de hiperparâmetros adequadamente

MODEL – Treinou e comparou adequadamente os modelos para selecionar o melhor modelo

PERF – Análise de desempenho do modelo. **NÃO SERÁ EXIGIDA NENHUMA PERFORMANCE EM PARTICULAR, ESTE PROJETO NÃO É UMA COMPETIÇÃO DE DESEMPENHO.**

BONUS – Deploy

- I – Insuficiente
 - Não entregou, ou entregou abobrinha
- D – Em desenvolvimento
 - Faltou um de FEAT, MODEL, PERF
- C – Mínimo aceitável
 - Fez FEAT, MODEL, PERF
 - Qualidade mínima
- B – Esperado
 - Fez FEAT, HYPER, MODEL, PERF
 - Qualidade mínima
- A – Excepcional
 - Fez FEAT, HYPER, MODEL, PERF
 - Qualidade de apresentação, clareza de código, tudo bem feito

BONUS aumenta 1 pt (e.g. de B para B+, etc).