

Mini-projeto de Aprendizagem de Máquina

Professor: Evandro Costa

Desenvolva, usando a biblioteca Scikit-Learn (ou alguma similar, de sua escolha), um Sistema de Predição para resolver um problema relativamente complexo, considerando alguma base de dados em repositórios, tais como: Kaggle ou UCI, ou eventualmente uma outra fonte, em uma das áreas seguintes:

Educação: (Ex.: Predict students' dropout and academic success),

Saúde: (Ex.: diabetes ou outra base no Kaggle),

Gestão de risco em projetos de software, informando entre risco alto, moderado e baixo (procurar uma base de dados).

Como parte da documentação da solução, apresente a descrição do problema a ser resolvido, indicando claramente os objetivos da aplicação, a especificação da base de dados, cumprindo as etapas seguintes: preparação de dados (análise exploratória e pré-processamento de dados):

1.1) Preparação de dados: análise exploratória e pré-processamento de dados,

1.2) Modelos classificadores: 4 algoritmos single, sendo 3 white-box (Árvore de Decisão (C4.5 ou CART), Regressão Logística, PRISM) e 1 black-box (escolhido entre SVM ou Rede Neural MLP ou alguma outra abordagem com Rede Neural) e 2 ensemble (combinação de classificadores): Random Forest, XGBoost,

1.3) Avaliação de desempenho de cada algoritmo usado,

1.4) Comparação entre os 2 modelos classificadores single que apresentem melhores indicadores de desempenho via acurácia ou F1-Score, pegando o melhor single e comparando com o melhor ensemble.

1.5) Pensando em melhoria na explicabilidade dos modelos, aplique os métodos: (i) SHAP (SHapley Additive exPlanations) e (ii) LIME (Local Interpretable Model-Agnostic Explanations), na saída dos modelos (melhor single e melhor ensemble, conforme 1.4), verificando e discutindo o que agregaria de mais esclarecimento nos aspectos de explicabilidade.

Obs.: (1) Em cada uma das etapas mencionadas, recomendamos, como apoio, atenção ao conteúdo do livro e ao que foi discutido no desenvolvimento da disciplina.

(2) O problema a ser resolvido, sendo diferente dos indicados, poderia ser proposto pela equipe, mas precisa ser aprovado pelos professores da disciplina.