

Universidade de São Paulo
Instituto de Ciências Matemáticas e de Computação
SCC0276 - Aprendizado de Máquina
Trabalho Prático 1 – Tarefas Preditivas

O trabalho consiste em escolher um conjunto de dados público para realizar uma tarefa de predição, seja ela classificação ou regressão. Escolham conjuntos de dados que sejam desafiadores, de preferências contendo vários atributos e muitos exemplos. Há vários conjuntos de dados disponíveis no Kaggle - <https://www.kaggle.com/datasets>, mas vocês também têm liberdade para escolher qualquer conjunto de dados que acharem interessante. Tentem também entrar em contato com pessoas que vocês conhecem em empresas para ver se as empresas podem fornecer algum conjunto de dados para solucionar tarefas preditivas que sejam do interesse delas.

Uma vez escolhido o conjunto de dados e definida a tarefa a ser executada, vocês devem escolher o melhor método de aprendizado de máquina para a realização da mesma. Façam uma análise do conjunto de dados, verificando desbalanceamento de classes no caso de classificação, tipos de variáveis e medidas estatísticas de cada variável. Façam o pré-processamento necessário de acordo com a necessidade do método escolhido.

Para a avaliação do método, escolham as medidas de avaliação mais adequadas, e também as técnicas de particionamento de dados. Não utilizem hold-out (única divisão em treino e teste), pois a divisão pode ser benéfica para seu método, fornecendo uma estimativa otimista da taxa de acerto. Utilizem técnicas mais robustas, como por exemplo o k-fold cross-validation.

Quanto aos hiperparâmetros do método escolhido, verifique a melhor maneira de escolher seus valores adequadamente. A escolha do método de aprendizado de máquina e de seus valores de hiperparâmetros faz parte do trabalho.

A entrega do trabalho deve ser realizada no e-disciplinas USP, e conter o seguinte material:

- Código bem documentado feito no Google Colaboratory;
- Um relatório de 8 a 10 páginas, em formato de artigo científico, contendo i) resumo e motivação, ii) introdução, iii) fundamentação teórica do método escolhido, iv) metodologia adotada nos experimentos com descrição das medidas de avaliação e conjunto de dados, v) descrição e análise dos experimentos, e vi) conclusão;
- Slides de uma apresentação de no máximo 20 minutos sobre o trabalho realizado;
- Vídeo de 10 minutos explicando todas as fases de desenvolvimento do projeto e resultados.

Templates de artigo para Overleaf e .docx. Vocês devem seguir esses templates.

- Overleaf: <https://www.overleaf.com/latex/templates/sbc-conferences-template/blbxwjwzdng>
- Docx: <https://www.sbc.org.br/wp-content/uploads/2024/07/modelosparapublicaodeartigos.zip>

A avaliação consistirá na análise do material entregue: 1/3 da nota para o código e documentação interna, 1/3 da nota para o relatório, e 1/3 da nota para a apresentação e vídeo.

O trabalho deve ser feito em grupos de 2 a 3 pessoas. **Prazo de entrega: 15/05/2025.**