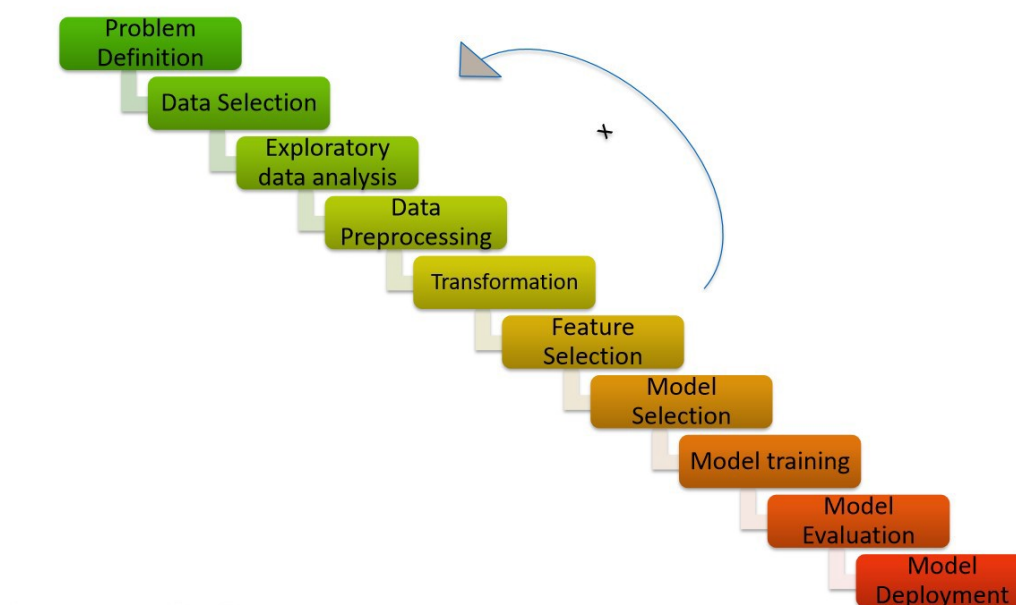
	Instituto Federal de Educação Ciência e Tecnologia – Sudeste de Minas Gerais		
	Disciplina: Machine Learning		Data: 26/06/2023
	Professor: Gustavo Willam Pereira		
	Trabalho Prático III	Grupo: 02 Alunos	Valor: 3,5 pontos
	Forma de Entrega: SIGAA		Data de Entrega: 07/07/2023

O ciclo de vida de um projeto de Machine Learning pode ser definido através das seguintes etapas:



Problem Definition: Definição do problema, qual o seu objetivo.

Data Selection: Aquisição dos dados, como eles serão coletados

Exploratory data analysis: Análise exploratória dos dados, verificar correlações, gerar gráficos, tentar descobrir ‘insights’.

Data Preprocessing: Pré processamento dos dados. Eliminar outliers, Tratar dados categóricos, Substituir/Eliminar dados ausentes (NaN)

Transformation: Verificar se precisa de realizar alguma padronização/escalamento nos dados.

Features Selection: Realizar seleção das features: Correlação de Pearson, Spearman, RFE, RFECV, KBestFeatures, Features Importances, dentre outras. (outras técnicas ver em: https://scikit-learn.org/stable/modules/feature_selection.html)

Model Selection: Realizar a seleção do modelo ideal com base na validação cruzada, testar hiperparâmetros, definir as métricas de performance para os modelos. (https://scikit-learn.org/stable/model_selection.html).

Model training: Realizar o treinamento a partir do modelo selecionado e das features selecionadas.

Model Evaluation: Acompanhar a evolução do modelo quando aplicado a dados de testes, se necessário realizar ajustes.

Model Deployment: Realizar o Deploy do modelo colocando-o em produção para utilização do cliente.

Objetivo desse trabalho:

Escolha uma base de dados de sua preferência (sua base de dados deve ser diferente de todas que foram utilizadas na disciplina) e aplique as etapas de um projeto de Machine Learning. Seu problema pode ser para: Classificação, Regressão ou Agrupamento.

Proponha 4 soluções para o problema e ao final informe a mais eficiente.

- a) Escolha um modelo de Machine Learning de sua preferência e aplique-o sem utilizar seleção de features ou redução de dimensionalidade. Simplesmente aplique o modelo e encontre o valor do R^2 (para Regressão) ou Acurácia (para Classificação).
- b) Utilize a técnica de PCA ou LDA para redução de dimensionalidade e aplique o mesmo modelo de Machine Learning utilizado na letra a.
- c) Utilize uma técnica de Seleção de Features (pode ser uma apresentada em sala de aula) ou qualquer outra de sua preferência e aplique o mesmo modelo de Machine Learning utilizado na letra a.
- d) Utilize a biblioteca Pycaret para a solução do seu problema.

Ao final informe qual foi a melhor estratégia (que produziu o melhor resultado) para a solução do seu problema.

Data da Entrega: 07/07/2023

Datas para apresentação: 07/07/2023, 12/07/2023.