

Esta atividade visa a utilização de um sistema IIoT em uma **suposta** planta de processo e separador trifásico de água, gás e petróleo (Figura 2). O fluido produzido no reservatório é uma mistura de hidrocarbonetos, água, areia e outras impurezas. Quando o poço começa a produzir, com a queda da pressão que vai ocorrendo até o hidrocarboneto que era quase todo líquido chegar à superfície, parte dele passa para a forma gasosa. Porém, o principal interesse na produção é o óleo e, por isso, é feita a separação. Separar o gás do óleo em condições controladas e remover água, sal, e outras impurezas até que o óleo fique livre desses e suficientemente estável para ser transferido é necessário. Isto é feito nos chamados sistemas de produção que são um conjunto de sistemas, tais como compressão

de gás e tratamento de água oleosa, que podem ser instalados em terra (onshore) ou no mar (offshore).



Figura 2 - Planta de Processo e Separador Trifásico com IIoT incorporado.

Base de Dados da Planta de Processo

A base de dados usada na atividade é gerada sinteticamente, simulando um sistema de aquisição de temperaturas em graus Celsius por toda a planta de processo.

São 50 dispositivos coletando 1 leitura por segundo durante 48 horas (2 dias) do processo, resultado em 172.800 intervalos com o total de 8.640.000 leituras.

Durante essas leituras ocorreram 50% de perda de rendimento instantâneo no processo, devidamente identificados como classe 1.

Características da base de dados:

- Faixa de temperatura de 0 a 100 graus Celcius;
- Cada dispositivo é um atributo (coluna), resultando em 50 atributos;
- Cada intervalo de tempo (linha) resulta em um caso de amostra, resultando 172.800 amostras;
- Existe um atributo alvo que será tutor do processo de identificação de rendimento, sendo 0 como normal e 1 como falha;
- São 8.640.000 totais para a base de dados.

Base de dados para baixar:

<https://drive.google.com/file/d/17XOsLTMn3B2F9odJxGysu0J8rW9tJj6S/view?usp=sharing>

Etapas da Atividade

A atividade será realizada com as seguintes etapas:

- (01/04/2024) Análise da Base de Dados
 - Características dos atributos
 - Dados faltantes
 - Correlação
 - Histogramas
 - Detectando anomalias nos dados
- (08/04/2024) Capacidade de Aprendizado sobre a Base
 - Utilizar a base de dados filtrada da etapa 1;
 - Separar o conjunto de dados em treinamento e testes;
 - Utilizar 10 metodologias de aprendizado classificador na base de dados;
 - Avaliar os resultados com métricas recomendadas com o problema
 - Ordenar decrescente os melhores modelos avaliados pelas métricas;
- (15/04/2024) Pipeline, pré-processamento e redução de dimensionalidade
 - Incorporar um pipeline no aprendizado;
 - Utilizar pré-processamento na base de dados;
 - Reduzir a dimensionalidade da base;
 - Incorporar a validação cruzada para melhor avaliar as técnicas;
 - Reavaliar as metodologias e modelos com essa nova abordagem.
- (29/04/2024) Seleção de Metodologias e Hiperparâmetros de Aprendizado de Máquina
 - Incorporar as metodologias XGBoost e LightGBM;
 - Utilizar as 5 melhores metodologias da etapa anterior;
 - Utilizar o recurso de grade de hiper-parâmetros e validação cruzada para melhorar resultados com métricas selecionadas;
 - Identificar os 2 melhores modelos obtidos.

Entrega no SIGAA

Cada etapa terá uma tarefa específica marcada no SIGAA, com pontuação de 0 a 10 conforme a descrição solicitada.

O arquivo entregue deverá ser o Notebook Python em formato IPYNB já executado.

Não vou aceitar PDF, pois tem alunos entregando apenas resultados sem código fonte.