**PROCESSO SELETIVO CCIA – HACKATON**

Objetivo: Realizar a criação de um modelo de previsão da velocidade do vento a 50 metros de altura para 3h à frente a partir de dados de torres anemométricas da plataforma SONDA – Estação de Petrolina.

Os dados para treinamento e teste serão disponibilizados. Serão os arquivos chamados “treino.csv” e “teste.csv”.

**1º ETAPA:**

* Leia o arquivo de dados;
* Altere os nomes das colunas “temratura-50m” e “velocidde-vento-25m” para “temperatura-50m” e “velocidade-vento-25m”;
* Exiba os 5 primeiros elementos do *dataset* de treino e de teste;
* Exiba a quantidade total de elementos nas tabelas e a média dos valores da coluna velocidade-vento-50m;
* Exiba qual Data apresentou o maior valor de velocidade-vento-50m;
* Realize análise exploratória dos dados a seu próprio critério. Apresente os gráficos que julgar interessante para o problema em questão e justifique a escolha deles;

**2º ETAPA:**

* Realize algum tipo de tratamento de dados para lidar com os dados faltantes;
* Considere as variáveis temporais: ano, dia, data-hora e minuto. Caso deseje, considere transformar as colunas para que elas sejam representadas por valores contínuos. Caso faça isto, disponibilize a função para o tratamento dos dados;
* Divida os dados de treinamento em X\_train e y\_train. Considere que:
  + X\_train: São os dados de treinamento;
  + y\_train: São os dados de velocidade-vento-50m do *dataset* de treinamento, porém a 3 *timesteps* à frente (para cada linha do
  + X\_train, o y\_train deverá representar a velocidade-vento-50m das 3 próximas horas);
* Separe uma parte dos dados de X\_train e y\_train para validação (X\_val e y\_val);
* Divida os dados de teste em X\_test e y\_test. Utilize o mesmo raciocínio da divisão de dados de treinamento.
* Escalone os dados de entrada para que fiquem dentro da faixa de valores entre 0 e 1;
* Salve o escalonador em arquivo (.pkl);

**3º ETAPA:**

* Desenvolva um modelo que aprenda a predizer a velocidade-vento-50m das 3 próximas horasa partir dos dados de treinamento;
* Apresente o gráfico de *loss* de treinamento e validação do modelo. Mostre que o modelo não apresenta over/underfitting;
* Avalie quantitativamente o desempenho final do modelo desenvolvido a partir dos dados de teste, utilizando as métricas *Mean Squared Error* (MSE), *Mean Absolute Error* (MAE), coeficiente de correlação Pearson r e o coeficiente de determinação R²;
* Apresente gráficos e tabelas relacionados ao desempenho do modelo desenvolvido, explicando os resultados;
* Apresente em uma tabela os valores médios das métricas para cada horizonte de previsão, bem como os valores médios finais de cada métrica;
* Salve o modelo desenvolvido em arquivo (.h5).

**Observações:**

1. Utilizar Python, Tensorflow e Keras;
2. Utilizar Jupyter notebook para criação, apresentação e entrega do trabalho;
3. Comentar e estruturar o trabalho com elegância e qualidade de forma a permitir a sua correta interpretação e avaliação.
4. Enviar o notebook desenvolvido para o E-Mail [ia@fieb.org.br](mailto:ia@fieb.org.br). O título do E-Mail deverá estar no formato:
   * + [HACKATON-Jul2021] Seu\_nome\_completo
5. Caso realize qualquer alteração (*reshape*) no formato dos dados, disponibilize a função para tal;
6. Não é permitido realizar *Data Augmentation;*
7. Qualquer processo de *Feature Augmentation* deve ser detalhado e tratado como uma função que recebe os dados originais e devolve os dados transformados;
8. Ao final, o modelo criado será testado em um *dataset* não disponibilizado para avaliação final do Hackaton. Portanto, tenha isto em mente ao desenvolver o trabalho de modo a permitir utilizar a sua solução facilmente para este processo de avaliação final;
9. Envios realizados fora do prazo não serão considerados. A data e hora de referência será a de recebimento do E-Mail em nossa caixa de entrada. Caso a sua solução não esteja completa no prazo, sugerimos que envie mesmo assim, pois ela pode vir a ser considerada.