

Pós graduação

Desafio:

Fazer setorização de mercado por correlação de alta frequência (minuto a minuto) .
Implementar artigos de reamostragem de séries temporais caóticas.

LUZ

Desafio: Fazer parte do time nos sprints de entrega de versão, sem ter atuado no Front End e C# antes.

Stattus4

- Tratamento de séries temporais audio de hidrometro:
 - Series reais x mock de osciladores stochasticos: Spectrograma, Interpolação, Reamostragem com Kernel Density Function.
 - Time Series Clustering tomando o inverso da covariancia fitando Kmeans, e vendo quantas séries caiam em cada cluster.
- SQL inserir series em Bancos de dados
- Filtro inverso usando ruido como baseline e audios com ruido para retirar ruido, com Tensorflow 1.0
- Rede Neural para tratamento de audio no Tensorflow 1.0.
- Desafio: Dados insuficientes, base de dados pequena.

Portfólio de projetos (Em andamento)

- Metodologia cíclica de trabalho:
 - Planejamento levando em conta dados de entrada, produto de saída e Tasks.
 - Passar rapidamente pelos estágios do projeto e entregar valor de negócio desde o primeiro ciclo.
 - Premissas do negócio.
 - Descrição dos dados
 - Análise de valores faltante, valores inconsistentes e valores inconsistentes em relação as premissas do negócio.
 - Filtragem de Valores.
 - EDA, identificar outliers, colinearidade e variabilidade dos dados, insights de negócio, mindmap de hipóteses de negócio. Uso de métodos automatizados como pandas profiling e análise.
 - Engenharia de features, adicionar features ou retirar features.
 - Identificar features relevantes, uso de algoritmos como Extra Trees para identificar feature importance.
 - Modelagem do espaço de features, projeções multi-dimensionais
 - Modelo baseline.
 - Métricas do modelo.
 - Modelos candidatos, cross validation e hyperparameter fine tuning.

- Apresentar valor de negócio para a modelagem.
 - Implementação de ETL e colocar o algoritmo em produção na nuvem.
-
- Cross Validation para séries temporais
 - Algoritmo de seleção de features Boruta.
 - Hyperparameter fine tuning com método bayesiano.
 - Modelagem de séries temporais: AR, MA, ARMA, ARCH e GARCH.
 - Espaço de embedding Random Forest + UMAP.
 - Métodos de clusterização, Kmeans e Clusterização Hierárquica.
 - Rankeamento de clientes.
 - Segmentação de clientes.
 - Dashboard Streamlit.
 - Random Forest, XGBoost, Redes Neurais.

Especialização Engenharia de Dados e Big Data (Em andamento)

- Gestão da qualidade de dados.
- Modelagem ER de bancos de dados.