

# Hackathon JPA Agro 2021

## Data Science Research Group - DSRG Universidade Federal de Lavras - UFLA

### Identificação da equipe

**Nome da equipe:** CB Tech

**Integrante 1:** Cecília Aparecida Santos Silva.

**Integrante 2:** Lucas Alexandre Alvarenga Cardoso.

**Integrante 3:** Pedro de Oliveira Annoni Farah.

### Descrição da solução

#### 1. Entendimento do negócio

Para desenvolver a solução, foram buscadas vários exemplos de outras séries temporais e suas principais características como a sazonalidade, tendência e estacionaridade. Além disso, algumas referências do mercado da Polpa Cítrica foram consultadas para que fosse conhecida o comportamento do seu preço ao longo dos anos. Também foi visto o comportamento do preço do mês de Agosto de outros anos para que houvesse uma comparação com os valores obtidos para os próximos 30 dias da série temporal disponibilizada (correspondente ao mês de agosto de 2019)

#### 2. Pré-processamento dos dados

Durante o processo de busca do modelo que melhor se ajustasse aos dados fornecidos, várias técnicas foram utilizadas para o pré-processamento dos dados, como por exemplo, a busca por valores fora do padrão (outliers), a retirada das componentes de sazonalidade e tendência e a suavização da série temporal. Essas técnicas foram utilizadas em varios testes e mostraram-se de qualidade variada, dependendo do modelo treinado. Porém o modelo escolhido para a previsão (WaveNet) se ajustou melhor aos dados em sua forma pura, levando a um erro menor nos testes que foram feitos. Análises gráficas como a decomposição da série em suas componentes sazonais, tendencionais e residuais, além de testes de estacionaridade também foram aplicados e ajudaram na escolha do algoritmo final, que se provou como o melhor dentre as outras opções considerada.

#### 3. Enriquecimento dos dados

Utilizamos o banco de dados disponibilizado com a variável "sold-price" correspondente ao valor de venda da Polpa Cítrica.

## 4. Modelos

Foram utilizados vários modelos diferentes para o teste das previsões. Os principais utilizados foram o ARIMA (Auto-Regressive Integrated Moving Average), Prophet, Holt-Winter, XGBoost, Random Forest, LSTM, GRU, RNN simples, e Wavenet, que foi escolhido como o método utilizado pois apresentou menor erro (RMSE) para a previsão de valores já conhecidos dentro da série. Alguns modelos se saíram melhores que outros por conseguirem se ajustar aos dados não-estacionários da série, o que também foi determinante para que o erro encontrado com a WaveNet fosse o menor. A WaveNet é uma rede neural convolucional com a arquitetura construída para previsão de séries temporais. Originalmente foi construída para a geração de áudio mas acabou se mostrando bastante efetiva para previsão de valores.

## 5. Avaliação da solução

O modelo foi avaliado ao separar os dados em treino e teste e assim calculado a métrica RMSE com a previsão e o valor real esperado. O método utilizado para separação de dados em treino e teste foi baseado em janelamento. Para cada janela de dados utilizada como previsores, havia uma janela alvo com os dados a serem preditos. Foram feitos vários testes e os melhores resultados foram obtidos com janelas de previsores de tamanho 151 e janelas de valores a serem preditos de tamanho 30. A capacidade preditiva do modelo foi avaliada com a métrica RMSE e a capacidade de ajuste aos dados. Foi observado que os resultados previstos dos testes se assemelhavam bem aos esperados.

## Referências

A. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. Wavenet: A generative model for raw audio. arXiv:1609.03499, 2016.

Géron, Aurélien. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media, 2019.

REIS, Felipe. Carta Insumos - A polpa cítrica, o milho e o custo de produção em 2020. [S. l.], 27 dez. 2019. Disponível em:

<https://www.scotconsultoria.com.br/noticias/cartas/51795/carta-insumos—a-polpa-citrica-o-milho-e-o-custo-de-producao-em-2020.htm>. Acesso em: 23 fev. 2021.