



## Atividade 1: Competição de Previsão

**Danilo Kawaguchi (12558763), Felipe Adrian (12559020), Henrique Felix (11554214), Hércules Akira (13686880), Kaique Albuquerque (11544860), Lucas Alexandre (11322900), Luiz Felipe (12693455)**

### 1 Descrição

O trabalho teve como objetivo tomar as 11 séries temporais fornecidas, realizar uma análise de seus comportamentos e a partir disso, implementar diferentes métodos de previsão para obter os próximos 12 dados com maior acurácia possível. Para esse propósito, o RMS foi o teste estatístico usado para arbitrar a escolha de melhor modelo.

Cada membro utilizou um tipo de tratamento estatístico, treinando e testando seus modelos com diferentes porcentagens dos dados e usando diferentes modelos de previsão, entre as técnicas testadas estavam: interpolação de dados, diferentes regressões lineares, aplicação de log retorno aos dados, k-nearest neighbors e a biblioteca scikit learn. Os dados foram divididos em conjuntos de treino e teste, com proporções variando entre 80/20 e 95/5 dependendo do modelo. Reservou-se os últimos doze dados para fins de comparação entre os modelos, comparando os resultados de cada abordagem assim três deles se sobressaíram.

## 2 Modelos Utilizados

### 2.1 Rede Neural

Um dos modelos com menor RSME numa das séries temporais foi uma RNA criada a partir do modelo apresentado em aula. Esse modelo teve o melhor desempenho na série temporal #11.

Foi definida uma topologia de 7 camadas, na primeira é feita a normalização dos números recebidos, as 5 camadas seguintes são densas, tem 100 neurônios cada e utilizam a função de ativação ReLU. A última camada serve para a saída de dados, ela possui 12 neurônios e função de ativação sigmoide. Foi usado 80% do conjunto de dados para teste e a previsão foi realizada a partir dos últimos 12 dados das séries temporais.

### 2.2 Métodos ARIMA / SARIMA, Theta e Holt-Winters

Para garantir que cada teste estatístico e que cada parâmetro fosse bem ajustado foi feito um código que manualmente checava e garantia a validade de cada passo do processo, assim adequadamente maximizando/minimizando os coeficientes desejados. Partindo dos dados tratados dessa maneira 4 métodos de previsão de séries temporais foram empregados e comparados.

Inicialmente checou-se a estacionariedade através do método ADF, diferenciando a série sazonalmente e regularmente até que seja garantido que a série resultante seja estacionária. Em seguida foram analisados os resíduos usando o teste Ljung-Box para verificar se cada ponto tem valor independente dos outros (portanto qualificando os resíduos como ruído) ou se os resíduos seguem uma tendência.

Além disso também foram plotados gráficos Q-Q para garantir que os resíduos seguissem uma distribuição normal e que o modelo ARIMA pode ser devidamente aplicado. Por fim foi feito o diagrama de autocorrelação para determinar se era possível encontrar efeitos de sazonalidade nas séries e então foram fixados manualmente os hiperparâmetros para o ARIMA e SARIMA.

Tendo feito todo esse processamento, treinou-se os modelos mencionados no cabeçalho e eles foram usados para prever os 12 elementos finais das séries temporais. Para cada um deles foi determinado seu AIC, MAPE e RMSE para determinar qual deles ge-

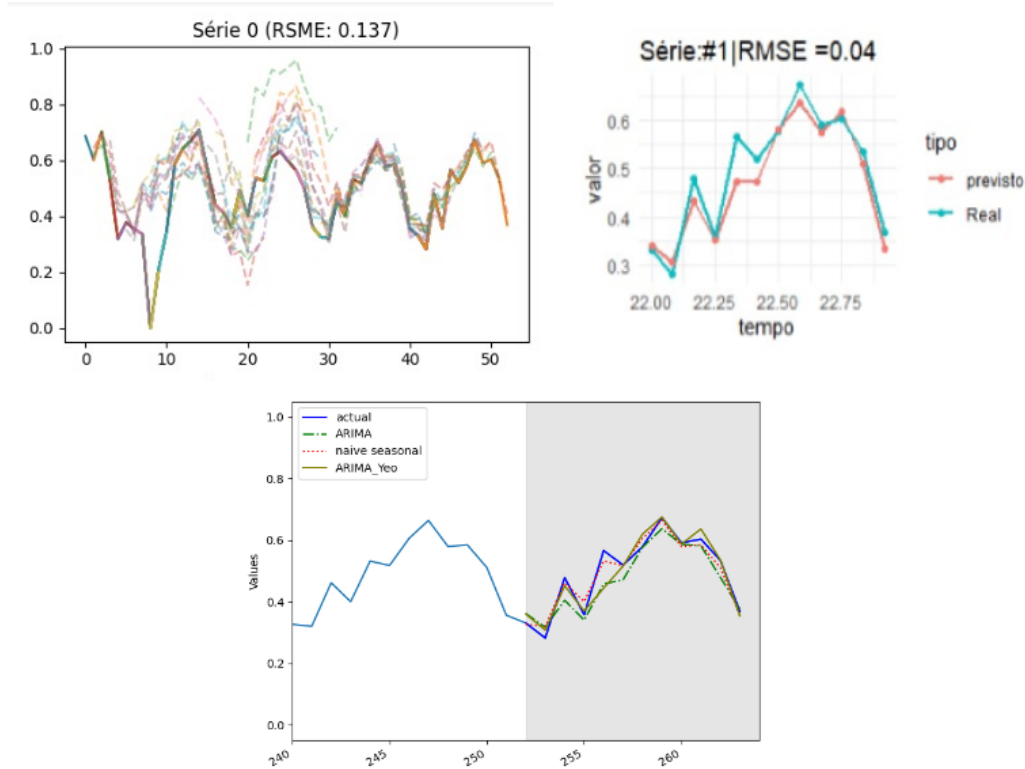


Figura 1: Comparação entre os resultados obtidos pelos 3 diferentes métodos de previsão para uma mesma série

rava melhor resultado para cada conjunto de dados, que resultou no melhor modelo entre os 3 discutidos, tendo menor RMSE nas séries 1,3,6,7,8,10 e menor WRMSE em geral.

### 2.3 Auto ARIMA Sazonal

Semelhantemente ao último método, esse método também resulta da aplicação do modelo ARIMA com observações de sazonalidade, entretanto enquanto o modelo antigo construiu seus testes e veri-

ficou os parâmetros manualmente, essa abordagem utiliza uma das bibliotecas padrão do R que automatiza o treinamento do modelo. Foi usado 95.5% dos dados no treino e após serem testados com os dados restantes, os resultados por esse método produziram o menor RMSE entre todos os modelos 36% das vezes, tendo tido o menor RMSE para as séries 9, 4, 2 e 5.

Como a biblioteca forecast automatiza o processo ARIMA, dispensou-se a necessidade de manualmente escolher os hiperparâmetros (p,d,q).