



UNICAMP

Universidade Estadual de Campinas

Faculdade de Engenharia Elétrica e de Computação

IA048 – Aprendizado de Máquina

23 de março de 2024

Docentes: Levy Boccato & Romis Attux

Discente:

– Gabriel Toffanetto França da Rocha – 289320

Atividade 1 – Regressão Linear

Sumário

1	Apresentação da série temporal	2
2	Modelo de predição da série temporal	4
2.1	Utilizando 70% dos dados para treinamento	4
2.2	Utilizando 80% dos dados para treinamento	6
	Referências	9
	Anexos	10

1 Apresentação da série temporal

De acordo com a base de dados *U.S Airline Traffic Data*, carregada por meio do arquivo `air_traffic.csv`, é possível obter diversas informações sobre o tráfego aéreo norte-americano no período de 2003 até 2023, com uma atualização mensal, e dados como número de voos e de passageiros.

Extraindo desses dados a série temporal do número total de voos em cada mês, é possível observá-la por meio do gráfico da Figura 1. Analisando visualmente os dados, observa-se que a distribuição dos voos não é constante, e que tem um aspecto periódico, apresentando um formato parecido que se repete ano a ano. Esse comportamento é esperado, uma vez que nas férias de verão, as famílias tendem a viajar mais a turismo, e no inverno tendem a voar menos devido ao mal tempo.

Porém, mesmo com uma característica regular, é possível observar faixas que se distinguem da demais, causando irregularidades no padrão da série temporal. Isso pode ser notado principalmente em dois pontos, destacados na Figura 2, em Agosto de 2008, onde o padrão da série decaí, e em Janeiro de 2020, onde o número de voos despenca drasticamente, com um transitório lento de recuperação, divergindo totalmente o que se observava no comportamento do número de voos.

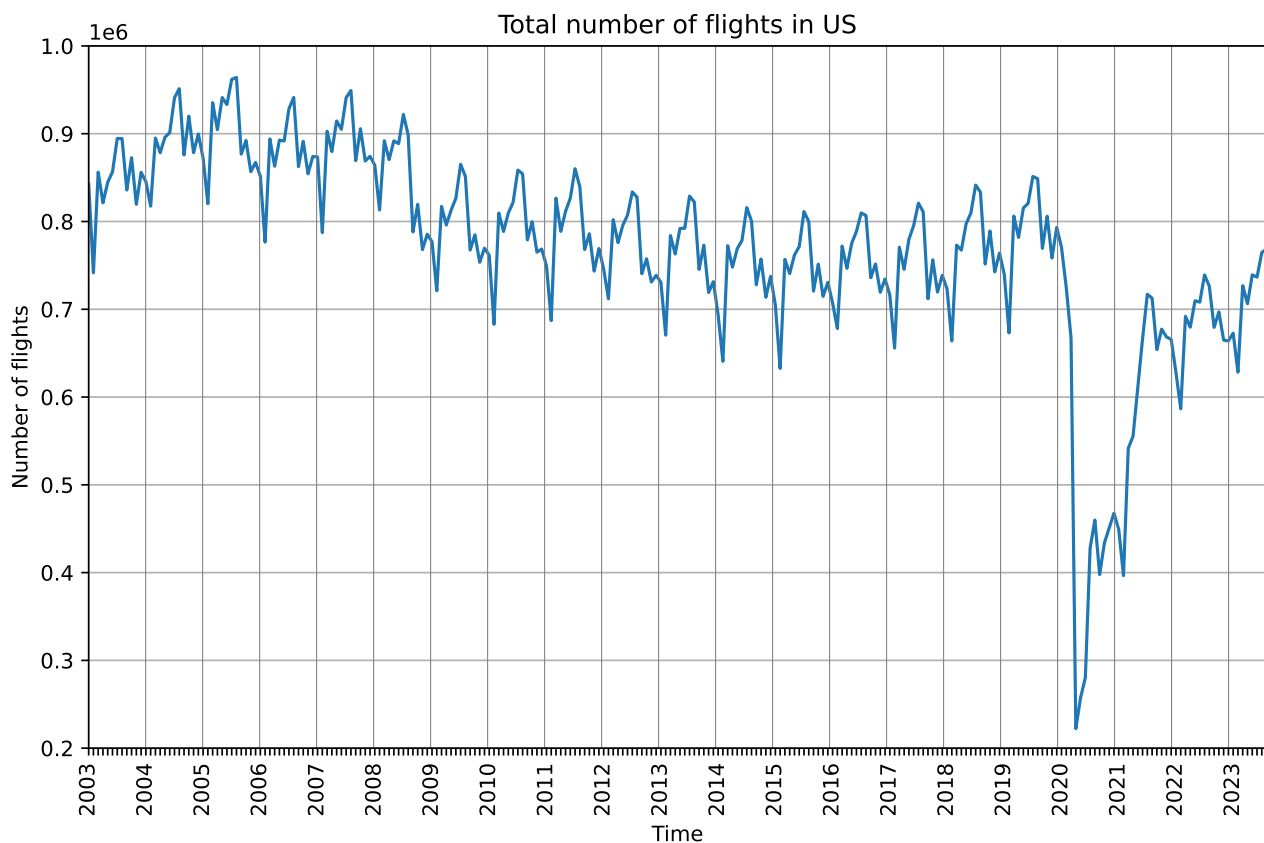


Figura 1: Série temporal do número total de voos nos EUA entre 2003 e 2023.

A primeira mudança de regularidade da série se dá em Agosto de 2008, época que o mundo enfrentou grande crise econômica, e os EUA estavam em estado de recessão.

A queda brusca do número de voos em Janeiro de 2020 até aproximadamente Julho de 2021 é facilmente explicada, ocorrendo devido a pandemia do vírus Covid-19 enfrentada pelo Mundo na época, que levou ao fechamento de inúmeros aeroportos até que fosse possível buscar formas de prevenir a doença, e principalmente a imunização da população com a vacina. Dessa forma, se observa uma primeira queda muito forte, ocorrida com o início da pandemia e o fechamento do aeroportos. Porém, entre o meio de 2020 e o início de 2021, os aeroportos passaram a ser reabertos, principalmente para voos domésticos, resultando em uma elevação da série. Com a vacina em 2021, ocorreu um salto no número de voos, com a abertura de fronteiras e o retorno a normalização do padrão da série a partir do meio de 2021, com o número de voos se elevando em 2022 e 2023.

Com isso, observa-se que mesmo podendo detectar padrões na série temporal do número de voos nos EUA, que se repetem com os anos, existem fatores externos, imprevisíveis, como crises econômicas e pandemias globais que influenciam diretamente na variável que pode ser de interesse sua previsão futura. Dessa forma, durante o comportamento regular da série, se espera que a previsão seja mais fácil, porém, ao ocorrer eventos imprevisíveis, a estimação tende a ser mais errônea.

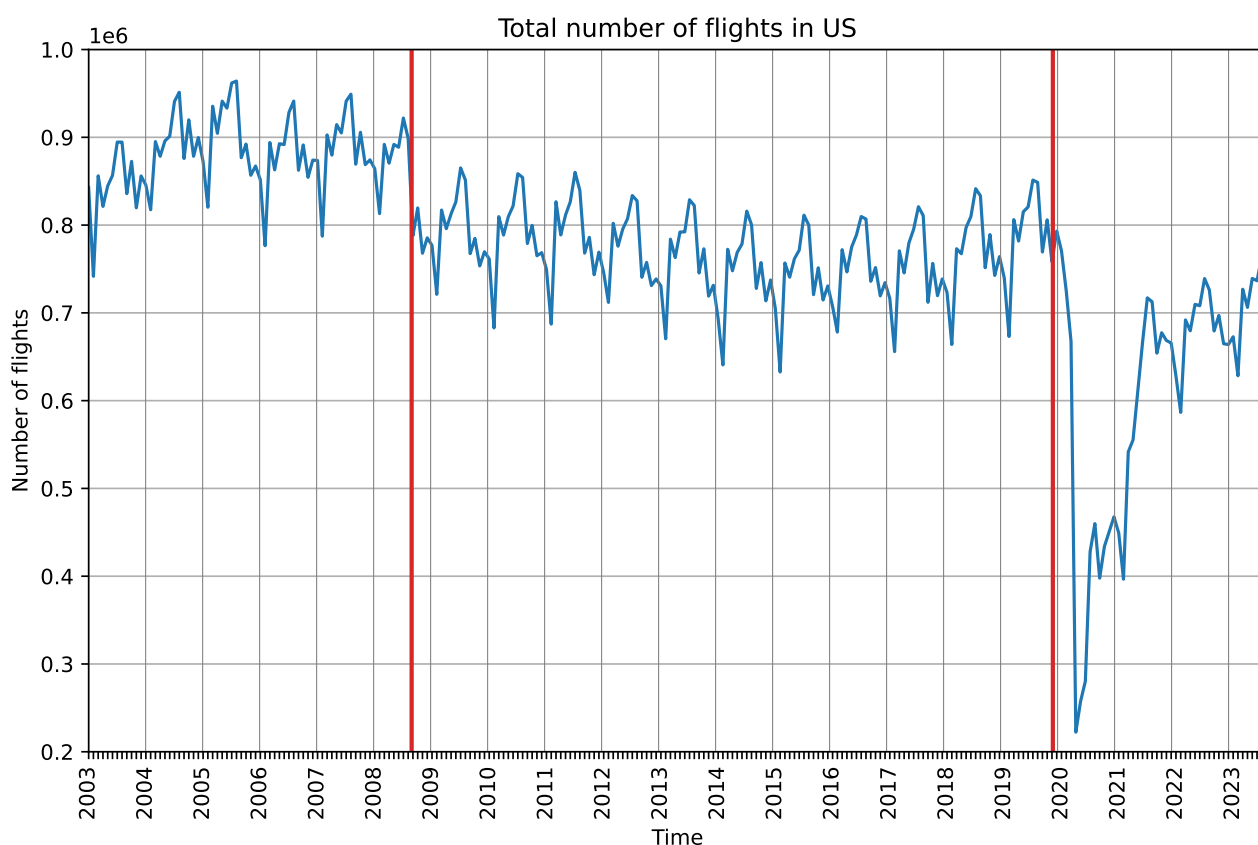


Figura 2: Série temporal do número total de voos nos EUA entre 2003 e 2023.

2 Modelo de predição da série temporal

$$y(n) = \mathbf{w}^T \mathbf{x}(n) + w_0 \quad (1)$$

2.1 Utilizando 70% dos dados para treinamento

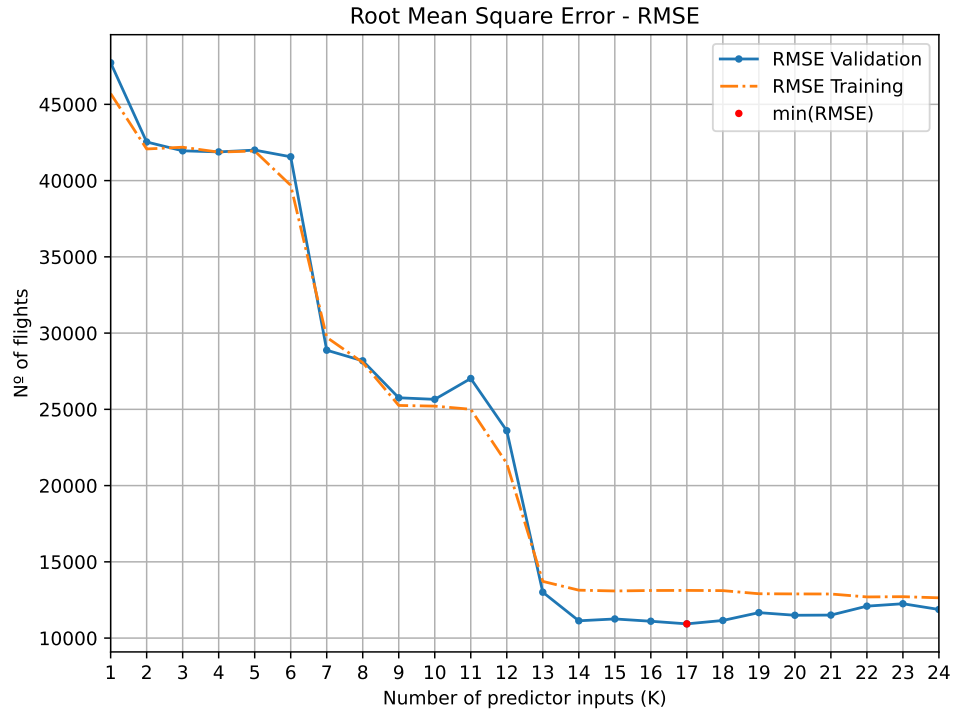


Figura 3: Progressão da raiz quadrada do erro quadrático médio de treinamento e validação em função do número de entradas do preditor (K).

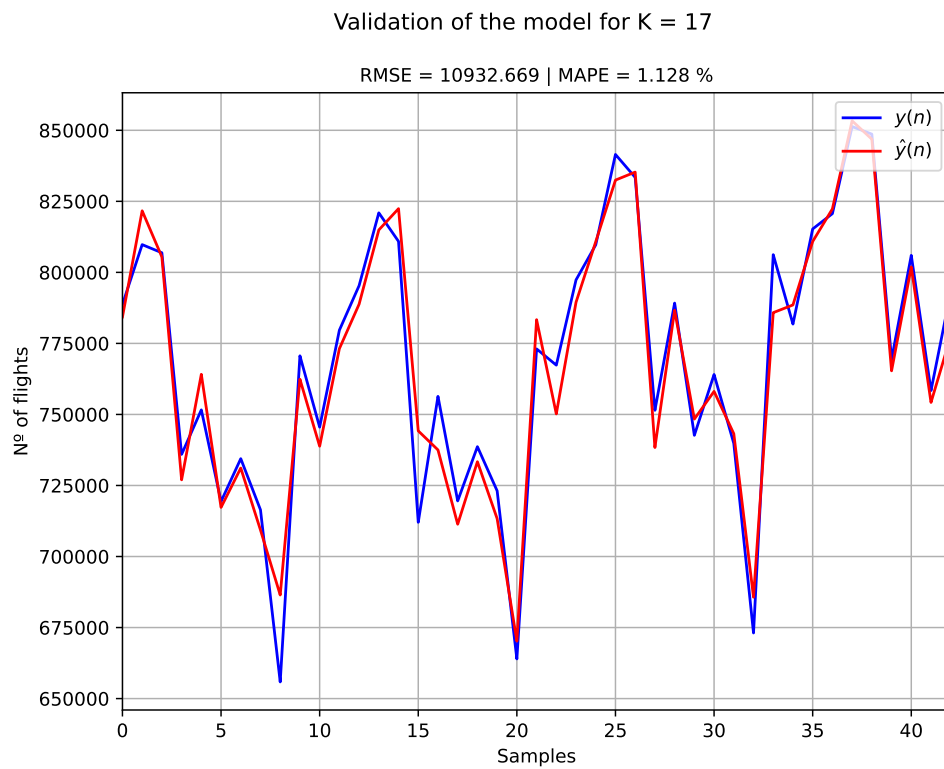


Figura 4: Validação da predição da série temporal para o preditor com menor RMSE.

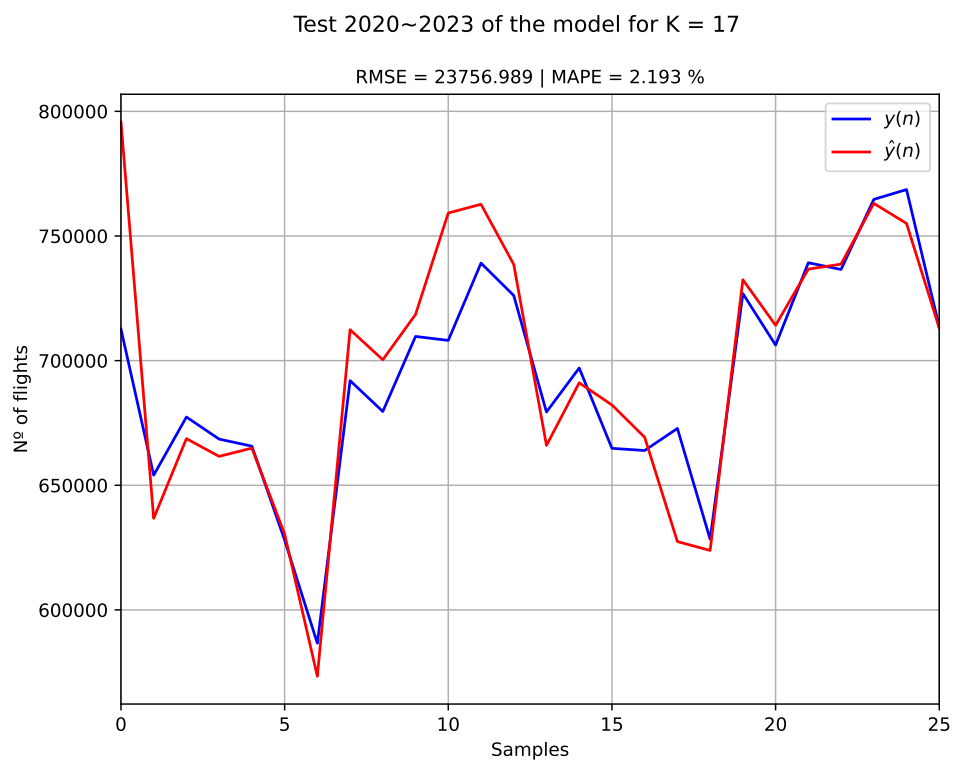


Figura 5: Teste da predição da série temporal para o preditor com menor RMSE utilizando todos os dados de teste.

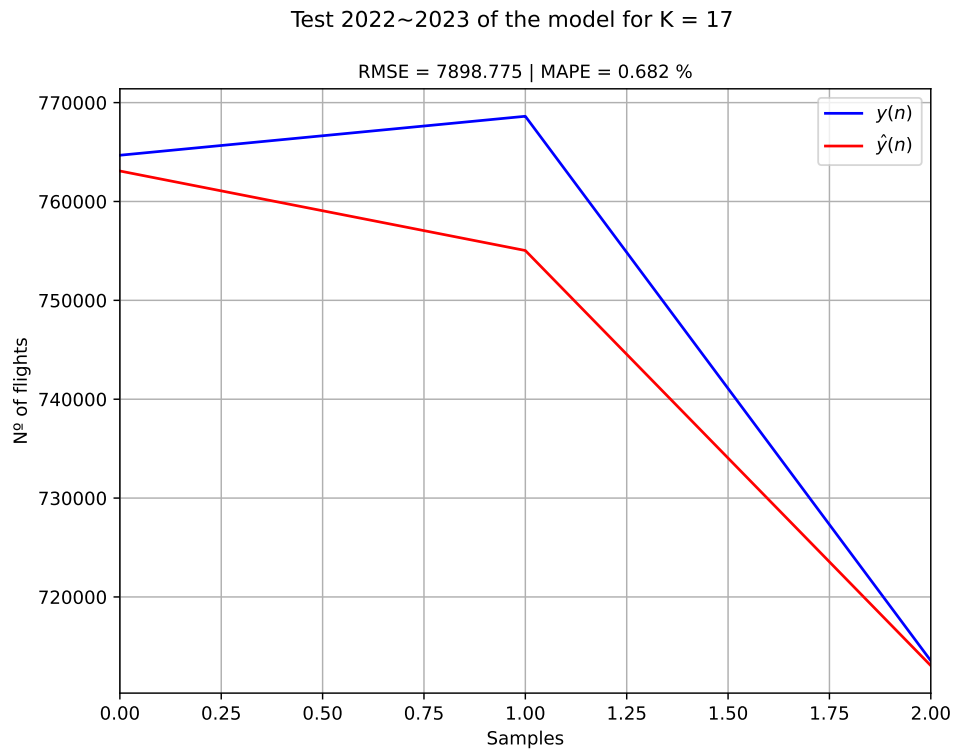


Figura 6: Teste da predição da série temporal para o preditor com menor RMSE utilizando os dados de teste entre 2022 e 2023.

2.2 Utilizando 80% dos dados para treinamento

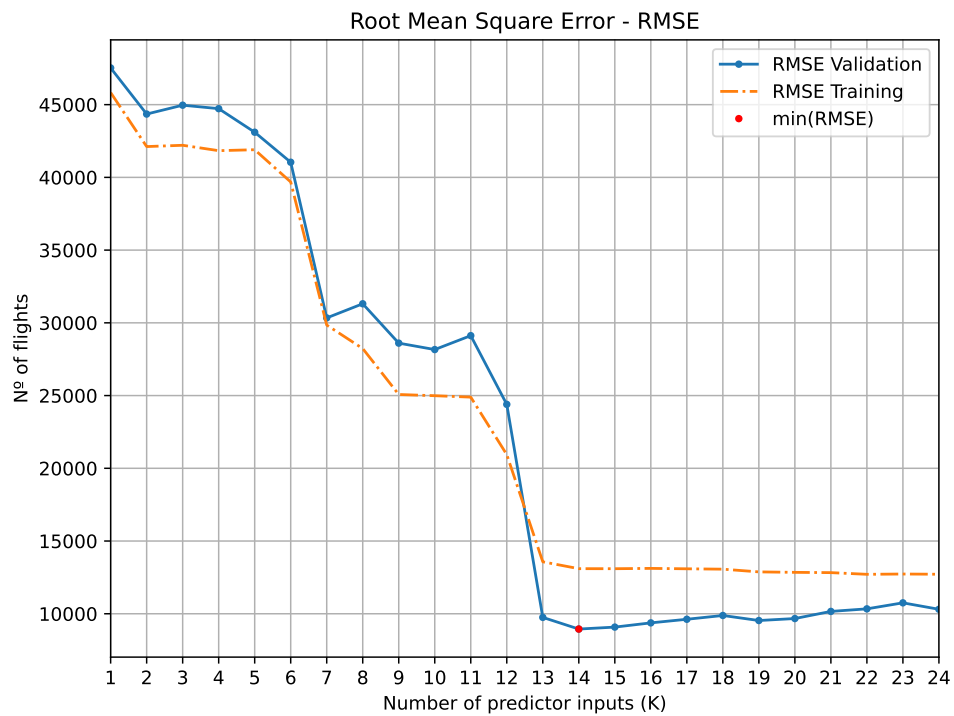


Figura 7: Progressão da raiz quadrada do erro quadrático médio de treinamento e validação em função do número de entradas do preditor (K).

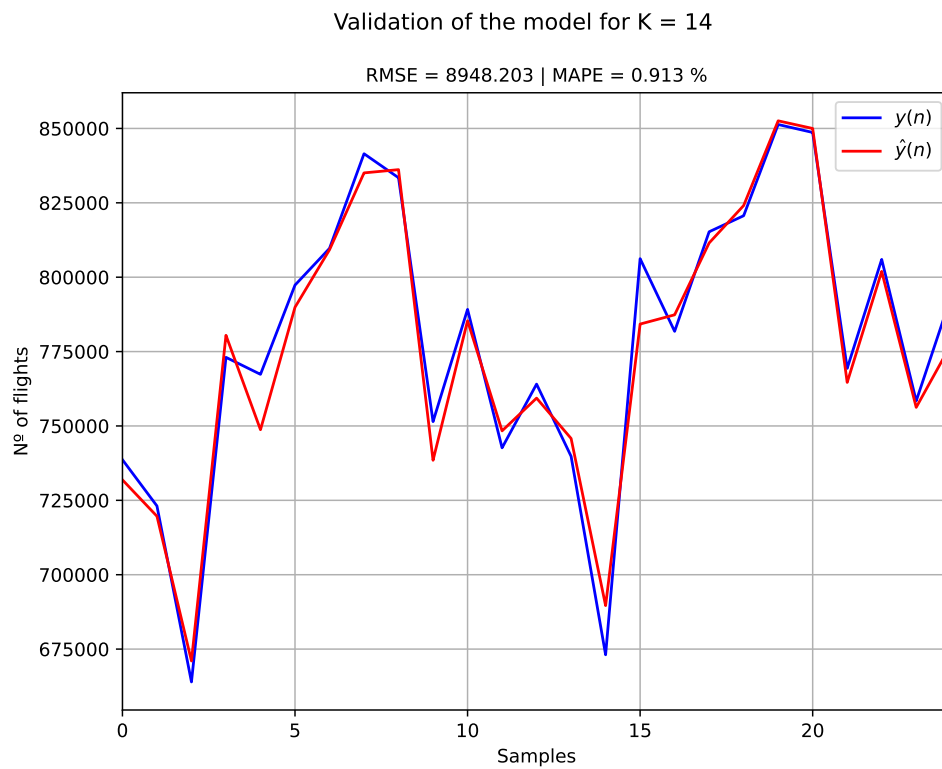


Figura 8: Validação da previsão da série temporal para o preditor com menor RMSE.

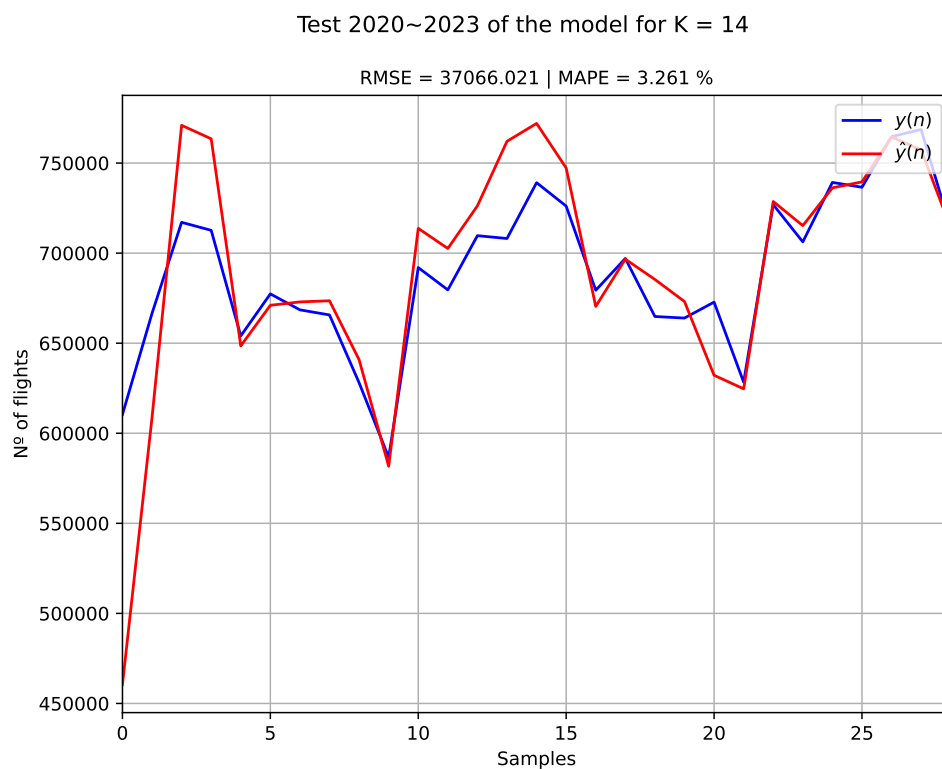


Figura 9: Teste da previsão da série temporal para o preditor com menor RMSE utilizando todos os dados de teste.



Figura 10: Teste da previsão da série temporal para o preditor com menor RMSE utilizando os dados de teste entre 2022 e 2023.

Referências

Anexos