
Previsão Epidemiológica na Plataforma JFST e a Utilização de Pipeline de Dados para sua Automatização

Bolsista: Gustavo Almeida Silva

Orientador: Prof. Dr. Marcel de Toledo Vieira



A Plataforma





A Venn diagram consisting of two concentric ellipses. The outer ellipse is a dark red color and contains the text "SRAGs". The inner ellipse is a lighter red color and contains the text "COVID -19". The inner ellipse is positioned such that it is entirely contained within the outer ellipse, illustrating that COVID-19 is a subset of SRAGs.

SRAGs

COVID -19



Análise Descritiva

Gráficos e Dados
Disponibilizados



Análise Exploratória

Relatórios Automatizados e
Boletins Informativos Mensais



Previsão Epidemiológica

Prever Dados
Confirmados para o
Futuro

Coleta Limpeza e
Armazenamento dos
Dados

Etapa 1

Deployment

Etapa 3

Etapa 2

Modelagem

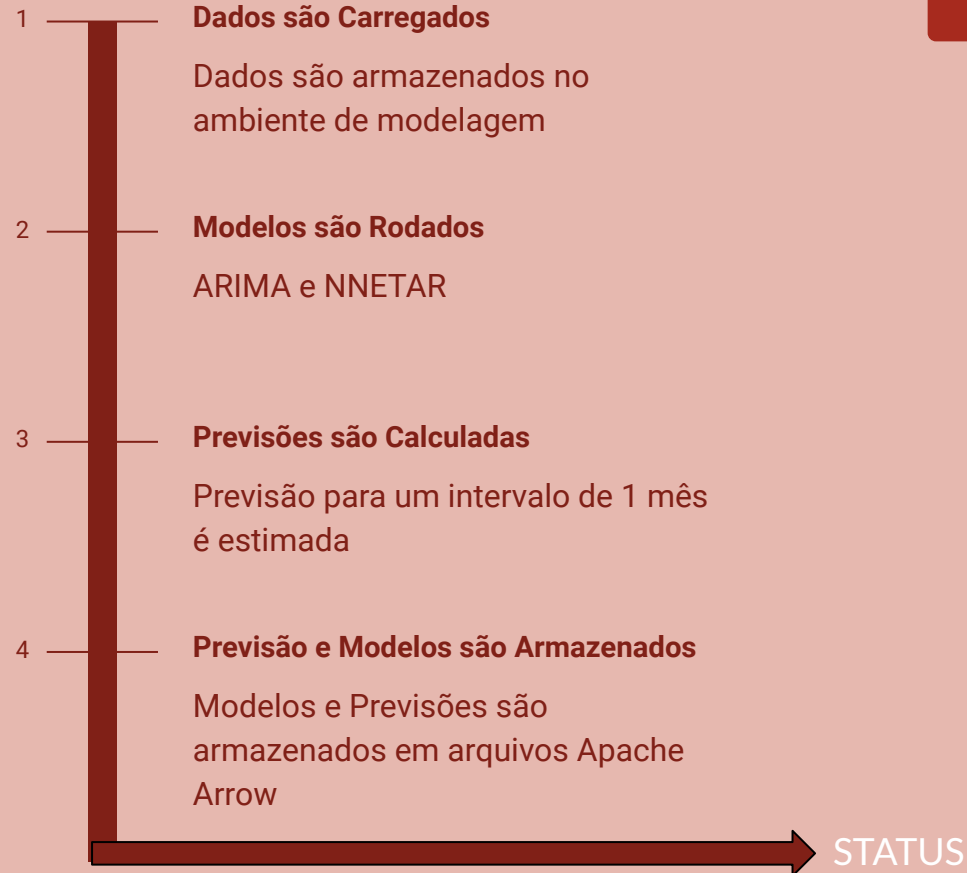
Coleta Limpeza e Armazenamento dos Dados

Etapa 1



- 1 Trigger é Disparada**
Semanalmente, as 23 hrs do Domingo uma trigger é disparada via Github Actions
- 2 Dados são Coletados**
Dados são coletados via RSelenium e Rvest
- 3 Dados são Limpos e Organizados**
- 4 Dados são Armazenados**
Os dados recém limpos e organizados são armazenados no Banco de Dados da Plataforma

STATUS



Etapa 2



Deployment

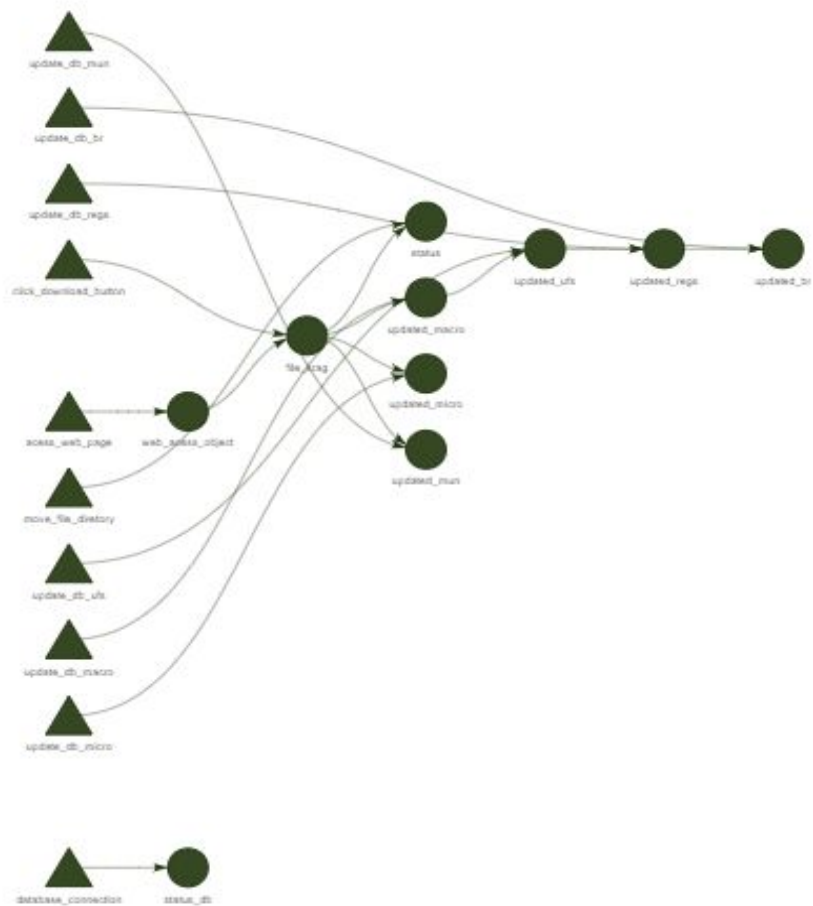
Etapa 3



O Workflow

Como o pipeline realmente funciona

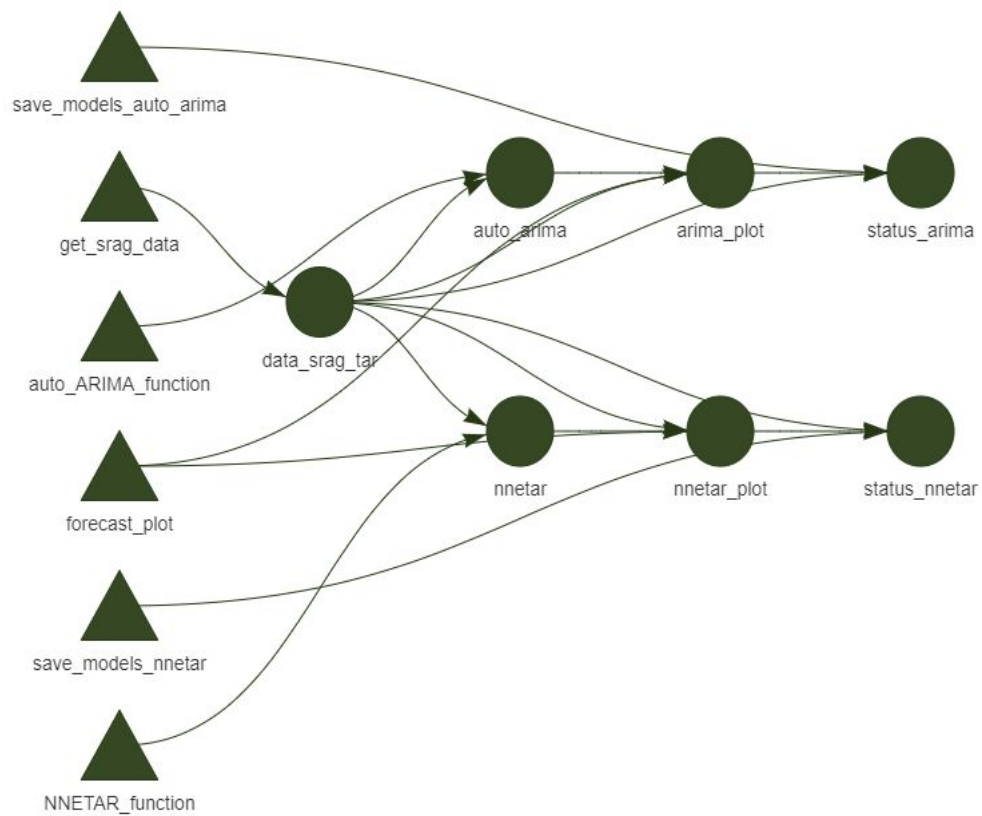


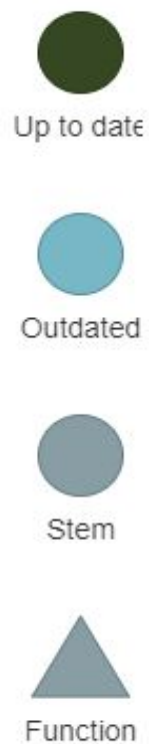
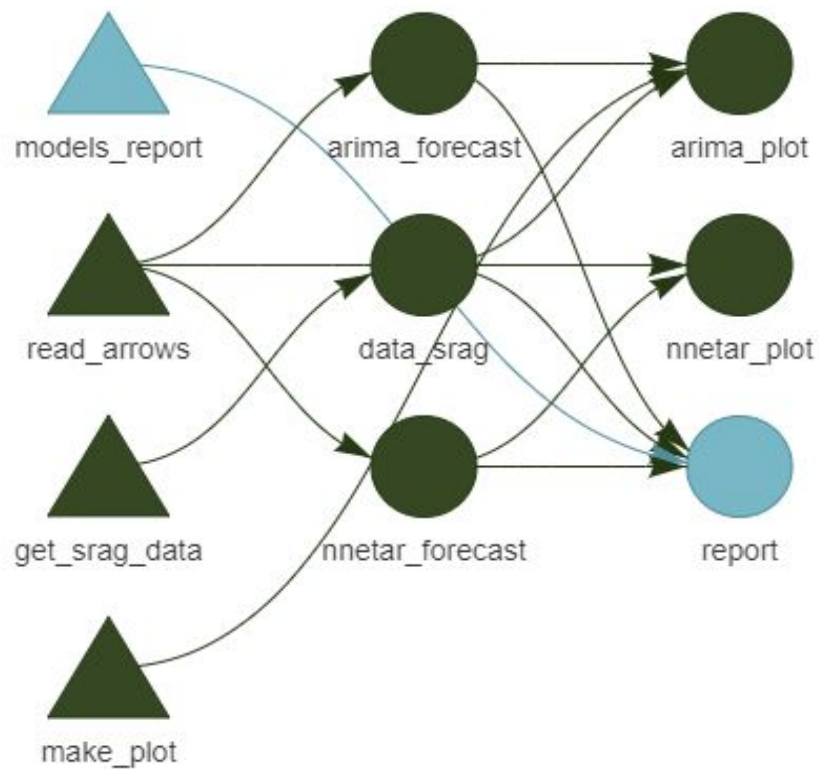


Up to date

Stem

Function





Modelos Utilizados

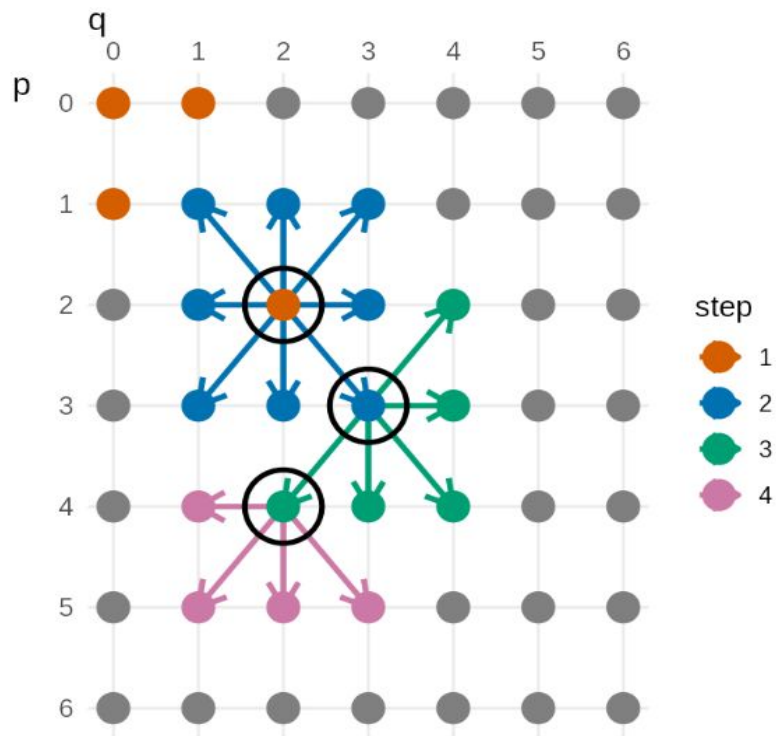
Assim como o processo, os modelos precisam ser automatizados



ARIMA/SARIMA

- Algoritmo de Hyndman & Khandakar, 2008
- Um stepwise para busca de parâmetros que minimizam o AICc
- Busca de tais parâmetros ocorre em uma malha
- Os modelos podem ser: $ARIMA(p, d, q)$ ou $SARIMA(p, d, q)(P, D, Q) [m]$

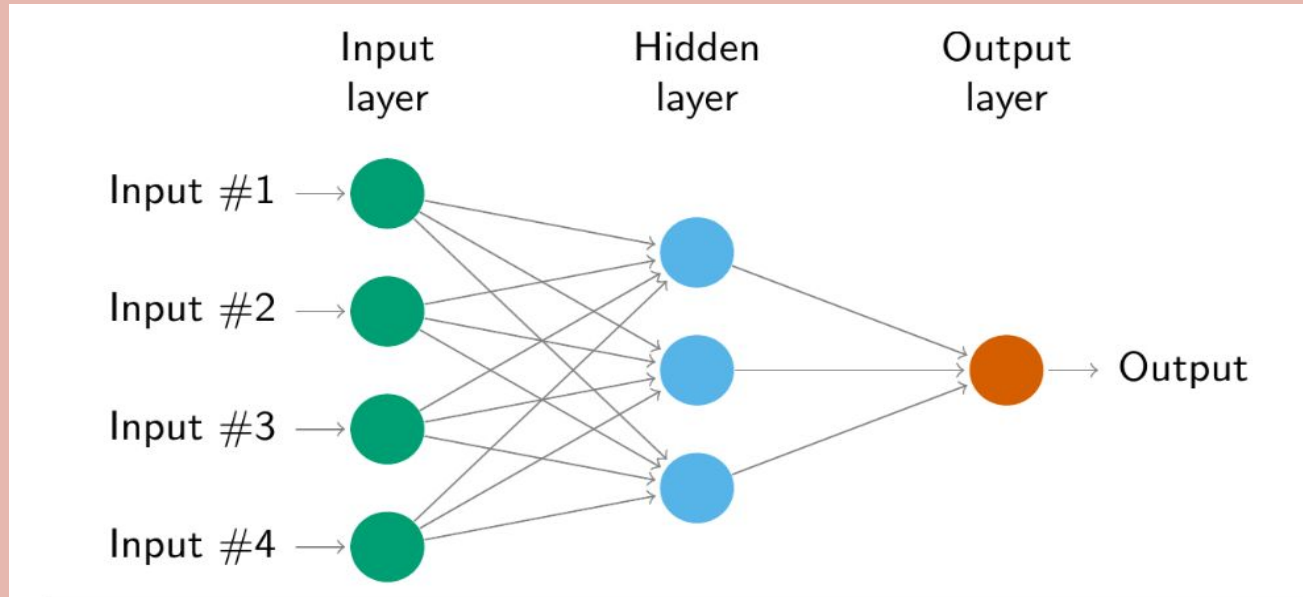
Algoritmo de busca



NNETAR

- Autoregressivo de Redes Neurais (Neural Network Auto Regressive)
- Os parâmetros autoregressivos e *hidden layers* são ajustados via tuning
- Um modelo $NNETAR(p, P, k)[m]$ é calculado

Rede Neural



Relatório de Acurácia - Modelos de Previsão Epidemiológica

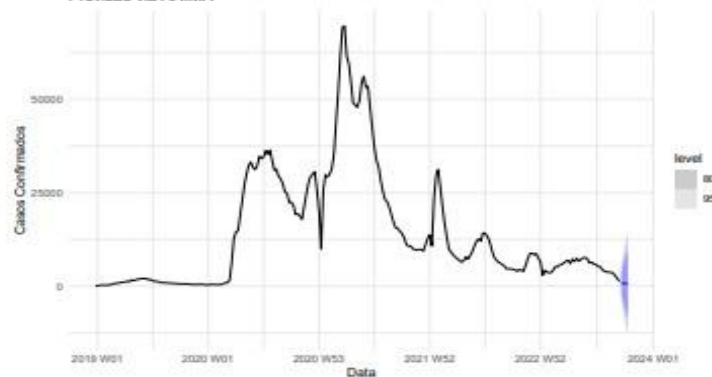
Dados da SRAGs

JF Salvando Todos
<http://jfsalvandotodos.ufjf.br/#!/>

Detalhes

• ARIMA

Previsão via ARIMA

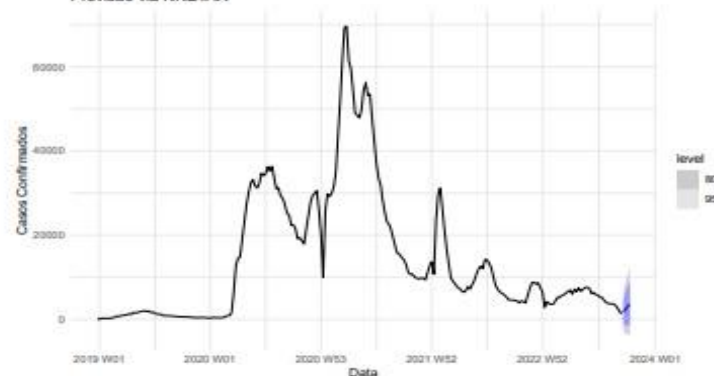


Model	MAE	RMSE	MAPE	MASE
ARIMA	1,401.99	2,458.36	12.73	0.08

Model	MAE	RMSE	MAPE	MASE
-------	-----	------	------	------

• NNETAR

Previsão via NNETAR



Model	MAE	RMSE	MAPE	MASE
NNETAR	1,086.81	1,905.7	9.87	0.06

Time difference of -35.81621 secs

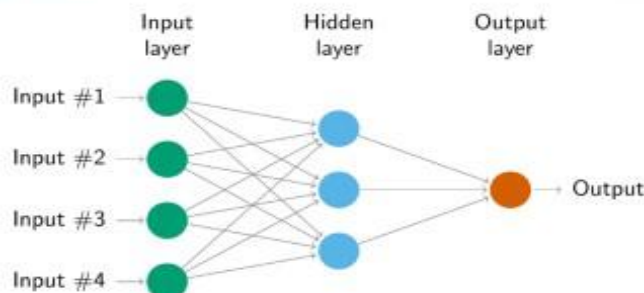
Metodologia

1. Seleção de Modelo

- Tópico busca explicar como os modelos disponibilizados são construídos

2. Cálculo da Acurácia

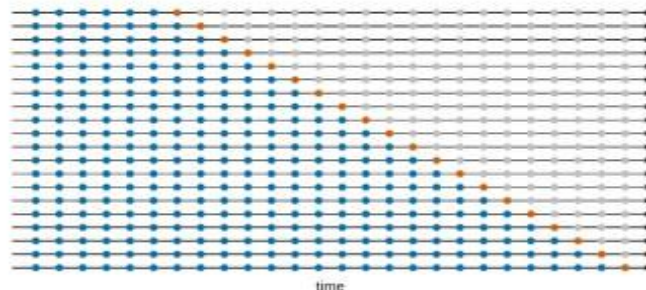
- Tópico busca explicar como a acurácia dos modelos é calculada



Cálculo da Acurácia

Os modelos construídos possuem o mesmo procedimento para o cálculo de suas acurácias.

O procedimento se baseia na divisão do conjunto de dados em k conjuntos de treino e teste, onde a cada conjunto novo de treino e teste feito, uma unidade adicional de tempo é adicionada para o treino. A seguinte figura ilustra o processo:



Tal procedimento é conhecido como *Time Series Cross Validation* ou *Evaluation on a Rolling Forecasting Origin*

As seguinte estatísticas dos erros são disponibilizadas:

- Média do Erro em Valor Absoluto (Mean Absolute Error): $MAE = \text{mean}(|e_t|)$
- Raiz do Erro Quadrático Médio (Root Mean Squared Error): $RMSE = \sqrt{\text{mean}(e_t^2)}$
- Média do Erro em Valor Absoluto em Porcentagem (Mean Absolute Percentage Error): $MAPE = \text{mean}(|p_t|), p_t = 100 \times \frac{e_t}{y_t}$
- Média do Erro Dimensionado em Valor Absoluto (Mean Absolute Scaled Error): $MASE = \text{mean}(|q_j|), q_j = \frac{e_j}{\frac{1}{T-1} \sum_{t=1}^{T-1} |y_t - y_{t-1}|}$

Referências

- Hyndman, Rob J., and George Athanasopoulos. Forecasting: principles and practice 3. OTexts, 2018.
- Landou, W. The {targets} R package user manual. ROpenSCI, 2022

OBRIGADO !