Wolt Data Science 2024

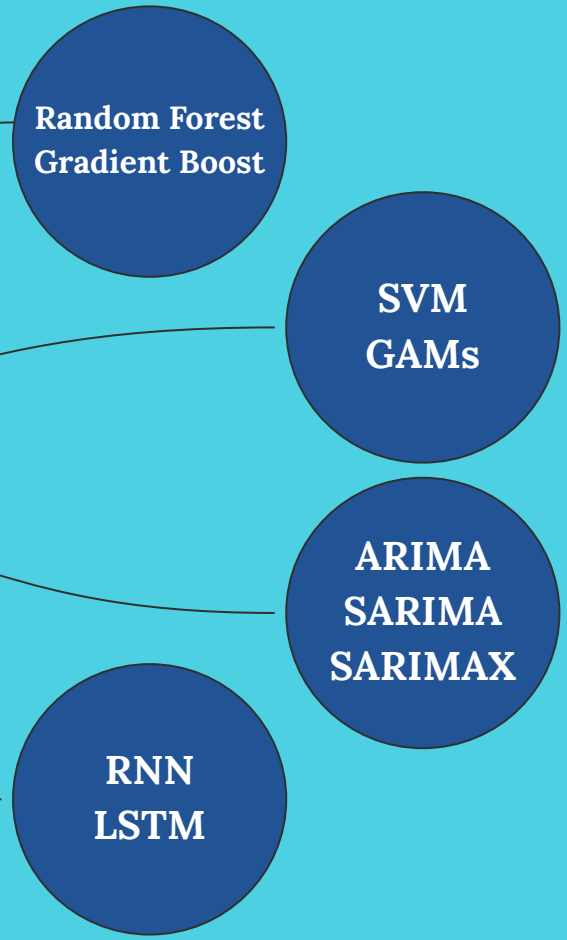
Presentation by Ahmad Baig

Problem Statement

Forecasting the amount of courier partners that will go online the following days

Possible Techniques

1. Tree-based Models
2. Regression Models
3. Autoregressive Models
4. Neural Networks



Before selection - Exploratory Data Analysis

15 Missing Values

Outliers in Target Variable (Courier Partners Online)

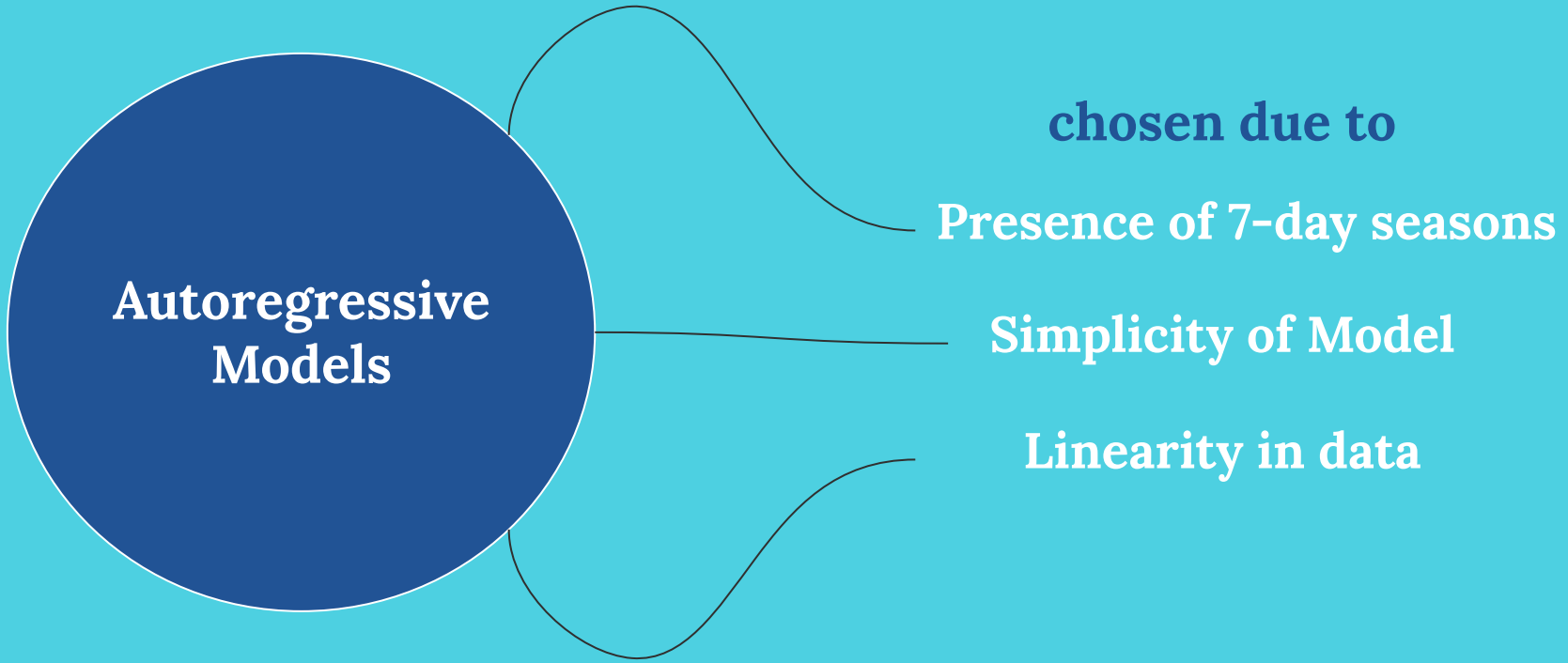
Positively Skewed Distribution for Precipitation

Humidity ↑ = ↓ Courier Partners Online

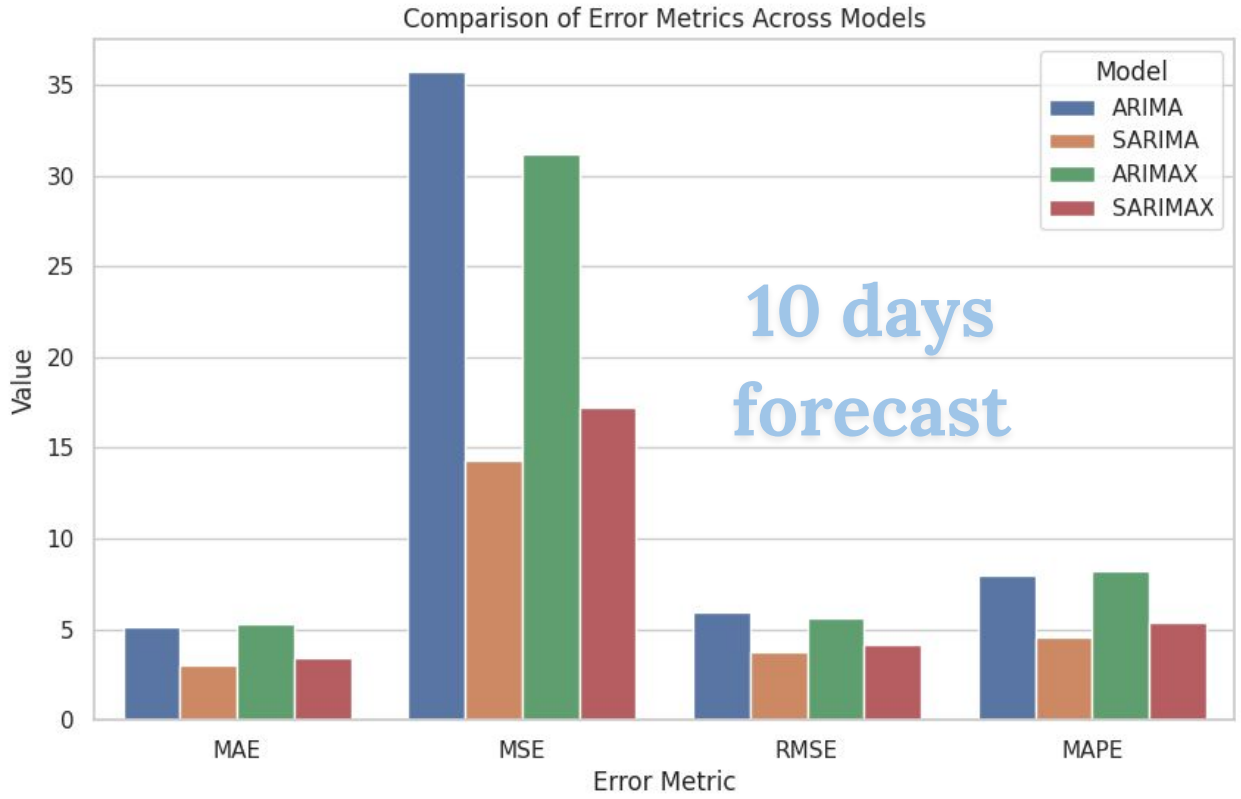
Precipitation ↑ = ↓ Courier Partners Online

Weekly fluctuations in courier activity

Model Selection



Model Performance



Insights

SARIMA and SARIMAX had the least error.

Seasonality appears to be a crucial element here.

Adding exogenous variables also shown slight improvement

i.e. temperature, humidity and precipitation

Future steps may include hypertuning and/or training other models

Ciao