

M5 – Accuracy

Problem Statement:

How much camping gear will one store sell each month in a year? To the uninitiated, calculating sales at this level may seem as difficult as predicting the weather. Both types of forecasting rely on science and historical data. While a wrong weather forecast may result in you carrying around an umbrella on a sunny day, inaccurate business forecasts could result in actual or opportunity losses. In addition to traditional forecasting methods you're also challenged to use machine learning to improve forecast accuracy. You will use hierarchical sales data from Walmart, the world's largest company by revenue, to forecast daily sales for the next 28 days. The data, covers stores in three US States (California, Texas, and Wisconsin) and includes item level, department, product categories, and store details. In addition, it has explanatory variables such as price, promotions, day of the week, and special events. Together, this robust dataset can be used to improve forecasting accuracy.

Background:

Prediction, expectation or gauge, is an announcement about a future occasion. They often yet not generally, founded on understanding or information. When making predictions, we imagine what will come next in the content, based on their earlier information or prior knowledge. Foreseeing urges us to effectively think ahead and pose inquiries. It likewise permits us to understand the actual and future's situation. We are going to understand how to predict the sales of products which are varying day to day, time to time using multivariate time series models and some machine learning algorithms. Time series forecasting is a significant territory of machine learning. Ordered sequence of values that are usually equally spaced over time is nothing but time series. To understand better we will take some data related to the sales and applied different techniques mentioned above on this data and estimated the sales for unknown data and found accuracy and loss function.

Methodology:

The models we will be used are as follows:

- 1) Prophet
- 2) ARIMAX
- 3) SARIMAX

4) LSTM

5) Stacked RNN models

Dataset:

- In this we are predicting item sales at stores of Walmart in various locations for two 28-day time periods .
- Locations included – Texas, Chicago, West Indies.
- calendar.csv - Contains information about the dates on which the products are sold.
- sales_train_validation.csv - Contains the historical daily unit sales data per product and store .
- sell_prices.csv - Contains information about the price of the products sold per store and date.
- sales_train_evaluation.csv - Includes sales

Evaluation Measures:

WMRSE will be used as the primary evaluation metric.

Software and Hardware Requirements:

Python based stats and forecasting libraries will be exploited for the development and experimentation of the project. Tools such as Anaconda Python, and libraries such prophet, auto_arima, Tensorflow, and Keras will be utilized for this process. Training will be conducted on NVIDIA GPUs, Kaggle notebooks and colab notebooks for training.