

Project Proposal

CS-579: Cloud Computing

Using Amazon Forecast to Build Forecasts – Comparative Analysis

Course Instructor:

Muhammad Usman Awais

Submitted By:

Ali Nauman – 18L1863

Hammad Akram – 18L1808

Introduction:

In this semester project, we will utilize one of the prominent machine learning services made available by Amazon Web Services, Amazon Forecast. Amazon Forecast is a fully managed service that uses machine learning to deliver highly accurate forecasts. This service will be utilized to evaluate a particular sales data-set to evaluate future predictions. Moreover, a comparative analysis will be carried out with respect to multiple algorithms to see how performance metrics vary.

Amazon Forecast:

Around the globe, the focus is surrounded on getting an insight on future predictions. For this scenario, Amazon Forecast comes into scope. It allows to accurately forecast future business outcomes such as product demand, resource needs or financial performance.

These tools allow us to build forecasts by looking at historical or forward looking data, which is called time-series data. The idea behind Amazon Forecast is that it uses historical data plus additional data which would impact the forecast. Once the data is provided to Amazon Forecast, it automatically examines it and creates a forecasting model which is capable of producing predictions.

Amazon Forecast works on three simple steps,

- **Importing Data-sets** – Data-sets are taken as an input,
- **Training Predictors** – These create custom models, which are trained on the data,
- **Generating Forecasts** – Generate time-series forecasts, and it can be queried as well through the *QueryForecast* API or visualized as well.

Algorithms:

An Amazon Forecast predictor uses an algorithm to train a model with the time series data-sets. In return, the trained model can be utilized to evaluate the results through numerous metrics available. Amazon Forecast has an option of *AutoML*, which finds the optimal algorithm as per the data-set and evaluates. However, we can manually select the algorithm to use as well.

Algorithms available are listed in the table below,

Algorithm	Data-set
ARIMA	Stationary Time-series
CNN-QR	Historical Time-series & Forward Looking Time-series
DeepAR+	Forward Looking Time-series
ETS	Seasonality/Prior Assumption
NPTS	Seasonality/Prior Assumption/Stationary Time-series
Prophet	Forward Looking Time-series

Objectives:

- Utilize a time-series data-set for future predictions.
- Make use of the *AutoML* feature in Amazon Forecast to select the most optimum model as per the data-set.
- Comparative Analysis – Amazon Forecast vs Implementation in Sequential mode.

Proposed Solution:

We will be utilizing a sales data taken from *Kaggle*, Walmart. Using this data-set, we will pre-process the data as per the need of the input format on Amazon Forecast. Using the *AutoML* feature, we will select the appropriate algorithm and make predictions.

On the other side, we shall implement a traditional model on the same data-sets. In both cases, we shall calculate the *Root Mean Square Error*, and use it to compare the models. It will give us an insight on the performance of Amazon Forecast and how good the web service is for generating forecasts.

Root Mean Square Error, RMSE, is a frequently used measure of the differences between values (sample or population values) predicted by a model or an estimator and the values observed. It is an accuracy measure, used to compare forecasting errors of different models for a data-set. Its value is between 0 (almost perfect, never achievable) to 1 (worst-case scenario).