**Airlines Revenue**

**By**

**Abdul Wahed**

**Nischay Mallikarjunappa Gowda**

**Pranay Manikanta Narava**

**Sailesh Venigalla**

**University of Maryland Baltimore, County**

Data 602 Project Report

Fall 2022

Abstract:

Introduction:

Forecasting demand is critical to the operation of airline pricing and revenue management systems. With so many moving aspects and problems in the market, such as capacity, scheduling, and competition a comprehensive and powerful revenue management system is still required. With the help of AI, airlines can understand their revenue and help them to determine the right price and time for their customers or passengers. Our datasets consists of revenue data from both passenger and cargo. When it comes to the passenger it shows an airline's intention to get the most amount of money feasible from the flying customer by convincing them to pay a fare equivalent to their economic willingness to pay. And coming to the cargo business, this will help air carriers to maximize by optimizing and distributing available capacity to the relevant customers and items. While the number of variables influencing an airline network's cost-efficiency grows by the year, reliable travel demand forecasts have become more important in airline network design procedures.

The need for AI in the aviation sector to leverage multiple data sources to capture future demand more precisely, both in its immediate and long-term changes, is maybe even more pressing now than it was previously. With the help of data sources such as online searches, economic fluctuation in the market, networks, customer feedback, and with booking data required information can be forecasted using a machine learning process. In this process, we sanitize the data dealing with null values, reducing outliers, ensuring data consistency, and so on, followed by some exploratory data analysis (EDA) and model fitting. The output from this will help the airline operators/industry to make decisions which intern helps them to increase overall revenue.

**LITERATURE REVIEW:**

1. **Join** **Reinforcement learning applied to airline revenue management**

Bondoux, N., Nguyen, A.Q (2020) have investigated the performance of reinforcement learning (RL) in the field of airline RM. RL takes a radically different approach compared to the traditional RMS. They discovered that QL takes millions of departures of training data to achieve even acceptable levels of performance, but DQL can achieve the theoretical optimal policy with just roughly three years of departure data. They discovered that when competing against RMS, DQL takes advantage of being model-free and outperforms RMS. When DQL fights against itself, however, they detect a form of spiral-down. While in practice, RL may never completely replace RMS, it is possible to envision a next version of RMS that incorporates RL. They have also showed that by initializing the DQN network using the assumed Q-function from RMS, they may get improved revenue performance and learning speed.

**Machine Learning Outcomes:**

**Naive**

**Chart, line chart

Description automatically generated**

**Autoregressive Model**

Autoregressive model is used here since we can see there is correlation between the values of the time series and the values that came before and after them and very good for predicting future values based on past values. Since, It uses past values to predict future behaviour they can prove inaccurate under some circumstances like market changes or change in technologies. Using this model the autocorrelation function may be used to determine whether there is a lack of randomness and capable of forecasting recurring patterns in data.

Autoregressive model process of regressing a time series' value on earlier values from the same time series.

yt = 𝛽0 + 𝛽1yt-1 + 𝜖𝑡

The model says, yt predictor in t time period and yt-1 is previous period values, coefficient 𝛽1 represents the numeric constant multiplied by the lagged variable (Xt-1).

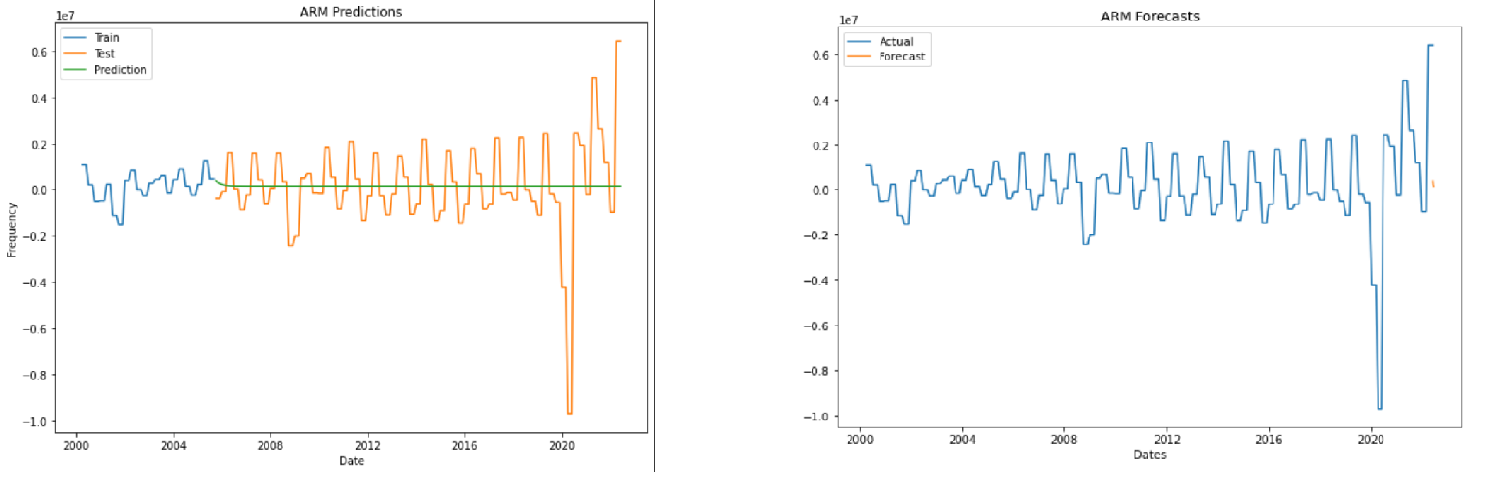


Fig: AR Forecast and Prediction

For Autoregressive the model is trained for 75%

Based on the forecast for the Airline Revenue data Predictive performance using Auto Regression model:

Graphical user interface, text, application

Description automatically generated

**Moving Average Model**

Moving average model (MA) is a regression model based on past forecasting errors. A statistical technique for predicting long-term trends is the moving average. The method entails shifting the range while averaging a group of data from a certain range. It is used here with time-series data to smooth out short-term fluctuations and long-term trends.

The data set of 22 years is divided into training set and test set. Under the premise of the same input of historical price time series, it is compared with the experimental results of autoregressive model (AR), Moving Average (MA).

Based on the forecast for the Airline Revenue data Predictive performance using Moving Average model:

Text, application

Description automatically generated

**ARIMA MODEL (Autoregressive Integrated Moving Average):**

Regression analysis that evaluates the effectiveness of one dependent variable in relation to multiple fluctuating variables is known as an autoregressive integrated moving average model (HAYES, 2021). Based on historical values, autoregressive integrated moving average (ARIMA) models forecast future values, moreover, lagged moving averages are used by ARIMA to smooth time series data furthermore, they are often used in quantitative data to assess future asset price trends (HAYES, 2021).

The underlying premise of autoregressive models is that the future will follow the past and as a result, they may turn out to be incorrect under specific market circumstances, such as economic collapse or times of fast technical advancement. In other words, instead of using actual values, the model looks at variations between values in the series to forecast future securities or financial market movements.

It is possible to comprehend an ARIMA model by detailing each of its parts as follows:

* A model known as autoregression (AR) depicts a variable that is altering and starts to decline on its own lagged, or previous, values.
* To enable the time series to become stable, the data values are replaced by the difference between the current values and the prior values in the Integrated (I) model.
* When a moving average model is applied to lagged data, the moving average (MA) includes the relationship between an observation and a residual error.

**ARIMA Parameters :**

Each element of the ARIMA algorithm performs the role of a parameter with a common nomenclature. Standard notation for ARIMA models is ARIMA with p, d, and q, where integer values are used in place of the parameters to denote the model's type. You may specify the parameters as follows:

* p : lag order, or the number of lag data in the model
* d : the number of differences applied to the original observations; often called the degree of differencing.
* q : the order of the moving average; usually referred to as the moving average window size.

In a nutshell, ARIMA(p,d,q) has three parameters and one of is for making a series stationary. The first and third parameters of ARIMA come from AR and MA components, respectively and as we can guess the second parameter represents the order of difference.

𝑋𝑡=𝛼+𝛽1𝑋𝑡−1+𝛽2𝜖𝑡−1+𝜖𝑡

𝑍𝑡=𝑋𝑡+1−𝑋𝑡

The model for an autoregressive process says that at time ‘t’ the data value, *Xt*, consists of a constant, α (alpha), plus an [autoregressive coefficient](https://www.sciencedirect.com/topics/mathematics/autoregressive-coefficient), β (beta), times the previous data value, *Xt*−1, plus random noise, ε*t*. Note that this is a linear regression model that predicts the current level (*X* = *Xt*) from the previous level (*X* = *Xt*−1).

**ARMA MODEL (Autoregressive Moving Average):**

Autoregressive Moving Average is known by the abbreviation ARMA. The Autoregressive (AR) and Moving Average (MA) models, two less complex models, were combined to create it (Mehandzhiyski, 2021). The "MA" section occurs second because, in analysis, we often add the residuals to the end of the model equation. Naturally, this will become clear when we look at the equation.

The ARMA model takes the following form when p and q is 1:

𝑋𝑡=𝛼+𝛽1𝑋𝑡−1+𝛽2𝜖𝑡−1+𝜖𝑡

If you are dealing with another ARMA model, say ARMA(2, 3), the equation takes the following form:

𝑋𝑡=𝛼+𝛽1𝑋𝑡−1+𝛽2𝑋𝑡−2+𝛽3𝜖𝑡−1+𝛽4𝜖𝑡−2+𝛽5𝜖𝑡−3+𝜖𝑡

The model for an autoregressive process says that at time ‘t’ the data value, Xt, consists of a constant, α (alpha), plus an [autoregressive coefficient](https://www.sciencedirect.com/topics/mathematics/autoregressive-coefficient), β (beta), times the previous data value, Xt−1, plus random noise, εt. Note that this is a linear regression model that predicts the current level (X = Xt) from the previous level (X = Xt−1).

**ARMA Parameters :**

It uses the following parameters :

* p : lag order, or the number of lag data in the model
* q : the order of the moving average; usually referred to as the moving average window size.

Conclusion:

With the

With the help of implementing stremlit, we can