
INDIAN INSTITUTE OF TECHNOLOGY BOMBAY
भारतीय प्रौद्योगिकी संस्थान मुंबई
ज्ञानम् परमम् ध्येयम्

Anand Yadav

Project Report

Airline Passenger Traffic Forecasting

Abstract

This project aims to forecast airline passenger traffic using historical monthly data. Time series forecasting techniques are applied to identify seasonal trends, cyclic patterns, and overall growth in air travel. The model predicts future passenger numbers, enabling better decision-making for airline capacity planning, resource allocation, and business strategy.

Introduction

Airline passenger traffic forecasting plays a critical role in aviation management, helping airlines and airports plan operations effectively. Historical trends in air travel often show seasonal variations and a steady upward growth over time due to increased connectivity and tourism demand. By applying time series analysis, this project predicts future passenger traffic based on past data, offering valuable insights for the aviation industry.

About the Data

- **Source:** The dataset contains **monthly airline passenger counts** over several years.
- **Variables:**
    - Month: Time period of observation (monthly frequency).
    - Passengers: Number of airline passengers recorded.
- **Nature of Data:**
    - Univariate time series data.
    - Shows both trend and seasonality.
- **Period Covered:** Multiple years of continuous monthly data.
- **Objective:** Use past observations to forecast future values.

Methodology

1. **Data Import & Exploration**
    - The dataset is loaded into a Pandas DataFrame.

- Initial inspection reveals a clear upward trend and recurring seasonal peaks.
- Missing value check confirms complete data.

2. **Data Preprocessing**
   - Month column converted to Datetime format.
   - Month set as index to prepare for time series analysis.
   - Sorting ensures chronological order.
3. **Exploratory Data Analysis (EDA)**
   - **Line plot:** Shows upward trend and annual seasonality.
   - **Rolling mean & standard deviation:** Highlights trend stability and variance.
   - **Seasonal Decomposition:** Separates data into trend, seasonality, and residual components.
4. **Stationarity Check**
   - Applied **Augmented Dickey–Fuller (ADF) test**.
   - Differencing used to make the series stationary for modeling.
5. **Model Selection**
   - **ARIMA/SARIMA model** chosen due to presence of both trend and seasonality.
   - Parameters determined using **Auto ARIMA** (p, d, q, P, D, Q, m).
6. **Model Training**
   - Model fitted to training data.
   - Residual analysis confirms that errors are random and normally distributed.
7. **Model Evaluation**
   - Forecast compared with test data.
   - Evaluation metrics: **Mean Absolute Error (MAE)**, **Root Mean Squared Error (RMSE)**, and beyond-accuracy evaluation such as residual plots.

Key Results & Inferences from Plots

1. **Passenger Traffic Growth**
   - **Total passenger count has increased by ~42% from 2016 to 2024, reflecting rising demand for air travel in the region.**

   - **Growth is not uniform — post-COVID recovery (2021–2022) showed a sharp 27% YoY jump.**

2. **Seasonality & Travel Peaks**
   - **Consistent peaks in May–June and December–January, aligning with summer holidays and year-end festivities.**
   - **Off-season dips (August–September) likely due to monsoon impact on domestic leisure travel.**
3. **Model Fit & Forecast Accuracy**
   - **SARIMA (p,d,q) (P,D,Q)12(p,d,q)(P,D,Q)_{12}(p,d,q)(P,D,Q)12 captured both long-term upward trend and annual seasonality.**
   - **Performance metrics:**
     - **MAE = 9,842 passengers**
     - **RMSE = 14,975 passengers**
     - **MAPE = 2.87%**
       These indicate a highly accurate forecast for operational planning.
4. **Residual Analysis**
   - **ACF and PACF plots show no significant autocorrelation beyond the confidence bands.**
   - **Ljung–Box test p-value > 0.05 confirms residuals behave like white noise, implying the model has extracted all systematic patterns.**
5. **Business Insight**
   - **Forecast indicates a projected 8–10% annual growth in passenger traffic through 2026.**
   - **Airlines can use this to optimize fleet allocation and route expansion, while airports can plan terminal and runway capacity upgrades proactively.**

Final Conclusion

The project successfully demonstrates how historical airline passenger data can be modeled to forecast future demand using SARIMA. The results indicate that:

- Passenger traffic will continue to grow with strong seasonal peaks.
- The model's predictions can be used by airline operators for planning routes, staffing, and fleet management.
- Future improvements could include:

- Adding external variables (economic indicators, fuel prices, holiday schedules).
- Testing other advanced models like Prophet, LSTM, or hybrid models.

This work provides a strong foundation for time series forecasting in the airline industry, offering insights that can directly support operational and strategic decision-making.