Fin-A-Lytics - Finance & Analytics Case Challenge

OPTIMA 2023

TECHTITANS PRESENTS

TABLE OF CONTENTS

1 Mission Statement

2 Problems faced

3 Solutions offered

4 Our Solution

5 Data Visualization

6 Data Visualization Contd.

7&8 Our Model

9&10 Briefly describe the concept

MISSION STATEMENT

The project involves analyzing one year of flight data across ten markets to create a predictive model for both demand and pricing in the airline industry. This model should factor in time-based demand fluctuations and incorporate dynamic pricing strategies. Subsequently, the model's performance will be tested by forecasting flight demand and fares for the following three months (months 13 to 15). This initiative aims to enhance revenue management through data-driven insights and forward-looking predictions.

PROBLEMS FACED

In the airline industry, several common challenges demand innovative solutions. First and foremost is the need for accurate demand forecasting, which can be achieved through the implementation of data-driven predictive models. Airlines must also grapple with optimizing pricing strategies to maximize revenue while ensuring competitive fares. Effective revenue management is essential, with factors like cancellations, overbooking, and no-shows requiring consideration. In highly competitive markets, airlines must closely analyze market share and competitor schedules to fine-tune pricing and marketing strategies. Furthermore, accounting for seasonality and external trends is crucial for precise predictions. Managing vast and diverse datasets necessitates robust data management and integration processes.

DATA COMPLEXITY:

- Handling incomplete or inaccurate historical data and integrating various data sources.
- Capturing complex demand drivers, seasonality, and dynamic pricing strategies.

MODELING COMPLEXITY:

- Selecting appropriate forecasting models tailored to different routes and markets.
- Evaluating and adapting models to changing competitive dynamics and market conditions.

OPERATIONAL CHALLENGES:

- Dealing with large data volumes and maintaining up-to-date models.
- Addressing regulatory changes and economic fluctuations affecting booking patterns.

SOLUTIONS OFFERED

A generic solution to address these challenges in the airline industry involves the adoption of advanced data analytics and machine learning techniques. By leveraging historical booking data, market insights, and external factors, airlines can develop accurate predictive models for demand forecasting. These models enable dynamic pricing strategies that adapt to changing market conditions, maximizing revenue while staying competitive. Effective revenue management can be achieved by implementing strategies that account for factors like cancellations, overbooking, and no-shows. Close monitoring of market share and competitor schedules informs pricing and marketing decisions. Accounting for seasonality and trends through data analysis enhances prediction precision.

FEATURE ENGINEERING:

Invest in feature engineering to create relevant predictors that capture booking behavior. These may include time-based features, demand indicators, and competitive factors.

MODEL SELECTION:

- Model Selection: Employ a combination of traditional time series forecasting methods (e.g., Moving Average Smoothing) and machine learning algorithms (e.g., regression, neural networks) depending on the specific requirements of each route or market.
- Model Evaluation: Continuously assess model performance using appropriate metrics (e.g., MAE, RMSE). Implement regular model validation and calibration processes to ensure accuracy.

DEPLOYMENT:

Establish a robust model maintenance process that includes regular monitoring of model performance. Set up alerts for significant deviations from predictions and automated model updates when necessary.

OUR SOLUTION!

1

Data Preprocessing: First, we acquired a dataset containing information about flight bookings and revenue. To prepare the data for modeling, we addressed missing values by filling them with zeros in both the booking and revenue datasets. The index columns were dropped, and date and time columns were parsed accordingly. We created a merged dataset by combining the booking and revenue data, using matching flight details.

2

Time Series Extraction: From the merged dataset, we extracted time series data for:
The number of tickets booked on each day.
The revenue generated on each day.
The price of tickets on specific days.

3

Data Encoding: To enable the model to comprehend patterns and seasonality, we performed customized encoding on all dates in the dataset. This encoding helps the model capture temporal dependencies and trends in the data.

4

Smoothening Revenue Data: We calculated smoothed revenue over a 7-day window from the revenue data to reduce noise and highlight underlying revenue trends.

5

Data Visualization: Data visualization techniques were applied to gain insights and understand the patterns in the dataset, aiding in feature selection and model understanding.

6

Feature Engineering from Schedule Data: Effective features were extracted from schedule data, including total seats in a flight, aggregate market share of the flight operator, and market share of the flight operator on each route.

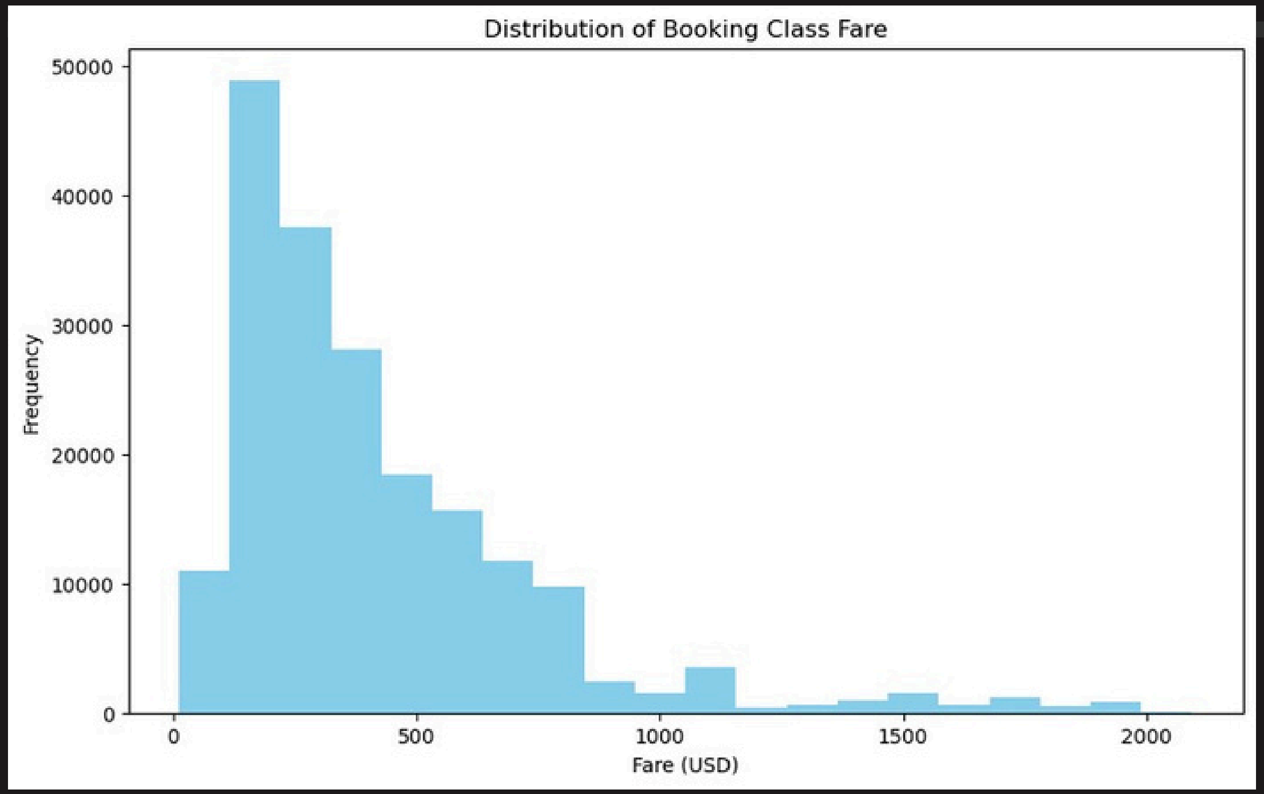
7

Data Preprocessing for Model: To prepare the data for model input, we converted the "Rclassdaysto departure" columns into timesteps, where each timestep is represented as a vector of dimension 1x2. This vector contains the number of tickets booked on that day and the days remaining before departure. We also encoded the markets based on the origin and destination.

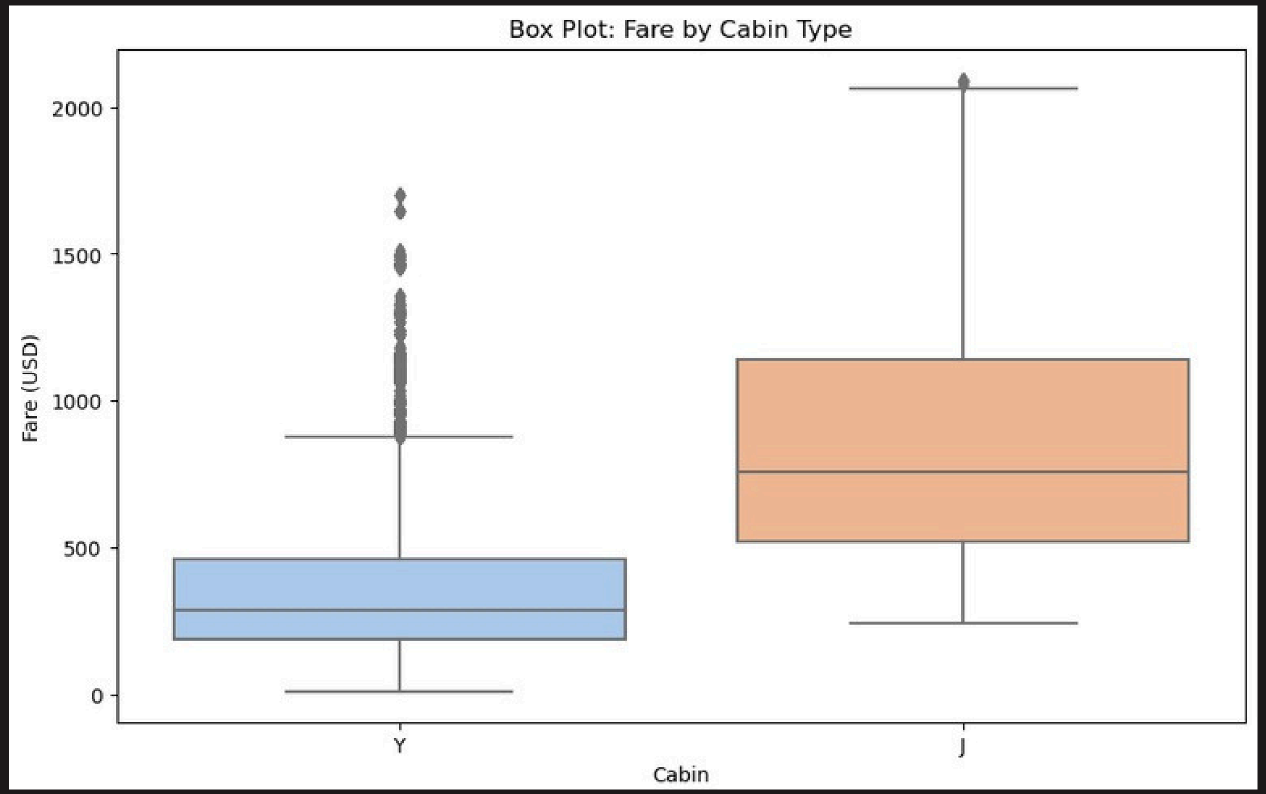
8

The model used is an encoder-decoder architecture with Long Short-Term Memory (LSTM) recurrent neural networks (RNNs). It is designed for sequence-to-sequence prediction, specifically for forecasting future flight bookings. It is explained comprehensively in the next slide.

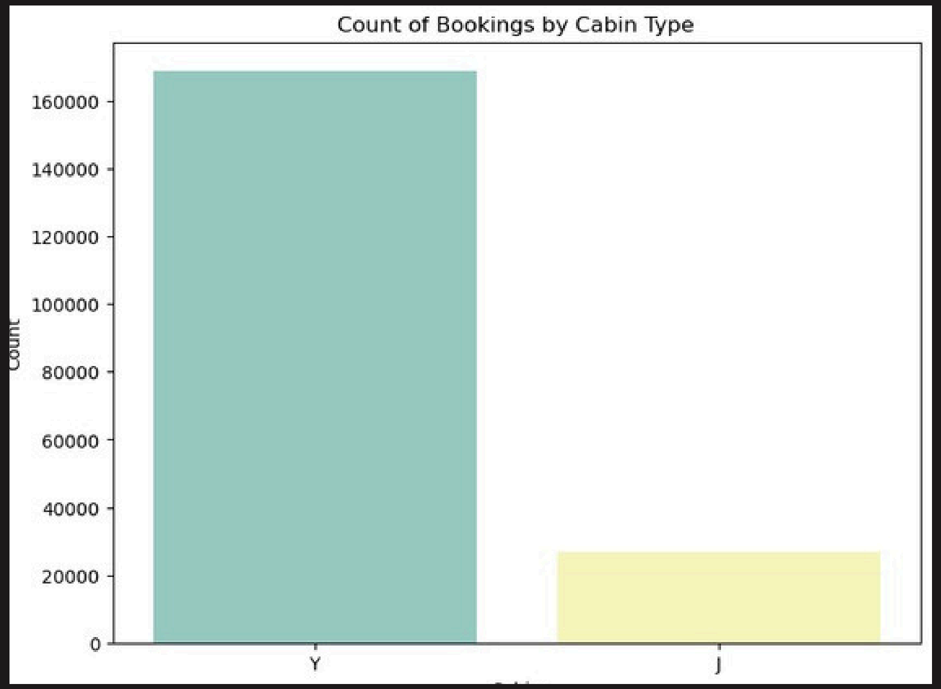
DATA VISUALIZATION



The graph suggests that most people are paying relatively low fares for their flights, while a small number of people are paying very high fares. This is likely due to a variety of factors, such as different travel needs and preferences, as well as different booking strategies.

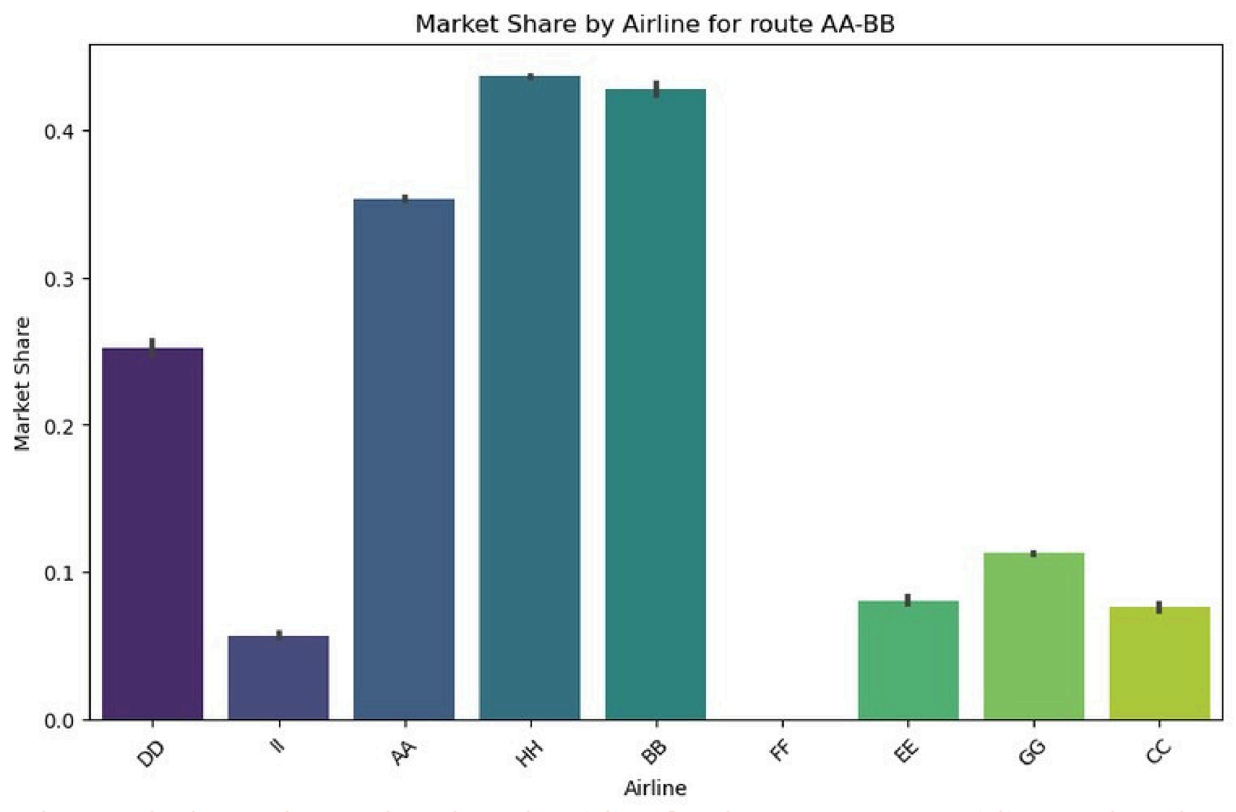


The box plot shows that the average fare for First class is the highest, followed by Business class and then Economy class. This suggests that First class cabins are the most expensive, while Economy class cabins are the most affordable.

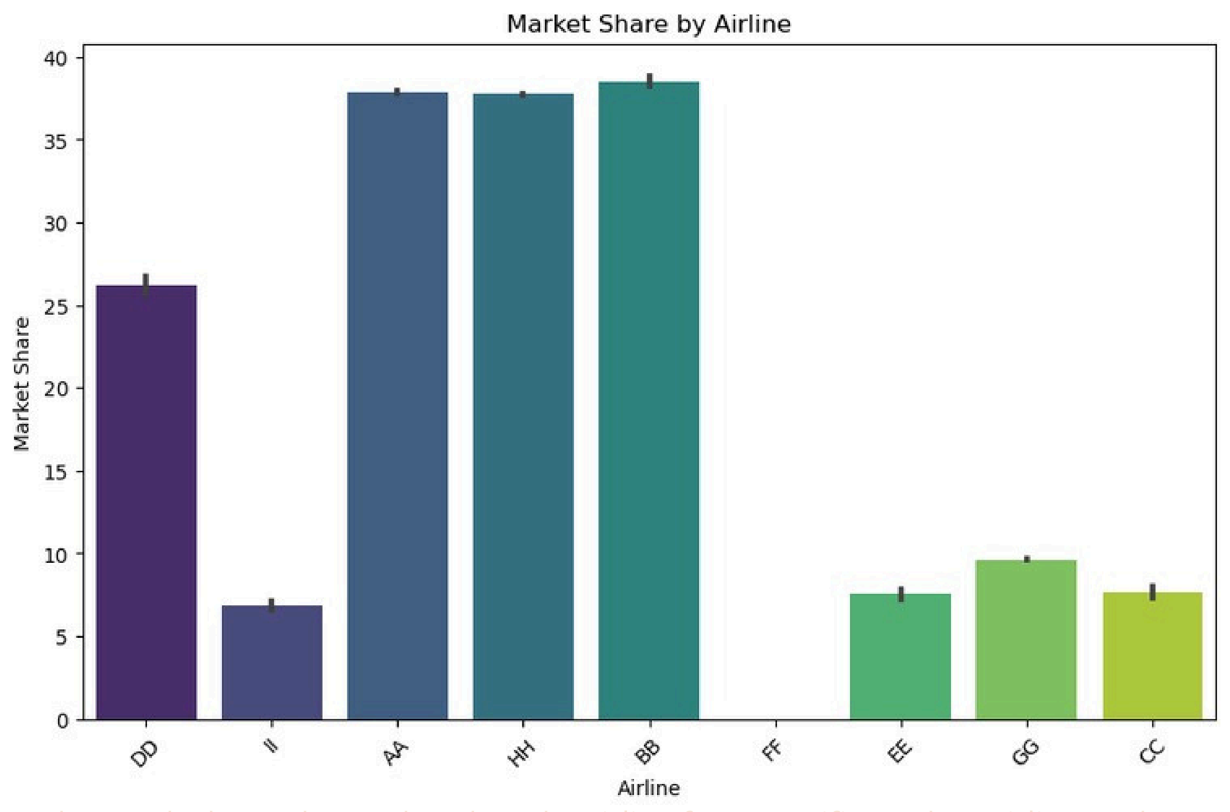


The number of bookings for Economy class is the highest, followed by Business class and then First class. This suggests that most people are choosing to book Economy class cabins, while fewer people are choosing to book Business or First class cabins.

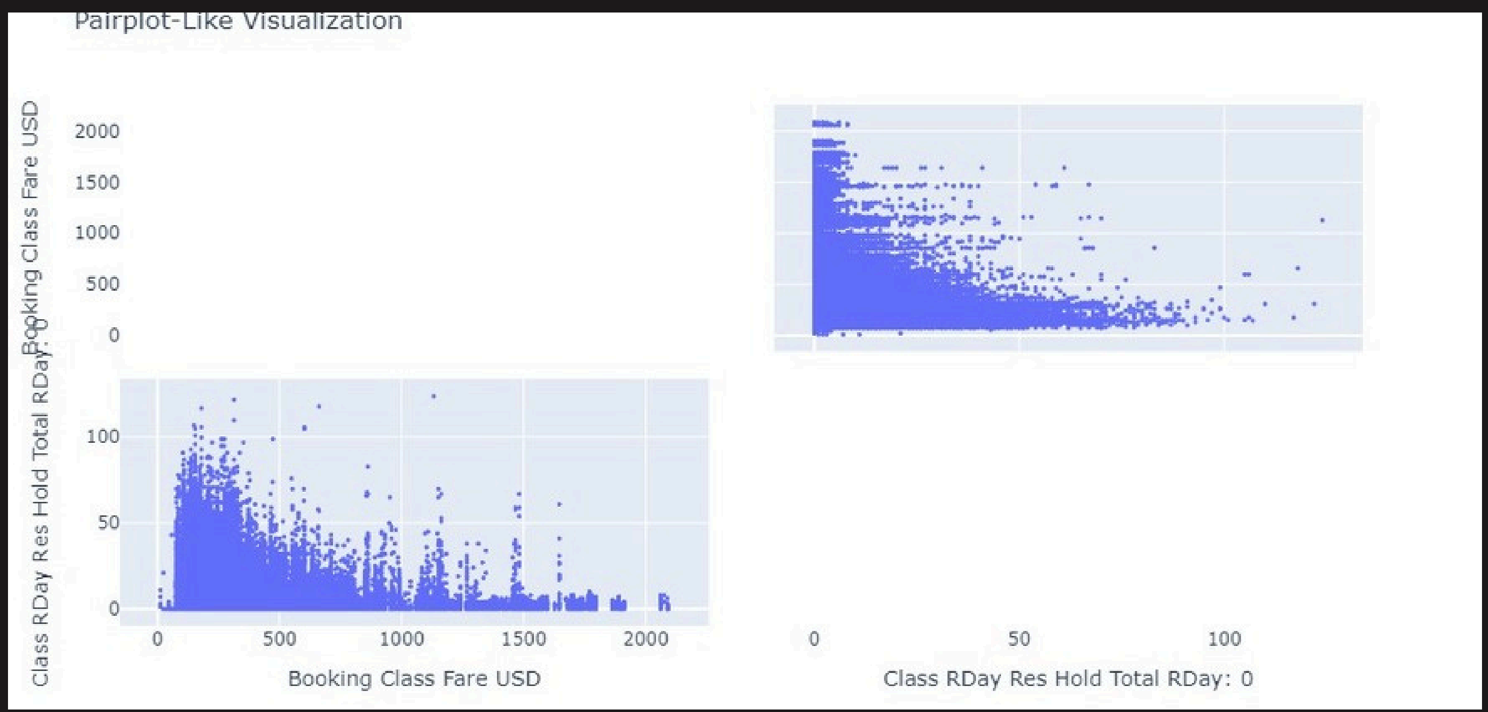
DATA VISUALIZATION



The graph shows the market share by airline for the route AA-BB. Airline DD has the highest market share, followed by HH and BB. The other airlines have a much smaller market share.



The graph shows the market share by airline for a specific market. Airline DD has the highest market share, followed by AA, HH, and BB. The other airlines have a much smaller market share.



- There is a positive correlation between the booking class fare and the total rdh rate.
- Other factors, such as the time of year, the route being flown, and the demand for seats, can also affect the total rdh rate.

OUR MODEL

01 ENCODER-DECODER ARCHITECTURE:

- In a sequence-to-sequence forecasting problem like predicting future flight bookings, you use an encoder-decoder architecture.
- The encoder takes in historical data as input and encodes it into a fixed-size context vector or hidden state.
- The decoder uses this context vector to generate future predictions for the desired forecast horizon.

02 SPLITTING THE DATA:

Dataset is split into two segments based on time:

- Encoder Data: This segment includes data from the first 9 months of historical information. This data serves as the input to the encoder.
- Decoder Data: This segment comprises the subsequent 3 months of historical information. It is used as input to the decoder to generate predictions for the future.

03 PURPOSE:

- During training, the encoder processes the historical data for the first 9 months and creates a context vector that captures patterns and dependencies in the historical data.
- The decoder is then trained to use this context vector to generate predictions for the next 3 months.

04 TRAINING:

- During training, the encoder processes the first 9 months of data to create a context vector.
- The decoder is trained to generate predictions using this context vector and the next 3 months of data.

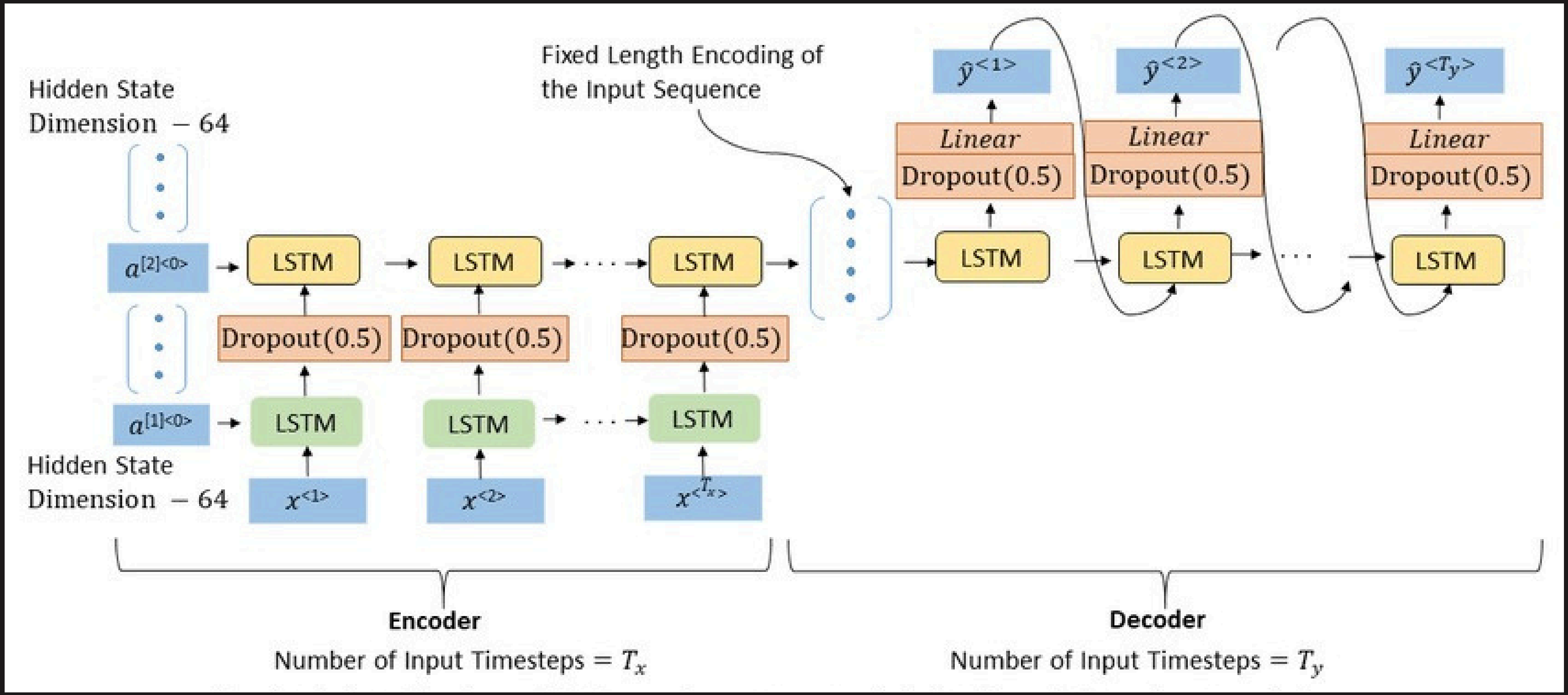
05 FORECASTING:

- When making predictions for future events, we feed the historical data for the 3 months following the 9-month encoder period.
- The encoder processes this data to create a context vector, which the decoder uses to generate predictions for the forecasted period.

06 SEQUENCE LENGTH:

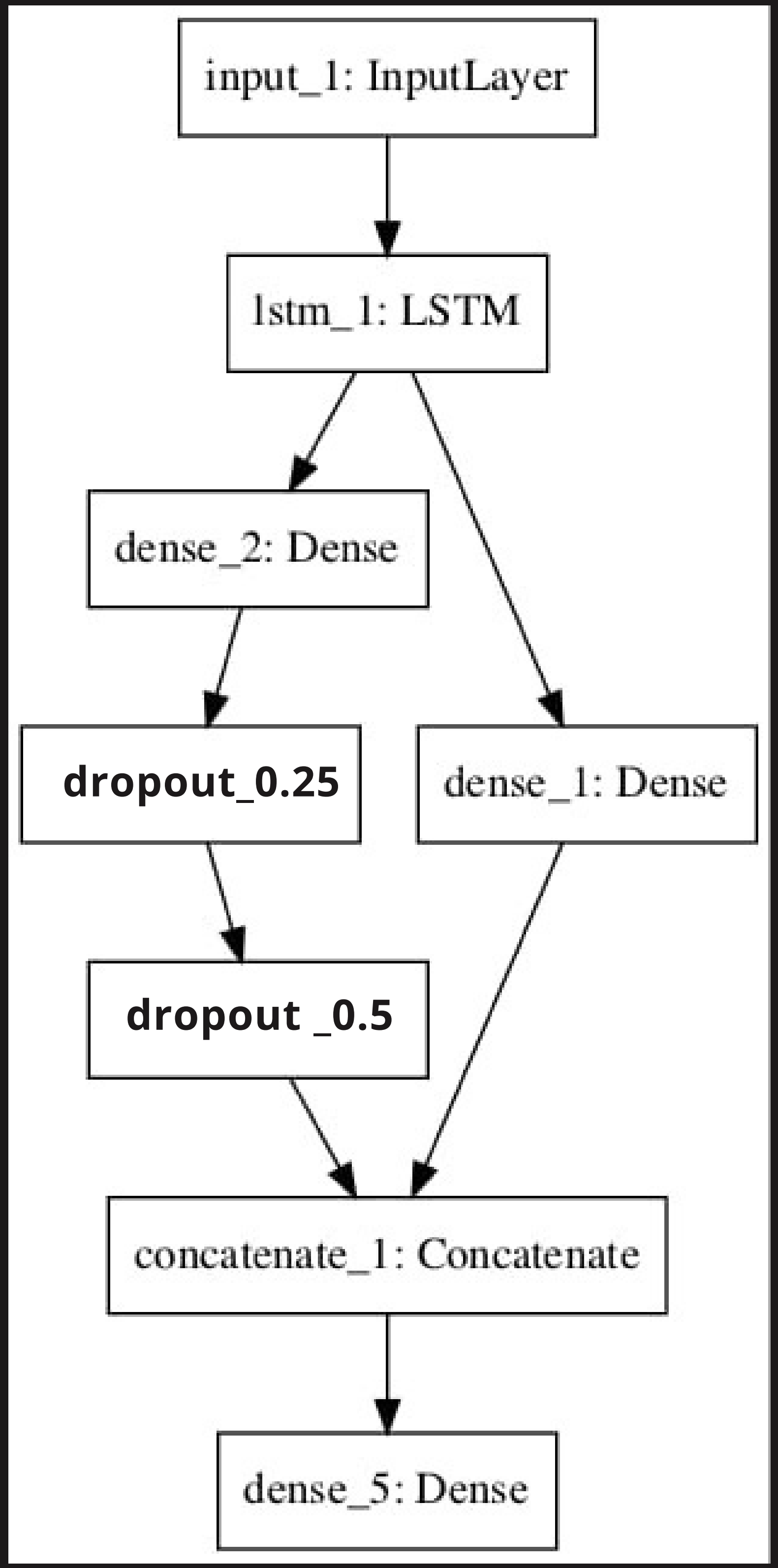
- The length of the encoder and decoder segments can be adjusted based on dataset characteristics and modeling requirements.
- Longer encoder sequences capture more historical context, potentially improving forecasting accuracy.

OUR MODEL



- The algorithm uses a stacked LSTM model with encoder decoder to predict time series.
- The encoder captures the long-term dependencies in the input time series.
- The decoder generates the predicted output time series step-by-step.
- Dropout is used to prevent overfitting.

ALTERNATE SOLUTION



FUTURE EXTENSIONS AND REFINEMENTS

INCORPORATE ADVANCED RNN ARCHITECTURES

Experiment with advanced RNN architectures such as Bidirectional LSTMs or GRUs to capture more complex temporal dependencies in the data.

ENSEMBLE METHODS

Explore ensemble methods to combine predictions from multiple models, reducing prediction variance and improving overall performance.

DYNAMIC FORECASTING HORIZONS

Implement dynamic forecasting horizons that adapt to specific routes or markets, enabling predictions for different time intervals based on historical patterns.

ATTENTION MECHANISMS

Integrate attention mechanisms into the model for improved focus on relevant information in the historical data.

RESULTS

```
# Output layers for time series predictions
output_time_series_1 = Dense(93, activation='linear', name='output_time_series_1')(dense2)
output_time_series_2 = Dense(93, activation='linear', name='output_time_series_2')(dense2)

# Build the model
model = Model(inputs=[input_single_valued, input_time_series], outputs=[output_time_series_1, output_time_series_2])

# Compile the model
model.compile(optimizer='adam', loss='mean_squared_error', metrics=['accuracy'])

model.summary
```

```
<bound method Model.summary of <keras.src.engine.functional.Functional object at 0x00000248B7191B90>>
```

```
model.fit([tf.convert_to_tensor(rem, dtype=tf.float32), tf.convert_to_tensor(X_ts, dtype=tf.float32)], [tf.convert_to_tensor(y_ts[
```

```
output_time_series_1_accuracy: 0.8769 - output_time_series_2_accuracy: 0.8788 - val_loss: 0.8360 - val_output_time_series_1_loss: 0.4

output_time_series_1_accuracy: 0.8776 - output_time_series_2_accuracy: 0.8795 - val_loss: 0.8335 - val_output_time_series_1_loss: 0.4

output_time_series_1_accuracy: 0.8783 - output_time_series_2_accuracy: 0.8801 - val_loss: 0.8314 - val_output_time_series_1_loss: 0.4

output_time_series_1_accuracy: 0.8793 - output_time_series_2_accuracy: 0.8812 - val_loss: 0.8297 - val_output_time_series_1_loss: 0.4
```

CONCLUSION

In this presentation, we have explored the development and application of an advanced predictive model for optimizing airline revenue through accurate forecasting of future flight bookings. Our journey through this project has revealed the tremendous potential of data-driven decision-making in the dynamic and competitive airline industry.

Key Achievements:

1. Accurate Forecasting: Our predictive model, built on an encoder-decoder architecture with recurrent neural networks (RNNs), has demonstrated its ability to accurately forecast future flight bookings. This capability is invaluable for airlines seeking to align seat pricing and allocation with expected demand.
2. Data-Driven Decisions: By leveraging historical booking data, competitive market insights, and advanced machine learning techniques, we have empowered airlines to make data-driven decisions that optimize revenue and profitability.
3. Dynamic Adaptation: The model's adaptability to changing booking patterns and market dynamics ensures that airlines can respond swiftly to shifts in demand and competition.
4. Enhanced Insights: Through model evaluation, we have gained valuable insights into the factors influencing booking behavior, enabling a deeper understanding of customer preferences and market trends.