Initial Project Proposal

PREDICTIVE MODELING OF CUSTOMER BOOKING A FLIGHT FOR AN AIRLINE COMPANY

MIS 637- Data Analytics and Machine Learning

By Prof. Mahmoud Daneshmand

Vanshika Nair

Sakshi Tonde

Shreya Melam

Medha Baddam

Table of contents

Sr.No	Title
1.	Project Title
2.	Problem Statement
3.	Project Objective
4.	Dataset
5.	Software Used
6.	Algorithm Used

1. Project Title :

PREDICTIVE MODELING OF CUSTOMER BOOKING A FLIGHT FOR AN AIRLINE COMPANY.

2. Problem Statement

The airline industry is very competitive, and airlines are constantly looking for new methods to attract consumers and raise revenue. One method is to forecast client booking behavior and give them customized bargains and incentives that are likely to motivate them to buy a flight. Predicting client booking behavior, on the other hand, is a difficult process since it requires assessing a large quantity of data from many sources, such as customer demographics, past booking habits, and external factors such as economic trends and seasonal fluctuations.

This project aims to develop a predictive model that can accurately forecast customer booking patterns for an airline. This project's predictive model can assist the airline sector in optimizing operations, increasing revenue, and increasing customer pleasure. The airline can plan its operations, including pricing strategies, promotions, and seat inventories, by properly forecasting passenger booking trends. Furthermore, the model can assist the airline in providing individualized services to its consumers by gaining a knowledge of their preferences and behavior.

3. Project Objective

Customers are more powerful than ever before because they have instant access to a multitude of information. This is one of the reasons why the buying cycle is no longer what it once was.

With the use of data and prediction models, this is achievable. The quality of the data used to train the machine learning algorithms is the most significant aspect in a prediction model.

The objective is to use machine learning algorithms and data analytics techniques to create a dependable prediction model that will assist the airline in optimizing its operations and making informed decisions about flight capacity, pricing, and

scheduling. The success of this initiative will have far-reaching consequences for the airline sector, resulting in better customer happiness, more revenue, and lower operating costs.

4. Data Set

Data Source: https://www.kaggle.com/datasets/ghassenkhaled/airways-customer data?datasetId=2914207&select=customer_airways_data.csv

Data Set Used:

Customer_booking

num_passengers	sales_channel	trip_type	purchase_lead	length_of_stay	flight_hour	flight_day	route	booking_origin	wants_extra_baggage	wants_preferred_seat	wants_in_flight_meals	flight_duration	booking_complete
2	Internet	RoundTrip	262	19	7	Sat	AKLDEL	New Zealand	1	0	0	5.52	0
1	Internet	RoundTrip	112	20	3	Sat	AKLDEL	New Zealand	0	0	0	5.52	0
2	Internet	RoundTrip	243	22	17	Wed	AKLDEL	India	1	1	0	5.52	0
1	Internet	RoundTrip	96	31	4	Sat	AKLDEL	New Zealand	0	0	1	5.52	0
2	Internet	RoundTrip	68	22	15	Wed	AKLDEL	India	1	0	1	5.52	0
1	Internet	RoundTrip	3	48	20	Thu	AKLDEL	New Zealand	1	0	1	5.52	0
3	Internet	RoundTrip	201	33	6	Thu	AKLDEL	New Zealand	1	0	1	5.52	0
2	Internet	RoundTrip	238	19	14	Mon	AKLDEL	India	1	0	1	5.52	0
1	Internet	RoundTrip	80	22	4	Mon	AKLDEL	New Zealand	0	0	1	5.52	0
1	Mobile	RoundTrip	378	30	12	Sun	AKLDEL	India	0	0	0	5.52	0
2	Internet	RoundTrip	185	25	14	Tue	AKLDEL	United Kingdom	1	1	1	5.52	0
1	Internet	RoundTrip	8	43	2	Sat	AKLDEL	New Zealand	1	1	1	5.52	0
4	Internet	RoundTrip	265	24	19	Mon	AKLDEL	New Zealand	1	0	1	5.52	0
1	Internet	RoundTrip	185	17	14	Fri	AKLDEL	India	0	0	0	5.52	0
1	Internet	RoundTrip	245	34	4	Tue	AKLDEL	New Zealand	1	1	1	5.52	0
1	Internet	RoundTrip	192	18	14	Thu	AKLDEL	India	1	0	0	5.52	0
1	Internet	RoundTrip	259	37	6	Sun	AKLDEL	India	0	0	0	5.52	0
1	Internet	RoundTrip	19	29	10	Sat	AKLDEL	New Zealand	0	0	1	5.52	0
1	Internet	RoundTrip	67	155	8	Sun	AKLDEL	New Zealand	1	0	0	5.52	0
1	Internet	RoundTrip	351	17	3	Sun	AKLHGH	China	0	0	0	5.07	0
2	Internet	CircleTrip	228	29	23	Wed	AKLHND	New Zealand	1	0	0	7.57	0
1	Internet	RoundTrip	21	88	9	Wed	AKLICN	South Korea	1	0	1	6.62	0
2	Internet	RoundTrip	23	19	8	Mon	AKLICN	New Zealand	1	0	1	6.62	0
2	Internet	RoundTrip	25	27	13	Fri	AKLICN	South Korea	1	0	1	6.62	0

Dataset Description

To provide more context, below is a more detailed data description, explaining exactly what each column means:

num_passengers = number of passengers traveling
sales_channel = sales channel booking was made on
trip_type = trip Type (Round Trip, One Way, Circle Trip)
purchase_lead = number of days between travel date and booking date
length_of_stay = number of days spent at destination
flight_hour = hour of flight departure
flight_day = day of week of flight departure
route = origin -> destination flight route
booking_origin = country from where booking was made
wants_extra_baggage = if the customer wanted extra baggage in the booking, denoted by 0 for no extra baggage and 1 for extra baggage.
wants_preferred_seat = if the customer wanted a preferred seat in the booking, denoted by 0 for no preferred seating and 1 for needed preferred seating.
wants_in_flight_meals = if the customer wanted in-flight meals in the booking, denoted by 0 for no flight meals and 1 for in flight meals.
flight_duration = total duration of flight (in hours)
booking_complete = flag indicating if the customer completed the booking

5. Softwares Used

- IDE- Jupyter Notebook

- Language- Python
  Packages:
    - Pandas
    - NumPy
    - Scikit
    - Seaborn
    - MatPlot

- Microsoft Excel(csv file for dataset)

6. Algorithms Used

The algorithm used in the project is Random Forest Algorithm. Random Forest is a well-known machine learning technique that may be used to anticipate consumer bookings for an airline. It is an ensemble learning approach that integrates many decision trees to increase the model's accuracy and resilience. Random Forest works by picking a subset of the training data and a subset of the features at random for each decision tree, ensuring that each tree is unique and has its own interpretation of the data. The output of each tree is pooled to form the final forecast during the prediction phase.

Random Forest may be used in the context of customer booking for an airline to detect patterns and estimate the possibility of a consumer booking a flight by analyzing multiple criteria such as flight dates, ticket pricing, customer demographics, flight routes, and previous booking data. The model's findings may be utilized to optimize pricing and marketing initiatives, improve customer experience, and increase revenue for the airline.