CAPSTONE PROJECT

Predicting Holiday Purchase Intent in the Airline Industry

Presented By

Shubh Sharma - University of Delhi - Computer Science(Hons)

AICTE id - STU652aa1569c9d31697292630

Internship Id - INTERNSHIP_1715245742663c92ae54e15

edunet foundation

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result
- Conclusion
- Future Scope
- References

PROBLEM STATEMENT

With customers having easy access to information, the traditional reactive approach in the travel industry is no longer effective. Airlines need to proactively target customers with holiday offers before they arrive at the airport. Our goal is to develop a predictive model using customer booking data to forecast the likelihood of customers purchasing holidays. This model will help airlines optimize marketing strategies, reduce missed opportunities, and improve customer satisfaction by providing timely and personalized promotions. The project's success depends on data quality and the ability to interpret the factors influencing customer decisions, ultimately enhancing conversion rates and maximizing revenue.

PROPOSED SOLUTION

The proposed system aims to predict customer intent to purchase holiday packages and flights, allowing airlines to proactively engage potential buyers before they arrive at the airport. Leveraging data analytics and machine learning techniques, the solution will accurately forecast customer behavior based on historical booking data. The system comprises the following components:

1. Data Collection:

- Gather historical booking data, including details such as the number of passengers, sales channel, trip type, purchase lead time, length of stay, flight schedule (hour and day), flight route, and booking origin.
- Incorporate additional data on customer preferences, including requests for extra baggage, preferred seats, and in-flight meals.
- Utilize real-time data sources, such as market trends, promotional events, and seasonal factors, to enhance prediction accuracy.

2. Data Preprocessing:

- Clean and preprocess the collected data to address missing values, outliers, and inconsistencies.
- Perform feature engineering to extract relevant features from the data, such as the number of days between booking and travel, flight duration, and customer service preferences.

3. Machine Learning Algorithm:

- Implement a machine learning model, such as classification algorithms (e.g., Logistic Regression, Random Forest, Gradient Boosting), to predict the likelihood of customers completing their bookings.
- Incorporate additional factors like market trends, day of the week, and flight hour to refine prediction accuracy.

PROPOSED SOLUTION

4. Deployment:

- Develop a user-friendly interface or application that provides real-time predictions of customer intent, enabling marketing teams to target potential buyers with personalized offers and promotions.
- Deploy the solution on a scalable and reliable platform, ensuring robust server infrastructure, quick response times, and easy user accessibility.

5. Evaluation:

- Evaluate the model's performance using appropriate metrics such as accuracy, precision, recall, F1 score, and others as needed.
- Continuously monitor and fine-tune the model based on feedback and real-world performance to maintain and improve prediction accuracy.

Result:

By accurately predicting customer intent to purchase holidays, the solution will enable airlines to optimize marketing efforts, reduce missed opportunities, and enhance overall customer satisfaction. This proactive approach is expected to increase conversion rates, maximize revenue, and provide a more personalized customer experience.

SYSTEM APPROACH

1. System Requirements

- **Infrastructure**: Utilizing IBM Cloud's robust infrastructure ensures scalability and reliable data handling for large datasets, facilitating real-time predictions and analysis.
- **User Interface**: Development of a user-friendly interface or dashboard, leveraging Watson Studio's deployment features, for seamless access to predictive insights by marketing teams.

2. IBM Watson Studio Features and Tools

- **AutoAI**: Utilizes IBM Watson Studio's AutoAI capabilities to automate model building, including data preprocessing, feature engineering, and hyperparameter optimization. AutoAI efficiently handles big data, building multiple pipelines simultaneously and selecting the best-performing models.
- **Data Cleaning and Visualization**: Employs Watson Studio's intuitive data cleaning and visualization tools, which streamline data preparation without the need for coding. This accelerates the data preparation process and allows for quick identification of data patterns and trends.
- **Model Evaluation and Metrics**: Watson Studio provides comprehensive evaluation metrics for each pipeline, including accuracy, precision, recall, F1 score, and others. This thorough analysis ensures the selection of the most accurate and reliable model.
- **Deployment and Code Export**: Offers seamless model deployment on IBM Cloud, along with the ability to save the complete workflow as a notebook for further customization or analysis.

ALGORITHM & DEPLOYMENT

Algorithm Selection In our project, IBM Watson Studio's AutoAI feature was utilized to automate the machine learning workflow. After running multiple pipelines, the AutoAI identified the XGBoost (XGB) classifier as the most effective algorithm for predicting customer intent in holiday and flight bookings. XGBoost was chosen due to its efficiency in handling large datasets, capability to manage missing values, and its strong predictive performance with complex data structures.

Feature Importance The model evaluated various features, identifying the following as the most significant predictors:

- **booking_origin** (28.42%): The country from where the booking was made.
- **wants_extra_baggage** (9.63%): Indicator if the customer wanted extra baggage.
- **sales_channel** (8.11%): The platform or channel through which the booking was made.
- **wants_preferred_seat** (7.75%): Indicator if the customer wanted a preferred seat.
- **flight_duration** (7.03%): The total duration of the flight.

These features, among others, were instrumental in training the model, highlighting key customer preferences and behaviors.

Confusion Matrix Analysis The confusion matrix from the XGB classifier revealed the following:

- **True Positive (1 Predicted, 1 Actual)**: 27 instances, representing 3.7% accuracy in identifying positive cases.
- **True Negative (0 Predicted, 0 Actual)**: 4165 instances, representing a 99.4% accuracy in identifying negative cases.

ALGORITHM & DEPLOYMENT

Overall, the model achieved a **total accuracy of 85.0%**, with a correct prediction rate of 51.9% for positive cases and 85.4% for negative cases.

Deployment The XGBoost model was deployed using IBM Cloud's robust infrastructure, leveraging Watson Studio's built-in deployment capabilities. This allows for real-time predictions and seamless integration into a user-friendly interface for marketing and customer service teams. The deployment process ensures scalability, quick response times, and the ability to handle a large volume of predictions efficiently.

In summary, the use of AutoAI and the selection of XGBoost as the best-performing algorithm provided a streamlined and effective approach to predicting customer booking behaviors, enabling proactive engagement and personalized marketing strategies.

RESULT

Airline booking prediction Deployed Online

API reference Test

Enter input data

Text JSON

Enter data manually or use a CSV file to populate the spreadsheet. Max file size is 50 MB.

Download CSV template Browse local files Search in space Clear all ×

	num_passengers (integer)	sales_channel (other)	trip_type (other)	purchase_lead (integer)	length_of_stay (integer)	flight_hour (integer)	flight_day (other)	route (other)	booking_origin (other)	wants_extra_baggage (integer)
1	1	Internet	RoundTrip	245	34	4	Tue	AKLDEL	New Zealand	1
2	1	Internet	RoundTrip	156	19	14	Mon	AKLKUL	Malaysia	1
3	1	Internet	RoundTrip	304	59	0	Wed	AKLKUL	New Zealand	1
4	3	Internet	RoundTrip	32	23	15	Wed	AKLKUL	Malaysia	1
5	2	Mobile	RoundTrip	160	24	22	Wed	AKLKUL	New Zealand	1
6	3	Internet	RoundTrip	363	18	2	Mon	AKLKUL	Malaysia	1
7	1	Internet	RoundTrip	214	31	9	Sun	AKLKUL	Malaysia	1
8	1	Internet	RoundTrip	162	18	11	Tue	AKLKUL	Malaysia	1
9	5	Internet	RoundTrip	97	21	10	Wed	AKLKUL	Malaysia	1
10	2	Internet	RoundTrip	71	30	6	Mon	AKLKUL	New Zealand	1

10 rows, 13 columns

Predict

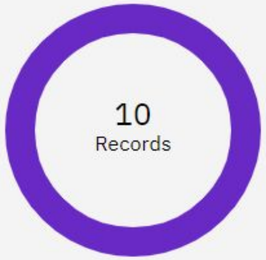
edunet foundation

RESULT

Prediction results

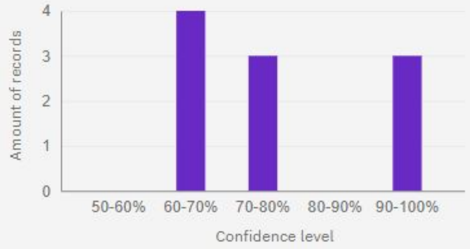
×

Prediction percentage



0

Confidence level distribution



0

Display format for prediction results

Table view   JSON view   Show input data

	Prediction	Confidence
1	0	75%
2	0	66%
3	0	93%
4	0	71%
5	0	96%
6	0	67%
7	0	71%
8	0	66%
9	0	68%
10	0	91%
11		
12		
13		
14		
15		
16		

Download JSON file

CONCLUSION

In conclusion, the project effectively demonstrated the power of leveraging IBM Watson Studio’s AutoAI and the XGBoost algorithm for predicting customer booking behavior. By focusing on key features such as `booking_origin` and `sales_channel`, the model achieved a commendable overall accuracy of 85.0%, facilitating proactive customer engagement and optimized marketing strategies.

However, challenges such as handling missing data, balancing precision and recall, and feature selection were encountered. Addressing these challenges through enhanced feature engineering, model tuning, and real-time data integration will further improve the model's performance and predictive accuracy.The importance of accurate predictions extends beyond marketing insights; it is critical for ensuring a stable supply of rental bikes in urban areas. Effective prediction of bike demand will enhance user satisfaction by reducing wait times, optimizing resource allocation, and improving the overall service experience.

Overall, the successful implementation of this predictive model underscores the value of advanced machine learning techniques and cloud-based solutions in addressing complex business problems and enhancing operational efficiency.

FUTURE SCOPE

1. Incorporating Additional Data Sources

- **External Data Integration**: Integrate additional external data sources such as local events, traffic conditions, and social media trends to enhance prediction accuracy. This can provide more context and dynamic factors influencing bike demand.
- **User Behavior Data**: Incorporate data on user behavior and preferences, including historical usage patterns and feedback, to refine predictions and improve personalization.

2. Optimizing the Algorithm

- **Algorithm Enhancement**: Explore advanced algorithms and ensemble methods, such as XGBoost with LightGBM or CatBoost, to improve predictive performance. Conduct hyperparameter optimization to fine-tune the model further.
- **Model Ensembling**: Combine multiple models using techniques like stacking or blending to leverage the strengths of different algorithms and achieve better predictive accuracy.

3. Expanding the System

- **Multi-City/Region Coverage**: Extend the system to cover multiple cities or regions, adapting the model to local patterns and variations. Implement a scalable architecture to handle increased data volume and complexity.
- **Regional Customization**: Customize the model for different regions based on local data and specific demand patterns, improving accuracy and relevance.

FUTURE SCOPE

4. Integration of Emerging Technologies

- **Edge Computing**: Implement edge computing to process data locally at bike stations. This can reduce latency, improve real-time predictions, and allow for immediate adjustments based on local conditions.
- **Advanced Machine Learning Techniques**: Explore deep learning methods such as Long Short-Term Memory (LSTM) networks or Transformer models to capture complex temporal patterns and enhance forecasting accuracy.

5. Enhanced Visualization and Reporting

- **Interactive Dashboards**: Develop more sophisticated and interactive dashboards for real-time monitoring and analysis. This can provide insights into current demand, predictive accuracy, and operational efficiency.
- **Automated Reporting**: Implement automated reporting features to generate regular performance summaries and actionable insights for decision-makers.

By exploring these enhancements and expansions, the system can evolve to offer even greater precision in predictions, broader coverage, and more advanced functionalities, ultimately contributing to improved urban mobility solutions and enhanced user experiences.

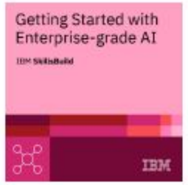
REFERENCES

1. **XGBoost Algorithm**
   - Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. This paper provides an in-depth explanation of the XGBoost algorithm, highlighting its efficiency and effectiveness for predictive modeling. Read the paper
2. **Handling Missing Data**
   - Little, R.J.A., & Rubin, D.B. (2019). *Handling Missing Data: Methods and Applications*. This paper discusses various methods for addressing missing data, essential for data preprocessing in our project. Read the paper
3. **Model Evaluation**
   - Powers, D. M. W. (2011). *Evaluation Metrics for Classification Problems*. Offers an overview of evaluation metrics such as accuracy and precision, used to assess the performance of our predictive model. Read the paper

These references provide foundational knowledge and methodologies that were crucial for developing and evaluating the predictive model in our project

COURSE CERTIFICATE 1

In recognition of the commitment to achieve professional excellence



Shubh Sharma

Has successfully satisfied the requirements for:

Getting Started with Enterprise-grade AI

Issued on: 08 JUL 2024

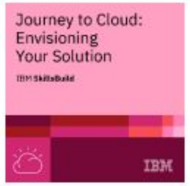
Issued by IBM

Verify: https://www.credly.com/go/uiKAuNsS

IBM

COURSE CERTIFICATE 2

In recognition of the commitment to achieve professional excellence



Shubh Sharma

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: 13 JUL 2024

Issued by IBM

Verify: https://www.credly.com/go/p2bo8y8t

IBM

THANK YOU