Travel Insurance Recommendation AI System Based on Flight Delay Predictions and Customer Sentiment

Proposal

Yuzhe Yang[1], **Haoqi Zhang**[1], **Zhidong Peng**[1], **Yilin Guo**[2], **Tianji Zhou**[3]
[1]School of Data Science, [2]School of Medicine, [3]School of Management and Economics

1 Introduction

As the economy and aviation industry develops, more and more people choose to travel by plane for leisure and business trips. As a result, flight delays have become a major source of conflict between passengers and airlines, delays always cause much loss to the airlines [1], and lead to negative emotions to passengers [2]. Flight delays disrupt passengers' travel plans. Consequently, flight delay insurance emerged, underwriting the airline's on-time departure credibility to compensate passengers for delays, mitigating their inconvenience. While sold as a standalone policy, it is often offered as a supplementary coverage to travel accident insurance. However, nowadays many airlines cannot provide great delay prediction values as a reference for passengers, but only calculate the average value based on historical records, which reveals that there exists an urgent need to develop better insurance strategies.

Therefore, Our project will design a new Al system to predict flight delays and price the insurance, enabling the airlines can provide personalized insurance recommendations for passengers. This Al system will analyze the rates of flight delays based on multiple relevant data, such as flight dynamic data, city weather, and special situations data. Besides, it will construct user profiles with sentiment analysis based on users' feedback to airlines. Finally, the system will offer a personalized and optimal insurance recommendations, contributing to supplying a suitable service for each customer and gaining more benefits for airlines. Through personalized insurance recommendations, our AI system can boost customer satisfaction and loyalty while reducing stress caused by flight delays. Our system could reshape customer service in the airline industry and promote a positive image and competitive edge for airlines, resulting from focusing on prediction accuracy and user interaction quality.

2 Dataset

The collected data can be classified into two main categories based on their usage. The first category of data is primarily used for analyzing and predicting flight delay rates, including flight dynamic data, city weather, special situations, from January 2018 to December 2019. The flight dynamic data is collected from United States Department of Transportation official database. The detailed variable description of flight dynamic is shown in Table 1.

Table 1: Brief Description of Flight Dynamic Data

Field Name	**Description**
Flight_Date	Scheduled flight date (yyyymmdd)
Flight_Number_Operating_Airline	Flight's unique ID
OriginCityName	Origin city
DestCityName	Destination city
Cancelled	Flight cancellation indicator (1=Yes)
CancelledCode	Cancellation reason (Carrier, Weather, National Air System, etc.)
CarrierDelay	Airline control delay (min)
WeatherDelay	Weather condition delay (min)
NASDelay	National Air System delay (min)
SecurityDelay	Security issue delay (min)
LateAircraftDelay	Late incoming aircraft delay (min)

The city weather is collected from U.S. Climate Normals by the National Centers for Environmental Information (NCEI). This data set includes the maximum temperature, minimum temperature, precipitation, wind of cities across the United States. The data will be mainly used to analysis and predict the flight delay caused by the weather conditions. The special situations, mainly refers to the Temporary Flight Restrictions (TFRs) in the U.S., is collected from the Federal Aviation Administration (FAA) website. Delays that are within the control of the NAS has a lot to do with TFRs. The data set includes states, types and description of the TFRs. Take the special situations into consideration will effectively reduce the delay forecast error due to unpredictable reasons.

The second category is intended for constructing customer profiles by examining data related to passengers' perceptions of airlines, flight experiences, and attitudes towards purchasing flight delay insurance. We collect two data sets from Kaggle [2] [3] to do sentiment analysis. The data sets includes airline name, overall rating, reviews, review dates and recommendation situation (whether to recommend the reviewed airline to others). These variables will contribute to constructing the customer profiles through sentiment analysis, which facilitates our recommendation of suitable delay insurance policies.

Furthermore, once the entire system is built, we will attempt a public release and collect user experience data to further enhance our system.

[1] https://www.airlines.org/dataset/u-s-passenger-carrier-delay-costs/
[2] https://www.kaggle.com/datasets/khushipitroda/airline-reviews/data
[3] https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment?resource=download&select=Tweets.csv

3 Methodology

Our designed framework of AI system is illustrated in Figure 1. It aims to create a comprehensive insurance recommendation system that can accurately push insurance products to customers, thereby enhancing their willingness to purchase flight delay insurance and improving the user experience.

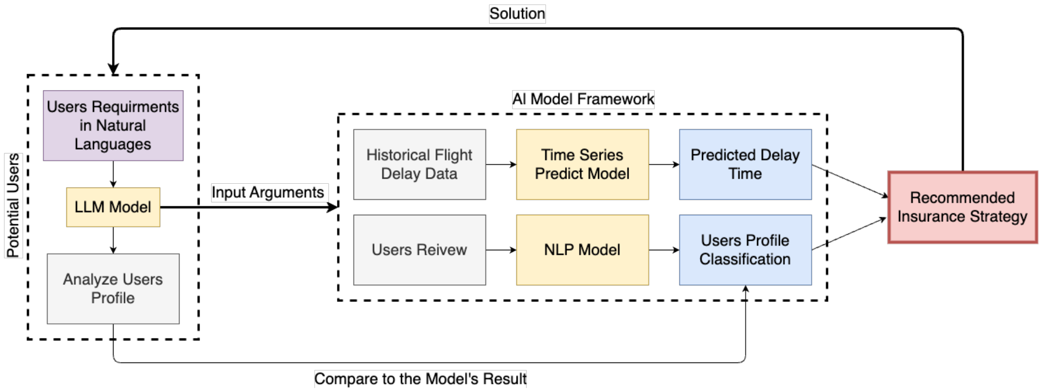

Figure 1: Framework of AI system

The entire system is broadly divided into two parts, with the first being the user requirements processing segment. Our vision is for users to articulate their needs in natural language during their interactions with the system, which would then profile users based on this input and compare it with existing user data. In this segment, to make the LLM to the specific demands of this task, for this purpose, we will try to allocate the LLM into our devices use the python lib of ollama [4], we have two possible choices for the LLM optimization. First, we can try to fine-tune a LLM optimized for insurance-related knowledge using a corpus from the insurance industry [1]; second, if we can't quantize the LLM to enable it to be fine-tuned, we can also try to do RAG for it using the insurance-related knowledge we have generated by ML methods below. Additionally, previous research results have also shown that flight delays can negatively affect passengers' emotions and behaviors, thereby reducing customer loyalty intention [2]. This context underscores the importance of our LLM, we will utilize a specific agent to detect the sentiment when customers chatting with our LLM, and the data we get from the review of previous air flight orders, which is customized and expanded to address these issues by providing targeted insurance product recommendations that can mitigate the negative impacts of such travel disruptions.

The second part is the core of the system, primarily focused on data processing and prediction. The AI model consists of two integrated models. First, the time series prediction model aims to input a time series and forecast future time series moments. Due to limitation of computational resources, we cannot train a very complex model. Therefore, we are aiming to strike a balance between performance and resource efficiency, by referencing some design ideas from the previous study, such as random forests [3] and LSTM [4]. Furthermore, in the NLP model, we analyze user profiles through their comments on flights, assessing the willingness of different user groups to purchase flight delay insurance based on this data. Combining the predictive outcomes from both the time series and NLP models, the system is designed to offer a well-defined range for flight delay insurance options. This integration allows for a nuanced assessment, taking into account the specific flight's potential delay predictions alongside the detailed user profiles derived from their interactions and comments about flights. By doing so, the system can tailor insurance recommendations to match the likelihood of flight delays with the individual needs and preferences of users, providing a personalized and informed insurance selection that enhances customer satisfaction and trust in the service.

To evaluate the performance and user experience of the natural language interaction component, a comparison with existing systems like Ctrip [5] will be conducted. This comparison will likely involve assessing qualitative aspects such as response accuracy, interaction intuitiveness, user satisfaction, and the breadth of understood queries. For the deep learning model, we will use $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$ and $MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$ to measure the performance of the model.

4 Conclusion

Our proposal provides a strategy for improving the airline industry's customer service by integrating advanced AI techniques into the development of personalized flight delay insurance. We seek to precisely predict flight delays and analyze passenger sentiments by utilizing a combination of time series prediction methods and natural language processing. Because of these two techniques, tailored insurance strategies can be developed that consider both emotional and experiential effects of flight delays in addition to the technical ones. By providing personalized flight delay insurance, our AI system aims to improve passenger experience and increase benefits of airlines .

References

[1] M. Feng, B. Xiang, M. R. Glass, L. Wang, and B. Zhou. Applying deep learning to answer selection: A study and an open task. *CoRR*, abs/1508.01585, 2015.

[2] S. Li, Y. Jiang, B. Cheng, and N. Scott. The effect of flight delay on customer loyalty intention: The moderating role of emotion regulation. *Journal of Hospitality and Tourism Management*, 47:72–83, 2021.

[3] J. J. Rebollo and H. Balakrishnan. Characterization and prediction of air traffic delays. *Transportation research part C: Emerging technologies*, 44:231–241, 2014.

[4] W. Zeng, J. Li, Z. Quan, and X. Lu. A deep graph-embedded lstm neural network approach for airport delay prediction. *Journal of Advanced Transportation*, 2021:1–15, 2021.

[4] `https://github.com/ollama/ollama?tab=readme-ov-file`
[5] `https://www.ctrip.com/`