Literature Review for Ongoing Airline Ticket Price Prediction using GLM

Group 8: Neal Xu, Sam Ding, Lulu Wang, Richard Yang

Using generalized linear models to build dynamic pricing systems:

Pricing is one of the most important aspects of business decision-making, as it directly impacts a firm's profitability. In recent years, dynamic pricing has become increasingly popular as it allows firms to adjust prices in real-time based on changes in demand and supply conditions. Generalized linear models (GLMs) have been used to build dynamic pricing systems as they are relatively easy to interpret and allow a clear understanding of how each predictor influences the outcome. This literature review explores the use of GLMs in building dynamic pricing systems, with a focus on the seminal work by Murphy, Brockman, and Lee (2000).

Murphy et al. (2000) proposed the use of GLMs for dynamic pricing, citing their flexibility and ability to handle complex pricing data. The authors point out that standard pricing in the insurance industry follows a set of one-way tables that ignores the different mix of businesses, whereas a multiple regression approach like GLM takes that into account, therefore producing more effective and accurate pricing that optimizes profitability.

The authors begin the paper with a brief overview of the family of GLMs and introduce their parameters. They then evaluate the uses of GLMs in the insurance industry, namely, assessing risk premiums or a policy as well as assessing price elasticity curves. The paper culminates with a proposed pricing algorithm to bring the cost and demand sides of the equation together. The research shows that GLM's model and variance structure reflects the processes in insurance and GLM works well with large data sets and produces robust parameter estimation.

The authors also highlighted some tips for their modeling that would be helpful for our research purposes. One of them is using categorical dummy variables for continuous variables in the original data. In some variables, breaking down numeric variables into several categories may help with interpreting model outcomes, for example, in the case of insurance, age. For our model predicting flight time prices, we have received departure and arrival times in our pricing data converted into categories like "early morning" or "afternoon", which would be helpful in identifying patterns. Another thing Murphy et al. mentioned is the feature selection process. The authors recommended that all the factors should be included in the baseline model, even though that could potentially cause overfitting and reduced accuracy. Doing so avoids prejudgments and will let analysis determine which factors are less important than the others.

In conclusion, even though this paper provides some applications of GLM outside of our research topic, given the year of publication is more than two decades ago, this paper serves as a groundbreaking integration of statistics and economics at their time and provided many insights for future statisticians. We think with this successful employment of GLMs in dynamic pricing for insurance, we can learn from their findings and experiences and build our own.

Dynamic Price Prediction for Amazon Spot Instances:

The article "Dynamic Price Prediction for Amazon Spot Instances" by V. K. Singh and K. Dutta explores a similar theme to the project design proposed. The article discusses a predictive model that was developed to forecast the price of Amazon Spot Instances, which is a type of cloud computing service offered by Amazon. The authors explain that the pricing of Spot Instances is influenced by mainly two components (1) global trend which is captured on a daily and monthly basis and, (2) local seasonality which mainly occurs due to sudden increases or decreases in demand for a particular spot instance at some point of time.

To develop the predictive model, the authors analyzed historical spot prices. The author provided the forecast on an hourly basis because there might be no changes on the user if the spot instance is terminated within one hour from its start. Then the author constructed a regression model to predict the spot price. The authors also used a gradient descent algorithm to estimate parameters on the training set to improve the accuracy of the predictions. The gradient descent algorithm is mainly used to solve a system of linear equations. This is achieved by reformulating the linear equation into a quadratic minimization problem. The author gradually optimized the gradient of the error with respect to parameters using a partial differential equation.

To evaluate the model, the author conducted both short-term and long-term price predictions. In the short-term prediction, the next hour's spot price was predicted using the proposed algorithm with an input of the last hour before data. In the long-term prediction, the model is implemented to forecast the spot price for a week ahead.

The prediction results showed that the predictive model is more effective in forecasting the shorter period spot prices than the longer period. The model is capable of predicting the spot price within 9.4% of MAPE for one day (24 hours) ahead of prediction. However, the MAPE score for the five-day prediction is roughly 20%.

This study effectively demonstrated the application of implementing a regression model and gradient descent method to predict the dynamic price. The conclusions can be well referenced in our project study for flight ticket price prediction. The accuracy differences in predicting longer and shorter periods can be further studied in our research. The model might become less effective if it can only be accurate in the short-term price prediction in real settings.

In addition, the author only implemented linear regression in the spot price model without further studying the non-linear relationships between response variables and the features. This should be thoroughly studied in our ticket price project as the nonlinear influential factors may affect the price changes significantly.

Airline ticket price and demand prediction: A survey:
The airline industry has been growing rapidly in recent years, and the competition among airlines has intensified. As a result, it has become increasingly important for airlines to accurately forecast the demand for airline tickets and to set ticket prices accordingly. Accurate demand and price predictions can help airlines optimize their revenue and profitability, and ultimately improve customer satisfaction. This literature review aims to provide an overview of a variety of machine learning and artificial intelligence techniques to predict airline ticket prices and demand.

In this article, Abdella first briefly introduced two main research areas, which are prediction models that help customers save money and those models that are designed to increase the revenue of airline companies. Then the author presented a comprehensive survey of various models for predicting airline ticket price and demand, including traditional regression models, machine learning models, hybrid models that combine the previous two, and deep learning models. The author found that machine learning and deep learning models have gained increasing popularity in recent years due to their superior performance in prediction accuracy and efficiency.

In addition, the study highlights the importance of considering demand factors, such as seasonality and market trends, when predicting airline ticket prices. The article further outlines the challenges and limitations of the existing models, such as data availability and the complexity of the airline industry.

Our project shares many similarities with the methods and techniques discussed in Abdella. Like the authors, we are using various features to predict airline ticket prices. A more specific example would be that the author proposes a linear quantile regression model for predicting airline ticket prices, which is a type of regression model. For our project, we will use GLM to predict the flight ticket price. Moreover, we are also considering the clustering algorithm to cluster the distance between departure and arrival cities which is similar to the use of a quantile regression model for segmenting prices.

In conclusion, our project aligns well with the methods and techniques discussed in the article by Abdella, and we believe that their insights will guide us in designing an accurate and effective prediction model. The challenges and limitations highlighted in the article will help us in addressing the complexities of the airline industry and improving our model's performance. This paper gives us a comprehensive view of how other kinds of models can be used to solve the same problem we have.

References:

Abdella, Juhar Ahmed, et al. "Airline ticket price and demand prediction: A survey." Journal of King Saud University-Computer and Information Sciences 33.4 (2021): 375-391.

Murphy, K. P., Brockman, M. J., & Lee, P. K. (2000). Using generalized linear models to build dynamic pricing systems. Casualty Actuarial Society Forum, Winter.

Singh, Vivek Kumar, and Kaushik Dutta. "Dynamic Price Prediction for Amazon Spot Instances." *2015 48th Hawaii International Conference on System Sciences*, 2015, https://doi.org/10.1109/hicss.2015.184.