# **EXPERIMENT REPORT**

|  |  |
| --- | --- |
| **Student Name** | Rohan Rocky Britto |
| **Project Name** | Assignment 3 |
| **Date** | 10-November-2023 |
| **Deliverables** | Group\_10-Britto\_Rohan-24610990-multiple\_algorithms.ipynb  Multiple algorithms |

|  |  |
| --- | --- |
| 1. **EXPERIMENT BACKGROUND** | |
| Provide information about the problem/project such as the scope, the overall objective, expectations. Lay down the goal of this experiment and what are the insights, answers you want to gain or level of performance you are expecting to reach. | |
| **1.a. Business Objective** | The business objective is to accurately predict and optimize flight ticket fares specific to airlines operating on a given route at a designated hour, empowering the airline industry to strategically adjust pricing strategies for increased revenue and enhanced customer satisfaction. |
| **1.b. Hypothesis** | Our hypothesis aims to test whether we can accurately predict flight ticket fares specific to airlines on a given route at a designated hour. This can help improve revenue management and customer satisfaction in the airline industry. This hypothesis is worth considering because precise fare predictions empower airlines to strategically adjust pricing strategies, responding to real-time market dynamics. Improved revenue management and competitiveness can be achieved by aligning fares with customer expectations, ultimately enhancing the overall customer experience and loyalty. |
| **1.c. Experiment Objective** | The experiment anticipates a positive impact, manifesting as increased revenue and heightened customer satisfaction through accurate fare predictions. If we are able to provide accurate predictions companies can achieve optimized pricing, driving revenue growth and positive customer feedback. |

|  |  |
| --- | --- |
| 1. **EXPERIMENT DETAILS** | |
| Elaborate on the approach taken for this experiment. List the different steps/techniques used and explain the rationale for choosing them. | |
| **2.a. Data Preparation** | I created a function to split all airlines in segmentsAirlineName and convert it to columns. Cabin codes have a hierarchy or order, i.e., coach is the least expensive cabin type and first is the most expensive. Hence, I have ranked them accordingly. As per my logic, [coach, coach] will be converted to (1+1)/2, which is like a single coach flight, but if the cabin code is [first, coach], it will calculate it as (4+1)/2, which will inform the system that the price difference might be due to a higher cabin code. Standard scaling was performed on numerical data and Target encoding was used on remaining categorical data using Pipelines. |
| **2.b. Feature Engineering** | Flight Year, month, week of the year, hour was extracted from flight datetime. |
| **2.c. Modelling** | For this model, we used various algorithms like linear regression, random forest, and gradient boosting. They were compared against each other and, also against a dummy baseline model that predicted the mean value of the fare. For linear regression, no specific hyperparameters were altered. In contrast, hyperparameters such as n\_estimators, max\_depth, and min\_samples\_leaf was adjusted for random forest and gradient boosting to enhance performance, address overfitting, and manage computational resources. |

|  |  |
| --- | --- |
| 1. **EXPERIMENT RESULTS** | |
| Analyse in detail the results achieved from this experiment from a technical and business perspective. Not only report performance metrics results but also any interpretation on model features, incorrect results, risks identified. | |
| **3.a. Technical Performance** | The following best scores were recorded for each of the algorithms used:   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Algorithm | Training MAE | Validation MAE | Training RMSE | Validation RMSE | | Base Model | 158.9194 | NA | 231.7206 | NA | | Linear Regression | 109.1267 | 113.6353 | 161.9104 | 157.8313 | | Random Forest | 95.9027 | 101.8221 | 137.0676 | 142.0360 | | Gradient Boosting | 76.7069 | 81.9992 | 112.9231 | 117.8925 |   Based on these scores, it was clear that Gradient Boosting algorithm with n\_estimators=50, max\_depth=8 and min\_samples\_leaf=3 performed best among all though there was slight overfitting that was noticed. |
| **3.b. Business Impact** | Accurate fare predictions positively impact the airline industry by optimizing revenue and enhancing customer satisfaction. Incorrect results, particularly in overpricing or underpricing scenarios, may have a high impact, leading to revenue loss, diminished competitiveness, and potential customer dissatisfaction. |
| **3.c. Encountered Issues** | Overfitting was one of the major issues that I faced with the Gradient Boosting and Random Forest model. However, I have adjusted some hyperparameters to achieve a good score with reduced overfitting. |

|  |  |
| --- | --- |
| 1. **FUTURE EXPERIMENT** | |
| Reflect on the experiment and highlight the key information/insights you gained from it that are valuable for the overall project objectives from a technical and business perspective. | |
| **4.a. Key Learning** | The experiment affirms the substantial benefits of accurate fare predictions, emphasizing the strategic role of precise pricing in revenue optimization and the importance of a customer-centric approach. Further experimentation is warranted to fine-tune models and explore additional variables, ensuring continued success and refinement of pricing strategies in the dynamic airline industry. |
| **4.b. Suggestions / Recommendations** | **Fine-Tune Prediction Models:** Prioritize model refinement for increased accuracy and immediate impact on revenue optimization.  **Implement Real-Time Updates:** Deploy dynamic pricing adjustments to maintain competitiveness and adapt to real-time market dynamics. |