# **EXPERIMENT REPORT**

|  |  |
| --- | --- |
| **Student Name** | Tahmidul Islam |
| **Project Name** | Binary Classification with Automated Hyperparameter Tuning |
| **Date** | 01-05-2023 |
| **Deliverables** | Attached File: Experiment 5 - Automated Hyperparameter Tuning.ipynb  Model: Random Forest |

|  |  |
| --- | --- |
| 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 goal is to do binary classification with Random Forest, but optimize the Hyperparameters using Algorithms like Grid Search and Random Search.  The accuracy scores will be evaluated, and depending on how the model performs on the training, validation and testing set, the effectiveness of the model will be judged.  The accuracy or correctness of results will influence marketing strategies and business decisions that want to determine if a person having a certain model of car driving a certain mile of distance, etc. will buy another car or not. |
| **1.b. Hypothesis** | The thing we find out in the experiment is which optimization algorithm gives us the best set of hyperparameters.  It is worthwhile considering it because so far we have implemented a lot of algorithms in our previous experiments and tuned many hyperparameters. Thus, performing Grid Search and Random search, and comparing between the two would provide us with valuable insights as to which works better overall, and how can they be implemented in future studies to make the most of our models. |
| **1.c. Experiment Objective** | The expected outcome is that the Tuning algorithms we are using will give an accuracy score that will help us to identify the success of the algorithm. It will help us to make a fair comparative analysis on which approach works better.  The score could be very high, indicating that the tuning algorithm is very successful in its optimization. The score could be lower indicating the opposite. |

|  |  |
| --- | --- |
| 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** | Before data preparation, a copy of the dataset was made and all the modifications were made on them so that the actual copy is preserved. Some columns were dropped because of having too many null values. Then we encoded the columns that were categorical. Once done, we found out that half the features of the dataset had 0 and 1 values while the other half had values that ranged upto 10. Thus we normalized the data to ground all the values.  Once all was done, we split the data into train, validation and test variables.  For future experiments, data scaling may be potentially important to ensure that values of only a handful of features get to express their significance in the model. |
| **2.b. Feature Engineering** | In the dataset, we have the ‘car\_model’ as a categorical value. Now the issue is that we can either consider it as an ordinal by assuming the model names refer to some form of chronology, and higher models mean newer brands. Or we can consider it as a nominal which means that the model names cannot be placed in any sequence. I have assumed it as nominal because it could be possible that model\_15 and model\_16 for example, have been introduced in the same year for that car brand. Thus in this case, we cannot put model\_15 and model\_16 in any sequential form.  Consequently, One-hot encoding was performed on the column which produced 19 more columns (or features) of boolean values.  We also had to remove some columns. We found out that there were a high number of missing values (85%) of the age\_band column and the gender (52%) column.  An initial thought was to apply mode imputation to the age\_band column, but since it would cover up 85% of the values, making such an assumption would be risky. Thus I decided to drop the column. Consequently, this will act as one of the limitations of the project that age being such a seemingly important factor was not considered, but this is for the sake of training our model correctly.  The same decision was made for the gender column since half the data was missing.  The features car\_segment (size) and ag\_of\_vehicle\_years might be important for future experiments as it has the highest chances of determining the decision of vehicle purchase. |
| **2.c. Modelling** | The model trained was Random Forest while the hyperparameter tunings were done with Grid Search and Random Search.  Random Forest was chosen over other models like SVMs, Decision Trees and Logistic Regressions because we can evaluate this algorithm in a robust and scalable manner. |

|  |  |
| --- | --- |
| 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 best set of hyperparameters from grid search algorithm:  Max\_depth = 25, Min\_Samples\_leaf =2, N\_Estimators = 90  Accuracy on Validation set: 0.991767  Accacy on Testing set: 0.9912822  The best set of hyperparameters from random search algorithm:  Max\_depth = 20, Min\_Samples\_leaf =4, N\_Estimators = 69  Accuracy on Validation set: 0.99091  Accacy on Testing set: 0.990026  From the results above we can see that grid search performs better than random search since it gives higher accuracy. The reason is that grid search is actually like a brute-force algorithm that tries out every combination until it find the best set of parameters. Random search, however, is more selective at random thus it has lesser chances of hitting the optimal spot. |
| **3.b. Business Impact** | The results tell us that grid search will help us find the peak conditions in the algorithm to predict whether a person will buy a car or not, with the highest accuracy. Here the impacts of incorrect results would not affect the business very much since the default accuracy score is very high. This means the wrong set of hyperparameters would not cost us much either. |
| **3.c. Encountered Issues** | One of the major issues during the experiment is the age\_band and gender column as mentioned before. The workaround was to drop these columns entirely, but this may not be a good solution as these two were very important parameters. I wanted to mode imputations to the rows, but then I would run the risk of making assumptions for 85% of the data. This issue may persist in future experiments. |

|  |  |
| --- | --- |
| 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 outcome satisfies our interest pretty well since we have obtained an optimal hyperparameter. Having tried out several algorithms before, and finding out the way to optimize the hyperparameters, we do not need further experimentation. |
| **4.b. Suggestions / Recommendations** | The next steps of the project may be to re-define the dataset by obtaining values for the age\_band and gender parameters. Model-wise, the experiment is doing well but we just need to include the mentioned features to be sure that we have a strong model. This model can be deployed into production. This can be done by Containerizing the application like in Docker and choosing an appropriate infrastructure for hosting the application like AWS or GCP. |