MACHINE LEARNING CE802

Investigation Report of Hotel Profit Prediction

Name: SHASHI LOKESH

Student id: sl21833

Registration number: 2100459

TABLE OF CONTENTS

1 INTRODUCTION

The main motivation behind the hotel business is to make a good profit and make ensured about the fact that the profit will be achievable. The application of predictive analytics will be a helpful aspect to the hotel management to design their future planning (Grama, et al., 2021). In this report, the profit production will be done and discussed with the application of machine learning and predictive analytics by designing two separate systems of prediction.

2 AIM OF INVESTIGATION

The aim of the investigation is to determine the profit for the hotel. To execute the scenario, the predictive analysis will be applied to the historical data of the hotel management regarding the business and the prediction will be done using machine learning. In this project, two systems will be designed and those are briefed below:

1. System-1: This system will predict whether the hotel will achieve the profit or not
2. System-2: This system will predict the annual profit by the hotel

3 PREDICTIVE ANALYTICS AND SYSTEM DESIGN

In this section, the discussion will be done based on the systems that have been designed to predict the profit of the hotel. In this context, two different historical data have been used based upon which the productions have been done.

3.1 HOTEL PROFIT PREDICTION

The first execution has been done to predict whether the hotel will attain the profit or not. To experiment, the 1st System has been designed. In this system. In this system, the training data has been provided to the designer regarding the historical records of the profit and losses of the hotel (Lv & Chi, 2017). The data has been obtained in two sets namely the training set and test set. To design the system, three classifiers have been chosen and the following steps have been taken into consideration:

Step-1: Read the historical data

Step-2: Set the selected classifiers in queue

Step-3: Partition the data into partition-1 and partition-2 where partition-1 will be applied to train each classifier and partition-2 will be applied to predict the profit-making.

Step-4: Train each classifier in the queue

Step-5: Predict the profit-making by each classifier

Step-6: Record the performance regarding the classification metrics such as Accuracy, Precision, Recall and F1-Scores (Ramesh, 2017).

Step-7: Compare the performances to get the most effective classifier

Step-8: Predict the profit-making using the most effective classifiers

Step-9: Store the result in the test data (provided by hotel management)

So, in this experiment, three classifiers have been chosen namely Logistic Regression, Decision Tree and MLP Classifier. To determine the profit-making of the hotels, all those classifiers have been applied ad step-1 to step-6 has been followed for all.

First, Logistic Regression has been trained and tested using the data partitions and the classification result is shown below:

Accuracy: 61.0%

Classification Report for Logistic Regression

	precision	recall	f1-score	support
False	0.59	0.67	0.63	49
True	0.64	0.55	0.59	51
accuracy			0.61	100
macro avg	0.61	0.61	0.61	100
weighted avg	0.61	0.61	0.61	100

Confusion Matrix for Logistic Regression

Predicted	False	True
True		
False	33	16
True	23	28

Fig-1: Classification Performance by Logistic Regression

So, as seen from the result, Logistic Regression is 61% correct to determine the profit-making. Next, the MLP classifier has been trained and tested using the data partitions and the classification result is shown below:

Accuracy: 53.0%

Classification Report for MLP Classifier

	precision	recall	f1-score	support
False	0.51	0.98	0.67	49
True	0.83	0.10	0.18	51
accuracy			0.53	100
macro avg	0.67	0.54	0.42	100
weighted avg	0.68	0.53	0.42	100

Confusion Matrix for MLP Classifier

Predicted	False	True
True		
False	48	1
True	46	5

Fig-2: Classification Performance by MLP Classifier

It can be seen that the MLP classifier is performing badly in detecting profit-making by 53% correctness. So, the nest classifiers that is the Decision Tree Classifier has been trained and tested using the data partitions and the classification result is shown below:

Accuracy: 86.0%

Classification Report for Decision Tree Classifier

	precision	recall	f1-score	support
False	0.91	0.80	0.85	49
True	0.82	0.92	0.87	51
accuracy			0.86	100
macro avg	0.87	0.86	0.86	100
weighted avg	0.86	0.86	0.86	100

Confusion Matrix for Decision Tree Classifier

Predicted	False	True
True		
False	39	10
True	4	47

Fig-3: Classification Performance by Decision Tree Classifier

Finally, using the Decision Tree Classifier, the prediction accuracy has been achieved as 86% which is far higher compared to the other two classifiers (Karcioğlu & Bulut, 2021). The result of the comparison is shown in the below table:

Table-1: Comparision of Classifier's Performances

Classifiers	Accuracy	Precision	Recall	F1-Score
Decision Tree Classifier	86	86.49	86	85.93
Logistic Regression	61	61.33	61	60.86
MLP Classifier	53	67.52	53	41.84

From the table-1, it can be observed that the Decision Tree Classifier is the most effective one through the historical data can be predicted for profit-making. So, finally, the prediction for profit-making by the hotel has been done using the Decision Tree Classifier. The overall comparison of performances is visualized below:

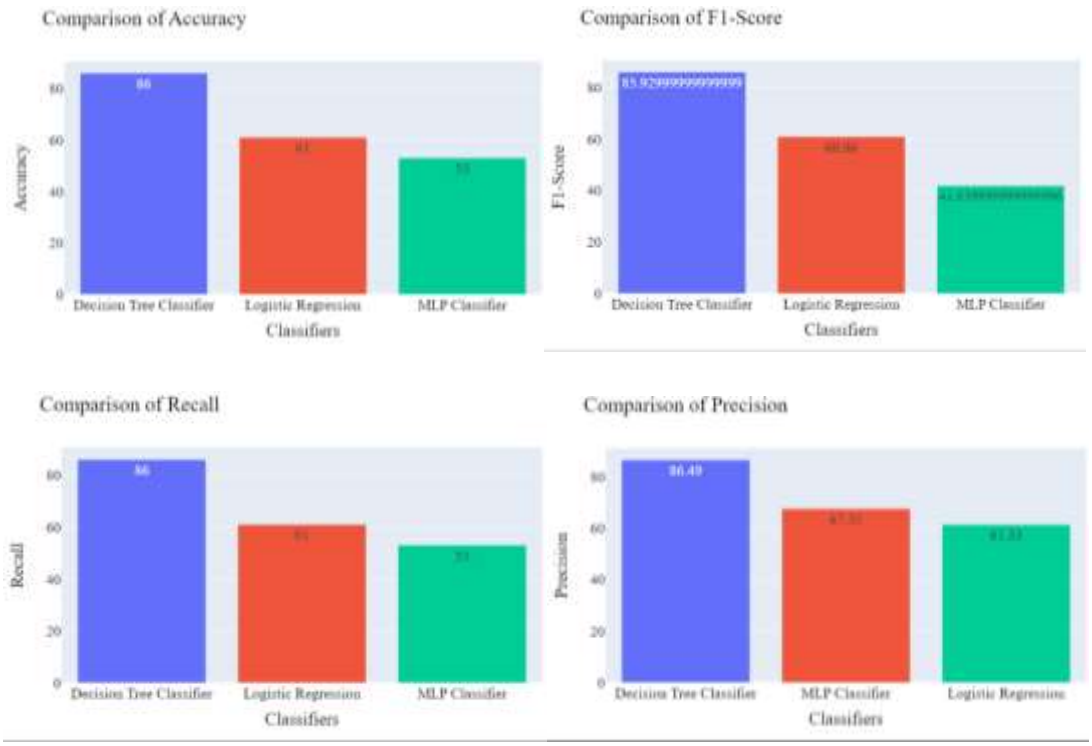

Fig-4: Comparison of Performances of Classifier

3.2 ANNUAL PROFIT PREDICTION

After the detection of whether the hotel will make a profit, the annual profit has been predicted suing System-2. In this experiment, regression algorithms have been used for predicting the annual profit. In this prediction, another historical data has been selected that contains several features and the annual profit (Srisa-An, 2021). Like System-1, the data has been obtained in two sets namely the training set and test set. To predict the annual profit, the following steps have been taken:

Step-1: Read the historical data

Step-2: Set the selected regressor in queue

Step-3: Partition the data into partition-1 and partition-2 where partition-1 will be applied to train each of the regression algorithms and partition-2 will be applied to predict the annual profit.

Step-4: Train each classifier in the queue

Step-5: Predict the annual profit by each regressor

Step-6: Record the performance regarding the regression metrics such as R2-Score, MSE etc.

Step-7: Compare the performances to get the most effective regressor

Step-8: Predict the annual profit using the most effective regressor

Step-9: Store the result in the test data (provided by hotel management)

In this system, three regression algorithms have been chosen namely Linear Regression, Random Forest and MLP Regressor (Kwong, et al., 2018). To determine the annual profit, all those algorithms have been used for the prediction and step-1 to step-6 has been followed for all.

Primarily, the prediction has been done using Linear Regression by training and testing it with the relevant partition of the data like System-1. The outcome of the prediction has been measured using R2-Score and Mean Squared Error (MSE) (Hirose, et al., 2017). The outcome is shown below:

```
------------------------------------------------
  Outcome of Prediction by: Linear Regression
  R2-Score: 58.0
  Mean Squared Error: 66283631.0
------------------------------------------------
```

Fig-5: Prediction Performance by Linear Regression

The outcome of the prediction for annual profit by MLP regression is shown below:

```
------------------------------------------------
  Outcome of Prediction by: Random Forest Regressor
  R2-Score: 100.0
  Mean Squared Error: 8962.0
------------------------------------------------
```

Fig-6: Prediction Performance by MLP Regression

The outcome of the prediction for annual profit by Random Forest regression is shown below:

```
------------------------------------------------
  Outcome of Prediction by: MLP Regressor
  R2-Score: 59.0
  Mean Squared Error: 64533382.0
------------------------------------------------
```

Fig-7: Prediction Performance by Random Forest Regression

So, with the application of the three chosen regression algorithms, the comparison has been done concerning the R2-Score and MSE and that is presented below:

Table-2: Comparision of Regressor's Performances

Regressor	R2 Score	Mean Squared Error
Random Forest Regressor	100	8962
Linear Regression	58	66283631
MLP Regressor	59	64533382

From the comparison table, it can be seen that Ransom Forest regression in predicting the annual profit with 100% R2-Score and lowest MSE value compared to others. So, random forest regression has been selected as the most effective predictor for annual profit prediction. Finally, using Random Forest Regression, the prediction has been applied to the test segment of the data and the predicted profit has been stored in the same data (Kwong, et al., 2018). The comparison of performances of the regression algorithms are visualized below:

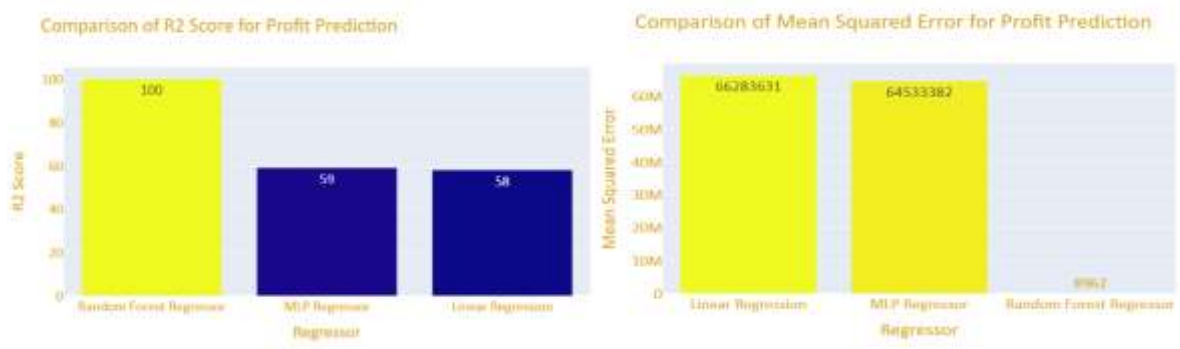

Fig-8: Comparison of Performances of Prediction

4 DISCUSSION AND CONCLUSION

In this assignment, the predictions of profit have been done using predictive analytics through machine learning algorithms. In this context, two systems have been designed for the purpose o profit-making prediction and annual profit prediction respectively. In the first system, it has been observed that the decision tree classifier has a classifier with higher accuracy (86%) and so, it can be considered to be the most effective one to predict profit-making. On the other hand, in the case of system-2, random forest regressor has performed best with the highest R2-Score and lowest MSE to predict annual profit. So, after the application of the algorithms have been done, the prediction results have been stored in two different datasets.

5 REFERENCES

Grama, L., Muscar, L. & Rusu, C., 2021. Sound Classification Algorithms for Indoor Human Activities. *16th International Conference on Engineering of Modern Electric Systems (EMES),* pp. 1-5.

Hirose, H., Soejima, Y. & Hirose, K., 2017. NNRMLR: A Combined Method of Nearest Neighbor Regression and Multiple Linear Regression. *IIAI International Conference on Advanced Applied Informatics,* pp. 1-7.

Karcioğlu, A. A. & Bulut, H., 2021. Performance Evaluation of Classification Algorithms Using Hyperparameter Optimization. *6th International Conference on Computer Science and Engineering (UBMK),* pp. 1-6.

Kwong, C. K., Chen, Y., Chan, K. Y. & Wong, H., 2018. The Hybrid Fuzzy Least-Squares Regression Approach to Modeling Manufacturing Processes. *IEEE Transactions on Fuzzy Systems,* pp. 1-4.

Lv, Y. & Chi, R., 2017. Data-driven adaptive iterative learning predictive control. *6th Data Driven Control and Learning Systems (DDCLS),* pp. 1-5.

Ramesh, R., 2017. Predictive analytics for banking user data using AWS Machine Learning cloud service. *2nd International Conference on Computing and Communications Technologies (ICCCT),* pp. 1-4.

Srisa-An, C., 2021. Guideline of Collinearity - Avoidable Regression Models on Time-series Analysis. *2nd International Conference on Big Data Analytics and Practices (IBDAP),* pp. 1-5.