Restaurant revenue prediction using deep learning.

Mokhutli Letsae

Introduction

Predictive analytics play a crucial role in the businesses strategies, allowing them to predict or forecast revenue and make informed decisions to mitigate potential losses and maximize profits. This project focuses on developing a predictive model to predict a restaurant's monthly revenue using a dataset from Kaggle. In building the models, a deep learning model that outperformed traditional ensemble machine learning model XGBoost hence the MLP neural network regressor was used.

1.Dataset Overview

The dataset from Kaggle (https://www.kaggle.com/datasets/mrsimple07/restaurants-revenue-prediction) with usability of 10 contains various features that could influence a restaurant's revenue. These include:

attribute	ML-model data role	data type	value count
Number_of_Customers	Predictor	int	1000
Menu_Price	Predictor	float	1000
Marketing_Spend	Predictor	float	1000
Cuisine_Type	Predictor	string	1000
Average_Customer_Spending	Predictor	float	1000
Promotions	Predictor	float	1000
Reviews	Predictor	float	1000
Monthly_Revenue	Target	float	1000

Data attributes definitions

Number_of_Customers: The count of customers visiting the restaurant.
Menu_Price: Average menu prices at the restaurant.
Marketing_Spend: Expenditure on marketing activities.
Cuisine_Type: The type of cuisine offered (Italian, Mexican, Japanese, American).
Average_Customer_Spending: Average spending per customer.
Promotions: Binary indicator (0 or 1) denoting whether promotions were conducted.
Reviews: Number of reviews received by the restaurant.
Monthly_Revenue: Simulated monthly revenue, the target variable for prediction.

2.Methodology

Crisp-DM framework was used as a methodology for this project. The framework acted as a guideline for optimal attempt of the project.
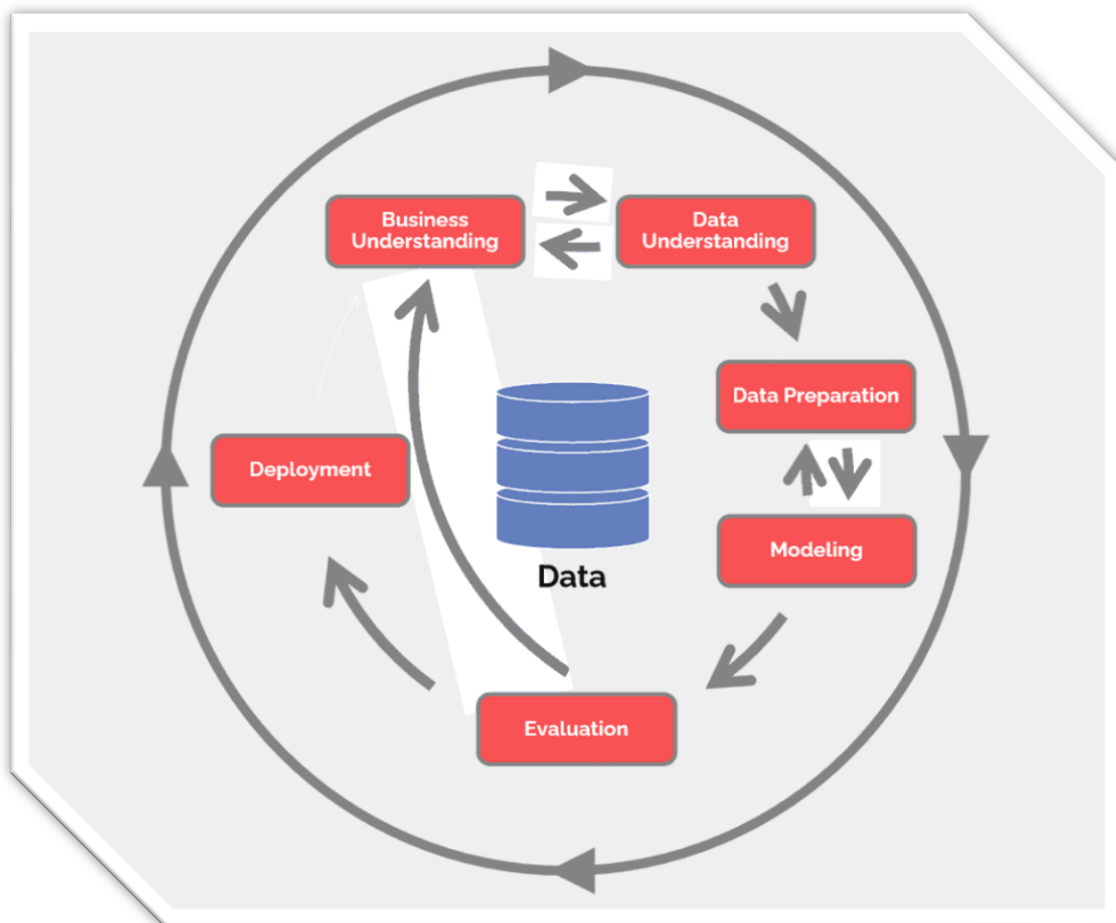

Figure 1

Figure 1 shows the stages of the framework, from the figure you can see that the framework follows a cyclic flow implying agility. The project was built in an agile manner.

a. **Business understanding:** The business problem in this project is predicting monthly revenue for a restaurant. Full view of the business problem can be accessed on the project article (https://medium.com/@mokhutliletsae/restaurant-revenue-prediction-using-deep-learning-b582cd3704f7)

b. **Data understanding (EDA):** While establishing facts extracted from the data, details extracted can be accessed on the project article (https://medium.com/@mokhutliletsae/restaurant-revenue-prediction-using-deep-learning-b582cd3704f7)

c. **Data Preparation:** For this stage, basic processes for data processing and preparation and cleaning: Handling missing values, correcting data types, and addressing outliers, Feature engineering and selection (using variance and correlation threshold), and normalizing values.

d. **Model Building:** Two models were developed for comparison: a Deep Learning model (multi-layer perceptron) and XGBoost regressor. The multi-layer perceptron regressor outperformed the XGBoost regressor and it was used for predictions.

e. **Evaluation:** Evaluation of the model's performance metrics.

The below performance metrics are for the MLP (multi-layer perceptron) neural network. They show that the model does not overfit as there's very little difference on all the training and test metrics.

The R-squared is 0.67 which defines how well the predictors well predict our target variable, the monthly revenue. The average mean absolute error is 44 which is acceptable but not optimal and our error percentage is 35%. All this metrics are acceptable to use and echo good performance of the model

	Mean Squared Error	Root Mean Squared Error	Mean Absolute Error	Mean Absolute Percentage Error	R-squared	Adjusted R-squared	Explained Variance Score	Median Absolute Error
Test								
	3,101.19	55.69	44.67	34.62	0.67	0.66	0.67	38.25
Train								
	3,343.92	57.83	46.17	26.93	0.69	0.69	0.69	38.99

f. **Deployment:** The model artifacts can be deployed on on-prem or cloud platforms, and it be utilized using a web pass, desktop app or just a environment it is used in.

3.Future Work

Augmenting data (features and row count) will be the next step to improve the model performance metrics.

4.Conclusion

The project was undertaken as a capstone project for completing a nanodegree from Udacity. It was undertaken by Mokhutli Letsae(https://www.linkedin.com/in/mokhutliletsae/), has the code repo on GitHub(https://github.com/letsaemokhutli/capstone_project.git). The MLP neural network model demonstrates good performance with an acceptable level of accuracy. The metrics suggest that the model does not overfit, as evidenced by the minimal difference between training and test metrics. While the error percentage indicates there is room for improvement, the model's predictions are reliable enough for practical use in decision-making processes for the restaurant.