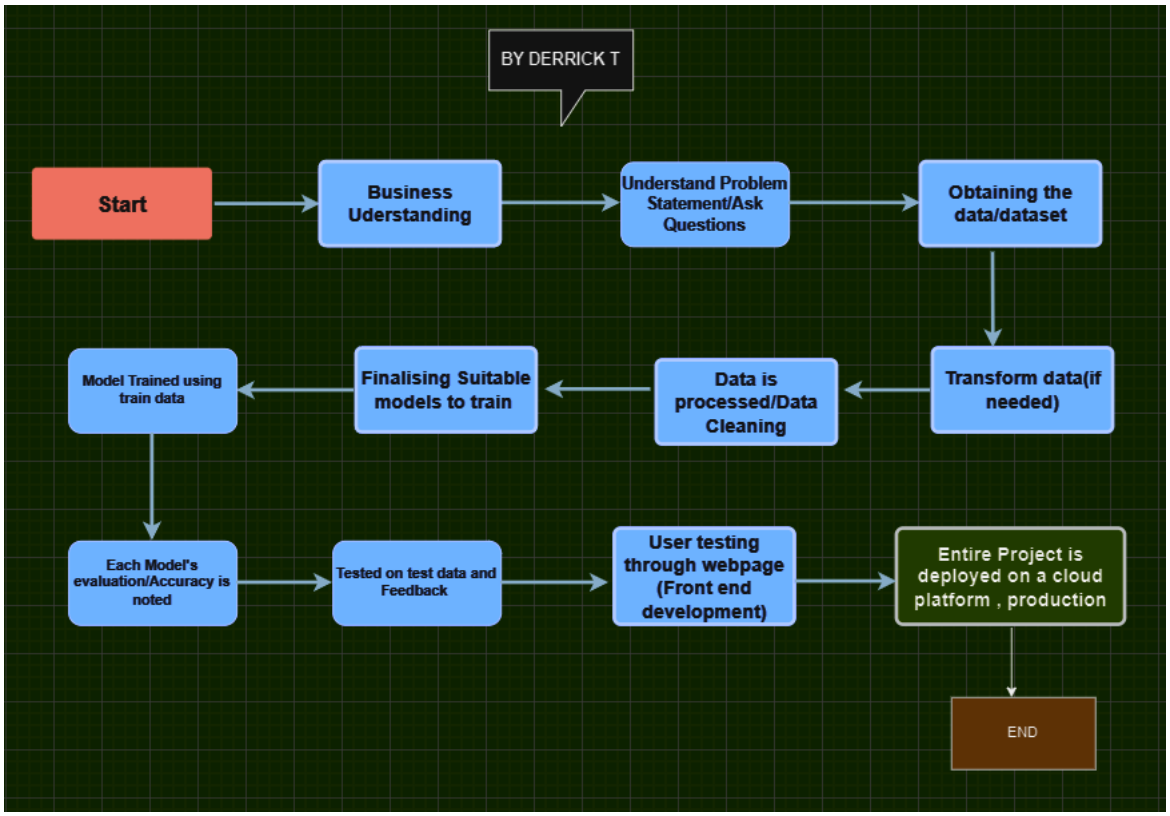
Stores Sales Prediction

Architecture Document

-Derrick T

Contents

1. Architecture



2. Architecture Description

Data Description

Stores Sales Dataset, which contains information related to sales transactions across multiple stores. The dataset encompasses a comprehensive record of sales data, including various attributes associated with each transaction, such as date, store location, products sold, and corresponding sales metrics.

Web Scrapping

Was Not needed as the dataset was available as a .csv file

Data Transformation

Transformation was not required as the data was in .csv format when obtained.

Data Pre-processing

Data preprocessing involves cleaning, transforming, and organizing the data for analysis. Steps include handling missing data, addressing outliers, normalizing or standardizing numerical variables, encoding categorical variables, selecting relevant features, and handling skewed data. Additionally, data integration may be necessary when working with multiple datasets. Balancing imbalanced classes and splitting the dataset into training, validation, and test sets are important. Data visualization aids in understanding the pre-processed data. Specific steps may vary based on the dataset, objectives, and algorithms used.

Model Selection/Model Building

As the task was to predict, regression models and random forest models were chosen and trained.

Random forest model yielded better accuracy

Model Trained and Tested

Here the random forest model was tested again and was ready for the test data.

Model Evaluation

Random Forest model trained and tested and ready for deployment and yield results for new data

Designed a webpage which accepts user data (front-end)

Used HTML and Flask API to create an interface for the model and the user data. (to predict)

Deployment

The project was deployed on a cloud platform for public sharing/production