Stores Sales Prediction

Objective:

Development of a predictive model for Item Outlet Sales . The model will predict the sales of the specified items in stores and Big Marts

Benefits:

- Estimates of Future Sales.
- Provides Production Planning.
- Helps in Purchasing the items.
- Helps in Sales Strategy .

Data Sharing Agreement :

- Sample file name (Train.csv)

- Number of Columns

- Column names

- Column data type

Data Schema

Feature name	Datatype	Null/Required
Item Weight	float	Required
Item Fat Content	object	Required
Item Visibility	float	Required
Item Type	object	Required
Item MRP	float	Required
Outlet Identifier	object	Required
Outlet Establishment Year	int	Required
Outlet Size	object	Required
Outlet Location Type	object	Required
Outlet Type	object	Required
Item Outlet Sales	float	Required

Architecture
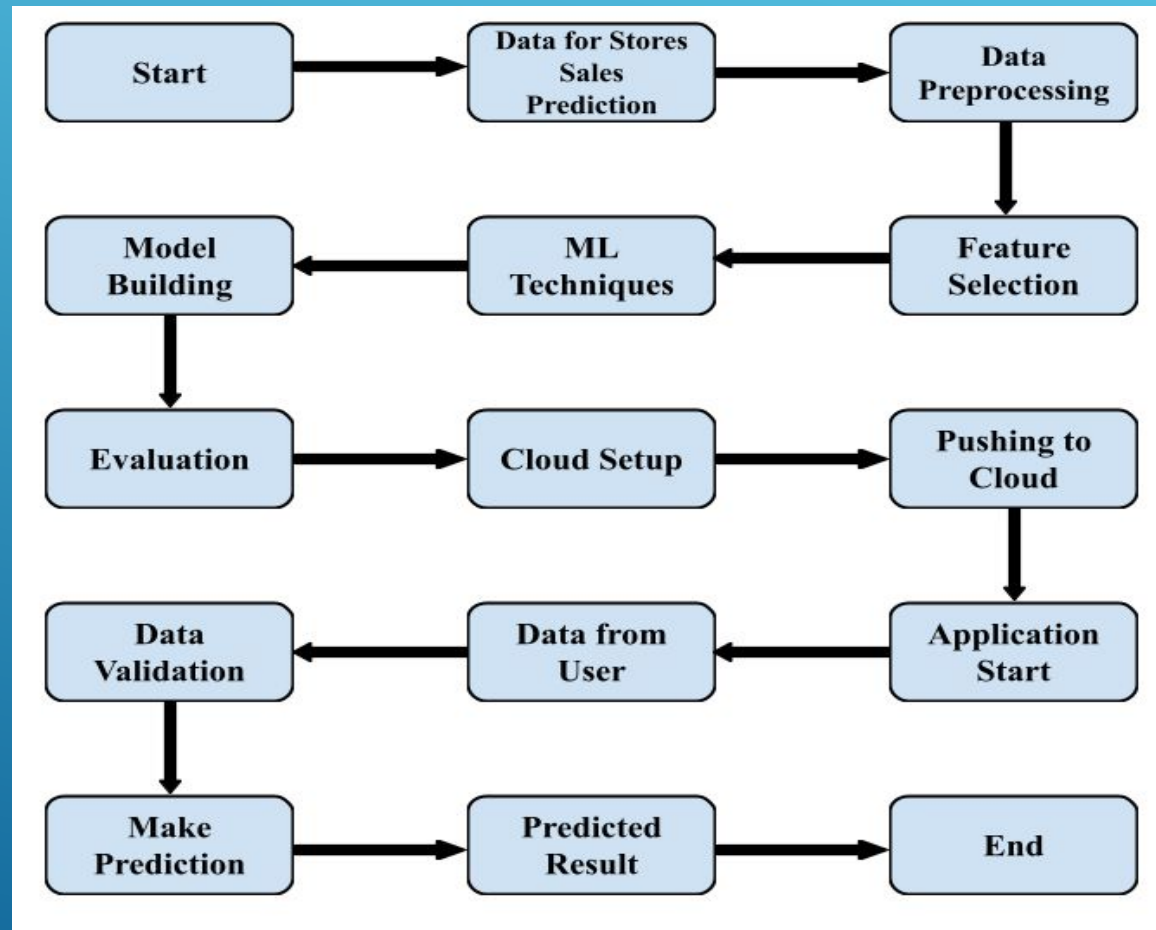
Architecture Description

Data Description

The dataset used in this project is collected from Kaggle. The dataset is divided into two sets of data. One is the test (5681) data and the other is the train (8523) data. The train dataset has both input and output variables.

Data Pre-processing

Data Pre-processing steps we could use are data cleaning, data integration, data reduction and data transformation.

Feature Selection

Feature Selection helps us to find the best set of features that allows us to build the necessary model for the project. This helps in selecting a subset of features from an initially large volume of features.

Machine Learning Techniques

Based on the problem statement and requirements we can use supervised or unsupervised technique which fits the project.

Model Building

Depending on the data type of the target variable we are either going to be building a classification or regression model. The main aspect of machine learning model building is to obtain actionable insights and in order to achieve that it is important to be able to select a subset of important features from the vast number.

Evaluation

The Evaluation of accuracy can be done using the test data. Mean Absolute error can be found using test data and prediction data.

Cloud Setup

Using Heroku as the cloud deployment platform, the platform is setup for deploying the virtual app.

Pushing to Cloud

Once the cloud is setup, the virtual app created will be pushed to the cloud and will finally be deployed into the cloud

Application Start

Once the virtual app is deployed in to the cloud we can open the web application using any web browser.

Data from user

Using a web browser we open the web application and provide the necessary information as the input for prediction.

Data Validation

Once the input is provided and we click on the submit button, the system will provide the output based on its requirements.

Result Prediction

Once the data validation is completed the prediction will be done for the type of product in Stores and Big Marts provided in the input.

Q & A

Q1) What's the source of data?

   The data for training is provided by the client in the form a link which takes us to the site containing the datasets.

Q 2) What was the type of data?

   The data was the combination of numerical and Categorical values.

Q 3) What's the complete flow you followed in this Project?

   Refer from slide 6 for better Understanding

Q 4) How logs are managed?

   We are using different logs as per the steps that we follow in validation and

   modeling like Data Validation log ,Info log, Error log , Data Insertion ,Model Training log , prediction log etc.

Q 5) What techniques were you using for data pre-processing?

- ▸ Removing unwanted attributes

- ▸ Visualizing relation of independent variables with each other and output variables

- ▸ Checking and changing Distribution of continuous values

- ▸ Cleaning data and imputing if null values are present.

- ▸ Converting categorical data into numeric values.

- ▸ Transforming the data based on the program requirements.

Q 6) How training was done or what models were used?

- ▸ Before diving the data in training and validation set we performed data pre-processing, exploratory data analysis and feature selection.

- ▸ Based on the client given dataset, the training and validation data were divided.

- ▸ The label encoder and one hot encoding was performed over training and validation data

- ▸ Algorithms like Linear Regression, Extra Tree Regression, Gradient Boosting Regression, Random Forest Regression, XGBoost Regression and K-neighbours Regression were used and we saved that model .

Q 7) How Prediction was done?

The test data was shared by the client .We performed pre processing, EDA and feature selection to the given test data, then with this test data, prediction was performed. In the end we get the accumulated data of predictions.

Q 8) What was the platform used for deploying the project ?

We used a cloud servicing platform named Heroku for deploying the project into the cloud.

Q 9) What are the advantages of the platform used for deploying the project?

▸ The advantages of Heroku are:

▸ It is free of cost.

▸ It is easy to use.

▸ Developer Centric.

▸ Easy to scale.

▸ Provides security.

▸ Powerful CLI

Q 10) What was the Framework used for doing the backend?

We used Flask Framework for completing the backend.

Q 11) How were the errors removed from the program?

There were no errors left by the end of the program execution because all of the errors were identified and debugged during the execution of the program itself.

Q 12) What is the future scope of the project?

► Use multiple algorithms.

► Optimize flask app.py and Stores Sales Prediction.ipynb

► The front end can be developed even more.