**Sales Store Prediction**

**Architecture Design**

**Document Version - 1.0**

**Last Revision date - 30/01/2023**

**Written By – Arpit Kumar Maurya**

**Contents**

**1. Introduction -------------------------------------------------------------- 3**

**1.1 What is Architecture -Level design document? ------------ 3**

**1.2 Scope --------------------------------------------------------------------- 3**

**2. Architecture ----------------------------------------------------------- -- 4**

**3. Architecture Description -------------------------------------------- 5**

**3.1 Data Gathering -------------------------------------------------------- 5**

**3.2 Data Description ------------------------------------------------------ 5**

**3.3 Tool Used --------------------------------------------------------------- 6**

**3.4 Data Pre-processing ------------------------------------------------- 6**

**3.5 Model Building -------------------------------------------------------- 6**

**3.6 Hyperparameter tuning --------------------------------------------- 7**

**3.7 Deployment -------------------------------------------------------- --- 7**

1. **Introduction** 
   1. **What is architecture design?**

Any software needs the architectural design to represent the design of the software. IEEE defines architectural design as “the process of defining a collection of hardware and software components and their interfaces to establish the framework for the development of a computer system.” The software that is built for computer-based systems can exhibit one of these many architectures.

Each style will describe a system category that consists of:

1. A set of components (eg: a database, computational modules) that will perform a function required by the system.
2. The set of connectors will help in coordination, communication, and cooperation between the components.
3. Conditions that how components can be integrated to form the system.

Semantic models help the designer to understand the overall properties of the system

* 1. **Scope**

Architecture Design Document (ADD) is an architectural design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the design principles may be defined during requirement analysis and then refined during architectural design work.

1. **Architecture**

**Modeling**

Linear Regression

Polynomial Regression

Decision Tree Regression

Random Forest Regression

AdaBoost Regression

Gradient Boost Regression

XGBoost Regression

Predict sales

Preprocessing

Best Model

Training Set

Testing Set

HyperParameter Tuning

Fit and Transform

Best score

Fit

Performance

1. **Architecture Description**

This project is to make associate interface for the stores to predict their approximate sales.

Worth, additionally to the present, in would like of obtaining the important time project expertise.

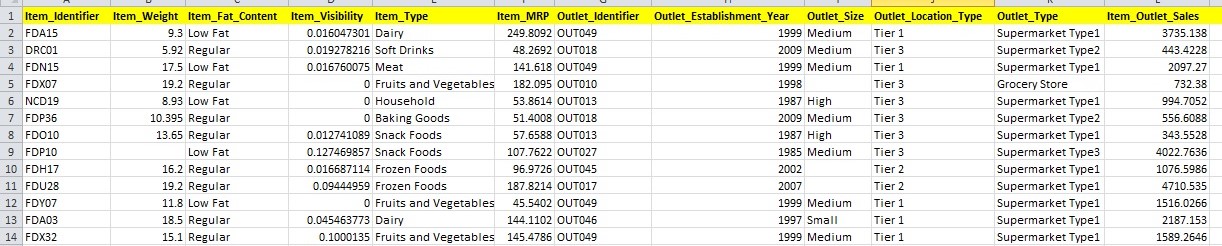
I have a tendency to square measure mercantilism the gathered information into our own information then begin the project from the scratch

**3.1 Data Gathering**

The data for the current project is being gathered from Kaggle dataset, the link to the data is:

<https://www.kaggle.com/brijbhushannanda1979/bigmart-sales-data>

**3.2 Data Description**

We have train (8523) and test (5681) data set, train data set has both input and output variable. In dataset thee are columns such that Item\_Weight ,Item\_Fat\_Content, Item\_Visibility, Item\_Type, Item\_MRP, Outlet\_Identifier, Outlet\_Establishment\_Year ,Outlet\_Size Outlet\_Location\_Type, Outlet\_Type , Item\_Outlet\_Sales.

**3.3 Tool Used**

* Python 3.9 is used as the programming language and Python frameworks like numpy, pandas, streamlit, skitlearn and alternative modules for building the model.
* Jupiter Notebook and Spider is used as IDE.
* For visualizations seaborn and components of matplotlib are getting used.

**3.4 Data Preprocessing**

There are some steps performed in preprocessing –

* Data Cleaning- In this step , we clean the data which have null values and duplicate values.
* Data Transformation- In this step , we transform data which are in skewed distribution to normal distribution.
* Encoding - In this step, we transform text data into numerical value.
* Train-Test split – In this step, we split data into training and testing part.

**3.5 Model Building**

In this phase, various regression modelling techniques are selected and applied and their parameters are calibrated to optimal values. There we evaluate which model is best for our dataset in both testing and training purpose. This is important phase of data science life cycle. Here we check training and testing data on various regression models. And examine the score of different models. Which model have more testing accuracy without overfitting that is most suitable model.

**3.6 Hyperparameter Tuning**

After getting the more suitable model from previous step we do hyperparameter tuning. Where we check our model with different combination of possible parameters. Here we find the best model with hyper parameter. Then build a model with this parameter.

**3.7 Deployment**

This is the final step of model building. In this step we save our best model as a joblib file. Further this file used in building a user interface as like web app.