

High Level Document

Store Sales Prediction

Document Version Control

Date Issued	Version	Description	Author
1-10-2021	1.0	Initial HLD- V1.0	Nikhil Patil

Abstract

Machine learning is a category of algorithms that allows software applications to become more accurate in predicting outcomes without being explicitly programmed. The revolutionary applications of Machine Learning can be seen in every field including education, healthcare, engineering, sales, entertainment, transport and several more.

Sales forecasting is one of the major aspects of the retail industry. The retailers face a lot of new challenges with the increasing competition. The mentioned system design is to find the most frequent combinations of the items. So, to analyze the data and predict sales of certain products and their overall impact on the sales of certain stores we use several Machine Learning algorithms. A comprehensive sales prediction is done using Machine Learning models such as Linear Regression, Support Vector Machine, Decision Tree, Random Forest Regressor, AdaBoost Regressor, Gradient Boosting Regressor, XGBoost Regressor.

Contents

Document Version Control	1
Abstract.....	2
1.0 Introduction	4
1.1 Why this High-Level Document?.....	4
1.2 Scope.....	4
1.3 Definitions.....	4
2.0 General Description	6
2.1 Problem Statement.....	6
2.2 Proposed Solution.....	6
2.3 Further Improvements.....	6
2.4 Technical Requirements.....	6
2.5 Data Requirements	7
2.6 Tools used	7
3.0 Design Details.....	9
3.1 Process Flow.....	9
3.2 Event Log.....	9
4.0 Performance	10
4.1 Reusability.....	10
4.2 Application Compatibility.....	10
4.3 Deployment.....	10
5.0 Dashboards	11
6.0 Conclusion.....	12

1.0 Introduction

1.1 Why this High-Level Document?

The purpose of this high-Level document is to provide a complete description or architecture of the application. The HLD involves system architecture, database design, a description of systems and relationship among modules.

The HLD also includes detailed information about:

- All design aspects and define them in detail.
- The user interface.
- The hardware and software interfaces.
- Describe the performance requirements.
- Portability and compatibility.

1.2 Scope

The HLD document presents the structure of the system, such as the database architecture, application architecture, and technology architecture. The HLD uses non-technical to middle-technical terms which should be understandable to the stakeholders.

1.3 Definitions

Term	Description
Database	Collection of all the information
IDE	Integrated Development Environment
ML	Machine Learning
API	Application Programming Interface
VS Code	Visual Studio Code

XGB	Extreme Gradient Boosting
SVM	Support Vector Machine
EDA	Exploratory Data Analysis
KPI	Key Performance Indicator
AWS	Amazon Web Services

2.0 General Description

2.1 Problem Statement

Nowadays, shopping malls and Big Marts keep track of individual item sales data in order to forecast client demand and adjust inventory management. In the data warehouse, these data stores hold a significant amount of consumer information and particular item details. By mining the data store from the warehouse, more anomalies and common patterns can be discovered.

2.2 Proposed Solution

The proposed solution provides an efficient and simple to use application, that can be used without any prior knowledge.

2.3 Further Improvements

As a future scope, features like data visualizations after model training and setting up an auto-mail system to send email to respective domain experts for further analysis.

2.4 Technical Requirements

The solution can be a cloud-based or application hosted on an internal server or even be hosted on a local machine. For accessing this application below are the minimum requirements:

- Good internet connection.
- Web Browser.

For training model, the system requirements are as follows:

- +4 GB RAM preferred

- Operation System: Windows, Linux, Mac
- Visual Studio Code / Jupyter notebook

2.5 Data Requirements

Data requirements completely depends on out problem statement.

- Comma separated values (CSV) file.
- Input file feature/field names and its sequence should be followed as per decided.

2.6 Tools used

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Plotly, Flask are used to build the whole model.

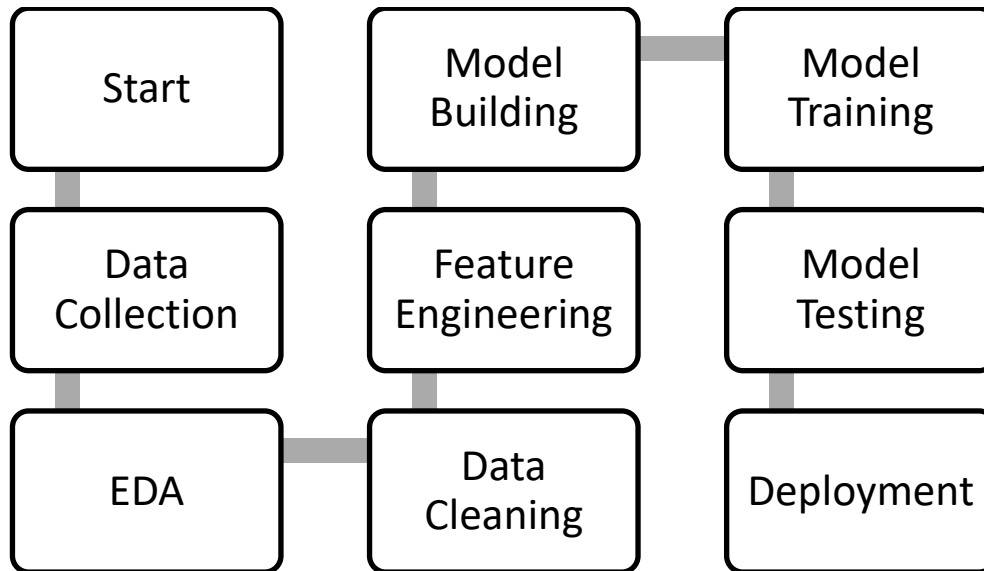


- Pandas is an open-source Python package that is widely used for data analysis and machine learning tasks.
- NumPy is most commonly used package for scientific computing in Python.

- Plotly is an open-source data visualization library used to create interactive and quality charts/graphs.
- Scikit-learn is used for a machine learning.
- Flask is used to build API.
- VS Code is used as IDE (Integrated Development Environment)
- GitHub is used as version control system.
- Front end development is done using HTML/CSS.
- Heroku is used for deployment of the model.

3.0 Design Details

3.1 Process Flow



3.2 Event Log

The system should log every event so that the user will know what process is running internally.

Initial Step-By-Step Description:

- The system identifies at what step logging required.
- The system should be able to log each and every system flow.
- Developer can choose logging method. You can choose database logging.
- System should not hang out even after using so many loggings.

4.0 Performance

Big Mart sales prediction is used for detection of outlet sales of the items. It will help Big Mart to analyze which type of the items sales were increasing or decreasing and, on that basis, Big Mart will inform concern authorities and take necessary action, so it should be increase sales.

4.1 Reusability

The entire solution will be done in modular fashion and will be API oriented. So, in the case of the scaling the application, the components are completely reusable.

4.2 Application Compatibility

The interaction with the application is done through the designed user interface, which the end user can access through any web browser.

4.3 Deployment



A dashboard is a data visualization and analysis tool that displays on one screen the status of key performance indicators (KPIs) and other important business metrics.

5.0 Dashboards

A dashboard is a data visualization and analysis tool that displays on one screen the status of key performance indicators (KPIs) and other important business metrics.



As a high-level reporting mechanism, dashboards provide fast 'big-picture' answer to critical business questions and assist and benefit decision making in several ways:

- Communicating how the sales vary with outlet location type.
- Display how outlet type affecting sales.
- Detects which type of items sales maximum.
- To refine their methodologies and strategies to increase the sales.

6.0 Conclusion

This system shows us that the different techniques that are used to forecast sales of the store. On implementation, the prediction results show the correlation among different attributes considered and how a particular location of medium size recorded the highest sales, suggesting that other outlet locations should follow similar patterns for improved sales. Multiple instance parameters and various factors can be used to make this sales prediction more accurate and successful. Accuracy, which plays a key role in prediction-based system. From the results I could see that Random Forest Regressor turned out to be best working model for this problem in terms of the accuracy. These predictions help user to refine their methodologies and strategies which in turn helps them to increase their profit.