Architecture

Stores Sales Prediction

Revision Number – 1.0

Last Date of Revision – 04-09-2024

Swapnil Shinde

Document Version Control

Date	Version	Description	Author
01-09-2024	1.0	Abstract	Swapnil Shinde
02-09-2024	1.1	Architecture	Swapnil Shinde
03-09-2024	1.2	Architecture Design	Swapnil Shinde

Contents

Abstract

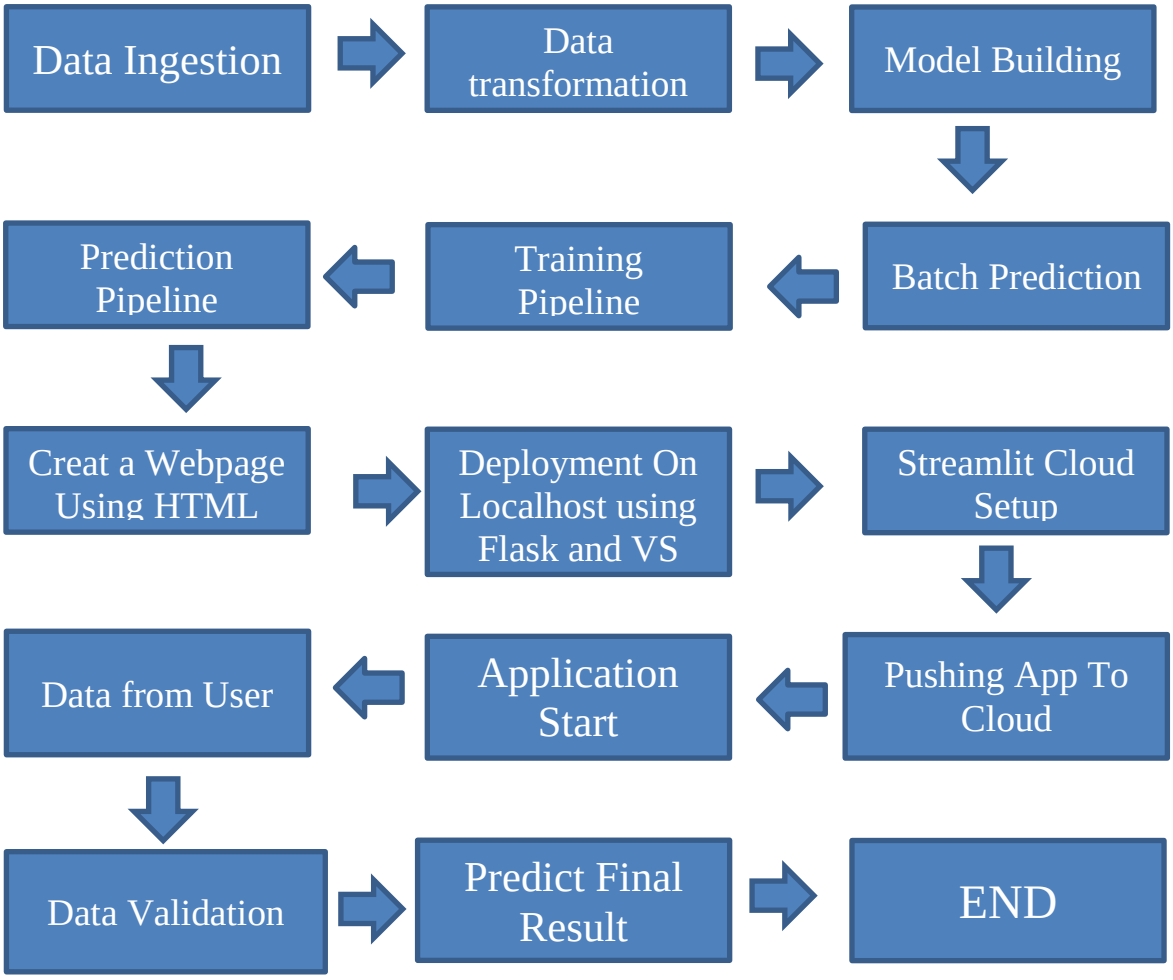
In the contemporary retail ecosystem, shopping malls and Big Marts meticulously collect and archive individual item sales data, yielding invaluable insights into consumer behavior and product specifics. These data repositories, securely stored within a data warehouse, serve as reservoirs of opportunity. Our system embarks on a transformative journey through the realms of data, expertly harnessing its power. The journey unfolds through meticulously orchestrated steps: data ingestion from Kaggle datasets, data transformation for cleanliness and relevance, model building to extract meaningful patterns, and the establishment of an efficient batch prediction pipeline. We don't stop there; we extend this journey to the end-users with a well-crafted, user-friendly interface, bridging the gap between data and actionable insights. This document, a testament to our technical prowess, delves deep into the modular architecture, interfaces, algorithms, and visualizations that underpin this transformative solution, setting the stage for a future where anomalies and common patterns emerge as strategic assets in the world of retail..

Introduction

Why this Architecture Design documentation?

The main objective of the Architecture design documentation is to provide the internal logic understanding of the flight fare prediction code. The Architecture design documentation is designed insuch a way that the programmer can directly code after reading each module description in the documentation.

1 Architecture



2 Architecture design

The architecture for the Store Price Prediction project comprises data collection and preprocessing from various sources, storage in a database system, data analysis and feature selection using machine learning frameworks, model development with algorithm selection and training, real-time data integration for up-to-date predictions, and model deployment through containerization and microservices. A user-friendly web interface will enable users to input product details and receive accurate price predictions, facilitating informed shopping decisions. This comprehensive architecture ensures the project's ability to provide real-time and reliable price predictions for store products, enhancing the overall shopping experience.

3 Data gathering from main source

The data for the current project is being gathered from Kaggle dataset, the linkto the data is: BigMart Sales Data | Kaggle

3.1 Data description

We have train (8523) and test (5681) data set, train data set has both input and output

Columns Are :
variable(s). We need to predict the sales for test data set.
Item_Identifier: Unique product ID
Item_Weight: Weight of product
Item_Fat_Content: Whether the product is low fat or not
Item_Visibility: The % of total display area of all products in a store allocated to the particular product
Item_Type: The category to which the product belongs
Item_MRP: Maximum Retail Price (list price) of the product
Outlet_Identifier: Unique store ID
Outlet_Establishment_Year: The year in which store was established
Outlet_Size: The size of the store in terms of ground area covered
Outlet_Location_Type: The type of city in which the store is located
Outlet_Type: Whether the outlet is just a grocery store or some sort of supermarket
Item_Outlet_Sales: Sales of the product in the particulat store. This is the outcome variable to be predicted.

3.2 Data Ingestion

The cornerstone of our data-driven project was established through a systematic process of data acquisition and ingestion. Utilizing Kaggle, a reputable platform renowned for its high-quality datasets, we identified and acquired the crucial data required for our price prediction project. This dataset, integral to our goal of accurate price forecasting, was meticulously downloaded and securely stored within our local system infrastructure. Subsequently, we initiated the data ingestion phase, where the dataset seamlessly integrated into our project's data pipeline. This meticulous approach ensures that our project is built upon a solid foundation, setting the stage for robust and precise price prediction models and analysis.

3.3 Data pre-processing

Steps performed in pre-processing are:

- First read data from Artifact folder
- Checking unnecessary columns
- One column has product id which is unique for every product so I deleted that column.
- Checked for null values
- there are too many null values are present in two columns that's why I deleted them
- Performed one-hot encoder on categorical columns.
- Perform Ordinal Encoder on Ordinal Columns.
- Scaling is performed for needed information.
- And, the info is prepared for passing to the machine learning formula

3.4 Modelling

The pre-processed information is then envisioned and every one the specified insights are being drawn. though from the drawn insights, the info is at randomunfold however still modelling is performed with completely different machinelearning algorithms to form positive we tend to cowl all the chances. and eventually, Gradient Boosting performed well

3.5 UI integration

Both CSS and HTML files are being created and are being integrated with the created machine learning model. All the required files are then integrated to theapp.py file and tested locally

3.6 Data from user

The data from the user is retrieved from the created HTML web page and Streamlit webpage.

.

3.7 Data validation

The data provided by the user is then being processed by app.py and application.py file and validated. The validated data is then sent for the prediction.

.

3.8 Rendering the results

The data sent for the prediction is then rendered to the web page.

3.9 Deployment

The tested model is then deployed Streamlit Cloud and local machine. So, users can access the project from any internet devices.