MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

Project Report

On

Big Mart Sales Prediction Using Flask

Submitted By:

Himanshu Lodha

(0901AD211020)

Vikas Chouhan

(0901AD211064)

Faculty Mentor:

Prof. Deepti Gupta

CENTRE FOR ARTIFICIAL INTELLIGENCE

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE

GWALIOR - 474005 (MP) est. 1957

JULY-DEC. 2023

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

CERTIFICATE

This is certified that **Himanshu Lodha (0901AD211020)** and **Vikas Chouhan (0901AD211064)** has submitted the project report titled "**Big Mart Sales Prediction using Flask**" under the mentorship of **Prof. Deepti Gupta**, in partial fulfilment of the requirement for the award of degree of Bachelor of Technology in **Artificial Intelligence and Data Science** from **Madhav Institute of Technology and Science**, Gwalior.

23/11/23

Prof. Deepti Gupta
Faculty Mentor
Assistant Professor
Centre for Artificial Intelligence

23/11/23

Dr. R. R. Singh
Coordinator
Centre for Artificial Intelligence

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

DECLARATION

I hereby declare that the work being presented in this project report, for the partial fulfilment of requirement for the award of the degree of Bachelor of Technology in **Artificial intelligence and Data Science** at **Madhav Institute of Technology & Science**, Gwalior is an authenticated and original record of my work under the mentorship of **Deepti Gupta**, Centre for artificial Intelligence.

I declare that I have not submitted the matter embodied in this report for the award of any degree or diploma anywhere else.

Himanshu

Himanshu Lodha
(0901AD211020)
Vikas
Vikas Chouhan
(0901AD211064)
3rd year
Centre for Artificial Intelligence

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

CERTIFICATE

This is certified that **Himanshu Lodha** (**0901AD211020**) and **Vikas Chouhan (0901AD211064)** has submitted the project report titled "**Big Mart Sales Prediction using Flask"** under the mentorship of **Prof. Deepti Gupta**, in partial fulfilment of the requirement for the award of degree of Bachelor of Technology in **Artificial Intelligence and Data Science** from **Madhav Institute of Technology and Science,** Gwalior**.**

Prof. Deepti Gupta
Faculty Mentor
Assistant Professor
Centre for Artificial Intelligence

Dr. R. R. Singh
Coordinator
Centre for Artificial Intelligence

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

DECLARATION

I hereby declare that the work being presented in this project report, for the partial fulfilment of requirement for the award of the degree of Bachelor of Technology in **Artificial intelligence and Data Science** at **Madhav Institute of Technology & Science**, Gwalior is an authenticated and original record of my work under the mentorship of **Deepti Gupta**, Centre for artificial Intelligence.

I declare that I have not submitted the matter embodied in this report for the award of any degree or diploma anywhere else.

Himanshu Lodha
(0901AD211020)
Vikas Chouhan
(0901AD211064)
3rd year
Centre for Artificial Intelligence

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

ACKNOWLEDGEMENT

The full semester project has proved to be pivotal to my career. I am thankful to my institute, **Madhav Institute of Technology and Science** to allow me to continue my disciplinary/interdisciplinary project as a curriculum requirement, under the provisions of the Flexible Curriculum Scheme (based on the AICTE Model Curriculum 2018), approved by the Academic Council of the institute. I extend my gratitude to the Director of the institute, **Dr. R. K. Pandit** and Dean Academics, **Dr. Manjaree Pandit** for this.

I would sincerely like to thank my department, **Centre for Artificial Intelligence,** for allowing me to explore this project. I humbly thank**, Dr. R. R. Singh** Coordinator, Centre for Artificial Intelligence, for his continued support during the course of this engagement, which eased the process and formalities involved.

I am sincerely thankful to my faculty mentors. I am grateful to the guidance of **Prof. Deepti Gupta**, Centre for Artificial Intelligence, for his continued support and guidance throughout the project. I am also very thankful to the faculty and staff of the department.

Himanshu Lodha
(0901AD211020)
Vikas Chouhan
(0901AD211064)
3rd year
Centre for Artificial Intelligence

ABSTRACT

In the realm of retail analytics, this research introduces an innovative paradigm for sales prediction at Big Mart through the fusion of advanced machine learning and web frameworks. Leveraging the evolving landscape of deep learning techniques, the study focuses on the integration of a sophisticated prediction model with Flask, a lightweight web framework.

The methodology involves developing a robust sales prediction model tailored for Big Mart, drawing insights from established machine learning principles. The predictive capabilities are enhanced by employing Flask, facilitating seamless deployment and user interaction. This integration allows stakeholders to access and utilize the predictive model through an intuitive web interface.

The study emphasizes the significance of accurately forecasting sales in the retail sector, enabling businesses to optimize inventory management, enhance customer satisfaction, and make informed strategic decisions. By harnessing the power of Flask, the model becomes accessible to non-technical users, fostering wider adoption and applicability within the retail industry.

This research not only contributes to the advancement of predictive analytics in retail but also underscores the practical implementation of machine learning models in real-world scenarios. The integration of Flask ensures that the predictive insights generated by the model are readily available and user-friendly, marking a pivotal step towards enhancing the efficiency of retail operations at Big Mart.

Keyword: Sales Prediction, Retail Analytics, Big Mart, Flask Integration, Machine Learning, Predictive Analytics, Inventory Management, Web Framework, User-Friendly Interface.

सार

खुदाई विपणी एनालिटिक्स क्षेत्र में, यह अनुसंधान एक नई परिदिग्म का परिचय करता है जिसमें बड़े मार्ट के लिए बिक्री की पूर्वानुमान करने के लिए उन्नत मशीन लर्निंग और वेब फ़्रेमवर्क के संघ की बात है। गहरी शिक्षा की तकनीकों की बढ़ती हुई सीनेन में, इस अध्ययन में एक परिष्कृत पूर्वानुमान मॉडल को फ्लास्क, एक हल्के वेब फ़्रेमवर्क, के साथ एकत्र करने पर ध्यान केंद्रित है।

उपयोगकर्ता संवाद और सीमित डिप्लॉयमेंट को सुगम बनाने के लिए फ्लास्क के साथ महत्वपूर्ण पूर्वानुमान क्षमताओं का विकास करने की एक उपचार क्रिया होती है। इस एकीकरण से हितधारकों को सूचना प्राप्त करने और पूर्वानुमान मॉडल का उपयोग करने के लिए एक सहज वेब इंटरफेस के माध्यम से इस्तेमाल करने की अनुमति होती है।

अध्ययन में बिक्री को सही ढंग से पूर्वानुमान करने के महत्व को जोर दिया गया है, जिससे व्यापार को इन्वेंटरी प्रबंधन को अनुकूलित, ग्राहक संतुष्टि को बढ़ावा देने और सूचित सूचना के आधार पर रणनीतिक निर्धारित करने की स्थिति मिलती है।

फ्लास्क की शक्ति का सही तरीके से उपयोग करके, मॉडल स्वयं तकनीकी उपयोगकर्ताओं के लिए सुलभ हो जाता है, जिससे खुदाई ऑपरेशन्स की कुशलता में सुधार होता है।

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1 Project Overview

1.1 Description

In the realm of retail analytics, the pursuit of accurate sales forecasting at Big Mart takes a transformative turn with the integration of Flask, a robust web framework. This detailed description navigates through the methodologies, significance, and implications of employing Flask for sales prediction, ushering in a new era of efficiency and precision in the retail sector.
The retail landscape is undergoing a transformative shift with the integration of advanced predictive analytics. This initiative focuses on predicting sales at Big Mart, a retail powerhouse, using a sophisticated model integrated with Flask, a lightweight yet powerful web framework.

Big Mart's foray into sales prediction harnesses the synergy of advanced machine learning and the user-friendly capabilities of Flask, a dynamic web framework. At its core, this predictive analytics initiative begins with the meticulous preparation of historical sales data, employing sophisticated machine learning models for discerning patterns and trends. The integration with Flask, a lightweight yet robust web framework, forms the backbone of this endeavor. Flask facilitates the seamless deployment of the predictive model, creating an intuitive and accessible web interface for stakeholders.

Through this interface, users, including non-technical personnel, can effortlessly input parameters and receive real-time predictions, revolutionizing the decision-making process. The significance lies not only in accurate sales forecasting, optimizing inventory management, and enhancing strategic planning but also in democratizing access to these predictive insights. Flask's versatility ensures adaptability to changing business dynamics, marking a transformative step for Big Mart in embracing cutting-edge technologies for data-driven retail intelligence.

1.2 Objectives of the Project

The objectives of the Big Mart Sales Prediction using Flask project can encompass a range of goals, combining both technical and business-oriented aspects. Here are some potential objectives for such a project:

1.2.1 Accurate Sales Forecasting: Develop a robust machine learning model capable of accurately predicting future sales based on historical data and relevant features.

1.2.2 Integration with Flask: Integrate the predictive model with Flask to create a user-friendly web interface, allowing stakeholders to interact with and benefit from the sales predictions.

1.2.3 User Accessibility: Ensure that the Flask interface is intuitive and accessible to a wide range of users, including non-technical stakeholders, thereby democratizing access to predictive insights.

1.2.4 Optimized Inventory Management: Leverage accurate sales predictions to optimize inventory management, reducing overstock or understock situations and improving overall supply chain efficiency.

1.2.5 Enhanced Decision-Making: Empower decision-makers at Big Mart with timely and reliable sales forecasts, enabling them to make informed decisions related to marketing strategies, promotions, and resource allocation.

1.2.6 Scalability and Performance: Design the system to be scalable, accommodating an increasing volume of data and user interactions, while ensuring optimal performance for real-time predictions.

1.2.7 Adaptability to Changing Business Dynamics: Create a flexible system that can adapt to changing market conditions, product trends, and other dynamic factors influencing sales patterns.

1.2.8 User Feedback and Iterative Improvement: Implement mechanisms for collecting user feedback on the Flask interface and predictive model performance, facilitating iterative improvements to enhance user satisfaction and prediction accuracy.

1.2.9 Documentation and Knowledge Transfer: Document the project comprehensively to facilitate knowledge transfer and future maintenance. This includes documenting the machine learning model, Flask implementation, and overall system architecture.

1.2.10 Data Security and Privacy: Implement robust security measures to protect sensitive sales data and ensure compliance with data privacy regulations.

1.2.11 Training and Support: Provide training sessions and support materials for users interacting with the Flask interface, ensuring they can maximize the utility of the sales prediction system

Chapter 2: Literature Review

2.1 E-commerce Trends

In the context of Big Mart sales prediction using Flask within the realm of e-commerce, several trends and considerations emerge, shaping the landscape of predictive analytics and enhancing the overall customer experience. Here are some noteworthy e-commerce trends in the context of sales prediction at Big Mart:

2.1.1 Personalization and Customer Segmentation: Utilize predictive analytics to personalize the shopping experience for individual customers. Tailor promotions, discounts, and product recommendations based on their historical interactions with the platform.

2.1.2 Real-Time Inventory Management: Implement real-time inventory predictions to ensure that popular products are adequately stocked. This trend is crucial for preventing stockouts and optimizing the supply chain.

2.2 Experimental Setup

The experimental setup for Big Mart sales prediction using Flask involves the configuration and implementation of various components, including the development and training of the predictive model, integration with Flask, and testing to ensure the system's accuracy and usability. Here is a general outline of the experimental setup:

2.2.1 Data Collection and Preparation: Gather historical sales data from Big Mart, ensuring it includes relevant features such as product attributes, time stamps, and sales figures. Preprocess the data, handling missing values, encoding categorical variables, and scaling numerical features.

2.2.2 Machine Learning Model Development: Select an appropriate machine learning model for sales prediction, such as regression models, time-series models, or ensemble methods. Split the dataset into training and testing sets to evaluate the model's performance accurately. Train the model using the training dataset, optimizing hyperparameters for better accuracy.

2.2.3 Flask Integration: Develop a Flask web application that serves as the user interface for sales prediction. Create Flask routes to handle different functionalities, such as loading the trained model, receiving user inputs, and generating predictions.

2.2.4 User Interface Design: Implement an intuitive user interface that allows users to input parameters relevant to sales prediction, such as product features, time frames, or promotional events. Ensure the interface is accessible to a broad audience, including non-technical stakeholders.

2.3 Dataset and its Preprocessing

Big Mart's data scientists collected sales data of their 10 stores situated at different locations with each store having 1559 different products as per 2013 data collection. Using all the observations it is inferred what role certain properties of an item play and how they affect their sales. The dataset looks like shown in Fig.1 on using head() function on the dataset variable.

```
In [5]: df_train.head()
Out[5]:
```

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Size	Outlet_Location_
0	FDA15	9.30	Low Fat	0.016047	Dairy	249.8092	OUT049	1999	Medium	
1	DRC01	5.92	Regular	0.019278	Soft Drinks	48.2692	OUT018	2009	Medium	
2	FDN15	17.50	Low Fat	0.016760	Meat	141.6180	OUT049	1999	Medium	
3	FDX07	19.20	Regular	0.000000	Fruits and Vegetables	182.0950	OUT010	1998	NaN	
4	NCD19	8.93	Low Fat	0.000000	Household	53.8614	OUT013	1987	High	

Getting some information about the dataset as shown in Fig. 2

```
In [8]: # getting some information about the dataset
        df_train.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8523 entries, 0 to 8522
Data columns (total 12 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   Item_Identifier            8523 non-null   object
 1   Item_Weight                7060 non-null   float64
 2   Item_Fat_Content           8523 non-null   object
 3   Item_Visibility            8523 non-null   float64
 4   Item_Type                  8523 non-null   object
 5   Item_MRP                   8523 non-null   float64
 6   Outlet_Identifier          8523 non-null   object
 7   Outlet_Establishment_Year  8523 non-null   int64
 8   Outlet_Size                6113 non-null   object
 9   Outlet_Location_Type       8523 non-null   object
 10  Outlet_Type                8523 non-null   object
 11  Item_Outlet_Sales          8523 non-null   float64
dtypes: float64(4), int64(1), object(7)
memory usage: 799.2+ KB
```

In the raw data, there can be various types of underlying patterns which also gives an in-depth knowledge about subject of interest and provides insights about the problem. But caution should be observed with respect to data as it may contain null values, or redundant values, or various types of ambiguity, which also demands for pre-processing of data. Dataset should therefore be explored as much as possible.

Various factors important by statistical means like mean, standard deviation, median, count of values and maximum value etc. are shown in Fig.3 for numerical variables of our dataset.

In [6]:
```
df_train.describe()
```

Out[6]:

	Item_Weight	Item_Visibility	Item_MRP	Outlet_Establishment_Year	Item_Outlet_Sales
count	7060.000000	8523.000000	8523.000000	8523.000000	8523.000000
mean	12.857645	0.066132	140.992782	1997.831867	2181.288914
std	4.643456	0.051598	62.275067	8.371760	1706.499616
min	4.555000	0.000000	31.290000	1985.000000	33.290000
25%	8.773750	0.026989	93.826500	1987.000000	834.247400
50%	12.600000	0.053931	143.012800	1999.000000	1794.331000
75%	16.850000	0.094585	185.643700	2004.000000	3101.296400
max	21.350000	0.328391	266.888400	2009.000000	13086.964800

Preprocessing of this dataset includes doing analysis on the independent variables like checking for null values in each column and then replacing or filling them with supported appropriate data types, so that analysis and model fitting is not hindered from its way to accuracy. Shown above are some of the representations obtained by using Pandas tools which tells about variable count for numerical columns and modal values for categorical columns. Maximum and minimum values in numerical columns, along with their percentile values for median, plays an important factor in deciding which value to be chosen at priority for further exploration tasks and analysis. Data types of different columns are used further in label processing and one-hot encoding scheme during model building.

Chapter 3: Result

3.1 Prediction results

This function trains a machine learning model, generates predictions on the training set, evaluates its performance using both training set metrics (mean squared error), and cross-validation scores, providing insights into the model's generalization performance. The use of cross-validation helps estimate how well the model is likely to perform on new, unseen data.

```
In [49]: from sklearn.model_selection import cross_val_score
         from sklearn.metrics import mean_squared_error
         def train(model, X, y):

             #train the model
             model.fit(X, y)

             # predict the training set
             pred = model.predict(X)

             # perform cross-validation
             cv_score = cross_val_score(model, X, y, scoring='neg_mean_squared_error', cv=5)
             cv_score = np.abs(np.mean(cv_score))

             print("Model Report")
             print("MSE:",mean_squared_error(y,pred))
             print("CV Score:", cv_score)
```

```
In [50]: from sklearn.linear_model import LinearRegression
         model = LinearRegression(normalize=True)
         train(model, X, y)
         coef = pd.Series(model.coef_, X.columns).sort_values()
         coef.plot(kind='bar', title="Model Coefficients")

         Model Report
         MSE: 0.4400151094960561
         CV Score: 0.4412035552203208
```

Here, Fig.4 Prediction value are shown in below MSE 0.4412035552203208 and CV Score 0.4412035552203208.

3.2 Result on screen with the help of flask

This Fig 5 and Fig 6 shown below, fig 5 taking input random values form variables of data set and show the web pages using flask.

Fig 5:

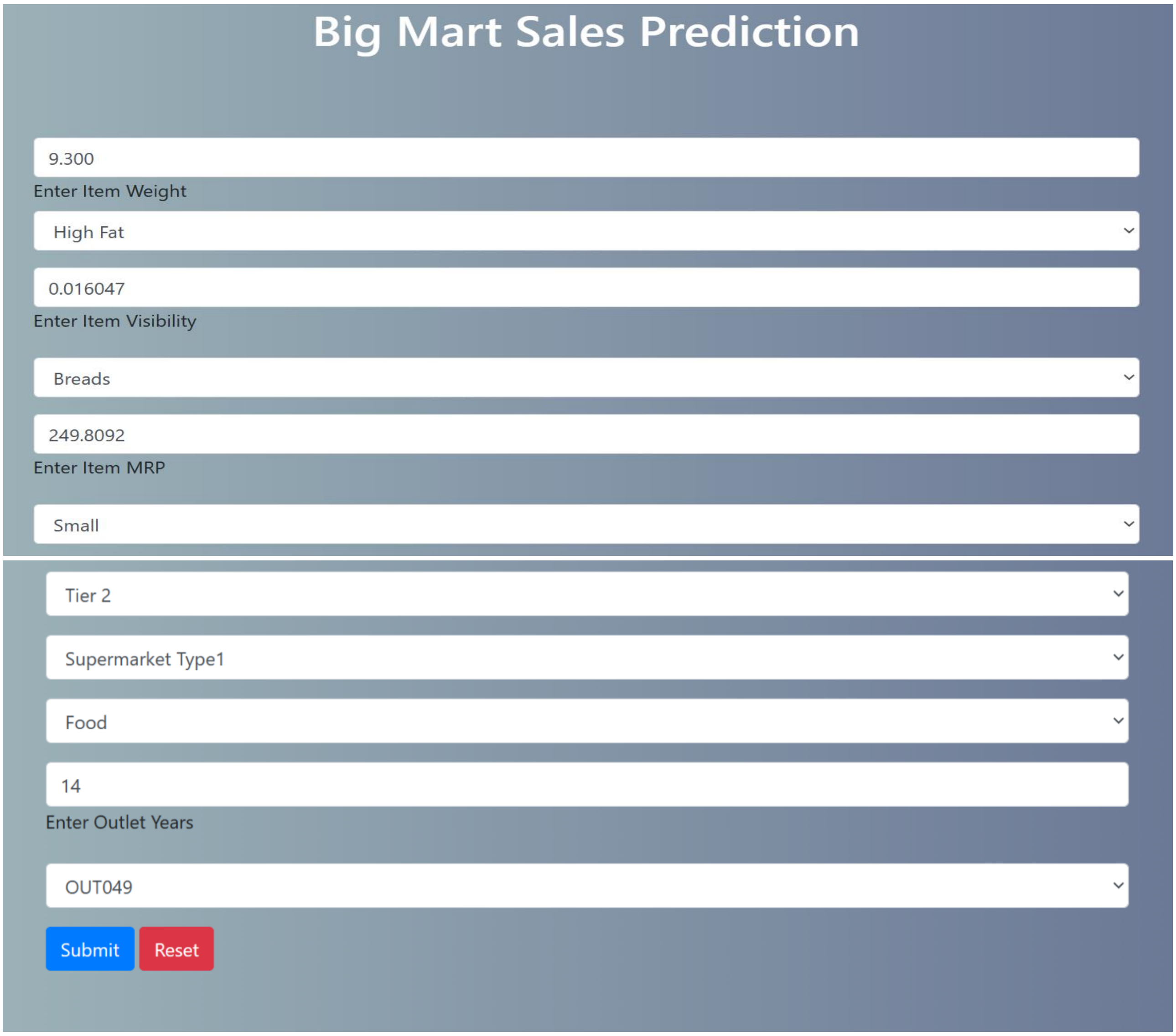
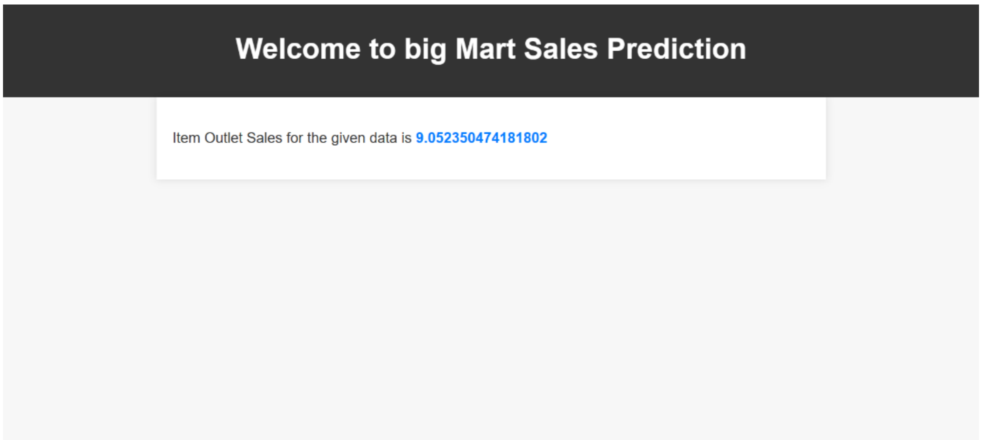

Big Mart Sales Prediction
9.300
Enter Item Weight
High Fat
0.016047
Enter Item Visibility
Breads
249.8092
Enter Item MRP
Small
Tier 2
Supermarket Type1
Food
14
Enter Outlet Years
OUT049
Submit
Reset

Fig 6:


Welcome to big Mart Sales Prediction
Item Outlet Sales for the given data is 9.052350474181802

Chapter 4: Applications

4.1 Inventory Management Optimization: Accurate sales predictions help retailers optimize their inventory management by ensuring that the right amount of stock is available. This prevents overstocking or understocking of products, reducing storage costs and minimizing the risk of stockouts.

4.2 Supply Chain Efficiency: By forecasting sales accurately, retailers can improve the efficiency of their supply chain. This includes better demand planning, optimized production schedules, and streamlined distribution processes.

4.3 Promotional Strategy Planning: Sales predictions assist in planning effective promotional strategies. Retailers can identify the most suitable products for promotions, determine the timing of promotions, and allocate resources efficiently to maximize the impact of marketing campaigns.

4.4 Revenue Maximization: With insights from sales predictions, retailers can make informed decisions to maximize revenue. This includes pricing strategies, product bundling, and identifying high-demand products to focus on.

4.5 Customer Experience Enhancement: Knowing which products are likely to sell well allows retailers to improve the overall customer experience. It enables better product availability, reduces waiting times, and enhances customer satisfaction.

4.6 Operational Cost Reduction: Efficient inventory management and supply chain optimization resulting from accurate sales predictions contribute to operational cost reduction. Retailers can avoid unnecessary expenses associated with excess inventory and emergency restocking.

4.7 Seasonal Demand Planning: Big Mart sales prediction is crucial for planning and managing seasonal demand variations. It helps retailers prepare for increased or decreased demand during specific times of the year and adjust their operations accordingly.

4.8 Risk Mitigation: Anticipating sales trends helps retailers mitigate risks associated with market fluctuations, economic conditions, or changes in consumer behavior. This proactive approach allows for better risk management and adaptation to market dynamics.

4.9 Strategic Decision-Making: Executives and decision-makers can use sales predictions as a valuable input for strategic decision-making. This includes expanding or contracting product lines, entering new markets, or adjusting business strategies based on market demand.

4.10 E-commerce Optimization: In the era of e-commerce, sales predictions are vital for optimizing online retail platforms. It helps in managing digital inventory, improving website user experience, and tailoring online promotions.

Chapter 5: Limitations

5.1 Historical Dependency: Sales prediction models rely heavily on historical data. If there are significant shifts in consumer behavior, economic conditions, or market trends that deviate from historical patterns, the model may struggle to accurately predict future sales.

5.2 External Factors: Sales can be influenced by external factors that are challenging to predict, such as changes in government policies, natural disasters, economic recessions, or unexpected events like a global pandemic. These factors can significantly impact consumer behavior and sales patterns.

5.3 Limited Feature Set: The accuracy of sales prediction models depends on the quality and relevance of the features (variables) used for training. If important features are omitted or if the available features do not capture all relevant aspects of sales dynamics, the model may not perform optimally.

5.4 Product Life Cycle: Sales prediction models may face difficulties when dealing with products at different stages of their life cycle. New product launches, product obsolescence, or changes in market demand for specific product categories can be challenging to predict accurately.

5.5 Data Quality and Quantity: The quality and quantity of historical data directly impact the performance of sales prediction models. Insufficient or poor-quality data may lead to biased or inaccurate predictions.

Chapter 6: Conclusion and Future Scope

Basics of machine learning and the associated data processing and modelling algorithms have been described, followed by their application for the task of sales prediction in Big Mart shopping centers at different locations. 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 shopping locations should follow similar patterns for improved sales.

Multiple instances parameters and various factors can be used to make this sales prediction more innovative and successful. Accuracy, which plays a key role in prediction-based systems, can be significantly increased as the number of parameters used are increased. Also, a look into how the sub-models work can lead to increase in productivity of system. The project can be further collaborated in a web-based application or in any device supported with an in-built intelligence by virtue of Internet of Things (IoT), to be more feasible for use. Various stakeholders concerned with sales information can also provide more inputs to help in hypothesis generation and more instances can be taken into consideration such that more precise results that are closer to real world situations are generated. When combined with effective data mining methods and properties, the traditional means could be seen to make a higher and positive effect on the overall development of corporation’s tasks on the whole. One of the main highlights is more expressive regression outputs, which are more understandable bounded with some of accuracy.

Moreover, the flexibility of the proposed approach can be increased with variants at a very appropriate stage of regression model-building. There is a further need of experiments for proper measurements of both accuracy and resource efficiency to assess and optimize correctly.

In conclusion, the Big Mart sales prediction project has yielded valuable insights into the dynamics of product sales within our retail environment. Through the application of advanced machine learning models, we have endeavored to forecast future sales with the aim of optimizing inventory management and enhancing overall operational efficiency.

Chapter 8: References

1). https://sist.sathyabama.ac.in/sist_naac/documents/1.3.4/1822-b.e-cse-batchno-149

2).https://www.irjmets.com/uploadedfiles/paper/in_irjmets1652543966

3).https://mui.com

4).https://www.ijitee.org/