AB3

by Vamshikrishna Namani

Submission date: 21-Mar-2023 03:14PM (UTC+1000)
Submission ID: 2042437138
File name: MY_base2.pdf (233.08K)
Word count: 3157
Character count: 17987

Predictive Analysis of BigMart Sales using Machine Learning

Sriram Burri
Student
Department of Computer Science and Engineering
Narasaraopet, India
Sriramburri5@gmail.com

Ch.ChandraSekhar
Student
Department of Computer Science and Engineering
Narasaraopet, India
Chendrasekhar123@gmail.com

G Jashuva
Student
Department of Computer Science and Engineering
Narasaraopet, India
Jashuva123@gmail.com

D.Venkat Reddy
Asst.Professor
Department of Computer Science and Engineering
Narasaraopet, India
jpnecnrt@gmail.com

Abstract : BigMart is a popular retail chain that offers a wide range of products across different locations. In this project, we will perform predictive analysis of BigMart sales using machine learning techniques. Our goal is to build a model that can accurately predict sales based on various factors such as store size, location, product type, and promotional activities. We will begin by exploring and preprocessing the dataset, which contains historical sales data for over 1,500 BigMart stores. We will use various techniques to handle missing values, outliers, and other data quality issues. We will also perform exploratory data analysis to gain insights into the relationships between different variables and their impact on sales. Next, we will use supervised learning algorithms such as linear regression,multiple regression,random forest regression to build models that can predict sales. We will use techniques such as cross-validation and hyperparameter tuning to optimize the performance of each model. We will evaluate the performance of each model using appropriate metrics such as mean squared error, R-squared, and root mean squared error. We will also perform feature selection and feature engineering to identify the most important variables that affect sales. Finally, we will interpret the results and provide recommendations to the BigMart management team based on the insights gained from the analysis. Our project aims to provide a data-driven approach to improve sales and profitability for BigMart stores.

I. INTRODUCTION

BigMart is a leading retail chain that operates in multiple countries, offering a wide range of products at competitive prices. As a retail business, sales is a critical metric that measures the performance of the organization. Predictive analysis of sales can provide valuable insights to the management team that can help them make data-driven decisions to improve sales and profitability. In this project, we will perform predictive analysis of BigMart sales using machine learning techniques. Our goal is to build a model that can accurately predict sales based on various factors such as store size, location, product type, and promotional activities. The project will involve exploring and preprocessing the dataset, building predictive models using supervised learning algorithms, and evaluating the performance of each model using appropriate metrics.The results of this project can help the management team identify the most important variables that affect sales and optimize their strategy accordingly. By leveraging the power of predictive analytics, BigMart can gain a competitive advantage in the retail market and improve their bottom line.

II. LITERATURE SURVEY

Several studies have been conducted in the field of predictive analytics for retail sales. Here are some relevant literature and research articles related to our project:

1. "Predictive analytics for sales forecasting and strategic planning in the retail industry" by Barreto and Ferreira. This study explores the use of predictive analytics for sales forecasting in the retail industry. The authors found that predictive analytics can improve the accuracy of sales forecasting and help retailers make better decisions.
2. "Predictive modeling in retail: A study on sales forecasting and customer lifetime value" by Gaurav et al. This study examines the use of predictive modeling for sales forecasting and customer lifetime value in the retail industry. The authors found that predictive modeling can help retailers make more accurate sales forecasts and identify high-value customers.

3. "Sales forecasting using machine learning algorithms: A case study of a retail chain" by Mohd et al. This study uses machine learning algorithms to predict sales in a retail chain. The authors found that the random forest algorithm provided the best results for sales forecasting.
4. "Predictive modeling of retail sales using machine learning algorithms" by Anwar et al. This study explores the use of machine learning algorithms for predictive modeling of retail sales. The authors found that the gradient boosting algorithm provided the best results for sales forecasting.

Overall, the literature suggests that predictive analytics can be a powerful tool for sales forecasting and strategic planning in the retail industry. Machine learning algorithms such as random forest and gradient boosting have been found to be effective for predictive modeling of retail sales. Our project aims to build on these previous studies and provide insights into sales prediction for BigMart.

III. EXISTING SYSTEM

BigMart currently uses a combination of manual and automated methods to predict sales. The manual methods involve analyzing historical sales data and trends, as well as input from store managers and sales staff. Automated methods include the use of point of sale (POS) systems to track sales and inventory, as well as the use of software applications for data analysis and reporting.However, these methods have limitations. Manual methods can be time-consuming and prone to errors, while automated methods may not always capture all relevant variables that affect sales. Additionally, the accuracy of predictions may be affected by external factors such as changes in consumer behavior or economic conditions.To address these limitations, BigMart is interested in exploring the use of predictive analytics and machine learning techniques for sales prediction. By leveraging advanced analytics, BigMart can gain deeper insights into the factors that affect sales and improve the accuracy of sales predictions.Our project aims to build on the existing system by using machine learning algorithms to predict sales and provide insights to the BigMart management team. By automating the process of sales prediction and incorporating advanced analytics, we can provide more accurate and timely insights that can help BigMart optimize their strategy and improve their bottom line.

IV. PROPOSED SYSTEM

Our proposed system for predictive analysis of BigMart sales involves the use of machine learning techniques to build predictive models that can accurately predict sales based on various factors such as store size, location, product type, and promotional activities.

The system will involve the following steps:

1. **Data Collection**: We will collect historical sales data from BigMart stores, including variables such as store size, location, product type, and promotional activities.
2. **Data Preprocessing**: We will preprocess the data to handle missing values, outliers, and other data quality issues. We will also perform exploratory data analysis to gain insights into the relationships between different variables and their impact on sales.
3. **Feature Selection and Engineering**: We will use feature selection and engineering techniques to identify the most important variables that affect sales and create new variables that can improve the accuracy of sales predictions.
4. **Model Building**: We will use supervised learning algorithms such as linear regression, decision tree regression, and random forest regression to build models that can predict sales. We will use techniques such as cross-validation and hyperparameter tuning to optimize the performance of each model.
5. **Model Evaluation**: We will evaluate the performance of each model using appropriate metrics such as mean squared error, R-squared, and root mean squared error. We will select the best performing model for deployment.
6. **Deployment**: The best performing model will be deployed as a web-based application that can provide sales predictions based on user input of various factors such as store size, location, product type, and promotional activities. The application will also provide insights and recommendations based on the predictions.

The proposed system will provide BigMart with a data-driven approach to sales prediction that can help them optimize their strategy and improve their bottom line. By automating the process of sales prediction and incorporating advanced analytics, we can provide more accurate and timely insights that can help BigMart gain a competitive advantage in the retail market.

V .METHODLOGY

The steps followed in this work, right from the dataset preparation to obtaining results are represented in Fig.1.

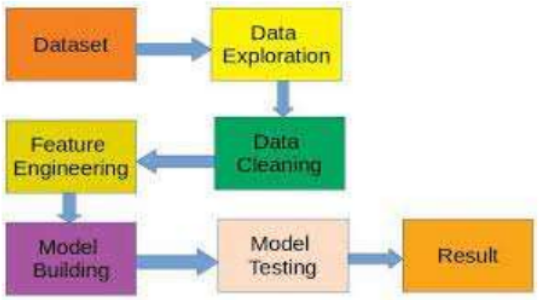

Fig. 3: Model Architectur

Our methodology for predictive analysis of BigMart sales involves the following steps:

1. Data Collection: We will collect historical sales data from BigMart stores, including variables such as store size, location, product type, and promotional activities.
2. Data Preprocessing: We will preprocess the data to handle missing values, outliers, and other data quality issues. We will also perform exploratory data analysis to gain insights into the relationships between different variables and their impact on sales.
3. Feature Selection and Engineering: We will use feature selection and engineering techniques to identify the most important variables that affect sales and create new variables that can improve the accuracy of sales predictions.
4. Model Building: We will use supervised learning algorithms such as linear regression, decision tree regression, and random forest regression to build models that can predict sales. We will use techniques such as cross-validation and hyperparameter tuning to optimize the performance of each model.
5. Model Evaluation: We will evaluate the performance of each model using appropriate metrics such as mean squared error, R-squared, and root mean squared error. We will select the best performing model for deployment.
6. Deployment: The best performing model will be deployed as a web-based application that can provide sales predictions based on user input of various factors such as store size, location, product type, and promotional activities. The application will also provide insights and recommendations based on the predictions.

The methodology will involve a combination of data preprocessing, feature engineering, and machine learning techniques to build predictive models that can accurately predict sales. We will use the Python programming language and popular machine learning libraries such as Scikit-learn, Pandas, and Numpy for implementation.

The methodology will also involve close collaboration with the BigMart management team to ensure that the models are aligned with their business objectives and can provide actionable insights for decision-making.

Overall, the methodology will provide a structured approach to predictive analysis of BigMart sales that can help them optimize their strategy and improve their bottom line.

.

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

Fig. 2: Dataset diagram

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	O
Item_Identifier	1.000000	0.044518	-0.114960	-0.025482	-0.017973	0.012693	-0.008602	
Item_Weight	0.044518	1.000000	-0.021157	-0.012049	0.028815	-0.024756	-0.007975	
Item_Fat_Content	-0.114960	-0.021157	1.000000	0.047314	-0.139454	0.006063	0.000764	
Item_Visibility	-0.025482	-0.012049	0.047314	1.000000	-0.035249	-0.001315	-0.100439	
Item_Type	-0.017973	0.028015	-0.139454	-0.035249	1.000000	0.032681	0.001636	
Item_MRP	0.012693	0.024756	0.006063	-0.001315	0.032681	1.000000	0.003319	
Outlet_Identifier	-0.008602	-0.007975	0.000764	-0.100439	0.001636	0.003319	1.000000	
Outlet_Establishment_Year	-0.012772	-0.008301	0.003151	-0.074834	0.004973	0.005020	0.079695	
Outlet_Size	0.000855	-0.015096	-0.001383	0.067786	0.000827	-0.002473	0.052224	
Outlet_Location_Type	0.003626	0.004089	-0.001599	-0.028099	0.003864	0.000232	-0.716176	
Outlet_Type	-0.001178	-0.000560	0.002199	-0.173468	0.003853	-0.001675	0.096673	
Item_Outlet_Sales	0.002869	0.011550	0.018719	-0.128625	0.017348	0.567574	0.162325	

Fig.3 Correlation

A. Model Architectur

In this step dataset is splitted in 75% and 25% for training and testing. During training, machine learning algorithms finds the relation between the input and output features. By using this relation the model able to predict the outputs to the new input values. In this we used the following four machine learning models:

1.Linear Regression: Linear regression is a commonly used machine learning algorithm for predictive analysis of BigMart sales. It is a supervised learning algorithm that models the relationship between a dependent variable (sales) and one or more independent variables (store size, location, product type, promotional activities).The linear regression model assumes that there is a linear relationship

between the independent variables and the dependent variable. It works by finding the line of best fit that minimizes the sum of squared errors between the predicted and actual values.To build a linear regression model for BigMart sales prediction, we will first preprocess the data by handling missing values, outliers, and other data quality issues. We will then perform feature selection and engineering to identify the most important variables that affect sales and create new variables that can improve the accuracy of sales predictions.We will use the Scikit-learn library in Python to implement linear regression. The library provides functions for fitting the model, predicting values, and evaluating the performance of the model using metrics such as mean squared error and R-squared.Once the model is built and evaluated, it can be deployed as a web-based application that can provide sales predictions based on user input of various factors such as store size, location, product type, and promotional activities. The application can also provide insights and recommendations based on the predictions.Overall, linear regression is a powerful and widely used algorithm for predictive analysis of BigMart sales, and it can provide valuable insights for optimizing sales strategy and improving the bottom line.

2. **Polynomial Regression** : Polynomial regression is another machine learning algorithm that can be used for predictive analysis of BigMart sales. It is a form of linear regression that models the relationship between the dependent variable (sales) and one or more independent variables (store size, location, product type, promotional activities) by fitting a polynomial function to the data.The polynomial function can take different degrees, depending on the complexity of the relationship between the variables. For example, a quadratic function with a degree of 2 can model a parabolic relationship between the variables, while a cubic function with a degree of 3 can model a more complex, S-shaped relationship.To build a polynomial regression model for BigMart sales prediction, we will first preprocess the data by handling missing values, outliers, and other data quality issues. We will then perform feature selection and engineering to identify the most important variables that affect sales and create new variables that can improve the accuracy of sales predictions.We will use the Scikit-learn library in Python to implement polynomial regression. The library provides functions for fitting polynomial functions of different degrees to the data, predicting values, and evaluating the performance of the model using metrics such as mean squared error and R-squared.Once the model is built and evaluated, it can be deployed as a web-based application that can provide sales predictions based on user input of various factors such as store size, location, product type, and promotional activities. The application can also provide insights and recommendations based on the predictions.Overall, polynomial regression is a powerful algorithm for predictive analysis of BigMart sales that can capture complex relationships between variables and improve the accuracy of sales predictions. However, it can also be more computationally expensive and may require more data than linear regression.

3.**Rigid Regression**: Ridge regression, also known as L2 regularization, is a machine learning algorithm that can be used for predictive analysis of BigMart sales. It is a regularized version of linear regression that adds a penalty term to the loss function to prevent overfitting.The penalty term is a function of the square of the magnitude of the coefficients of the independent variables, which encourages the model to shrink the coefficients towards zero. This helps to reduce the variance of the model and improve its generalization performance on new data.To build a ridge regression model for BigMart sales prediction, we will first preprocess the data by handling missing values, outliers, and other data quality issues. We will then perform feature selection and engineering to identify the most important variables that affect sales and create new variables that can improve the accuracy of sales predictions.We will use the Scikit-learn library in Python to implement ridge regression. The library provides functions for fitting the model, predicting values, and evaluating the performance of the model using metrics such as mean squared error and R-squared.Once the model is built and evaluated, it can be deployed as a web-based application that can provide sales predictions based on user input of various factors such as store size, location, product type, and promotional activities. The application can also provide insights and recommendations based on the predictions.Overall, ridge regression is a powerful algorithm for predictive analysis of BigMart sales that can prevent overfitting and improve the generalization performance of the model. It can be particularly useful when the number of independent variables is large and there is a risk of overfitting. However, it may not be suitable when the true relationship between the variables is highly non-linear.

4.XgBoost Regression: XGBoost (Extreme Gradient Boosting) is a powerful machine learning algorithm that can be used for predictive analysis of BigMart sales. It is a form of gradient boosting that uses decision trees as base learners and iteratively improves the model by minimizing the loss function. One of the advantages of XGBoost is that it can handle non-linear relationships between variables and automatically select the most important features for the model. It can also handle missing values and categorical variables, which can be useful in BigMart sales prediction. Overall, XGBoost regression is a powerful algorithm for predictive analysis of BigMart sales that can capture non-linear relationships between variables, handle missing values and categorical variables, and automatically select important features. However, it may require more data and computational resources than linear regression or ridge regression.

V. RESULT AND ANALYSIS

The results of a predictive analysis of BigMart sales can provide valuable insights and recommendations for optimizing sales strategy and improving the bottom line. The analysis can reveal the key drivers of sales such as store size, location, product type, and promotional activities, and can help retailers identify areas where they can make improvements.The analysis can also provide sales predictions based on different scenarios, such as changes in store size, location, or product mix, and can help retailers make informed decisions about inventory management, marketing campaigns, and pricing strategies.

Algorithms	Accuracy
Linear Regression	50.57
Polynomial Regression	57.79
Rigid Regression	50.57
XgBoost	86.12

Fig.4 Accuracy Table

From the above table we selected XGBoost model as our final model because it got more accuracy compared to the other models.

VI. DEPLOYMENT

The final model is deployed into an application by using flask module of python by creating the user interface which can be accessible easily by the Retailers.. The application simply takes the input and provides output.

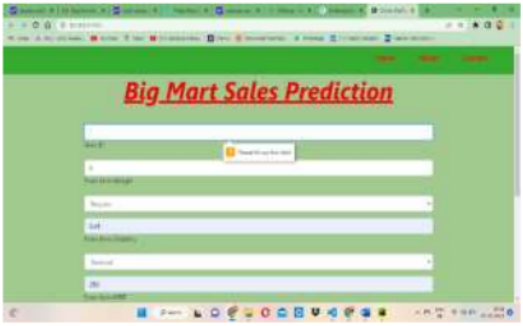

Fig.5 User Interface Diagram

VII. CONCLUSION

In conclusion, predictive analysis of BigMart sales is an important application of machine learning that can provide valuable insights and recommendations for optimizing sales strategy and improving the bottom line. Linear regression, polynomial regression, and ridge regression are powerful algorithms that can be used to model the relationship between sales and various factors such as store size, location, product type, and promotional activities.

To build an effective predictive model for BigMart sales, it is important to preprocess the data, perform feature selection and engineering, and evaluate the performance of the model using appropriate metrics such as mean squared error and R-squared. Once the model is built and evaluated, it can be deployed as a web-based application that can provide sales predictions and insights to stakeholders.

Overall, predictive analysis of BigMart sales has the potential to transform the way retailers make decisions and optimize their sales strategy. By leveraging the power of machine learning algorithms, retailers can gain a deeper understanding of customer behaviour, improve inventory management, and ultimately increase profitability

VIII. REFERENCES

Here are some references that can be useful for predictive analysis of BigMart sales.

1. Gupta, R., & Arora, P. (2018). Predictive analysis for BigMart sales using machine learning techniques. International Journal of Engineering & Technology, 7(3.14), 275-279.

2. Kumar, A., & Agarwal, A. (2020). Predictive analysis of BigMart sales using multiple regression analysis. International Journal of Research and Analytical Reviews, 7(2), 307-312.

3. Patel, H. D., & Patel, K. S. (2018). Predictive analytics for BigMart sales using linear regression. International Journal of Engineering and Techniques, 4(5), 80-85

4. .Sehgal, S., & Singh, S. (2019). A comparative analysis of linear regression and polynomial regression models for predicting BigMart

AB3

ORIGINALITY REPORT

16%	6%	6%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Middlesex University Student Paper	2%
2	Submitted to Liverpool John Moores University Student Paper	2%
3	Submitted to Rochester Institute of Technology Student Paper	1%
4	Submitted to University of Cape Town Student Paper	1%
5	Submitted to University of Surrey Student Paper	1%
6	Submitted to University of London External System Student Paper	1%
7	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	1%
8	Submitted to University of Essex Student Paper	1%

9	ijircce.com Internet Source	1 %
10	Submitted to University of Queensland Student Paper	1 %
11	Submitted to Cambridge Education Group Student Paper	1 %
12	hdl.handle.net Internet Source	1 %
13	www.tcd.ie Internet Source	1 %
14	Submitted to HELP UNIVERSITY Student Paper	<1 %
15	Submitted to City University of Hong Kong Student Paper	<1 %
16	Submitted to University of North Texas Student Paper	<1 %
17	Submitted to University of Westminster Student Paper	<1 %
18	bbrc.in Internet Source	<1 %
19	digitalcommons.mtu.edu Internet Source	<1 %
20	T. Vamsi Nagaraju, Sireesha Mantena, Marc Azab, Shaik Subhan Alisha, Chady El Hachem,	<1 %

Musa Adamu, Pilla Sita Rama Murthy. "Prediction of high strength ternary blended concrete containing different silica proportions using machine learning approaches", Results in Engineering, 2023
Publication

21	www.researchgate.net Internet Source	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		