

Big Mart Sales Analysis Using Machine Learning

Amisha Kothekar, Monali Bodhale, Pratibha Satapure, Rutuja Sarode

Department of Computer Engineering

Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Abstract: Day by day competition among different shopping malls as well as big marts is getting more serious and aggressive only due to the rapid growth of the global malls and on-line shopping. Every mall or mart is trying to provide personalized and short-time offers for attracting more customers depending upon the day, such that the volume of sales for each item can be predicted for inventory management of the organization, logistics and transport service, etc. Present machine learning algorithm are very sophisticated and provide techniques to predict or forecast the future demand of sales for an organization, which also helps in overcoming the cheap availability of computing and storage systems. In this paper, we are addressing the problem of big mart sales prediction or forecasting of an item on customer's future demand in different big mart stores across various locations and products based on the previous record. Taking various aspects of a dataset collected for Big Mart, and the methodology followed for building a predictive model, results with high levels of accuracy are generated, and these observations can be employed to take decisions to improve sales.

Keywords: Gait analysis, Machine learning

I. INTRODUCTION

Nowadays, shopping malls and Big Marts track sales data for each and every individual item in order to forecast future customer demand and update inventory management. In a data warehouse, these data stores basically contain a large number of customer data and individual item attributes. Anomalies and common patterns are also discovered by mining the data warehouse's data store. For retailers like Big Mart, the resulting data can be used to predict future sales volume using various machine learning techniques. In this paper, exploratory machine learning techniques are used to predict big-box store sales. In general, sales forecasting is critical for marketing, retailing, wholesale, and manufacturing, and this plays important role for marketing, retailing, wholesaling and manufacturing and this is done in different companies. This proposed system will allow the companies to plan better strategy and achieve revenue sales and will give better growth for the future in companies. Compared to the other learning methods, this machine learning method gives accurate result.

II. KEY TAKE AWAYS OF DEEP LEARNING

- Deep learning is an AI work that mirrors the activities of the human mind in preparing information for use in basic leadership.
- Deep learning can learn from the data that is from both unstructured and unlabeled source. Deep learning is machine learning's subset that can be used to help to detect fraud or money laundering.

III. LITERATURE SURVEY

“yangsen chen”, [1]. BIG MART SALES PREDICTION USING MACHINE LEARNING -Nowadays shopping malls and Big Marts keep the track of their sales data of each and every individual item for predicting future demand of the customer and update the inventory management as well. These data stores basically contain a large number of customer data and individual item attributes in a data warehouse. Further, anomalies and frequent patterns are detected by mining the data store from the data warehouse. The resultant data can be used for predicting future sales volume with the help of different machine learning techniques for the retailers like Big Mart. In this paper, we propose a predictive model using XG boost Regressor technique for predicting the sales of a company like Big Mart and found that the model produces better performance as compared to existing models.

Jian Liu.[2] Big Mart Sale Prediction using Machine Learning The sales forecast is based on Big Mart sales for various outlets to adjust the business model to expected outcomes. The resulting data can then be used to prediction potential sales volumes for retailers such as Big Mart through various machine learning methods. The estimate of the system proposed



should take account of price tag, outlet and outlet location. A number of networks use the various machine-learning algorithms, such as linear regression and decision tree algorithms, and XGBoost regressor, which offers an efficient prevision of Big Mart sales based on gradient. At last, hyperparameter tuning is used to help you to choose relevant hyperparameters that make the algorithm Shine and produce the highest accuracy

Ahmad M. Hamad. [3] Machine Learning is a category of algorithms that allows software applications to become more accurate in predicting outcomes without being explicitly programmed. The basic premise of machine learning is to build models and employ algorithms that can receive input data and use statistical analysis to predict an output while updating outputs as new data becomes available. These models can be applied in different areas and trained to match the expectations of management so that accurate steps can be taken to achieve the organization's target. In this paper, the case of Big Mart, a one-stop-shopping center, has been discussed to predict the sales of different types of items and for understanding the effects of different factors on the items' sales. Taking various aspects of a dataset collected for Big Mart, and the methodology followed for building a predictive model, results with high levels of accuracy are generated, and these observations can be employed to take decisions to improve sales.

Jinyoung Moon. [4] The sales forecast is based on BigMart sales for various outlets to adjust the business model to expected outcomes. The resulting data can then be used to prediction potential sales volumes for retailers such as BigMart through various machine learning methods. The estimate of the system proposed should take account of price tag, outlet and outlet location. A number of networks use the various machine learning algorithms, such as linear regression and decision tree algorithms, and an XGBoost regressor, which offers an efficient prevision of BigMart sales based on gradient. At last, hyperparameter tuning is used to help you to choose relevant hyperparameters that make the algorithm shine and produce the highest accuracy.

Elias Kougianos. [5] In this paper the prediction of big mart sales using exploratory machine learning techniques are implemented. Basically, the forecasting of sales plays important role for marketing, retailing, wholesaling and manufacturing and this is done in different companies. This proposed system will allow the companies to plan better strategy and achieve revenue sales and will give better growth for the future in companies. Compared to the other learning methods, this machine learning method gives accurate result. In the entire system the feature selection, data transformation and data exploration will plays important role and gives effective output in terms of accuracy.

Wenqing Cheng .[6] Nowadays shopping malls and Big Marts keep the track of their sales data of each and every individual item for predicting future demand of the customer and update the inventory management as well. These data stores basically contain a large number of customer data and individual item attributes in a data warehouse. Further, anomalies and frequent patterns are detected by mining the data store from the data warehouse. The resultant data can be used for predicting future sales volume with the help of different machine learning techniques for the retailers like Big Mart. In this paper, we propose a predictive model using Xgboost technique for predicting the sales of a company like Big Mart and found that the model produces better performance as compared to existing models. A comparative analysis of the model with others in terms performance metrics is also explained in details.

Basant Kumar .[7] Sales forecasting is an important aspect of different companies engaged in retailing, logistics, manufacturing, marketing and wholesaling. It allows companies to efficiently allocate resources, to estimate achievable sales revenue and to plan a better strategy for future growth of the company. In this paper, prediction of sales of a product from a particular outlet is performed via a two-level approach that produces better predictive performance compared to any of the popular single model predictive learning algorithms. The approach is performed on Big Mart Sales data of the year 2013. Data exploration, data transformation and feature engineering play a vital role in predicting accurate results. The result demonstrated that the two-level statistical approach performed better than a single model approach as the former provided more information that leads to better prediction.

IV. RELATED WORK

4.1 Machine Learning

Machine learning is an artificial intelligence (AI) application that allows systems to automatically learn and improve from experience without being explicitly programmed. Machine learning is concerned with the creation of computer programmers that can access data and use it to learn on their own. Machine learning is used in search engines, email filters to filter out spam, websites to make personalized recommendations, banking software to detect unusual transactions, and



many apps on our phones, such as voice recognition.

4.2 Artificial Intelligence

The simulation of human intelligence processes by machines, particularly computer systems, is known as artificial intelligence. Expert systems, natural language processing, speech recognition, and machine vision are some of the specific applications of AI. The fundamental goal of AI is to enable computers and machines to perform cognitive tasks such as problem solving, decision making, perception, and comprehension of human communication. Artificial intelligence (AI) is the foundation for simulating human intelligence processes by developing and deploying algorithms in a dynamic computing environment. Simply put, AI is an attempt to make computers think and act like humans.

V. PROPOSED METHODOLOGY

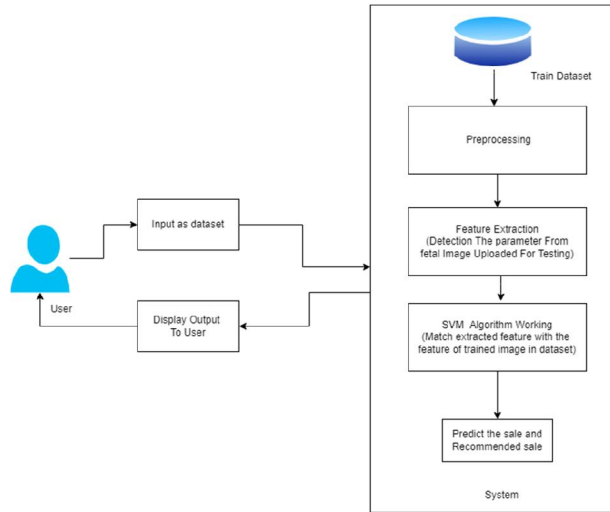


Fig 4.1 Proposed System Architecture

4.3 Algorithms

SVM Support Vector Machine(SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its SKNCOE, Department of Computer Engineering 2022 17 best suited for classification. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points SVM is a supervised machine learning algorithm which can be used for classification or regression problems. It uses a technique called the kernel trick to transform your data and then based on these transformations it finds an optimal boundary between the possible outputs.SVM can be used for classification (distinguishing between several groups or classes) and regression (obtaining a mathematical model to predict something). They can be applied to both linear and non linear problems. Until 2006 they were the best general purpose algorithm for machine learning.

V. RESULTS





Registration Form

Full Name :

Address :

Email :

Phone number :

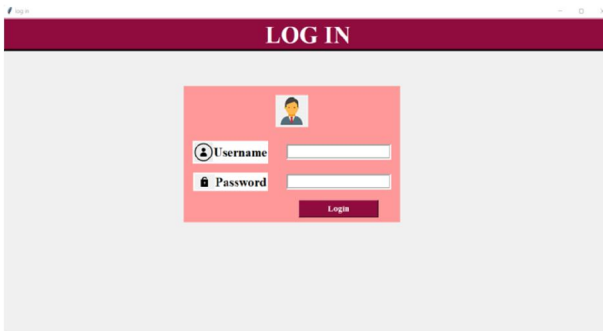
Gender : Male Female

Age :

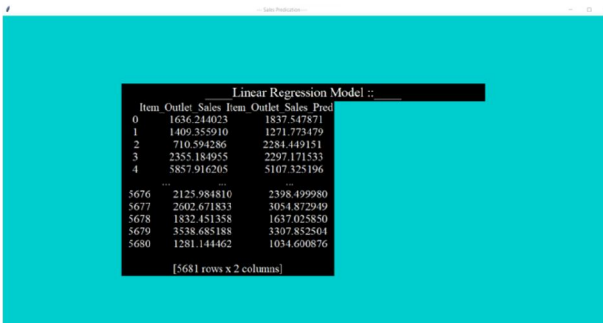
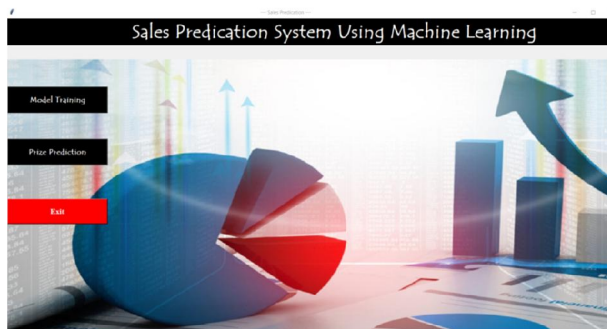
User Name :

Password :

Confirm Password :



LOG IN



Linear Regression Model ::

Item	Outlet	Sales	Item	Outlet	Sales	Pred
0	1636	244023		1837	547871	
1	1409	355910		1271	778179	
2	710	594286		2284	449151	
3	2355	184955		2297	171533	
4	5857	916205		5107	325196	
5676	2125	981810		2398	499980	
5677	2602	671833		3054	872949	
5678	1832	451358		1637	025850	
5679	3538	685188		3307	852504	
5680	1281	144462		1034	600876	

[5681 rows x 2 columns]



5.1 Test Cases

Test Case ID	Test Case	Test Case I/P	Actual Result	Expected Result	Test case criteria(P/F)
001	Store Xml File	Xml file	Xml file store	Error Should come	P
002	Parse the xml file for conversion	parsing	File get parse	Accept	P
003	Attribute identification	Check individual Attribute	Identify Attributes	Accepted	P
004	Weight Analysis	Check Weight	Analyze Weight of individual Attribute	Accepted	P
005	Tree formation	Form them-Tree	Formation	Accepted	P
006	Cluster Evaluation	Check Evaluation	Should check Cluster	Accepted	P
007	Algorithm Performance	Check Evaluation	Should work Algorithm Properly	Accepted	P
008	Query Formation	Check Query Correction	Should check Query	Accepted	P

Test Case ID	Test Case	Test Case I/P	Actual Result	Expected Result	Test case criteria(P/F)
001	Enter the number in username, middle name, last name field	Number	Error Comes	Error Should Comes	P
001	Enter the character in username, middle name, last name field	Character	Accept	Accept	P
002	Enter the invalid email id format in email id field	Kgmail.com	Error comes	Error Should Comes	P
002	Enter the valid email id format in email id field	kk@gmail.com	Accept	Accept	P
003	Enter the invalid digit no in phone no field	99999	Error comes	Error Should Comes	P
003	Enter the 10 digit no in phone no field	9999999999	Accept	Accept	P

Test Case ID	Test Case	Test Case I/P	Actual Result	Expected Result	Test case criteria(P/F)
001	Enter the number in username, middle name, last name field	Number	Error Comes	Error Should Comes	P
001	Enter the character in username, middle name, last name field	Character	Accept	Accept	P
002	Enter the invalid email id format in email id field	Kgmail.com	Error comes	Error Should Comes	P
002	Enter the valid email id format in email id field	kk@gmail.com	Accept	Accept	P
003	Enter the invalid digit no in phone no field	99999	Error comes	Error Should Comes	P
003	Enter the 10 digit no in phone no field	9999999999	Accept	Accept	P

Test case	Login Screen- Sign up
Objective	Click on sign up button then check all required/ mandatory fields with leaving all fields blank
Expected Result	All required/ mandatory fields should display with symbol “*”. Instruction line “* field(s) are mandatory” should be displayed
Test case	Create a Password >>Text Box Confirm Password >>Text Box
Objective	Check the validation message for Password and Confirm Password field
Expected Result	Correct validation message should be displayed accordingly or “Password and confirm password should be same” in place of “Password mismatch”.

VI. CONCLUSION

1. In this paper, the SVM machine learning algorithm is used to predict big-box store sales. The proposed system will improve efficiency and precision.
2. The profit made by the company is directly proportional to the prediction. This proposed system will help businesses plan better strategies and increase revenue.
3. Sales will improve in the future, allowing businesses to grow. In comparison to the others,
4. Machine learning is one of the most accurate learning methods.

REFERENCES

- [1]. Shashua, A. (2009). Introduction to machine learning: Class notes 67577. arXiv preprint arXiv:0904.3664.
- [2]. MacKay, D. J., Mac Kay, D. J. (2003). Information theory, inference and learning algorithms. Cambridge university press.
- [3]. Cerrada, M., Aguilar, J. (2008). Reinforcement learning in system identification. In Reinforcement Learning. IntechOpen.
- [4]. Welling, M. (2011). A first encounter with Machine Learning. Irvine, CA.: University of California, 12. 5
Quinlan, J. R. (2014). C4.
- [5]. Programs for machine learning. Elsevier.
- [6]. Zone-Ching Lin, Wen-Jang Wu, “Multiple Linear Regression Analysis of the Overlay Accuracy Model Zone”, IEEE Trans. on Semiconductor Manufacturing, vol. 12, no. 2, pp. 229 – 237, May 1999.
- [7]. X. Yua, Z. Qi, Y. Zhao, Support Vector Regression for Newspaper/Magazine Sales Forecasting, Procedia Computer Science 17 (2013) 1055–1062.
- [8]. D. Fantazzini, Z. Toktamysova, Forecasting German car sales using Google data and multivariate models, Int. J. Production Economics 170 (2015) 97- 135.