



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



Sales Forecasting using Predictive Learning Algorithms

Dr.R.Nagarajan¹, Manoj raj S²

¹Assistant Professor, PG & Research Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore, India

²UG Student, PG & Research Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore, India

ABSTRACT: In this paper, prediction of sales of a product from an 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 Departmental store. The proposed approach was organized into six stages, first is data collection, which includes collecting data and dataset, second is hypothesis definition, which used to analyse the problems, third is data exploration which used to explore the uniqueness of the data, fourth is data cleaning, which is used to detect and correct the inaccurate dataset, fifth is data modelling, which is used to predict the data using machine learning techniques, sixth is feature engineering, which is used to import the data from machine learning algorithm. And, finally validating and implementation of our results using precision and accuracy techniques. The result is demonstrated in two-level statistical approach to perform better than a single model approach as provided with more information that leads to better prediction

KEYWORDS: Sales, Predictive, Machine learning, Optimization

I. INTRODUCTION

Sales are a life blood of every company and sales forecasting plays a vital role in conducting any business. Good forecasting helps to develop and improve business strategies by increasing the knowledge about the marketplace. A standard sales forecast looks deeply into the situations or the conditions that previously occurred and then, applies inference regarding customer acquisition, identifies inadequacy and strengths before setting a budget as well as marketing plans for the upcoming year. In other words, sales forecasting is sales prediction that is based on the available resources from the past. An in-depth knowledge of the past resources allows preparing for the upcoming needs of the business and increases the likelihood to succeed irrespective of external circumstances. Businesses that treat sales forecasting as the primary step, tend to perform better than those data mining predictive techniques via stacking is considered a two-level statistical approach. It is named as two-level because stacking is performed on two layers in which bottom layer consists of one or more than one learning algorithms and top layer consists of one learning algorithm.

Stacking is also known as Stacked Generalization. It basically involves the training of the learning algorithm present in the top layer to combine the predictions made by the algorithms present in the bottom layer. In the first step, all the learning algorithms are trained using the departmental store dataset and in the second step, Stacking performs better than any single model because a stacking involves more information for prediction. In this paper approach has been done under six stages. In first stage, data is collected from dataset. In second stage, problems are analyzed from the data collection. In third stage, uniqueness of the data is explored. In fourth stage, data cleaning is done to detect and correct the dataset. In fifth stage, data modelling techniques is used to predict the data. In sixth stage, the feature engineering is used to import the data from the machine learning algorithm. Sales prediction is done accurately by using machine learning algorithms.

Weather expectation is a difficult issue in meteorological division since years. Indeed, even after mechanical and logical progression, the exactness in forecast of weather is adequate. Indeed, even in ebb and flow date this space stays as an examination point in which researchers to create a model or a calculation that will precisely foresee weather. There tremendous enhancements in the sensors that are answerable for recording data from climate and drop noise present in them alongside which incorporate various characteristics connected with weather to make precise expectation.

It offers an approach to dissecting data genuinely separate or infers such standards that for expectations. As of now it is being utilized in numerous areas like stock market, sports, banking segment, and so forth. The fundamental substance of data mining is data itself. Data can be put away in a coordinated way which is known as database.

II. RELATED WORK

Chammas M (2019) et.al -One of key factors for coordinating systems and making homes and urban communities wiser is the energy reserve funds that can be accomplished. In this work, they propose to demonstrate an energy prediction system in view of a Multilayer Perceptron neural network (MLP). A few data classifications have been utilized, i.e.; light energy, temperature, humidity, day of the week, and so on, gathered from a Wireless Sensor Network introduced in a two story building. They looked at the presentation of system against four order algorithms, specifically: LR, SVM, Gradient Boosting Machine and Random Forest. They evaluated the significance of transient information and light energy as extra highlights for the prediction model. The transient highlights slightly worked on the exhibition for all systems, while the lights energy diminished the presentation. For the strategy, they note a diminishing by 10% in relative execution while the thing that matters was unimportant for different classifiers (under 1% for Gradient Boosting Machine). They system outflanked the other four classifiers in all situations. In light of the led experiments, they infer that climate information would be to the point of anticipating energy consumption. This empowers minimal expense answers for energy predictions. Further experiments ought to be performed on various datasets to approve they results.

III. PROPOSED METHODOLOGY

DATASET COLLECTION

A data set is a collection of data. Departmental store data has been used as the dataset for the proposed work. Sales data has Item Identifier, Item Fat, Item Visibility, Item Type, Outlet Type, Item MRP, Outlet Identifier, Item Weight, Outlet Size, Outlet Establishment Year, Outlet Location Type, and Item Outlet Sales.

HYPOTHESIS DEFINITION

This is a very important step to analyse any problem. The first and foremost step is to understand the problem statement. The idea is to find out the factors of a product that creates an impact on the sales of a product. A null hypothesis is a type of hypothesis used in statistics that proposes that no statistical significance exists in a set of given observations. An alternative hypothesis is one that states there is a statistically significant relationship between two variables.

DATA EXPLORATION

Data exploration is an informative search used by data consumers to form true analysis from the information gathered. Data exploration is used to analyse the data and information from the data to form true analysis. After having a look at the dataset, certain information about the data was explored. Here the dataset is not unique while collecting the dataset. In this module, the uniqueness of the dataset can be created.

DATA CLEANING

In data cleaning module, is used to detect and correct the inaccurate dataset. It is used to remove the duplication of attributes. Data cleaning is used to correct the dirty data which contains incomplete or outdated data, and the improper parsing of record fields from disparate systems. It plays a significant part in building a model.

DATA MODELLING

In data modelling module, the machine learning algorithms were used to predict the sales. Linear regression and K-means algorithm were used to predict the sales. The user provides the ML algorithm with a dataset that includes desired inputs and outputs, and the algorithm finds a method to determine how to arrive at those results. Linear regression algorithm is a supervised learning algorithm. It implements a statistical model when relationships between the independent variables and the dependent variable are almost linear, shows optimal results. This algorithm is used to show the sales prediction with increased accuracy rate. K-means algorithm is an unsupervised learning algorithm. It deals with the correlations and



relationships by analyzing available data. This algorithm clusters the data and predict the value of the dataset point. The train dataset is taken and are clustered using the algorithm. The visualization of the clusters is plotted in the graph.

FEATURE ENGINEERING

In , feature engineering module, the process of using the import data into machine learning algorithms to predict the accurate sales. A feature is an attribute or property shared by all the independent products on which the prediction is to be done. Any attribute could be a feature, it is useful to the model.

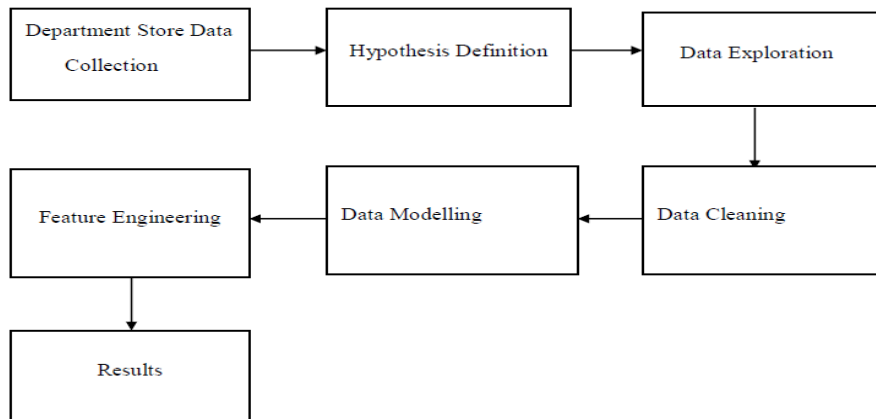


Fig-1 System Architecture

IV. RESULT AND DISCUSSION

Machine Learning algorithm such as Linear regression and K-means clustering is been used to predict the sales of the departmental store and it is implemented in this project. The use of algorithm enables to increase the accuracy of the sales. The accuracy of the sales can be reached up to the value and it is plotted in the graph format.

Item Identifier	Item_Wei	Item_Fat	Item_Visib	Item_Typr	Item_MRF	Outlet_Ids	Outlet_Es	Outlet_Siz	Outlet_Lo	Outlet_Ty	Item_Outlet_Sales
FDA15	9.3	Low Fat	0.016047	Dairy	249.8092	OUT049	1999	Medium	Tier 1	Supermar	3735.138
DRC01	5.92	Regular	0.019278	Soft Drink	48.2692	OUT018	2009	Medium	Tier 3	Supermar	443.4228
FDN15	17.5	Low Fat	0.01676	Meat	141.618	OUT049	1999	Medium	Tier 1	Supermar	2097.27
FDX07	19.2	Regular	0	Fruits and	182.095	OUT010	1998		Tier 3	Grocery St	732.38
NCD19	8.93	Low Fat	0	Householi	53.8614	OUT013	1987	High	Tier 3	Supermar	994.7052
FDP36	10.395	Regular	0	Baking Go	51.4008	OUT018	2009	Medium	Tier 3	Supermar	556.6088
FDO10	13.65	Regular	0.012741	Snack Foo	57.6588	OUT013	1987	High	Tier 3	Supermar	343.5528
FDP10		Low Fat	0.12747	Snack Foo	107.7622	OUT027	1985	Medium	Tier 3	Supermar	4022.764
FDH17	16.2	Regular	0.016687	Frozen Foo	96.9726	OUT045	2002		Tier 2	Supermar	1076.599
FDU28	19.2	Regular	0.09445	Frozen Foo	187.8214	OUT017	2007		Tier 2	Supermar	4710.535
FDY07	11.8	Low Fat	0	Fruits and	45.5402	OUT049	1999	Medium	Tier 1	Supermar	1516.027
FDA03	18.5	Regular	0.045464	Dairy	144.1102	OUT046	1997	Small	Tier 1	Supermar	2187.153
FDX32	15.1	Regular	0.100014	Fruits and	145.4786	OUT049	1999	Medium	Tier 1	Supermar	1589.265
FDS46	17.6	Regular	0.047257	Snack Foo	119.6782	OUT046	1997	Small	Tier 1	Supermar	2145.208
FDF32	16.35	Low Fat	0.068024	Fruits and	196.4426	OUT013	1987	High	Tier 3	Supermar	1977.426
FDP49	9	Regular	0.069089	Breakfast	56.3614	OUT046	1997	Small	Tier 1	Supermar	1547.319
NCB42	11.8	Low Fat	0.008596	Health ani	115.3492	OUT018	2009	Medium	Tier 3	Supermar	1621.889
FDP49	9	Regular	0.069196	Breakfast	54.3614	OUT049	1999	Medium	Tier 1	Supermar	718.3982
DR111		Low Fat	0.034238	Hard Drinl	113.2834	OUT027	1985	Medium	Tier 3	Supermar	2303.668
FDU02	13.35	Low Fat	0.102492	Dairy	230.5352	OUT035	2004	Small	Tier 2	Supermar	2748.422

Fig-2 Sample Dataset

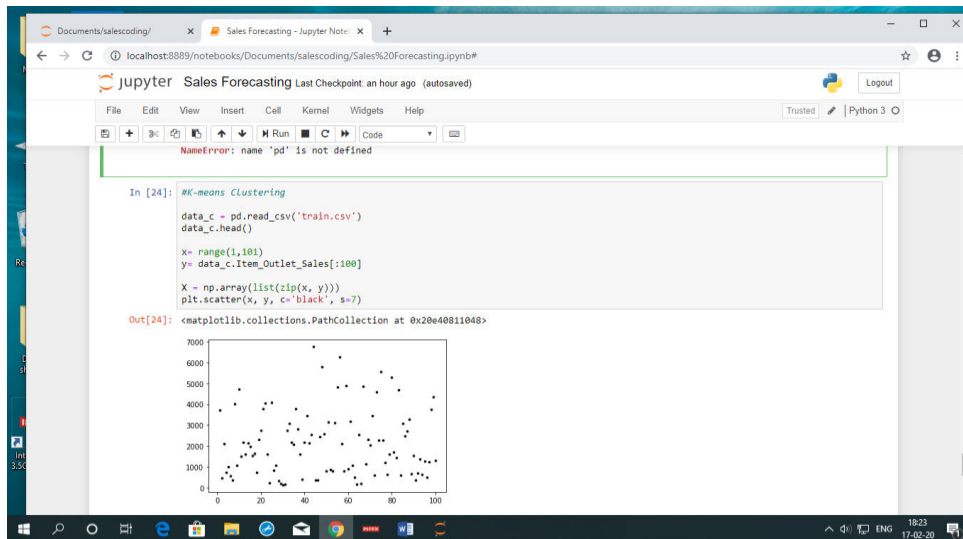


Fig-3 K-Means Clustering Visualization

V. CONCLUSION

Every company desires to know the demand of the customer in any season beforehand to avoid the shortage of products. As time passes by, the demand of the store to be more accurate about the predictions will increase exponentially. So, huge research is going on in this sector to make accurate predictions of sales. Better predictions are directly proportional to the profit made by the departmental store. The purpose of measuring accuracy was to validate our prediction with the actual result. In this project, an effort has been made to predict sales of the product from an outlet accurately by using a two-level statistical model that reduces the mean absolute error value. The two-level statistical model performed than the other single model predictive techniques and contributed better predictions to the departmental store dataset.

REFERENCES

- [1] I. Ahmad, A. Millie, P. Chang, and W. Ahn, "SVD based fragile watermarking scheme for tamper localization and self-recovery," *Int. J. Mach. Learn. Cybern.*, 2015.
- [2] H. Rhayma, A. Makhloufi, A. Ben Hamida, and H. Hamam, "Self-Embedding Authentication Scheme based on Data Representation through Combination with Recover Capability," *13th ACS/IEEE Int. Conf. Comput. Syst. Appl. AICCSA*, pp. 1–5, 2016.
- [3] C. Li, R. Yang, Z. Liu, J. Li, and Z. Guo, "Semi-fragile self-recoverable watermarking scheme for face image protection," *Comput. Electr. Eng.*, no. February, 2015.
- [4] H. Rhayma, A. Makhloufi, and A. Ben Hmida, "Self-Authentication Scheme Based on Semi-Fragile Watermarking and Perceptual Hash Function," pp. 1–6, 2014.
- [5] P. P. Augustus and D. S. George, "Watermarking Technique for Self Authentication and Recovery," pp. 157–165, 2013.
- [6] J. Said, R. Souissi, and H. Hamam, "A New Representation of Image Through Numbering Pixel Combinations," *J. Inf. Secur. Res.*, vol. 4, no. 1, 2013.
- [7] S. Kiatpapan and T. Kondo, 2015, "An image tamper detection and recovery method based on self-embedding dual watermarking," 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Hua Hin, pp. 1-6.
- [8] Valeriu Codreanu, Feng Dongy, et al., 2013, "GPU-ASIFT: A Fast Fully Affine-Invariant Feature Extraction Algorithm" 978-1-4799-0838-7/13©2013 IEEE pp. 474-481.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details