IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 11, Issue 6, June 2024

DOI: 10.17148/IARJSET.2024.11643

Sales Forecasting and Inventory Optimization at Big Mart using Time Series Technique: A Survey

Anubhav Sanket[1], Shipra Sinha[2], Yamini Mandagiri[3], Dr. Shivaprasad Ashok Chikop[4]

Student, Dept. of AIML, Dayananda Sagar Academy of Technology and Management, Bengaluru, India[1]

Student, Dept. of AIML, Dayananda Sagar Academy of Technology and Management, Bengaluru, India [2]

Student, Dept. of AIML, Dayananda Sagar Academy of Technology and Management, Bengaluru, India[3]

Assistant Professor, Dept. of AIML, Dayananda Sagar Academy of Technology and Management, Bengaluru, India[4]

Abstract: This study presents a feature-based approach for predicting and optimizing stock levels in a retail chain. Utilizing an extensive database of algorithms, interpretable features are extracted from historical stock level time series data. The method, which includes correlation structure, distribution, entropy, stationarity, and scaling properties, reduces dimensionality significantly. Through a forward feature selection process and a linear classifier, the approach autonomously learns distinctions between stock level classes, outperforming conventional classifiers. The selected features not only enhance classification accuracy but also provide crucial insights for optimizing inventory management in diverse retail locations. This research contributes a scalable and interpretable solution to the dynamic challenges of stock prediction in a retail setting.

Keywords: Time Series Analysis, Forecasting, Machine Learning, Big mart sales.

I. INTRODUCTION

In the dynamic landscape of retail, the ability to predict and optimize stock levels stands as a pivotal factor influencing operational efficiency, customer satisfaction, and ultimately, the bottom line. Accurate forecasting empowers supply chain management, enabling proactive decision-making to meet demand fluctuations effectively. Conversely, inadequate predictions may lead to product backlogs, inventory shortages, and dissatisfied customers, underscoring the critical need for precise stock level forecasting models.

Prior research in sales forecasting has demonstrated the efficacy of various methodologies across diverse domains. For instance, computational intelligence methods have been applied to predict book sales in editorial business management [1], while clustering-based forecasting schemes utilizing extreme learning machines have shown promise [2]. Hybrid approaches, such as combining seasonal autoregressive integrated moving averages and quantile regression, have been proposed for food sales forecasting [3]. As the retail landscape continues to evolve, the demand for tailored forecasting models remains imperative.

This literature paper addresses the unique challenges of predicting stock levels for different items across various stores within a retail chain. The aim is to develop a robust forecasting framework capable of adapting to the intricacies of individual products and locations. Leveraging a case study approach, our research focuses on implementing forecasting models using a dataset obtained from Big

Mart, covering a span from the past few years. The dataset's distinctive feature lies in its comprehensive representation of the entire retail market since the past few years, offering a rich source for exploration. With each time series comprising approximately 5 monthly samples, our approach incorporates automatic learning algorithms for the implementation of Nonlinear Auto-Regressive (NAR) forecasting models. The significance of this research extends beyond prediction accuracy; it delves into providing actionable insights that can guide strategic inventory management decisions in the intricate realm of retail.

A. Background

In the bustling retail sector, effective sales forecasting and inventory optimization stand as linchpins for sustained success. The intricate dance of supply and demand within the retail landscape necessitates precise predictions of future sales, allowing businesses to strategically manage inventory, minimize stockouts, and meet customer expectations seamlessly.

A failure in this predictive orchestration can lead to a cascade of challenges, from overstocking and excess inventory costs to the dissatisfaction of customers facing unmet demands. Big Mart, as a prominent player in the retail industry, recognizes the critical importance of mastering the delicate balance between sales forecasting and inventory management. As consumer preferences shift, market trends evolve, and external factors influence purchasing behaviour, the need for a sophisticated and adaptive forecasting framework becomes increasingly apparent.

The advent of time series techniques has ushered in a new era of forecasting capabilities, offering nuanced insights into temporal patterns and dependencies within sales data. Leveraging the power of time series analysis, Big Mart aims to enhance its forecasting accuracy and operational efficiency. This endeavour seeks to not only predict sales trends but also optimize inventory levels across the myriad of products and diverse stores that characterize the expansive reach of Big Mart.

The choice of time series techniques aligns with the inherent temporal nature of retail sales data. The methodology embraces the dynamism of sales patterns, capturing seasonality, trends, and other time-dependent factors that contribute to the ebb and flow of consumer demand.

As Big Mart embarks on this journey of Sales Forecasting and Inventory Optimization, the objective is twofold: to bolster the resilience of the supply chain by anticipating demand with precision and to streamline inventory management, ensuring that each store is stocked with the right products at the right time. Through the lens of time series techniques, this project endeavours to harness the potential for data-driven decision-making, steering Big Mart towards a future where sales predictions and inventory optimization become synonymous with operational excellence.

II. LITERATURE REVIEW

Sales forecasting and inventory optimization are critical components of successful retail operations, influencing supply chain efficiency, customer satisfaction, and overall profitability. A review of the literature reveals a myriad of approaches and methodologies aimed at addressing the complexities inherent in predicting consumer demand and optimizing inventory levels, with a particular focus on time series techniques.

In the realm of sales forecasting, computational intelligence methods have garnered attention for their ability to handle complex, non- linear relationships within retail datasets. Noteworthy work by [1] applied computational intelligence to predict book sales in an editorial business management environment. This approach highlights the adaptability of intelligent algorithms in capturing intricate patterns within sales data, providing a foundation for exploration in the retail sector.

The integration of clustering techniques with forecasting schemes has demonstrated promising results. [2] proposed a clustering-based sales forecasting scheme using extreme learning machines and ensemble linkage methods. This methodology, leveraging the power of clustering, offers a means to discern patterns and relationships within diverse product categories and store locations, addressing the challenge of heterogeneity in retail datasets. Hybrid Approaches for Diverse Product and Store Dynamics:

The dynamic nature of retail, characterized by diverse products and numerous store locations, necessitates specialized forecasting models. Hybrid approaches, such as combining seasonal autoregressive integrated moving averages (SARIMA) and quantile regression, have shown success in addressing this diversity [3]. By integrating multiple forecasting techniques, these hybrid models aim to provide granular insights tailored to different items and locations within the retail landscape.

Time series techniques have proven effective in addressing temporal challenges and seasonality inherent in retail sales data. The work of[4] on forecasting German car sales using Google data and multivariate models exemplifies the utilization of time series analysis to capture temporal dependencies. This approach recognizes the significance of accounting for seasonality and evolving trends, vital for accurate sales predictions in a rapidly changing retail environment.

Beyond sales forecasting, the literature emphasizes the critical role of inventory control and optimization. [5] conducted a simulation study concerning demand forecasting and inventory control of automotive spare parts, shedding light on the interplay between accurate predictions and effective inventory management. The study underscores the importance of aligning inventory levels with forecasted demand to prevent operational inefficiencies.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 11, Issue 6, June 2024

DOI: 10.17148/IARJSET.2024.11643

In specific retail domains, adaptive network-based fuzzy inference systems have shown promise. [6] proposed an adaptive network- based fuzzy inference system for forecasting automobile sales, emphasizing the adaptability of fuzzy logic in capturing and interpreting complex relationships within sales data. This approach provides insights into the potential of fuzzy systems for nuanced forecasting in the automotive retail sector.

Despite the progress in sales forecasting and inventory optimization methodologies, there remains a gap in applying these techniques to the specific challenges faced by large-scale retailers, such as Big Mart.

The proposed project aims to contribute to the literature by implementing time series techniques in the context of predicting stock levels for different items across diverse stores. Leveraging a case study approach using data from the Census Bureau of the U.S., the research endeavours to provide practical insights for enhancing forecasting accuracy, inventory optimization, and decision-making within the dynamic landscape of retail.

In summary, the literature review underscores the diversity of approaches in sales forecasting and inventory optimization, emphasizing the need for specialized methodologies in the retail sector. The proposed project builds upon these foundations, aiming to advance the state-of-the-art by applying time series techniques to the specific challenges faced by Big Mart in managing stock levels across its extensive retail network. The common methodology most of the researches go through are:

A. Data Collection

The dataset was sourced from Kaggle.com for this project. The dataset consists of both a training dataset, comprising 8,000 entries, and a test dataset with 5,000 entries. Fig.1 illustrates the training data, while Fig.2 displays a sample from the test dataset.

B. Data Processing

The Kaggle-sourced dataset, integral to the project's objectives of Sales Forecasting and Inventory Optimization at Big Mart, undergoes a systematic data processing pipeline to ensure its readiness for modelling and analysis. The following steps outline the key components of this data processing endeavour:

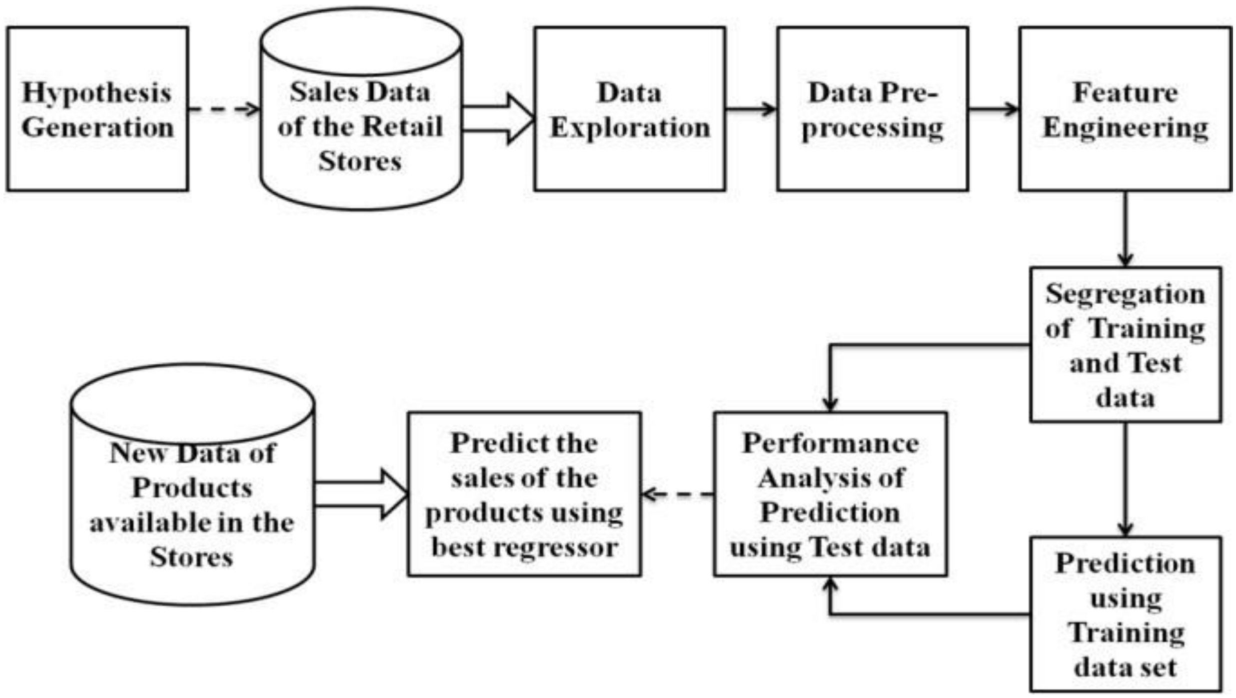

Fig.1 Data Processing [4]

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁓ Peer-reviewed & Refereed journal ⁓ Vol. 11, Issue 6, June 2024

DOI: 10.17148/IARJSET.2024.11643

Step 1. Data Exploration and Descriptive Statistics

Conduct exploratory data analysis (EDA) to gain a comprehensive understanding of the dataset's structure, including feature distributions, data types, and any evident anomalies. Generate descriptive statistics to obtain insights into the central tendencies and variability of the numerical features.

Step 2. Handling Missing Values

Identify and address missing values within the dataset, employing strategies such as imputation or removal based on the extent and nature of missing data. Document the rationale behind chosen imputation methods and communicate any potential impact on subsequent analyses.

Step 3. Feature Engineering

Engineer relevant features conducive to forecasting models, including temporal features, categorical encodings, and aggregations of historical sales data. Transform categorical variables into a format suitable for model training.

Step 4. Time Series Decomposition

Decompose time series features into components like trend, seasonality, and residual using methods such as seasonal-trend decomposition using LOESS (STL). Visualize the decomposed time series components to validate the separation of patterns.

Step 5. Data Splitting

Split the dataset into training and testing sets, preserving the temporal order of entries. Reserve a portion for testing to evaluate model performance on unseen data.

Step 6. Scaling and Normalization

Apply scaling or normalization to numerical features to prevent biases in model training. Common techniques include Min-Max scaling or Z-score normalization.

Step 7. Handling Categorical Variables

Employ appropriate techniques for handling categorical variables, such as one-hot encoding or label encoding, ensuring compatibility with machine learning models.

Step 8. Handling Outliers

Identify and address outliers within numerical features using statistical methodsor visualization techniques. Consider the impact of outliers on model performance.

Step 9. Multivariate Time Series Processing

If the dataset contains multivariate time series (e.g., sales data for different product categories), implement approaches like panel data processing to capture dependencies between variables.

Step 10. Data Format for Time Series Models

Structure the dataset in a time series-friendly format, organizing the data into sequences of time steps with corresponding target values for training forecasting models.

Step 11. Validation and Documentation:

Validate the processed dataset's integrity through thorough checks, ensuring that it aligns with the project's objectives. Document all data processing steps, transformations, and considerations to facilitate reproducibility and transparency.

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ※ Peer-reviewed & Refereed journal ※ Vol. 11, Issue 6, June 2024

DOI: 10.17148/IARJSET.2024.11643

The processed dataset, refined through these steps, lays the foundation for the subsequent stages of the project, encompassing the development, training, and evaluation of sales forecasting and inventory optimization models. The meticulous data processing efforts aim to maximize the dataset's utility, enabling meaningful insights and actionable recommendations for Big Mart's retail operations.

TABLE 1: Attributes Information

Attribute	Description
Item_Identifer	It is the unique product Id number.
Item Weight	It will include the product's weight.
Item_Fat_Content	It will mean whether the item is low in fat or not.
Item -Visibility	The percentage of the overall viewing area assigned to the particular item from all items in the shop.
Item -Type	To which group does the commodity belong
Item-MRP	The product's price list
Outlet-Identifier	a distinct slot number
Outlet-Establishment Year	The year that the shop first opened its doors.
Outlet-Size	The sum of total area occupied by a supermarket.
Outlet-Location	The kind of town where the store is situated.
Outlet-Type	The shop is merely a supermarket or a grocery store.
Item-Outlet-Sales	The item's sales in the original shop

Table 1: Attributes Information [2]

Train data set

	A	B	C	D	E	F	G	H	I	J	K	L
1	Item_Iden	Item_Wei	Item_Fat_	Item_Visil	Item_Typ	Item_MRF	Outlet_Id	Outlet_Es	Outlet_Siz	Outlet_Lo	Outlet_Ty	Item_Outlet_Sales
2	FDA15	9.3	Low Fat	0.016047	Dairy	249.8092	OUT049	1999	Medium	Tier 1	Supermar	3735.138
3	DRC01	5.92	Regular	0.019278	Soft Drink	48.2692	OUT018	2009	Medium	Tier 3	Supermar	443.4228
4	FDN15	17.5	Low Fat	0.01676	Meat	141.618	OUT049	1999	Medium	Tier 1	Supermar	2097.27
5	FDX07	19.2	Regular	0	Fruits and	182.095	OUT010	1998		Tier 3	Grocery St	732.38
6	NCD19	8.93	Low Fat	0	Househol	53.8614	OUT013	1987	High	Tier 3	Supermar	994.7052
7	FDP36	10.395	Regular	0	Baking Go	51.4008	OUT018	2009	Medium	Tier 3	Supermar	556.6088
8	FDO10	13.65	Regular	0.012741	Snack Foo	57.6588	OUT013	1987	High	Tier 3	Supermar	343.5528
9	FDP10		Low Fat	0.12747	Snack Foo	107.7622	OUT027	1985	Medium	Tier 3	Supermar	4022.764
10	FDH17	16.2	Regular	0.016687	Frozen Fo	96.9726	OUT045	2002		Tier 2	Supermar	1076.599
11	FDU28	19.2	Regular	0.09445	Frozen Fo	187.8214	OUT017	2007		Tier 2	Supermar	4710.535
12	FDY07	11.8	Low Fat	0	Fruits and	45.5402	OUT049	1999	Medium	Tier 1	Supermar	1516.027
13	FDA03	18.5	Regular	0.045464	Dairy	144.1102	OUT046	1997	Small	Tier 1	Supermar	2187.153
14	FDX32	15.1	Regular	0.100014	Fruits and	145.4786	OUT049	1999	Medium	Tier 1	Supermar	1589.265
15	FDS46	17.6	Regular	0.047257	Snack Foo	119.6782	OUT046	1997	Small	Tier 1	Supermar	2145.208
16	FDF32	16.35	Low Fat	0.068024	Fruits and	196.4426	OUT013	1987	High	Tier 3	Supermar	1977.426
17	FDP49	9	Regular	0.069089	Breakfast	56.3614	OUT046	1997	Small	Tier 1	Supermar	1547.319
18	NCB42	11.8	Low Fat	0.008596	Health an	115.3492	OUT018	2009	Medium	Tier 3	Supermar	1621.889
19	FDP49	9	Regular	0.069196	Breakfast	54.3614	OUT049	1999	Medium	Tier 1	Supermar	718.3982
20	DRI11		Low Fat	0.034238	Hard Drin	113.2834	OUT027	1985	Medium	Tier 3	Supermar	2303.668
21	FDU02	13.35	Low Fat	0.102492	Dairy	230.5352	OUT035	2004	Small	Tier 2	Supermar	2748.422
22	FDN22	18.85	Regular	0.13819	Snack Foo	250.8724	OUT013	1987	High	Tier 3	Supermar	3775.086
23	FDW12		Regular	0.0354	Baking Go	144.5444	OUT027	1985	Medium	Tier 3	Supermar	4064.043

Table 2: Train Dataset [2]

Test dataset

	A	B	C	D	E	F	G	H	I	J	K
1	Item_Iden	Item_Wei	Item_Fat_	Item_Visil	Item_Typ	Item_MRF	Outlet_Id	Outlet_Es	Outlet_Siz	Outlet_Lo	Outlet_Type
2	FDW58	20.75	Low Fat	0.007565	Snack Foo	107.8622	OUT049	1999	Medium	Tier 1	Supermarket Type1
3	FDW14	8.3	reg	0.038428	Dairy	87.3198	OUT017	2007		Tier 2	Supermarket Type1
4	NCN55	14.6	Low Fat	0.099575	Others	241.7538	OUT010	1998		Tier 3	Grocery Store
5	FDQ58	7.315	Low Fat	0.015388	Snack Foo	155.034	OUT017	2007		Tier 2	Supermarket Type1
6	FDY38		Regular	0.118599	Dairy	234.23	OUT027	1985	Medium	Tier 3	Supermarket Type3
7	FDH56	9.8	Regular	0.063817	Fruits and	117.1492	OUT046	1997	Small	Tier 1	Supermarket Type1
8	FDL48	19.35	Regular	0.082602	Baking Go	50.1034	OUT018	2009	Medium	Tier 3	Supermarket Type2
9	FDC48		Low Fat	0.015782	Baking Go	81.0592	OUT027	1985	Medium	Tier 3	Supermarket Type3
10	FDN33	6.305	Regular	0.123365	Snack Foo	95.7436	OUT045	2002		Tier 2	Supermarket Type1
11	FDA36	5.985	Low Fat	0.005698	Baking Go	186.8924	OUT017	2007		Tier 2	Supermarket Type1
12	FDT44	16.6	Low Fat	0.103569	Fruits and	118.3466	OUT017	2007		Tier 2	Supermarket Type1
13	FDQ56	6.59	Low Fat	0.105811	Fruits and	85.3908	OUT045	2002		Tier 2	Supermarket Type1
14	NCC54		Low Fat	0.171079	Health an	240.4196	OUT019	1985	Small	Tier 1	Grocery Store
15	FDU11	4.785	Low Fat	0.092738	Breads	122.3098	OUT049	1999	Medium	Tier 1	Supermarket Type1
16	DRL59	16.75	LF	0.021206	Hard Drin	52.0298	OUT013	1987	High	Tier 3	Supermarket Type1
17	FDM24	6.135	Regular	0.079451	Baking Go	151.6366	OUT049	1999	Medium	Tier 1	Supermarket Type1
18	FDI57	19.85	Low Fat	0.054135	Seafood	198.7768	OUT045	2002		Tier 2	Supermarket Type1
19	DRC12	17.85	Low Fat	0.037981	Soft Drink	192.2188	OUT018	2009	Medium	Tier 3	Supermarket Type2
20	NCM42		Low Fat	0.028184	Househol	109.6912	OUT027	1985	Medium	Tier 3	Supermarket Type3
21	FDA46	13.6	Low Fat	0.196898	Snack Foo	193.7136	OUT010	1998		Tier 3	Grocery Store
22	FDA31	7.1	Low Fat	0.10992	Fruits and	175.008	OUT013	1987	High	Tier 3	Supermarket Type1
23	NCJ31	19.2	Low Fat	0.182619	Others	239.9196	OUT035	2004	Small	Tier 2	Supermarket Type1

Table 3: Test Dataset [2]

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁓ Peer-reviewed & Refereed journal ⁓ Vol. 11, Issue 6, June 2024

DOI: 10.17148/IARJSET.2024.11643

C. The Proposed Architecture Diagram

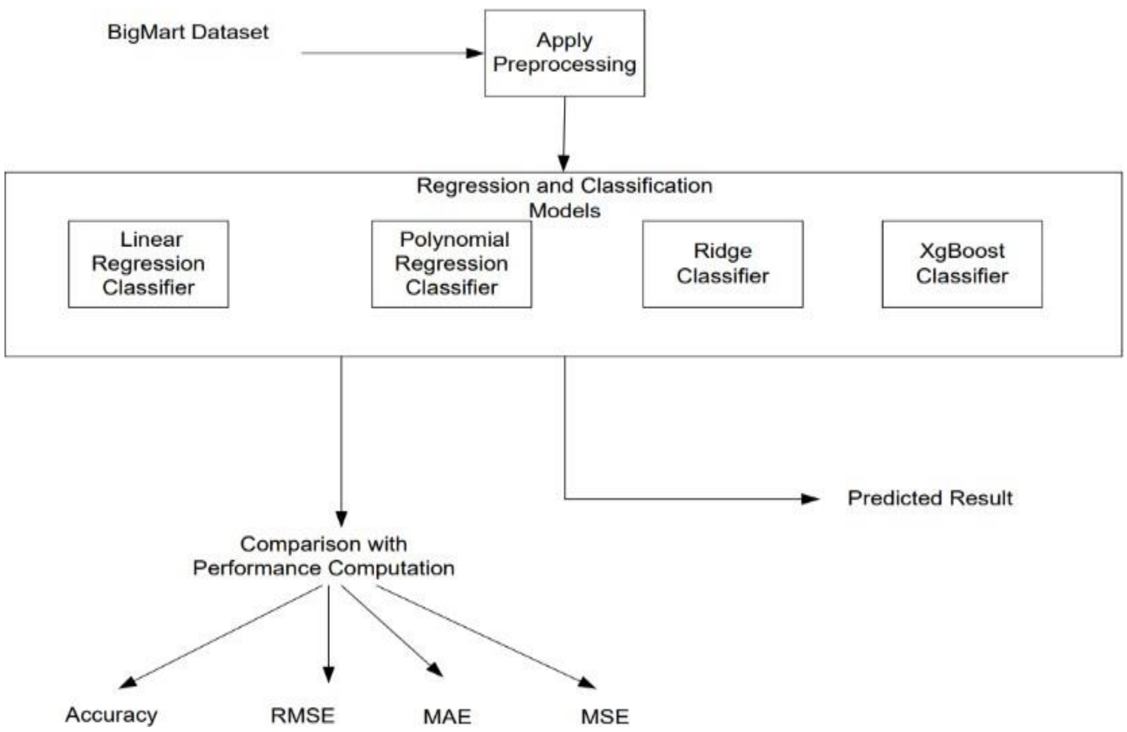

Fig.4 The Proposed Architecture Diagram [2]

1) Linear Regression

Linear regression, a core statistical method, models the connection between a dependent variable and one or more independent variables. The process commences with a visual examination of data patterns, aiming to discern linearity or non-linearity and identify outliers. Linear regression is considered for linear patterns, while non-linear patterns may prompt data transformation. The treatment of outliers depends on non-statistical justifications.

Next, data is linked to the least squares line, minimizing the sum of squared differences between observed and predicted values. Model assumptions are checked through residual and normal probability plots. Data transformation is employed when assumptions are unmet.

If needed, data undergoes transformation to align with linear regression assumptions. Common transformations include logarithmic or square root adjustments. A new regression line is constructed post-transformation, and the process iterates until a satisfactory model fit is attained.

Upon achieving a "good-fit" model, the least-squares regression line equation is established

Y=o1x1+ o2x2+⋯…… onxn

This equation encapsulates the relationship between the dependent variable (Y) and independent variables (x1, x2, ..., xn). The model summary incorporates relevant statistics, such as coefficient estimates, standard errors, and R-squared, which gauges the proportion of variance explained by the independent variables. This streamlined approach ensures a comprehensive analysis, leading to the development of an accurate linear regression model.This can be expressed mathematically as

$$R - Square = 1 - \frac{\sum(Y_{actual} - Y_{predicted})^2}{\sum(Y_{actual} - Y_{mean})^2} \quad x$$

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 11, Issue 6, June 2024

DOI: 10.17148/IARJSET.2024.11643

2) Polynomial Regression Algorithm

Polynomial Regression, a vital tool in our sales forecasting and inventory optimization project at Big Mart, models the intricate relationship between dependent (y) and independent (x) variables by incorporating polynomial terms. The equation captures non-linear patterns crucial in understanding the diverse dynamics of retail data.

y= b0+b1x1+ b2x1 2+ bnx1

Referred to as an exceptional case of multiple linear regression in machine learning, Polynomial Regression transforms the conventional model by adding polynomial terms. This adaptation is essential for improving accuracy, especially in scenarios where non- linear relationships prevail, as is often the case in our retail dataset. The algorithm excels in handling non-linear datasets, aligning perfectly with the nature of our project. It utilizes a linear regression foundation to accommodate complex functions, making it a suitable choice for our objective of forecasting sales and optimizing inventory. In a concise manner, Polynomial Regression proves to be a powerful tool, enhancing our ability to model and understand the nuanced patterns within the dynamic retail landscape at Big Mart.

3) Ridge Regression

In the realm of our sales forecasting and inventory optimization project at Big Mart, Ridge Regression emerges as a valuable model tuning tool. Specifically designed to address multicollinearity issues within the data, this method employs L2 regularization to enhance the stability of our predictive models.Ridge regression becomes indispensable when multicollinearity disrupts the traditional least squares approach, introducing bias and high variances. In such scenarios, where the expected values deviate significantly from actual values, Ridge Regression steps in to mitigate these challenges. The cost function for Ridge Regression is expressed as:

Min (||Y – X(theta)||^2 + λ||theta||^2)

Here, Y represents the dependent variable, X the independent variables, theta the coefficients, and λ the regularization parameter.

4) XGBoost Regression

In our sales forecasting and inventory optimization project at Big Mart, Extreme Gradient Boosting (XGBoost) proves to be a highly efficient enhancement to traditional gradient boosting systems. With both a linear model solver and a tree algorithm, "xgboost" outperforms current gradient boosting implementations, running multiple times faster. Its versatility extends to supporting various target functions, including regression, classification, and ranking. Despite its computational intensity, XGBoost is well-suited for competitive scenarios, offering quick and accurate predictions. Beyond speed, it provides features for cross-validation and identifying significant factors, enhancing its applicability in our project. XGBoost's accelerated computation speed and versatile capabilities make it a valuable asset for navigating the complexities of sales forecasting and inventory optimization at Big Mart.

III. CONCLUSION

In this study, we have delved into the effectiveness of various algorithms in predicting sales based on revenue and reviews. Our focus was on identifying the best-performing algorithm, and through rigorous analysis, we propose the utilization of a software employing regression approaches. Specifically, we explored the impact of linear regression, polynomial regression, ridge regression, and XGBoost regression on sales prediction accuracy.

Our findings underscore the efficacy of ridge and XGBoost regression in providing superior predictions compared to linear and polynomial regression. Notably, these algorithms outperform in terms of Accuracy, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). This suggests that the utilization of ridge and XGBoost regression methods can significantly enhance the precision of sales predictions, offering a more robust foundation for decision making in the realm of revenue forecasting.

Looking ahead, our study hints at the potential of leveraging sales forecasts to proactively manage cash flow, production, staffing, and financing requirements. This proactive approach can be instrumental in navigating unforeseen challenges and optimizing resource allocation. Additionally, our future work could explore the integration of the ARIMA model, offering insights into the time series aspects of sales data through graphical representation.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 6, June 2024

DOI: 10.17148/IARJSET.2024.11643

This holistic approach could further refine our understanding of sales dynamics, paving the way for more sophisticated and accurate forecasting models.

In conclusion, our investigation has illuminated the path toward enhanced sales prediction through the strategic application of regression algorithms. The superiority of ridge and XGBoost regression positions them as key players in optimizing sales forecasts, setting the stage for more informed and effective decision- making in the dynamic landscape of revenue and review-based sales predictions.

REFERENCES

[1]. Ben D. Fulcher and Nick S. Jones-'Highly Comparative Feature-Based Time-Series Classification', IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 12, DECEMBER 2014

[2]. Ranjitha P, Spandana M, Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms, Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021) IEEE Xplore Part Number: CFP21K74-ART; ISBN: 978-0-7381-1327-2

[3]. Ching-Seh (Mike) Wu, Pratik Patil, Saravana Gunaseelan, Comparison of Different Machine Learning Algorithms for Multiple Regression on Black Friday Sales Data, ©2018 IEEE/978-1-5386-6565-7118/$31.00

[4]. Akshay Krishna, Akhilesh V, Animikh Aich, Chetana Hegde, Sales-forecasting of Retail Stores using Machine Learning Techniques, 3rd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2018

[5]. Wang, Haoxiang. "Sustainable development and management in consumer electronics using soft computation." Journal of Soft Computing Paradigm (JSCP) 1, no. 01 (2019): 56.- 2. Suma, V., and Shavige Malleshwara Hills. "Data Mining based Prediction of D

[6]. Suma, V., and Shavige Malleshwara Hills. "Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics." Journal of Soft Computing Paradigm (JSCP) 2, no. 02 (2020): 101- 110

[7]. Giuseppe Nunnari, Valeria Nunnari, "Forecasting Monthly Sales Retail Time Series: A Case Study", Proc. of IEEE Conf. on Business Informatics (CBI), July 2017.

[8]. Zone-Ching Lin, Wen-Jang Wu, "Multiple LinearRegression Analysis of the Overlay Accuracy Model Zone", IEEE Trans. on Semiconductor Manufacturing, vol. 12, no. 2, pp. 229 – 237, May 1999.

[9]. O. Ajao Isaac, A. Abdullahi Adedeji, I. Raji Ismail, "Polynomial Regression Model of Making Cost Prediction In Mixed Cost Analysis", Int. Journal on Mathematical Theory and Modeling, vol. 2, no. 2, pp. 14 – 23, 2012.

[10]. C. Saunders, A. Gammerman and V. Vovk, "Ridge Regression Learning Algorithm in Dual Variables", Proc. of Int. Conf. on Machine Learning, pp. 515 – 521, July 1998.IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 56, NO. 7, JULY 2010 3561.

[11]. "Robust Regression and Lasso". Huan Xu, Constantine Caramanis, Member, IEEE, and Shie Mannor, Senior Member, IEEE. 2015 International Conference on Industrial Informatics- Computing Technology, Intelligent Technology, Industrial Information Integration."An improved Adaboost algorithm based

[12]. on uncertain functions".Shu Xinqing School of Automation Wuhan University of Technology. Wuhan, China Wang Pan School of the Automation Wuhan University of Technology Wuhan, China.

[13]. Xinqing Shu, Pan Wang, "An Improved Adaboost Algorithm based on Uncertain Functions", Proc. of Int. Conf. on Industrial Informatics – Computing Technology, Intelligent Technology, Industrial Information Integration, Dec. 2015.

[14]. Søren Kejser Jensen, Torben Bach Pedersen, Senior Member, IEEE, and Christian Thomsen, Time Series Management Systems: A Survey, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 29, NO. 11, NOVEMBER 2017

[15]. N. S. Arunraj, D. Ahrens, A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting, Int. J. Production Economics 170 (2015) 321- 335P

[16]. D. Fantazzini, Z. Toktamysova, Forecasting German car sales using Google data and multivariate models, Int. J. Production Economics 170 (2015) 97-135.

[17]. X. Yua, Z. Qi, Y. Zhao, Support Vector Regression for Newspaper/Magazine Sales Forecasting, Procedia Computer Science 17 (2013) 1055–1062.

[18]. E. Hadavandi, H. Shavandi, A. Ghanbari, An improved sales forecasting approach by the integration of genetic fuzzy systems and data clustering: a Case study of the printed circuit board, Expert Systems with Applications 38 (2011) 9392–9399.

[19]. P. A. Castillo, A. Mora, H. Faris, J.J. Merelo, P. GarciaSanchez, A.J. Fernandez-Ares, P. De las Cuevas, M.I. Garcia-Arenas, applying computational intelligence methods for predicting the sales of newly published books in a real editorial business management environment, Knowledge-Based Systems 115 (2017) 133-151.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 11, Issue 6, June 2024

DOI: 10.17148/IARJSET.2024.11643

[20]. https://www.kaggle.com/brijbhushannanda1979/bigmartsalesdata. [Accessed: Jun. 28, 2018.

[21]. Pei Chann Chang and Yen-Wen Wang, "Fuzzy Delphi and back propagation model for sales forecasting in PCB industry", Expert systems with applications, vol. 30, pp. 715-726, 2006.

[22]. R. J. Kuo, Tung Lai HU and Zhen Yao Chen "application of radial basis function neural networks for sales forecasting", Proc. of Int. Asian Conference on Informatics in control, automation, and robotics, pp. 325- 328, 2009.

[23]. R. Majhi, G. Panda, G. Sahoo, and A. Panda, "On the development of Improved Adaptive Models for Efficient Prediction of Stock Indices using Clonal-PSO (CPSO) and PSO Techniques", International Journal of Business Forecasting and Market Intelligence, vol. 1, no. 1, pp.50- 67, 2008.

[24]. R. Majhi, G. Panda, B. Majhi and G. Sahoo, "Efficient prediction of stock market indices using adaptive bacterial foraging optimization (ABFO) and BFO based techniques", Expert Systems with Applications, Elsevier, vol. 36, issue 6, pp. 10097-10104, August 2009.

[25]. K. Chakrabarti, E. Keogh, S. Mehrotra, and M. Pazzani, "Locally adaptive dimensionality reduction for indexing large time series databases," ACM Trans. Database Syst., vol. 27, pp. 188– 228, 2002.

[26]. J. Shieh and E. Keogh, "iSAX: Indexing and mining terabyte sized time series," in Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2008, pp. 623–631.

[27]. J. Lin, E. Keogh, L. Wei, and S. Lonardi, "Experiencing SAX: A novel symbolic representation of time series," Data Mining Knowl. Discovery, vol. 15, no. 2, pp. 107–144, 2007.

[28]. H. Kantz and T. Schreiber, Nonlinear Time Series Analysis., 2nd ed. Cambridge, U.K.: Cambridge Unive. Press, 2004.

[29]. G. E. Batista, X. Wang, and E. J. Keogh, "A complexity- invariant distance measure for time series," in Proc. SIAM Int. Conf. Data Mining, vol. 31, 2011, pp. 699–710.

[30]. Manasvi Jaipurkar,Neha Ragit, Chetana Tambuskar, Pranay Deepak Saraf IPL Data Analysis and Visualization Using Microsoft Power BI Tool, 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)