Sales Prediction using ML algorithms(random forest) and Time Series(ETS-model)

SEMINAR-1 REPORT

Submitted by

SATYA PRAKASH(RA2111027020045)

TEJA (RA2111027020058)

GOVIND RAO (RA2111027020035)

Under the guidance of

Dr. R.Pavithra Guru, M.E., Ph.D.,

Ms. M.Thamizharasi, M.E.

(Assistant Professors, Department of Computer Science and Engineering)

In partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

with specialization in

with specialization in Big Data Analytics.

of

COLLEGE OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

RAMAPURAM, CHENNAI-600089.

NOVEMBER 2023

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Deemed to be University Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that the Seminar-I report titled "**SALES PREDICTION USING ML ALGORITHMS (RANDOM FOREST) AND TIME SERIES(ETS-MODEL)**" is the bonafide work of "**SATYA PRAKASH [RA2111027020045], TEJA [RA2111027020058] , GOVIND RAO [RA2111027020035]**" submitted for the course 18CSP103L Seminar – I. This report is a record of successful completion of the specified course evaluated based on literature reviews and the supervisor. No part of the Seminar Report has been submitted for any degree, diploma, title, or recognition before.

SIGNATURE

Dr. R. Pavithra Guru, M.E., Ph.D.,
Assistant Professor
Dept. of Computer Science & Engineering
SRM Institute of Science and Technology
Ramapuram, Chennai.

SIGNATURE

Dr. K. RAJA, M.E., Ph.D.,
Professor and Head
Dept. of Computer Science & Engineering
SRM Institute of Science and Technology
Ramapuram, Chennai.

Submitted for the Seminar-1 Viva Voce Examination held on.....at SRM Institute of Science and Technology, Ramapuram, Chennai-600089.

EXAMINER 1

EXAMINER 2

ABSTRACT

The most common issue sellers face in today's day-to-day life with respect to customers is being unable to utilize the feedback and suggestion data in an effective way. Our project will analyze these data in a systematic manner, which helps sellers make better decisions for better revenue or margin. The ETS model and Random forest model will analyze the previous trends and help sellers in many ways, such as by identifying the top performing categories of products and stocking up on them well in advance to meet supply and demand, investing the funds, and taking the right appropriate decisions in terms of sales and marketing, further increasing sales growth and profit margins at a segment or category level.The core of the project lies in predictive modeling ,where the ETS model focuses on time series data analysis, uncovering trends, and forecasting future demand. The Random Forest model, on the other hand, employs ensemble learning to provide robust and adaptable predictions based on historical patterns. These models, coupled with sophisticated data analysis techniques, enable the extraction of actionable insights from a sea of customer feedback.The results and discussions highlight the tangible benefits of the ML Model project. It empowers sellers to improve product and service quality, anticipate market trends, and enhance the customer experience. The decision support system integrated within the project serves as a guiding compass for sellers, enabling them to allocate resources efficiently and optimize profitability at a segment or category level.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ANCRONYMS AND ABBREVATIONS

OS - OPERATING SYSTEMS

DBMS - DATABASE MANAGEMENT SYSTEM

ML - MACHINE LEARNING

SQL - STRUCTURED QUERY LANGUAGE

DDL - DATA DEFINITION LANGUAGE

DML - DATA MANIPULATION LANGUAGE

HTML - HYPER TEXT MARKUP LANGUAGE

CSS - CASCADING STYLE SHEETS

UI - USER INTERFACE

UX - USER EXPERIENCE

Chapter 1

INTRODUCTION

The most common issue sellers face in today's day-to-day life with respect to customers is being unable to utilize the feedback and suggestion data in an effective way. OPR will analyze these data in a systematic manner, which helps sellers make better decisions for better revenue or margin. The ETS model and Random forest model will analyze the previous trends and help sellers in many ways, such as by identifying the top performing categories of products and stocking up on them well in advance to meet supply and demand, investing the funds, and taking the right appropriate decisions in terms of sales and marketing, further increasing sales growth and profit margins at a segment or category level.

1.1 Objective of the Project

The objective for this Forecasting sales is the process of predicting future sales based on past performance. Sales forecasting is critical for businesses that are entering new markets, providing new services or products, or expanding rapidly. The common problem that retailers face in day-to-day life is optimizing sales by investing in the right product or category at the right time. It is difficult to decide on the investment without any proper analysis. At times, there will be high returns from a few products, the demand for a few products will be high compared to other products, etc…..For example, if there is a cloth retail shop, there will be more demand for ethnic wear and other specific types of clothes during festival season. So, retailers must wisely make the investment to maximize the sales and stock up on products to meet the demand. Sometimes Amazon, Flipkart, and some other e-commerce websites announce year-end or festival sales; during this time, demand for some of the product's changes . during summer sales most people try to buy air conditioners, refrigerators, and other electronic products that are necessary; likewise, based on the time and location, a few products might be in high demand.

1.2 Problem Statement

1. Consumer specific music cue: -

Our ml model model will tackle these challenges by identifying the categories and segments that have performed well based on previous trends and helping sellers invest in and stock up more of the top-performing products for next time .By analysing the customer reviews, categories, segments, and some other information, ETS model can increase the sales growth and profit margin. It helps to get the stock in place to meet the demand.

The primary objective of this study is to develop a robust and accurate sales prediction using random forest algorithm using machine learning technique and time series analysis using error trend seasonality model:

1. **Data Collection**: Gather a diverse dataset consisting of multi-view data sources, including sales of all stores of near of our client.

2. **Data Preprocessing and Feature Extraction**: Implement MVDP techniques to process and extract relevant features from the multi-view data sources. This involves addressing data integration, noise reduction, feature selection, and dimensionality reduction.

3. **Machine Learning Classification Algorithms**: Apply a range of ML classification algorithms, such as Support Vector Machines, Random Forest, Logistic Regression, and Deep Learning models, to the preprocessed data to build predictive models.

4. **Model Training and Evaluation**: Train the ML models using appropriate training and testing datasets. Evaluate the models' performance using relevant metrics, such as accuracy, sensitivity, specificity, and ROC curves.

5. **Model Comparison and Selection**: Compare the performance of different ML algorithms and identify the most effective model for Parkinson's disease detection.

6. **Cross-Validation and Generalization**: Ensure that the selected model(s) generalize well to unseen data and is not overfitting to the training data.

7. **Ethical Considerations**: Address ethical and privacy concerns related to consumer data, and ensure that the research adheres to relevant regulations and guidelines.

1.3 Project Domain – DBMS W/ ML

Machine Learning (ML) is transforming the world with research breakthroughs that are leading to the progress of every field. We are living in an era of data explosion. This further improves the output as data that can be fed to the models is more than it has ever been. Therefore, prediction algorithms are now capable of solving many of the complex problems that we face by leveraging the power of data. The models are capable of correlating a dataset and its features with an accuracy that humans fail to achieve. Bearing this in mind, this research takes an in-depth look into the of the problem- solving potential of ML in the area of Database Management Systems (DBMS). Although ML hallmarks significant scientific milestones, the field is still in its infancy. Integrated ML is used for SQL developers who wish to deploy machine learning models in their work. It is a SQL syntax provides an efficient interface to communicate with SQL databases for machine learning algorithms to feed in input and get output prediction from ML algorithms. Its use of Auto ML lets users create efficient ML models without much extensive depth knowledge of it. Integrated algorithm exploits novel internal tuple batching schemes at time of query processing, enabling efficient query predictions, regardless of context in which scalar predictions are invoked. Furthermore, it is embedded deeply with a plating relational engine which ensures that performance is not affected by this abstraction but rather enables good model prediction based on query plan. Integrated ML is provided as an embed capability in Inter Systems IRIS Data platform, multi model DBMS that support DDL/ DML type SQL syntax.

1.4 Scope of the Project

The "ML Model" project boasts a comprehensive scope designed to empower sellers in making strategic, data-driven decisions. This multifaceted endeavor centers on several key objectives. Firstly, it delves into the realm of data analysis, systematically extracting valuable insights and trends from customer feedback and suggestions. Leveraging the ETS model and Random Forest model, the project's second objective involves predictive modeling to anticipate future trends and customer preferences based on historical data. The third key focus area is strategic decision support, providing sellers with a platform to make well-informed decisions related to product categories, resource allocation, and marketing strategies. Ultimately, the project aims to enhance profitability by aligning seller operations with data-driven insights.

In terms of target areas, the ML Model project concentrates on assessing product categories' performance, identifying high-performing segments, and recognizing those with growth potential. It also provides guidance on resource allocation, whether financial investments or human capital. Furthermore, the project assists sellers in refining their sales and marketing strategies, tailoring them to specific product categories or segments. Methodologically, the project encompasses data collection from various sources, data analysis employing the ETS and Random Forest models, predictive modeling for future forecasting, and the development of a decision support system for sellers. However, it's crucial to acknowledge limitations, including data quality, market dynamics' impact on predictions, resource constraints, and adherence to ethical guidelines and data privacy regulations in data collection and analysis.

Chapter 2

PROJECT DESCRIPTION

2.1 Existing System

Many machine learning algorithms can be used for detection. Many research shows several algorithms that can be used for sales prediction as false or genuine one. The algorithms used in various experiment were Random Forest, Naive Bayes, Decision Tree and Multilayer Perceptron. Results show that each algorithm can be used for sales prediction with high/low accuracy.

1. **Data Collection**: The first step is to collect relevant data, typically in the form of motor performance data. This data can be collected through various means, such as wearable devices, smartphone apps, or clinical assessments.

2. **Preprocessing**: Data preprocessing is crucial to clean and prepare the data for analysis. This may involve filtering out noise, handling missing values, and normalizing data.

3. **Feature Extraction**: Motor Variability Derived Parameters (MVDP) are specific features extracted from the motor performance data. These parameters capture variations in motor movements that can be indicative of Parkinson's Disease.

4. **Dataset Splitting**: The dataset is usually split into training and testing sets to evaluate the performance of the ML models.

5. **Feature Selectio**n: Depending on the dataset and the algorithm used, feature selection techniques may be employed to choose the most relevant MVDP features.

6. **Machine Learning Algorithms**: Various ML classification algorithms can be employed to build predictive models. Common algorithms include:

- Logistic Regression
- Support Vector Machines (SVM)
- Random Forest
- Gradient Boosting
- Neural Networks (Deep Learning)

7. **Model Training**: The selected ML algorithms are trained on the training dataset, using the MVDP features as input and the presence or absence of Parkinson's Disease as the target variable.

8. **Model Evaluation**: The models are evaluated on the testing dataset using performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to assess their effectiveness in sales prediction

9. **Hyper parameter Tuning**: Parameters of the ML models may be fine-tuned to optimize their performance.

10. **Cross-Validation**: To ensure the robustness of the models, cross-validation techniques like k-fold cross-validation may be applied.

11. **Results Interpretation**: The results are interpreted, and the model's ability to predict Parkinson's Disease is assessed. Researchers often create ROC curves and confusion matrices to visualize model performance.

12. **Deployment**: If the model performs well in the evaluation, it can be deployed for real-world applications, such as assisting in sales prediction or monitoring the progression of the disease.

13. **Continual Improvement**: The system may be continually improved by gathering more data and refining the model with additional features or more advanced ML techniques.

2.2 Literature Review

Sl. No.	TITLE	AUTHOR	JOURNAL NAME	TECHNOLOGY USED	ADVANTAGE	DISADVANTAGE
i	Sales prediction using Machine Learning Techniques: A Comparative Analysis	John O. Awoyemi, Adebayo O. Adetunmbi, Samuel A. Oluwadare	IEEE Xplore	Naïve Bayes and k-nearest Neighbour algorithms were used	It showed a optimal accuracy in case of k-nearest algorithm	K-nearest algorithm works slower than the logistic regression Algorithm.
ii	Sales forecasting using deep learning.	Ruttala Sailusha, V. Gnaneswar, R. Ramesh, G. Ramakoteswara Rao	IEEE Xplore	The algorithms used are random forest algorithm and the Adaboost algorithm. The results of the two algorithms are based on accuracy, precision, recall	Higher accuracy of fraud detection. Compared to rule-based solutions, machine learning tools have higher precision and return more relevant results as they consider multiple additional factors	Imbalance data or skewed distribution of data.

Sl. No.	TITLE	AUTHOR	JOURNAL NAME	TECHNOLOGY USED	ADVANTAGE	DISADVANTAGE
III	Sales prediction using ML algorithm,	Anuruddha Thennakoon,Chee Bhagyani , Sasitha Premadasa, Shalitha	IEEE Xplore	Predictive analytics was used and machine learning models were implemented and an API module to decide if a particular transaction is genuine or fraudulent. Also assess a novel strategy that effectively addresses the skewed distribution of data	Less manual work needed for additional verification. Enhanced accuracy leads reduces the burden on analysts.	Lack of real life data
IV	Sales prediction using deep learning	Munira Ansari, Hashim Malik, Zaiyyan Khan	IJERT	Machine learning algorithms are used to detect credit card fraud. To evaluate the model efficacy, a publicly available credit card data set is used.	Fewer false declines. False declines or false positives happen when a system identifies a legitimate transaction as suspicious and wrongly cancels it.	Overlapping of data

SI. No.	TITLE	AUTHOR	JOURNAL NAME	TECHNOLOGY USED	ADVANTAGE	DISADVANTAGE
V	Sales predictio n sales System	V. Filippov , L. Mukhan ov, B. Shchuki n	IEEE Xplore	The use of Bayesian Networks is suitable for this type of detection, but results from previous research showed that some input data representation method should be used for effective classification. For transaction monitoring by bank employees the clustering model was developed. This model allows provision of fast analysis of transactions by attributes	Ability to identify new patterns and adapt to changes. Unlike rule-based systems, ML algorithms are aligned with a constantly changing environm ent and financial conditions	Difficulty in dealing with categorical data
VI	Sales predictio n using Bayesia n and Neural Network	Sam Maes, karl tuyls, Bram Vansch oenwin kle	Research Gate	Artificial Neural Network and Bayesian belief Network were used and their significant results on real financial data was showed.	Compare d to rule-based solutions, machine learning tools have higher precision and return more relevant results as they consider multiple additional factors.	Imbalanced Data i.e most of the transactions (99.8%) are not sales predictoin which makes it really hard for detecting the predictoin ones

2.3 Issues in Existing System

1. **Data Quality and Quantity**:
   - Insufficient or imbalanced data can affect the model's performance.
   - Noisy or inconsistent data can lead to inaccurate results.

2. **Feature Selection**:
   - The choice of MVDP features may not always be straightforward, and selecting the most relevant ones can be challenging.

3. **Overfitting and Generalization**:
   - ML models may overfit the training data, resulting in poor generalization to new, unseen data.

4. **Interpretability**:
   - Some ML models, like deep neural networks, are not easily interpretable, making it difficult to understand the reasoning behind a diagnosis.

5. **Ethical and Privacy Concerns**:
   - Medical data, including data related to sales prediction, must be handled with care to ensure patient privacy and comply with regulations like HIPAA.

6. **Model Explainability**:

- It may be challenging to explain to healthcare professionals or patients why a model made a particular prediction, especially in complex models like neural networks.

7. **Robustness to Variability**:
   - Variations in patient demographics, data collection methods, and environmental conditions can affect model performance.

8. **Lack of Real-time Monitoring**:
   - The system might not be capable of real-time monitoring, which is essential for tracking the progression of Parkinson's Disease.

9. **Incorporating Clinical Expertise**:
   - The system may not effectively leverage the expertise of stores, who can provide valuable insights into sales prediction

10. **Cost and Accessibility**:
    - Implementing and maintaining such systems can be costly, limiting their accessibility to some sales prediction providers or regions.

11. **Model Updates**:
    - As medical knowledge evolves, the system may require frequent updates to stay relevant and accurate.

12. **Validation and Regulatory Approval**:
    - To use such a system in a sales setting, it may need to undergo rigorous validation and gain regulatory approval, which can be a lengthy and expensive process.

2.4 Software Requirements

iOS	iOS 13 or above
Android	Android OS 5.0 or above
Mac	OS X 10.11 or above
Windows	Windows 7 or above

Chapter 3

DESIGN

3.1 Proposed System

Based on the current problems and opportunities, we will predict the reach and demand of the products under each category well in advance, and with the help of this prediction, the customer can set the price of the products in an effective way so that he can generate profitable revenues by using the given model. Using the retail store's sales data, the proposed study suggested the following several processes for projecting the sales of various categories. Figure 1 depicts the suggested system's architecture diagram. The many steps in the process are outlined below.

1. **Data Sources**: The system likely gathers data sales over different periods of times.

2. **Machine Learning Algorithms**:
   - Random Forest are used as classifiers to detect sales. It's mentioned that various other classifier algorithms might be explored as well, providing flexibility for experimentation.
   - The use of ensemble learning through soft voting can enhance the accuracy of predictions.

3. **Data Preprocessing**:
   - Data preprocessing is employed to clean and prepare the input data, ensuring data quality and integrity.
   - Feature extraction is used to derive relevant information from the audio and imaging data.

4. **Data Splitting**: The dataset is divided into training and testing data, a common practice in machine learning to assess model performance.

5. **Performance Metrics**:
   - The system's performance is evaluated using metrics such as accuracy, where it achieves an best accuracy
   - The F1 score is mentioned, indicating the system's ability to balance precision and recall and yielding a best score.

6. **Objective Achievement**: The system's main goal is to contribute to the early detection of Parkinson's Disease, leveraging machine learning techniques to assist in this process.

7. **Quality and Efficiency**: Mention is made of handling data efficiently and minimizing detection errors while optimizing training time, which is essential for practical applications.

3.2 Architecture diagram

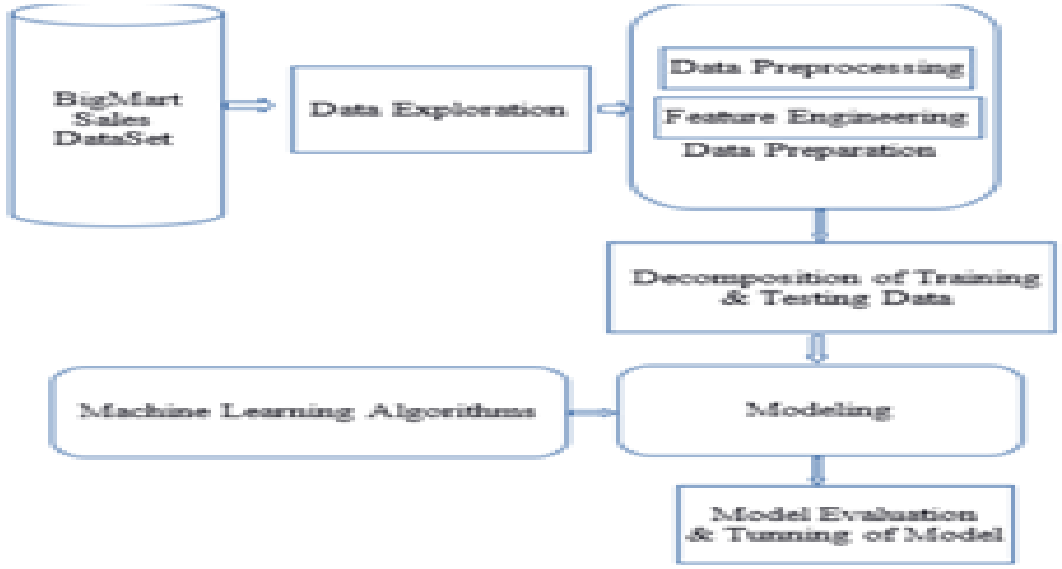


Figure 3.1 MODEL ARCHITECTURE

The above Fig 3.1 illustrates the central architecture of this project. First, the customer's data is put into the model after preprocessing data . The perfect datasets and the data is taken from the database. Further the data is preprocessed for the data analysis step. Then in the data analysis the process of inspecting cleaning, transforming the data and interpreting the data has done. Then the ML model is proposed for the model building and the training and testing of the data has done by the ML model. Finally, the evaluation of the of the trained and tested data has done in evaluation step.

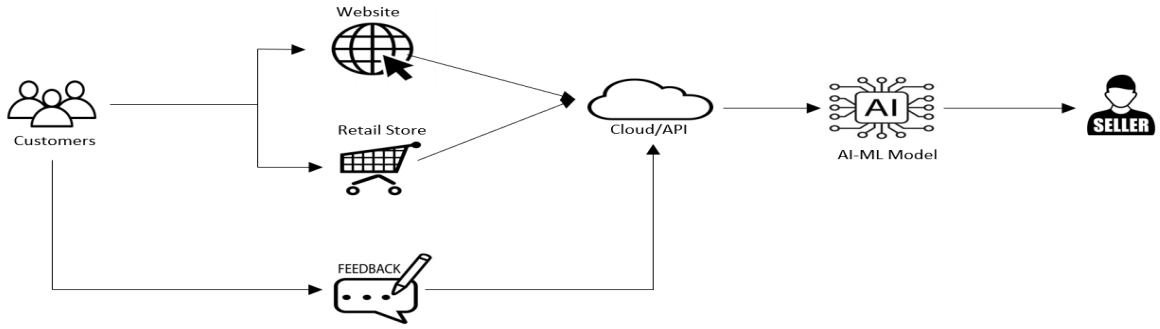


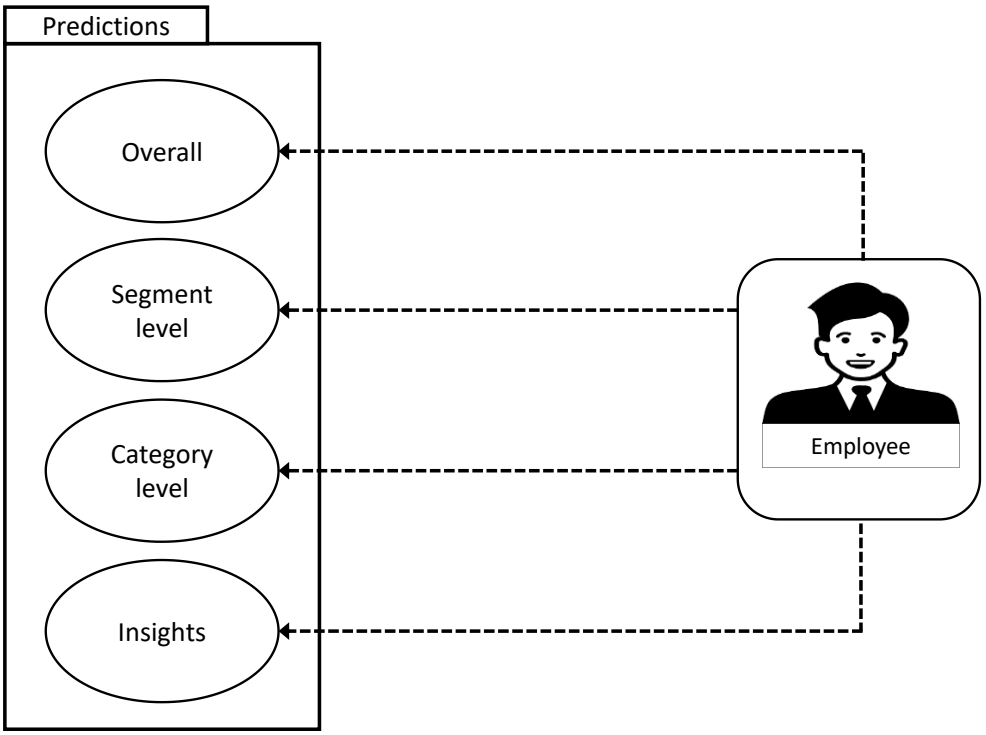
Figure 3.2 SYSTEM ARCHITECTURE

3.3 Design Phase

The Design Phase consists of the UML diagrams to design and construct the project.

1. Use Case Diagram
2. Data flow Diagram
3. Deployment Diagram

3.4 Use Case Diagram



,

Figure 3.3 Sales Prediction Use Case Diagram

The above figure 3.3 illustrates the Use case diagram of the project. Sales prediction is done based on the customer dataset, division is based on the different segment level and product category levels with some insights.

3.5 Data Flow Diagram

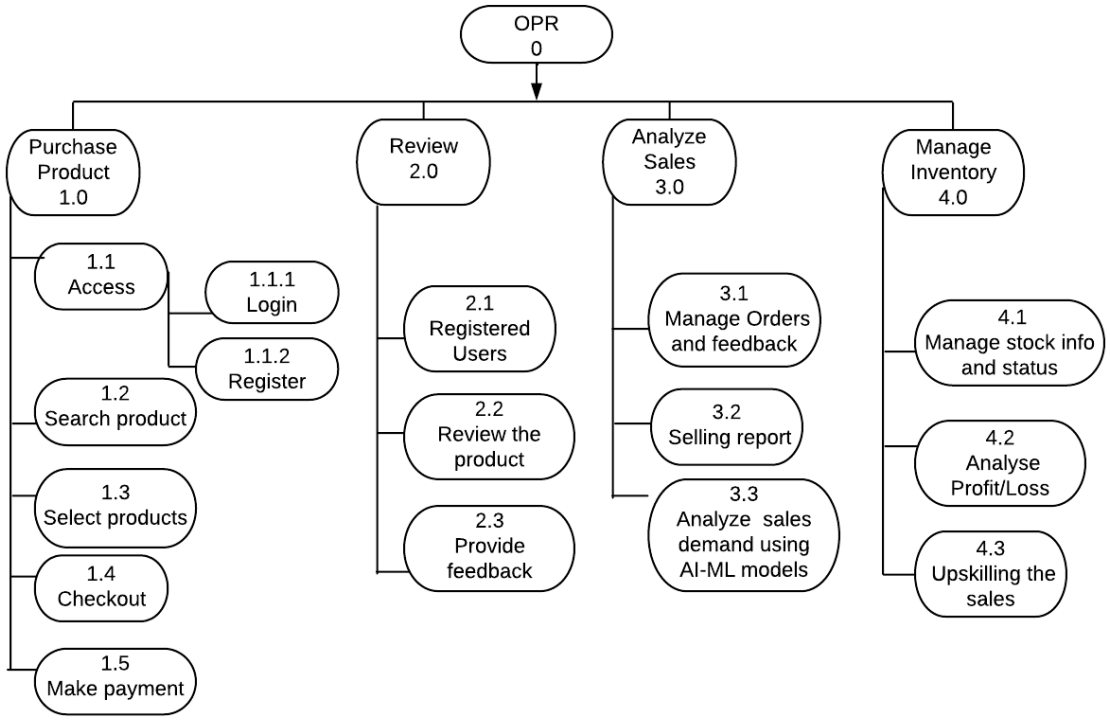


Figure 3.4 Music Application Data Flow

The above figure 3.4 illustrates the basic data flow of the project. Model will take the dataset and get into the model and predict the sales….above diagram shown clearly.

3.6 Deployment Diagram

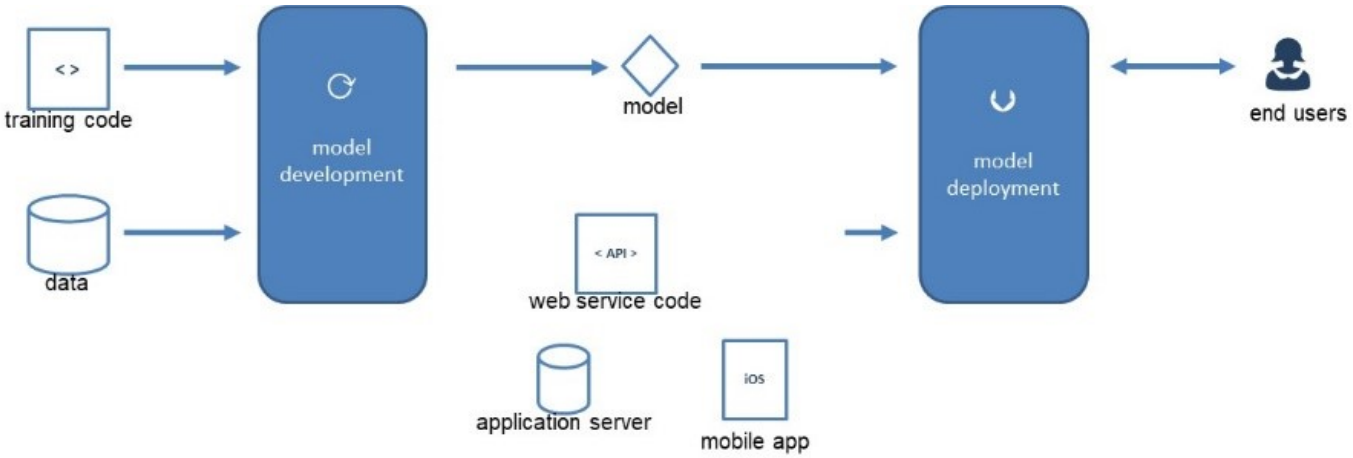


Figure 3.5 sales-model Database

The above figure 3.5 illustrates the deployment diagram of the project..

3.7 Module Description

Following are the main Modules of this sales Application

1. Login Module - At Admin Side
2. Process Module- At Admin Side
3. Database Module - AT Admin Side
4. Bot Module Management Module - At Admin Side
5. sales Recommendation Module – At Admin side (Database)
6. Sales Genre Identification Module – At Admin Side (ML)
7. User Management Module - At Admin Side
8. Register Module– At User Side
9. Login Module – At User Side

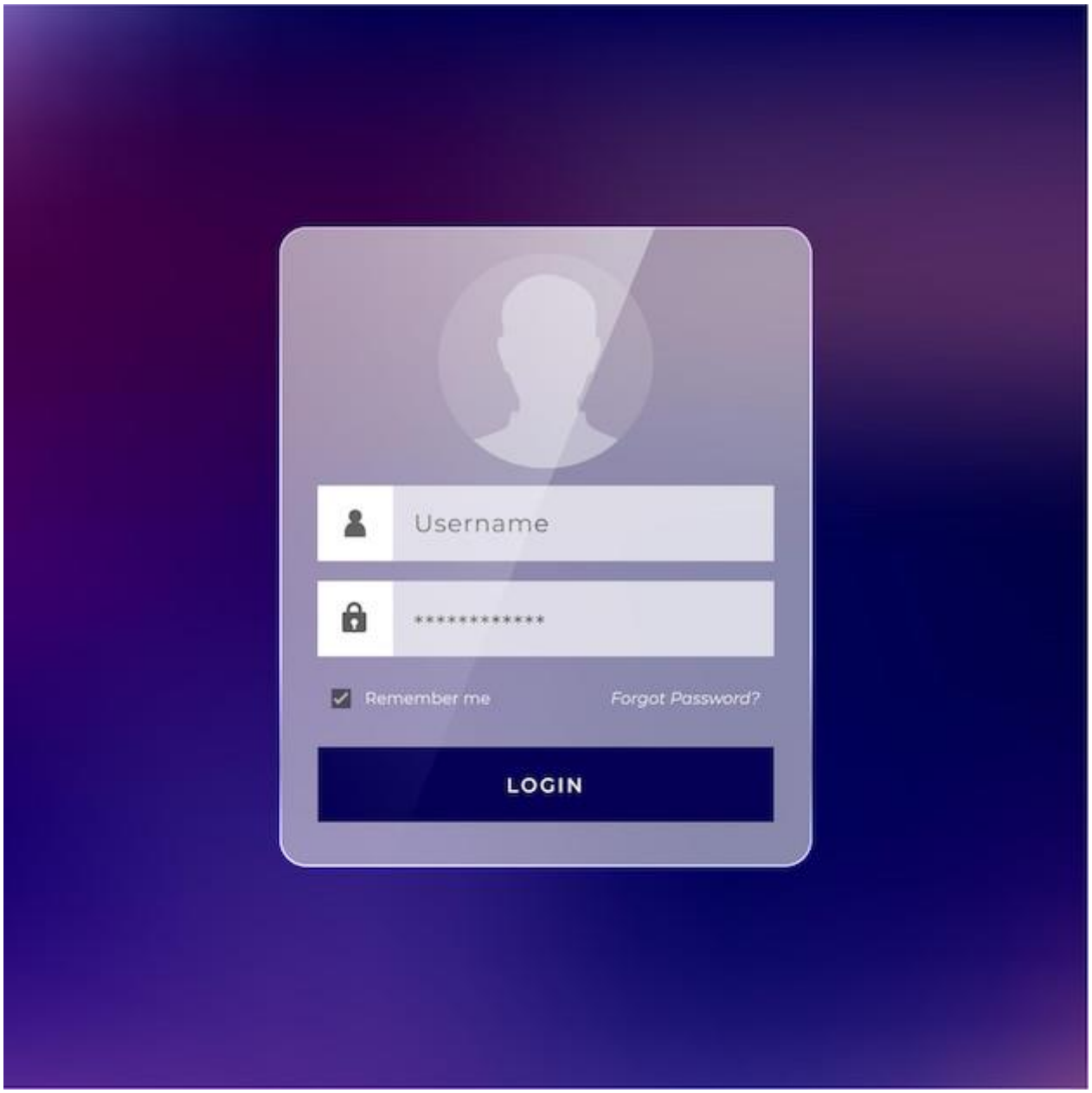
Features:

1. **Order ID:** This is a unique identifier for each order placed, which helps track and manage individual orders. It's crucial for order management and logistics.
2. **Order Date:** The date when the order was placed. It's essential for tracking the timing of orders and understanding sales trends.
3. **Ship Date:** The date when the order is shipped to the customer. It's significant for calculating shipping times and ensuring timely deliveries.
4. **Ship Mode:** This indicates the method of shipping chosen for the order, such as standard shipping, express delivery, or other options. It affects shipping costs and delivery speed.
5. **Customer Name:** The name of the customer who placed the order. It's essential for customer relationship management and personalization.
6. **Segment:** This categorizes customers into different segments based on specific criteria like demographics or purchase behavior. It helps in targeted marketing and service.
7. **State:** The state where the order is being shipped to. It's crucial for logistics and regional analysis.
8. **Country:** The country where the order is being shipped to. It's vital for international orders and understanding the geographic scope of sales.
9. **Market:** This indicates the market or industry segment to which the product belongs, such as technology, healthcare, or others. It's used for categorization and market analysis.
10. **Region:** The geographic region where the order is being shipped, which can be broader than just a state or country. It's valuable for regional sales analysis.

11. **Product ID:** A unique identifier for each product in the catalog. It's used to link orders to specific products and track inventory.
12. **Category:** The broad category to which a product belongs, such as electronics, furniture, or office supplies. It's crucial for categorizing and analyzing products.
13. **Sub-Category:** A more detailed category within the broader category. For example, under "Furniture," there can be sub-categories like "Chairs" or "Desks."
14. **Product Name:** The name of the specific product in the order. It helps identify the exact item purchased.
15. **Sales:** The total revenue generated from the order, often before any discounts or shipping costs. It's a key financial metric.
16. **Quantity:** The number of units of a product ordered in a given order. It's essential for managing inventory and assessing demand.
17. **Discount:** The discount applied to the order, which affects the final price. It's vital for understanding pricing strategies and customer behavior.
18. **Profit:** The profit generated from the order, which is calculated by subtracting the cost of goods sold from the sales. It's a fundamental financial metric.
19. **Shipping Cost:** The cost incurred for shipping the order to the customer. It's an expense that impacts overall profitability.
20. **Order Priority:** The priority level of the order, which can determine how quickly it's processed and shipped. It's essential for managing order fulfillment.
21. **Year:** The year in which the order was placed. It's crucial for tracking sales trends over time.
22. **Date_Diff:** This feature may represent the time it takes for an order to be delivered, calculated as the difference between the ship date and the order date.
23. **Customer Feedback:** This could include customer reviews, ratings, or comments about their shopping experience or the product received. It's important for assessing customer satisfaction and making improvements.

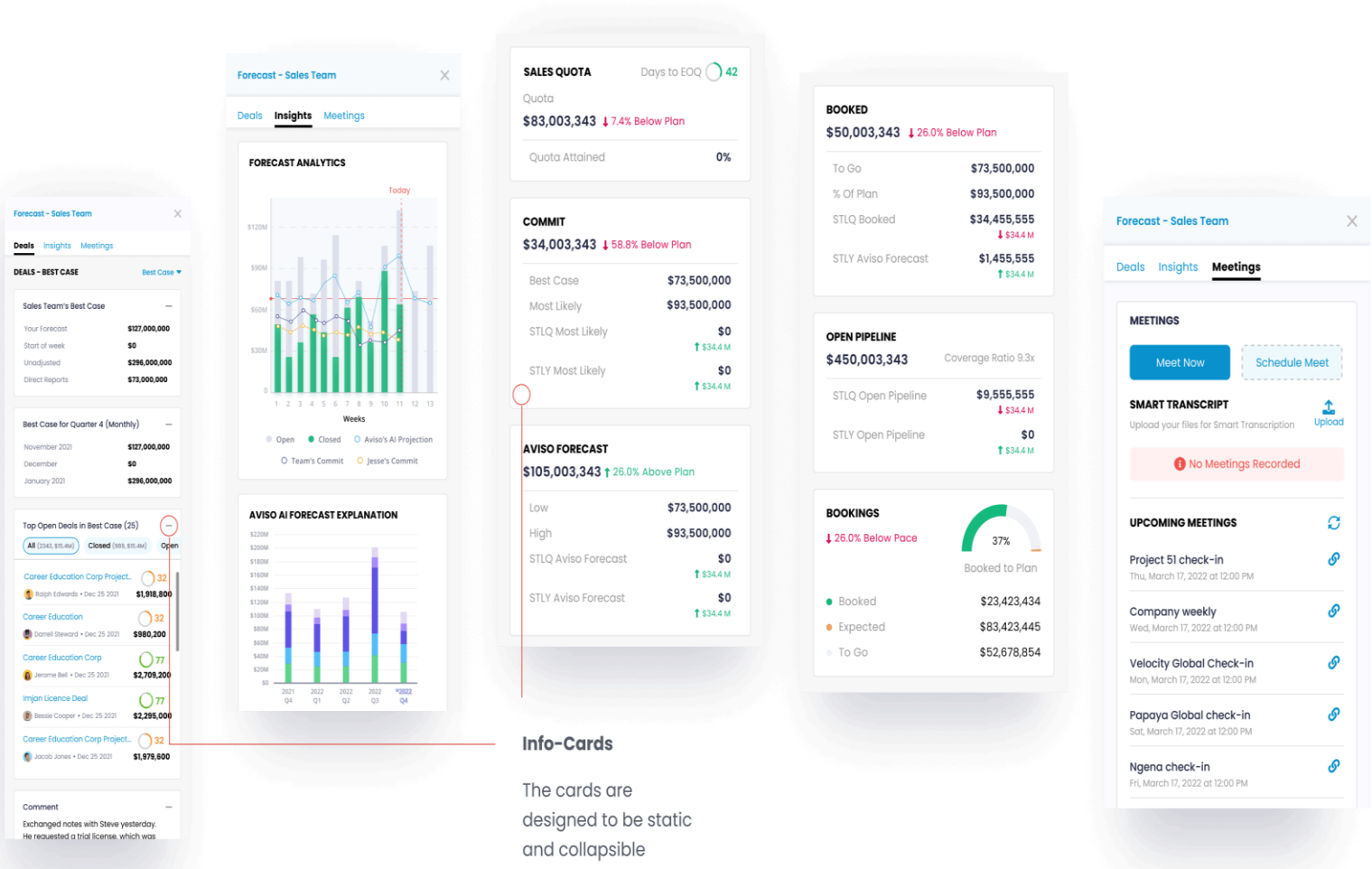
3.7.1 Login Module

The *Login Module* is a portal module that allows users to type a user name and password to log in. You can add this module on any module tab to allow users to log in to the system. In this component the artists can get an overhaul of their account and make sales prediction interfaces. The app provides the sales prediction in the app.



3.7.2 Process Module

Process modules are a central element in the project. They facilitate the application of the process model and form the basis of dataset they upload. They encapsulate work products, activities and roles. Process modules therefore contain all the components required to generate defined results.



3.7.3 Model explanation

A Random Forest is a versatile machine learning model known for its exceptional predictive performance. It's an ensemble learning method that leverages the power of multiple decision trees to make accurate and robust predictions. In this ensemble, each decision tree is built on a bootstrapped subset of the original dataset, ensuring diversity and independence. What sets Random Forest apart is its introduction of feature randomness at each node of a decision tree, where only a random subset of features is considered for splitting. This feature selection technique prevents certain attributes from dominating the model, making it more resilient. The splitting criteria in each decision tree rely on either Gini impurity or information gain, which measure the homogeneity of classes within the data. For binary classification, the Gini impurity calculates the degree of impurity, while information gain assesses the reduction in uncertainty achieved by splitting based on a particular feature. When it's time to make predictions, each decision tree casts a "vote" for the predicted class, and the final prediction is determined through majority voting. This ensemble approach combines the strength of individual trees to deliver reliable and accurate results. Random Forests are widely used for classification and regression tasks, known for their ability to mitigate overfitting and handle complex datasets effectively.

Impurity	Task	Formula	Description
Gini impurity	Classification	$\sum_{i=1}^{C} f_i(1-f_i)$	f_i is the frequency of label i at a node and C is the number of unique labels.
Entropy	Classification	$\sum_{i=1}^{C} -f_i \log(f_i)$	f_i is the frequency of label i at a node and C is the number of unique labels.
Variance / Mean Square Error (MSE)	Regression	$\frac{1}{N}\sum_{i=1}^{N}(y_i-\mu)^2$	y_i is label for an instance, N is the number of instances and μ is the mean given by $\frac{1}{N}\sum_{i=1}^{N} y_i$
Variance / Mean Absolute Error (MAE) (Scikit-learn only)	Regression	$\frac{1}{N}\sum_{i=1}^{N}\lvert y_i-\mu\rvert$	y_i is label for an instance, N is the number of instances and μ is the mean given by $\frac{1}{N}\sum_{i=1}^{N} y_i$

The ETS model, short for Error-Trend-Seasonality model, is a powerful tool for analyzing and forecasting time series data. It's particularly valuable in understanding and predicting patterns in data that exhibit systematic variations over time.

This model breaks down the time series data into three main components: error, trend, and seasonality. The error component accounts for random fluctuations or noise in the data that cannot be explained by the model. The trend component captures the long-term movement or direction in the data, whether it's increasing, decreasing, or remaining stable over time. Seasonality addresses periodic or repetitive patterns that occur at fixed intervals, such as daily, monthly, or yearly fluctuations.

The ETS model combines these components in a systematic way, making it a flexible and adaptable approach for different types of time series data. It's a state-space model, which means it continually updates and refines its estimates as new data becomes available. The model assigns weights to the different components, allowing it to adapt to changing patterns and adjust its forecasts accordingly.

Chapter 4

RESULTS AND DISCUSSION

4.1 sales prediction using random forest and ETS models:

The dataset was taken from the store owner and predicted the sales using random forest and ets model. The data analysis part discovered that the technology product sales are more when compare with the other two categories.And even in the technology products , the copiers and mobiles are more in sales.given below picture show the final predictoin of sales using the random forest algorithm.

EXPERIMENT:

The sales predicted in python module are show below.

```
In [93]: #RF-Trees: 25 -Depth: 2 -MaxF: auto -MinLeaf: 2 -MinSplit: 2
series = final_df2.iloc[:,0]

feedback = final_df2.iloc[:,1]

overall = random_forest(series,feedback,period, n_estimators = 25, max_depth = 2,max_features = 'auto', min_samples_leaf = 2, mi
# Meto_288MG100[str(Meto_200MG100.columns[0]+" Forecast Denormalised")] = (Meto_208MG100.iloc[:,1] dendre_indices.iloc[:,2]) .ro

# Forecast_All = pd.merge(Forecast_All, Meto_200MG100, left_index=True, right_index=True)
overall
```

2021-12-31	371269.0	364931.0
2022-01-31	268275.0	258062.0
2022-02-28	244179.0	221553.0
2022-03-31	347757.0	292809.0
2022-04-30	302140.0	307447.0
2022-05-31	304795.0	342755.0
2022-06-30	372608.0	371637.0
2022-07-31	278670.0	272784.0
2022-08-31	432742.0	399440.0
2022-09-30	405455.0	424389.0
2022-10-31	406676.0	387501.0
2022-11-30	508985.0	432392.0
2022-12-31	427759.0	432392.0

Figure 3.6 result of sales prediction of next year

Chapter 5

CONCLUSION AND FUTURE ENHANCEMENT

5.1 Conclusion

In conclusion, the "ML Model" project holds the promise of transforming the way sellers engage with customer feedback and data. By systematically analyzing this valuable information, leveraging predictive modeling, and offering a decision support system, the project equips sellers with the tools needed to make informed decisions. It empowers them to identify high-performing product categories, allocate resources strategically, and fine-tune their sales and marketing strategies. The ultimate goal is to boost profitability and foster growth at a segment or category level.

Limitation: The performance of the proposed model don’t applied for the days wise sales but for the monthly sales prediction.

Future Work: In future work, we intend to enhance the performance and take the security and privacy of the data in real time into consideration.

5.2 Future Enhancement

In future work, we intend to enhance the performance and take the security and privacy of the data in real time into consideration.
This ML model incorporates detection of sales using the acoustic features of the patients. However, sales can be detected using ml model with the time as a factor.so, we can create a model for those and can be incorporated into this for a perfect detection model.

5.3 Version 2.0 Enhancements

5.3.1 sales feature extraction:

Before applying machine learning approaches in sales information retrieval, an important stepis the determination of the features extracted from sales data. All the machine learning methods discussed in this chapter make use of the content features extracted from sales. In addition, for sales trend identification, we also make use of the text-based features from sales for next dates.

Content feature extraction:

There has been a considerable amount of work in extracting descriptive features from the dataset from the customer.

REFERENCES

1. 1.Akshay Krishna, Akhilesh V, Animikh Aich, Chetana Hegde (2018), " Sales-forecasting of Retail Stores using Machine Learning Techniques," IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions, IEEE Xplore,160-166. [CrossRef]

2. Sunitha Cheriyan, Shaniba Ibrahim, Saju Mohanan; Susan Treesa (2018), "Intelligent Sales Prediction Using Machine Learning Techniques," International Conference on Computing, Electronics & Communications Engineering (ICCECE),IEEE Xplore,53-58. [CrossRef]

3. Chenhui Lu, Shuo Feng, Jiahao Huang, Xiaojun Ye (2021), "A Multi-Task Prediction Framework for Sales Prediction," International Conference on Computing, Electronics & Communications Engineering (ICCECE),IEEE Xplore.194-19

4. I-Fei Chen, Chi-Jie Lu (2016), "Sales forecasting by combining clustering and machine-learning techniques for computer retailing," Neural Comput & Applications , Vol.18, 105–112

5. Xie dairu1,Zhang Shilong1 (2021), " Machine Learning Model for Sales Forecasting by Using Random forest" IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE 2021),IEEE, 480-483 [CrossRef]

6. Punam, K., Pamula, R., & Jain, P. K. (2018,),"A two-level statistical model for big mart sales prediction", In 2018 InternationalConference Computing, Power and CommunicationTechnologiesGUCON),IEEE, 617-620.machandra H V,Balaraju G,Rajashekar A, Harish Patil(2021),"Machine Learning Application for Black Friday Sales [CrossRef] Prediction Framework", 2021 International Conference on Emerging Smart Computing and Informatics (ESCI),IEEE,57-61.

7. Takashi Tanizaki, Tomohiro Hoshino, Takeshi Shimmura, and Takeshi Takenakam(2019),"Demand forecasting in restaurants using machine learning and statistical analysis," Procedia CIRP, 79:679– 683. [CrossRef]