DESIGNING AN INVENTORY MANAGEMENT SYSTEM USING DATA MINING TECHNIQUES

[1]A. A. Malik, [2]K. Amjad, [3]A. Saud, [4]A. Ali

[1,3,4]Faculty of Computer Science Lahore Garrison University Lahore
[2]Department of Computer Science University of Management and Technology
Email Corresponding Author email: dr_aftab_malik@lgu.edu.pk

ABSTRACT: Designing a POS system that enables the FBR to keep a record of the transactions made at the Tier-1 retailers who fall under the tax ruling of FBR. This will enable the FBR to check how much tax is being collected from the consumer. With that record, they will be able to estimate how many people are filled and what it enables to collect tax from the retailers with the help of FBR invoice number that will generate in real-time when the receipt is generated for the shopping consumer. Datamining techniques are effective in uncovering the previously hidden patterns for storing the data these techniques can be applied in every field such as banking, agriculture, and marketing. Several different data mining techniques have been used to improve the performance and predictions of exciting data. The novelty of the proposed system is that different data mining techniques get time stamps on yearly and monthly based transactions which gives us different variations-based graphs with precise working of data mining techniques.

Keywords: POS, FBR, Inventory Management System, Data Mining Techniques.

(Received 28.8.2023 Accepted 24.11.2023)

INTRODUCTION

Federal Board of Revenue (FBR) is planning to implement a Point of Sale (POS) system that will bring the entire retailer that falls under the Tier-1 retailers list [1]. This list defines who falls under the sellers' taxation list and tax will implied on that under what circumstances. In the past years, many of the retailers were evading this tax by not filling their income with FBR. These retailers were using their own POS which gives no record to the FBR. FBR decided to introduce a POS system to integrate into retailers' POS systems. This system will get an invoice number that will be generated in the FBR system and it's unique every time [2]. The motive behind this process is to integrate all the POS systems with FBR and submit sale invoices to FBR in real time. All retailers must be encouraged to use FBR invoices to collect a maximum of tax.

This system is cohesive with the FBR system when the consumer pays the bill, POS generates an invoice and gets a unique invoice number from the FBR system, and all the information of the invoice is saved on the FBR database [3]. The consumer will get a unique invoice number printed on their bill. The consumer can verify that bill's invoice by texting at FBR or by their mobile application. FBR system will generate an encrypted invoice number. It will synchronize data with the central FBR system periodically. POS will also generate. On verification, FBR will mark a 6% tax on verified invoice and 9% on non-verified invoices. The consumer is subjected to paying bills including tax. It will ensure that the tax, which consumer pay on their purchase, is propelled to FBR. Later on, it will help in filing returns from FBR [4].

The POS system must be designed to minimize any negative impact on the business. POS must be robust and it does not have any availability issues, so it must be available all the time to the retailers [5]. In addition, this process does not delay it; the process of getting the invoice number from FBR must done in microseconds. This software can be installed on the local machine or the cloud. The data controller of the POS system will be part of this. This system will have real-time integration with the FBR system. Data will be sent to the FBR system automatically [6]. The vendor must have a POS that is focused on FBR. In addition, it must be verified by FBR. FBR will include a certified retailers list on its website. FBR needs POS software that will push data to the Fiscal Soft Data Controller in the event of printing, The Fiscal Soft Data Controller will return SDC against the individual invoices, and the FBR logo along with the Fiscal Number and QR will get printed on Consumer invoice.

Datamining, originating from classical statistics, artificial intelligence, and machine learning, was introduced in the 1990s.

Many data mining techniques are related to machine learning techniques that have been developed in the past for exploring the limited amount of data. Large growth in the data collection of enterprises can be seen due to which data mining is a growing technology because it assists enterprises in understanding the

business better and in forecasting.

Actually the Datamining is based on knowledge based system which is a discovery process usually termed as (KDD) in Data Analysis; it is is a set of techniques or methods that are efficient for discovering previously unknown, valid useful, and understandable patterns that are actionable and can help companies in decision-making. In the proposed research data mining techniques have been used to make the POS system more effective.

Data mining is initiated with data pre-processing, in which the data gathered from different sources is cleaned and transformed into an understandable form that can use for the data mining process. The cleansed data is then integrated into a data warehouse which is a multidimensional database [7]. The next data mining phase, which follows data cleaning, involves transforming the data into a form that can be extracted from its raw form The last step is data evaluation in which the result obtained from the data mining process is analyzed The process is shown in Fig 1.

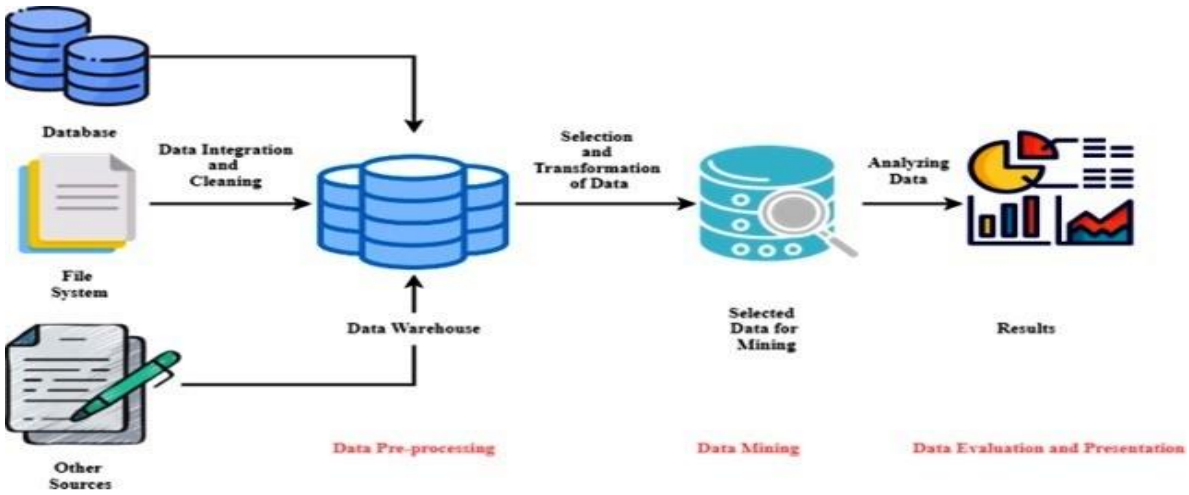

Figure 1. Datamining Flow

LITERATURE REVIEW

DBO Database System that will facilitate analytics for large groups of data. Instead of writing long queries, DOB will facilitate with such an algorithm that will cut off a long time and do the job in less time [8]. DOB will be competitive with other database systems. DOB will guess the final answer during query processing and will have constant meaning throughout. It will guess a query that will be accurate in a quite short period consisting of a few seconds or hardly a minute. [9]. With the massive increase in smartphone and electronic device usage, a large number of transactions are lodged every minute. Such database systems are required to meet the functional requirements of this kind of transaction. To meet the requirements Database systems also evolved. With that, technical things and hardware also has changed a lot. Such a system has become a need that meets these requirements and coops up with needed requirements [10]. Queuing model is required for database systems, which consists of three components; database server, computer network, and database requester. A model that is suitable for planning the capacity required for a database and can help in determining the exact capacity needed for it [11].

Stanford Public Information Retrieval System SPIRES designed a reliable, economical, virtual Database Management System (DBMS) to facilitate research. With this database design, SPIRES is supporting two hundred databases, which include virtual student records, their data, finding of surveys, and many others [12].

Data compression on the database is helpful for two basic reasons, one is it will take less physical space on the database and second, it will be easy to move around data. To use this feature, data compression, and expansion programs are required to be installed on the database. This article covers how this idea will be implemented on a database and what methodology is gainful [13].

For system performance measurements, the TP1 benchmark is utilized. TP1 benchmark consists of Debit-Credit, Scan, and Sort benchmark. Transaction processing system is a focus in this benchmark. These transactions focus on single-type transactions and work on large databases around 10GBs.This data consists of dummy data that shall work on banking models. These benchmarks work on a small number of transactions and update them. It was tested on a strict period to get more ideal output [14].Performance problems are common in databases. The database serves as a facilitator to write complex applications even if you are not very aware of

computer programming. Conversion of high- level language into low-level language is a work for the computer. The real aim of a database is how well it meets the requirements. What is necessary for the developer is to check how the input responds and how much time it takes while working on a specific task [15].

Table 1. Qualitative Analysis of Papers.

Paper	Author	Year	Methodology	Accuracy /Tools
The DBO Database System	F. Rusu	2019	Fundamental	Chartistc. Js
Evaluating Distributed Database Transaction System	P. Ma, K. Wei, and C. Jiang	2019	Exploratory	Fusion Charts
Modeling a Distributed Database system	N/A	2022	Descriptive	Jupiyter
Stanford's Generalized Database System	J. R. Schroeder, W. C. Kiefer, R.L. Guertin	2020	Applied	Datawra per
Data Compression on a Database system	G. V. Cormack	2018	Applied	Jupyter
Database System Performance Measurements	H. Boral, R. Bnce, D. DeWitt, D. Gawhck,	2019	Qualitative	Grafana
A view of Database system Performance Measures	J. Gray	2021	Qualitative	Google Charts
Command use in Relational Database system	J. D. Joyce, D. R. Warn	2019	Fundamental	D3.js
Extending a Database system with procedures	M. Stonebraker, J. Anton, E. Hanson	2017	Case Study	Chart Blocks
Deadlock prevention is a crucial aspect of ensuring security and integrity in a in distributed database system.	P.K Reddy, S. Bhalla	2020	Analytical	Infogram
The POS system is utilized for tracking and managing the transactions and cashier operations in supermarkets.	T. Antczak, R. Weron	2019	Quantitative	Google Charts

Analyzing commands in relational databases can help in building reliable structures in the future. General Motors Research Laboratories developed Regis (Relational General Information System), which was observed to be used in large databases to accommodate tasks. In a relational database queries like JOIN, PROJECTION, etc. must be efficient. Yet, INTERSECTION and "exclusive " must not limit the time required for it [16]. Following a full-fledged database procedure helps in creating powerful database management systems. Just like in object-oriented concepts, queries can be stored in database fields. To implement this, relations must be developed, and executed to make this achievable. This paper finds the particular query run time and its use across other databases [17].

Concurrence access to data from a database to different platforms can cause deadlocks. A reasonable algorithm must be adopted to prevent deadlock while data is shared across different platforms. This algorithm works by checking the wait block to be null, which is not null if the transaction already is in process at some other platform. It will propose such a system that is deadlock-free [18], [19].

To improve the waiting time in the queue, a system must be introduced to speed up the checkout process. It will require real data to work on and develop such a system. The unavailability of Pint of Sale representative data makes this task difficult. For this actual data must be collected from POS to work on it. The ultimate objective is to speed up checkout time and increase the satisfaction of consumers [20]. Level 0 of the inventory management system shows the internal and external behavior of the system. The boundary of the system is also defined at this level. Figure 2 shows the working of the system at the current level. The user enters the data into the system whereas the system provides an output according to the input provided by the user [21].

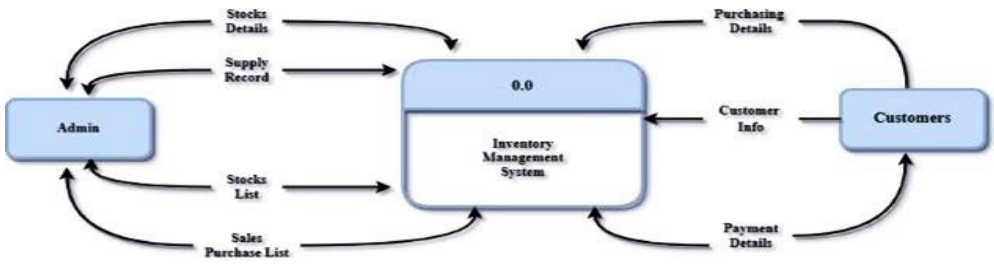

Figure 2. Level 0 of The Inventory Management System

Level 1 of the inventory management system provides a clear picture of the boundaries of the system. The process of data transformation from the input to the output is shown in **Figure 3** [22]. This level of the inventory management system contains information about the customers, sales, and stock payment, and all these attributes are managed in the current level of the inventory management system. The data flow starts from the admin and owner of the system then the data is stored in the database of the system. The data stored in the database of the system is used to provide the output accordingly [23].

Level 2 of the inventory management system shows the shown a borderer picture of the inventory management system as compared to level 1. At this level, the flow of all sub-processes of the system is shown in **Figure 4** [24].

The last level of the inventory management system shows how the data flows to different entities of the system. The customer's tax information is managed on a monthly and yearly basis depending upon the status of the customers (Filer or Non-Filer). The process is shown in **Figure 5** below [25].

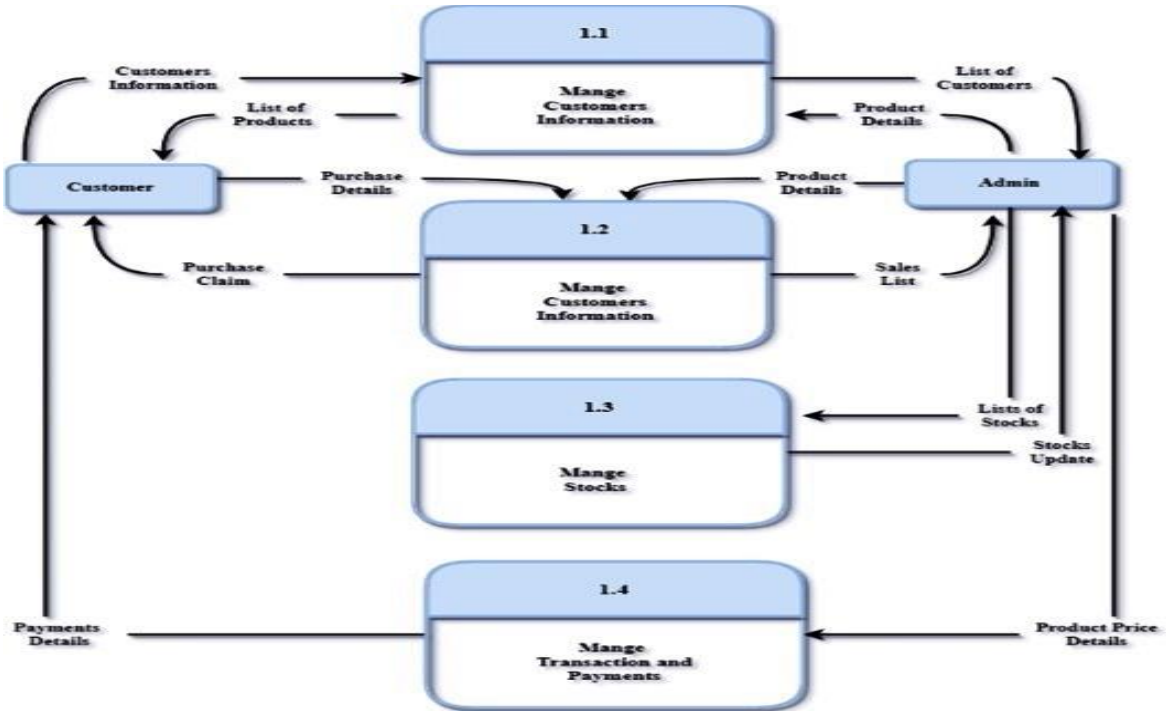

Figure 3. Level 1 of The Inventory Management System

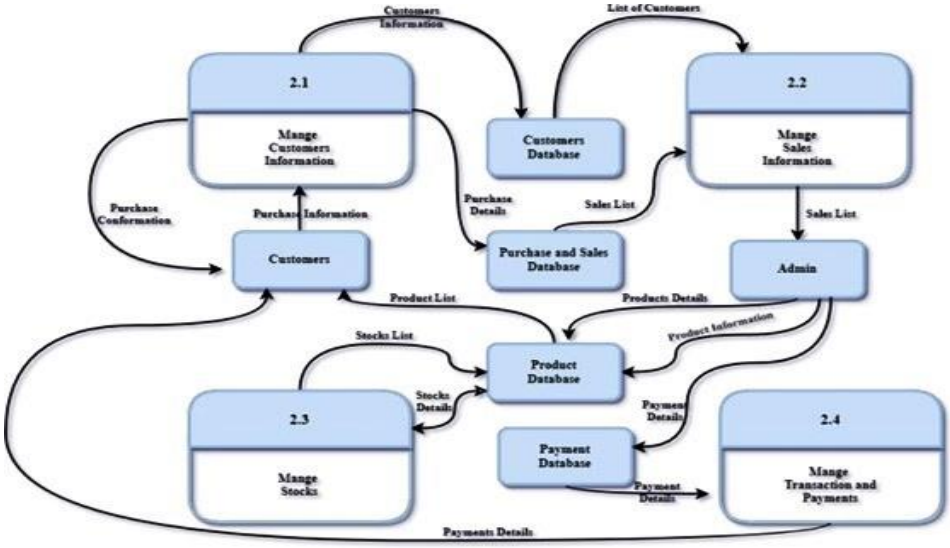

Figure 4. Level 2 of The Inventory Management System

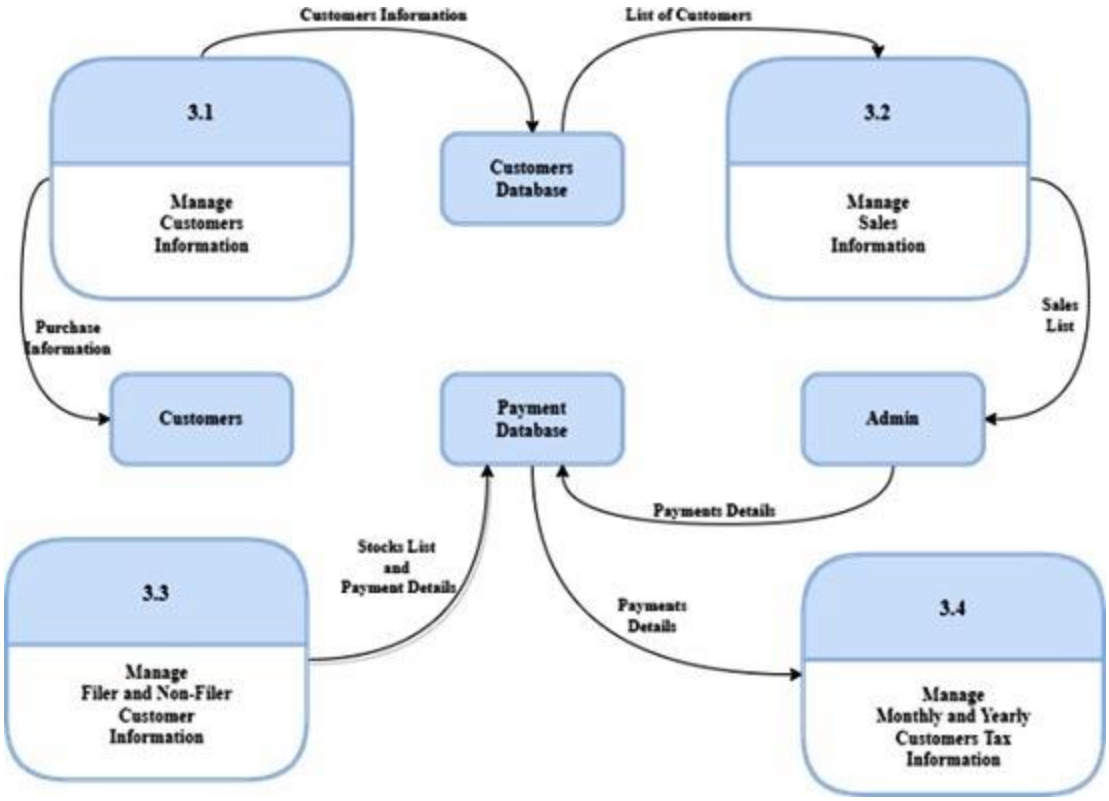

Figure 5. Level 3 of The Inventory Management System

METHODOLOGY

The methodology contains different data mining techniques through which we can pass out the data and get different results based on the applied techniques performed.

Datamining Techniques: The basic data mining techniques are divided into two parts:

- Predictive Datamining
- Descriptive Datamining

Predictive Datamining: The term "Predictive" means to predict something, so as the name suggests this data mining technique is used for forecasting or predicting future trends and making decisions accordingly.

Types of Predictive Datamining

- Classification
- Decision Tree
- Neural Networks
- K-Nearest Neighbor Classification
- Regression
- Time Series Analysis
- Prediction

Classification: In the classification data mining technique, the aim is to gather the most important and relevant information about the data. In this process, the data items are classified into different fields within the dataset for finding the class label of the unknown object.

Regression: Regression is also known as a "Continuous value classifier" It is a statistical modeling technique that is used to find the relationship between different variables. In this process, previously obtained data is used for predicting the continuous quantity for the new observations it is further divided into linear regression and multivariate regression.

Time Series Analysis: Time series analysis is a process in which data points are collected sequentially over an interval of time. In this process, the data is not just recorded randomly but at the same time, it is analyzed consistently at different time intervals.

Prediction: Prediction is another process of data analysis that provides numeric output like classification. In this process, the training dataset contains the input values and the corresponding numerical output values based on which the algorithm drives a model or predictor. The output provided by the model should be numerical whenever a fresh dataset is used.

Descriptive Datamining: The term "Descriptive" means to describe something, descriptive data mining aims to gather and convert the data into a meaningful form.

Descriptive data mining is a method that uses classification, association, and feature extraction to analyze the past behavior of data.

Types of Predictive Datamining

- Clustering
- Summarization
- Association Rules
- Sequence Discovery

Clustering: In this technique, the data is grouped into elements with similar characteristics. The elements can be simply categorized by identifying one or more attributes, clustering can be used to generate the class labels,-means clustering, Hierarchical clustering, and Gaussian clustering are some of the well-known clustering techniques.

Summarization: Summarization is the process of presenting and extracting information in a tabular or graphical format. Summarization is the most important step in data mining because it provides the key techniques to find a close description of a dataset. The data is analyzed by using a simple summarization method such as mean and devastation.

Association Rules: The Association rule is useful in discovering hidden patterns and the relationship or link between two items in a dataset. The probability of interaction between two items in large datasets or databases can be discovered by using the association if-then statements. Lift, support, and confidence are the three major measurement techniques.

Sequence Discovery: Sequence discovery also known as "Sequential pattern mining" is a data mining technique used to discover the statically relevant pattern in sequential data. Sequence discovery techniques evaluate the occurrence frequency duration, and values of sequence to find useful and interesting hidden patterns in this process the values are presumed as discrete.

Proposed System: The proposed architecture of the system is shown in Figure 6 in which the workflow of different phases of the system is described.

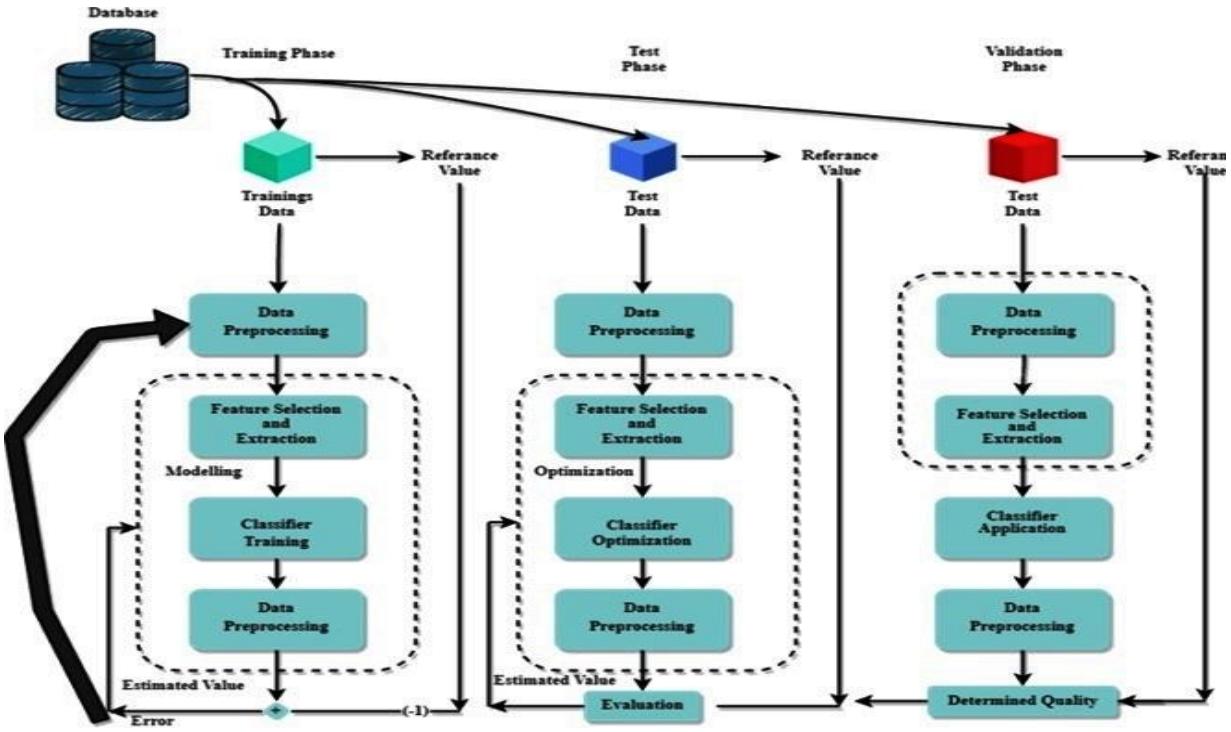

Figure 6. Architectural Diagram of the System

Data Description: This data is retrieved from an XML data set, which includes transactions of low-level data. After extraction, data files are converted to CSV for most relevant data like transactions, and cashier operations. This data is from a local grocery shop. Which operates manually and keeps no record of tax payments.

Data Ingestion: In the proposed architecture of the system ingestion of the data is the first step. The user or the owner of the shops can log in to their respective accounts to manage and monitor all the activities related to their shops. The owner of the shop can create a product catalog. The owner's information along with the store data is stored in the system database. The record of the sales made in the past is also stored in the system database. This inserted data by the user is used for the

training of the machine learning models.

Data Preprocessing: Data preprocessing is a very important and effective technique of data mining. Data preprocessing is the process of filtering a large amount of data and transforming this raw data into a more useful and efficient form. Data preprocessing is performed in three steps the first step is cleaning the data in which the missing and irreverent part of the data is handled. After cleaning the data, the next step is to transform the cleaned data into an understandable form that can be used for data mining. Data reduction is the last step of data preprocessing in which the data reduction method is used to increase efficiency and reduce the storage load and cost of analysis.

Storage Feature Extraction: The data of the industries is huge as it contains several different entities and each of these entities has several different fields. The accuracy of the system can be increased by extracting the right features. For training purposes, different fields can be selected respectively. For example, the training data may contain the customer's name, Product ID, and name of the product but for training, any field from these can be used.

Machine Learning (ML) Models: In the proposed research an algorithm known as XGBoost is used. A gradient Decision tree (DT) is used for the implementation of the XGBoost algorithm this framework also serves as a base for the Boost algorithms. Artificial Neural Networks (ANN) provide more effective and accurate results as compared to any other approach when the problem is related to unstructured data prediction whereas the decision tree (DT) approach is more preferred when the data is structured.

REPORT

In the proposed research for the prediction of the demanded values for future years, the XG boost model was used. The real values were rounded off to get the actual values. The obtained results can be used by the owners to increase the efficiency and accuracy of inventory management

Implementation: In the proposed research implementation is done by using an extreme gradient boosting XGBoost regression model. A Decision tree (DT) is used to ensemble the XGBoost. The proposed approach is more suitable and provides effective results as compared to the other approaches when the data is small and structured. The process is shown in Figure 7 below.

XGBoost Working: As mentioned above in the research the decision tree (DT) is used to ensemble the XGBoost model. The decision trees are constructed based on the error factor obtained from the last trees the only difference in the technique is the way the error factor is calculated from the lost function. The XGBoost constructs the trees by performing certain steps:

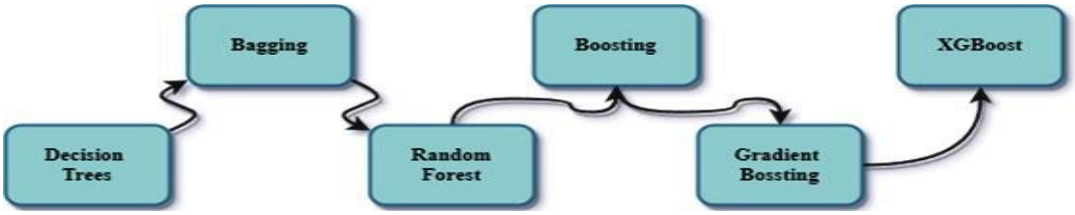

Figure 7. XGBoost Evolution

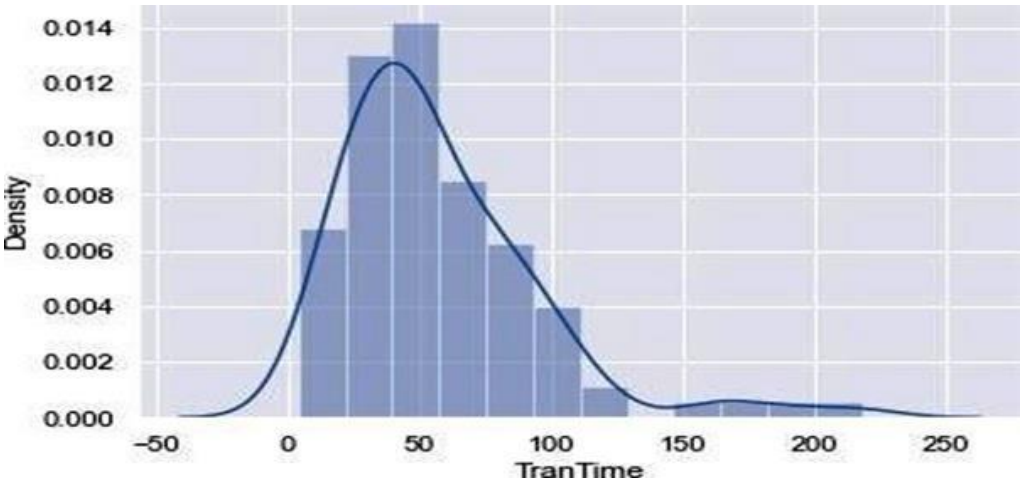

Figure 8. Average Time Taken for Transaction

Figure 9. Total Amount Paid On Cash and by Card.

- Filtering of the data
- Model fitting to the residual
- Forming a new model

This process continues until the error factor is improved.

Model Training: Model training is key for the prediction of the demanded values with more accuracy. The train-test-split function divides data arrays into two parts, one for training and the other for testing, ensuring efficient data management and accuracy The train-test-split function is reliable as the partition of the arrays is done digitally instead of manually so, the chances of human error are eliminated by using this approach. The partition of the arrays will be done randomly for the respective target values of A and B by default. To avoid overfitting an approach known as early- rounds is used the training is halted when the performance of the test dataset remains constant for the number of iterations. The train-test-split can automatically avoid overfitting by taking a point of inflection when there is a decrease in the performance of the test dataset and this process continues until there is an improvement in the training dataset towards the overfit.

RESULT AND DISCUSSION

The following data set is taken from the POS system that is already being used in a local store. This system does not include the FBR POS system. Originally, this data contained more than 60,000 entries, but we excluded all except 100. These random entries will be only used to show an overview of the POS system. Any entity that is not necessary will not be included in charts.

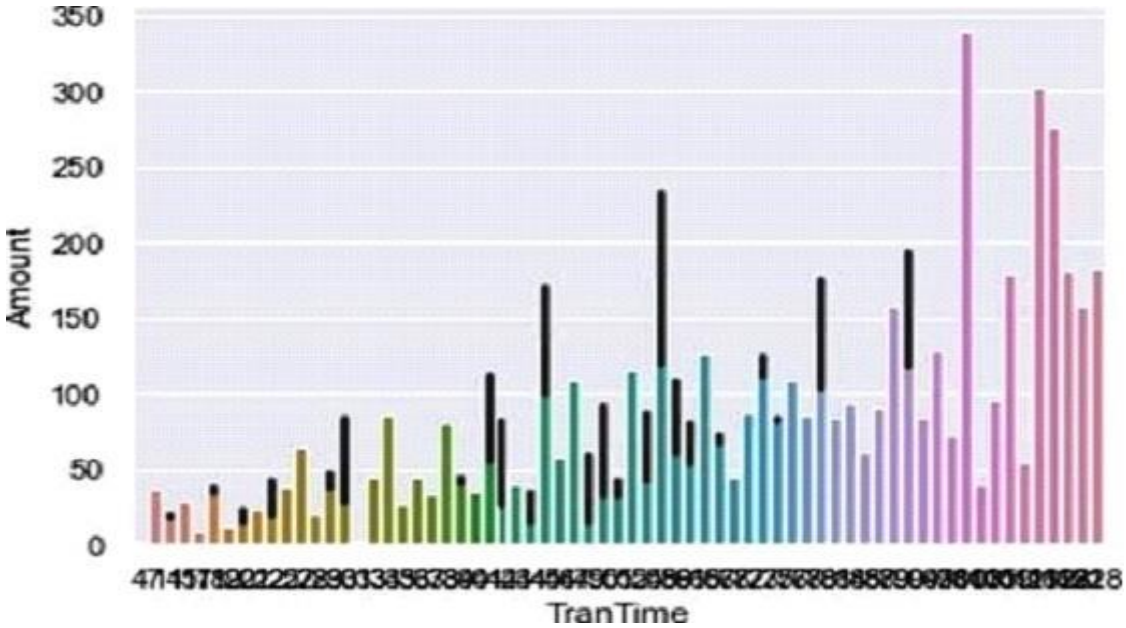

Figure 10. Time Taken by Different Price Ranges

Figure 8 depicts the average time taken by the system to perform a transaction. The total amount paid and the source are on the x-axis and y-axis respectively. Figure 9 shows the distribution of revenue. From Figure 9 it can be seen how much amount was paid in cash and how much was paid by card In Figure 10 two attributes of the system transaction time and density were used to find the total duration of different price ranges. The article numbers and transaction time are on the x-axis and y-axis respectively. This shows how much time is consumed for differ ent articles while processing as shown in Figure 11.

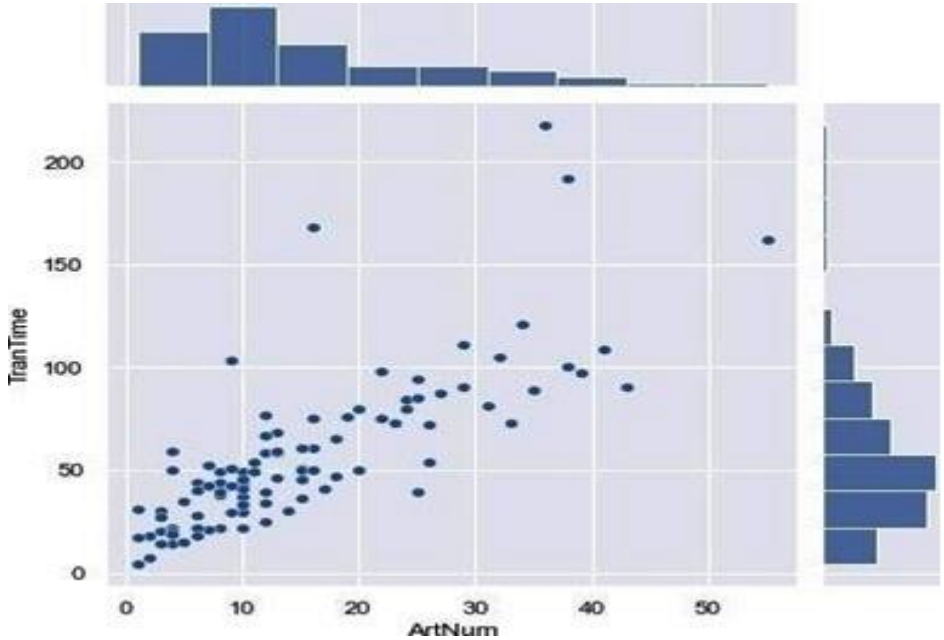

Figure 11. Duration of Different Articles

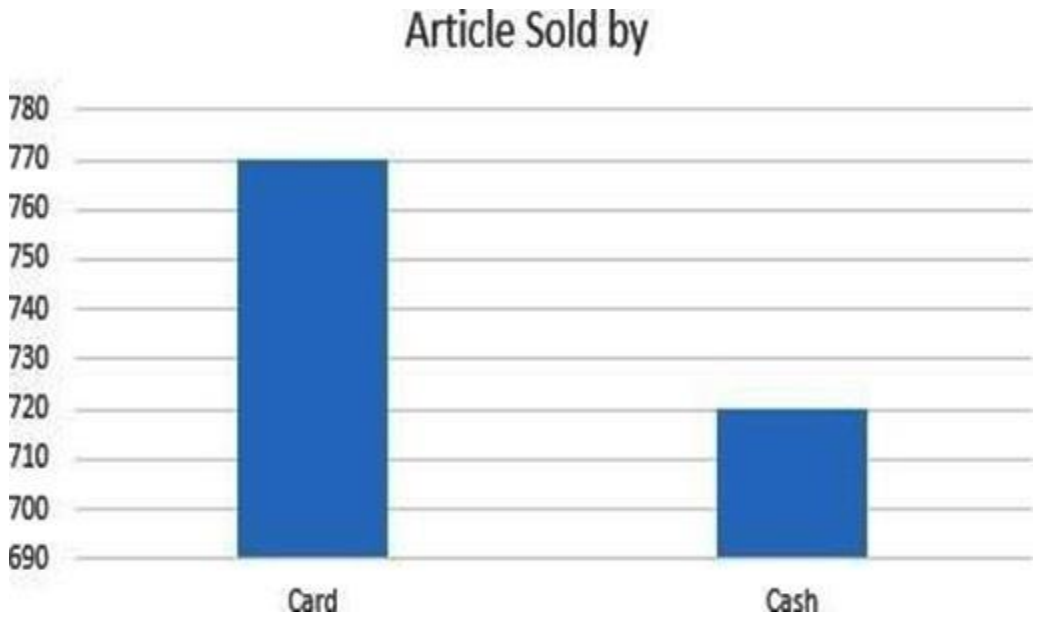

Figure 12. Total Sold Articles

Figure 12 shows the total number of articles sold in cash and by card. The number of articles and the source of the payment are taken on the x-axis and Y-axis respectively.

Conclusion and Results: The aim of developing this POS is to enable FBR to collect all the tax retailers are collecting from the consumer. There has been tax theft in which retailers collect tax on sales from consumers but

they do not pay it to the FBR. With this new point-of-sale system FBR will be able to collect all the tax with the help of an Invoice generation system. The invoice generation system will keep track of every transaction made by the retailer and how much tax has been collected on those sales and this will help the consumer to file his annual tax at the end of every year.

REFERENCES

[1] Arasa. (2020). Strategic Inventory Management Practices and the Performance of Supermarkets in Nairobi County. European Journal of Business and Management Research. doi: http://dx.doi.org/10.24018/ejbmr.2020.5.2.108

[2] Atakay. (2022). Spare Parts Inventory Management System in a Service Sector Company. Springer International Publishing.

[3] Baylosis, J. (2023). Web-based Inventory Management System. doi:https://doi.org/10.30534/ijsait/2023/021252023

[4] Chen. (2022). Development of an Inventory Management System Based on Ration Frequency Identification Technology. Sensors & Materials. doi:https://doi.org/10.18494/SAM3497

[5] Cristani, D. (2020). Inventory Management Through Mini-Dornes: Architecture and Proof of-concept Implementation. doi:: http://dx.doi.org/10.24018/ejbmr.2020.5.2.108

[6] Divya. (2021). A Study Of Inventory Management System In Tool Craft Engineering Company At Hosur.

[7] Farah. (2021). Medication inventory optimization and oral solution workflow implementation and integration into pharmacy inventory management system.

[8] Gammelli. (2022). . Predictive and prescriptive performance of bike-sharing demand forecasts for inventory management. Transportation Research Part C: Emerging Technologies.

[9] Gizaw. (2021). How is Information from ABC–VED–FNS Matrix analysis used to improve the operational efficiency of pharmaceutical inventory management? A Cross-Sectional Case Analysis. Integrated Pharmacy Research and Practice.

[10] Ho, G. T. (2021). A blockchain-based System to Enhance Aircraft Parts Traceability and trackability for Inventory management. Elsevier. doi:https://doi.org/10.1016/j.eswa.2021.115101

[11] Laping. (2023). Designing an Inventory Management System for Convenience Store X Using Design Thinking Approach. Proceedings of the International Conference on Industrial Engineering and Operations Management.

[12] Li. (2021). Research on the evaluation of real estate inventory management in China.

[13] Long. (2023). Intelligent decision support system for optimizing inventory management under stochastic events. Applied Intelligence.

[14] Menkar. (2023). Review of the Inventory Management System to Ensure the Optimization of Resources.

[15] Mishra. (2021). Optimum Sustainable Inventory Management with Backorder and Deterioration Under Controllable Carbon Emission. Journal of Cleaner Production. doi:https://doi.org/10.1016/j.jclepro.2020.123699

[16] Mohammad, S. (2020). Critical Analysis of Pharmaceuticals Inventory Management using the ABC-VEN Matrix in Dessie Referral Hospital. doi:https://doi.org/10.2147/IPRP.S265438

[17] Nguyen. (2020). Nguyen, Tam Dan Thi. "Creating an Inventory Management System using VBA for a small Vietnamese enterprise.

[18] Owori. (2022). Developing a manual inventory management system for the Uganda Institute of Banking and Financial Services (Doctoral dissertation, Makerere University).

[19] Panigrahi. (2021). Inventory management and performance of manufacturing firm. International Journal of Value Chain Management.

[20] Shirley, D. (2021). QR-Based Inventory Management System (QR-IMS) of Passenger Luggage Using Website. IEEE. doi: http://dx.doi.org/10.51505/ijaemr.2023.8317

[21] Silva-Aravena. (2020). Inventory management at a Chilean hospital pharmacy: A case study of a dynamic decision-aid tool. Mathematics.

[22] Sithiyopasakul. (2023). Inventory Management System based on IoT and Microservices Architecture Design. International Electrical Engineering Congress.

[23] Turaevich. (2023). Development of automatic storage and inventory management systems in engineering. Academia Science Repository.

[24] U, A. (2020). Improved MRO Inventory Management System in Oil and Gas Company: Increased Service Level and Reduced Average Inventory Investment Sustainability. MADPI. doi: https://doi.org/10.3390/su12198027

[25] Yusuf. (2022). Implementation of Cryptography and Access Control on IoT-Based Warehouse Inventory Management System. MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer.