

## A proposed model for enhancing e-bank transactions: an experimental comparative study

Wael Alzyadat<sup>1</sup>, Ameen Shaheen<sup>1</sup>, Ala'a Al-Shaikh<sup>2</sup>, Aysh Alhroob<sup>1</sup>, Ziyad Al-Khasawneh<sup>1</sup>

<sup>1</sup>Department of Software Engineering, Faculty of Science and Information Technology, Al-Zaytoonah University Amman, Jordan

<sup>2</sup>Department of Cyber Security, Faculty of Information Technology, Zarqa University, Zarqa, Jordan

### Article Info

#### Article history:

Received Jan 17, 2024

Revised Feb 7, 2024

Accepted Feb 20, 2024

#### Keywords:

Deep learning

Linear model

Machine learning

Random forest

SVM

### ABSTRACT

In this paper, we introduce a novel approach to address the dynamic prediction of customer activity in electronic payment transactions for individual clients. Our approach is founded on customer online payment transaction records from registered UK-based online retailers between 01/12/2009 and 09/12/2011. These retailers primarily specialize in unique gift items for various occasions, catering to a wide range of clients, including wholesalers. We used classification analysis based on the correlation coefficient to measure and describe a customer's electronic payment capability based on the quality of products they purchase. Furthermore, we trained multi-layered models (linear model, deep learning, random forest, and support vector machines (SVM)) to capture the dynamics of e-bank transaction reinforcement for retail customers using machine learning. Real transaction data from a UK online retailer was employed in our study. The experimental results consistently demonstrated the effectiveness of our proposed strategy.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



### Corresponding Author:

Ala'a Al-Shaikh

Department of Cyber Security, Faculty of Information Technology, Zarqa University

Zarqa, Jordan

Email: [ashaikh@zu.edu.jo](mailto:ashaikh@zu.edu.jo)

## 1. INTRODUCTION

In 1955, Herbert Simon, Allen Newell, and John Shaw initiated the "Theoretical Program" aimed at stimulating human problem-solving abilities, marking the inception of artificial intelligence (AI) in the realm of logic theory [1], [2]. Over the past two decades, this program has significantly contributed to the advancement of various AI technologies. As researchers gained a deeper understanding of computers and their increased storage capacity, it presented a tremendous opportunity for accountants, particularly those in the financial sector [3]. This advancement spared them the arduous task of sifting through massive volumes of data as part of their daily responsibilities [4].

The evolution of technology has catalyzed a profound transformation in the banking industry. Accounting professionals now employ information communication technology (ICT) tools, such as audit toolkits, checklists, and audit programs, to conduct in-depth data analysis and perform a majority of accounting tasks [5]. Expert models and internal controls are frequently used to assess system strengths and weaknesses. Additionally, integrated software procedures that continuously monitor real data and processing conditions have become integral [6].

The rising financial demands of society, coupled with rapid advancements in information technology, have ushered in the golden age of AI [7], [8]. The application of this technology is now an inevitable trend that promises to bring significant changes to the accounting industry [9]. Observations of

internet user behavior and consumer actions in developed countries have revealed shared characteristics [10], [11]. A dominant link has been identified between the decision to open an online account and the perceived level of security in online transactions [12].

The extensive utilization of AI in accounting-related fields like financial reporting and auditing stems from over 25 years of experience [13]. The management of technology will need to be incorporated as an independent facet of strategic planning to keep pace with evolving requirements, necessitating a deeper understanding of this discipline. It is also reasonable to expect that the demand for change will need to be met more swiftly than in the past [14].

From this perspective, this study aims to explore the role of machine learning in enhancing electronic bank transactions and its impact on these institutions. The study is titled "Enhancing electronic bank transactions using machine learning" and seeks to elucidate the pivotal role machine learning plays in bolstering electronic banking, a cornerstone of private sector financing.

The necessity detail phase represents a critical stage in the prerequisite engineering process. Visual displays serve as a means to convey semantic information in a format that can be readily comprehended and modified by the designer. The objective is to automate the modeling of requirement specifications by employing a linear model, deep learning, random forest, and support vector machines (SVM), with the inclusion of four benchmark models. The use of use cases is practical for eliciting both functional and non-functional requirements. A use case, a social model that outlines the general behavior of a system, is introduced into the requirement specifications to automatically identify erroneous relationships among use cases and actors. This automated process reduces the need for human intervention and leverages the linear model, deep learning, random forest, and SVM for relationship classification. This research serves as a valuable guide and reference for subsequent investigations, empowering experts within this industry to make more efficient algorithmic selections [15]. The banking industry has recently experienced widespread adoption of electronic banking services, where the quality, quantity, and speed of these services have become crucial elements in the competitive landscape among commercial banks, as they strive to attract and retain customers [16]. However, the utilization of electronic banking services still falls short of the desired level, even though the benefits of AI, particularly in business contexts, significantly outweigh the associated risks [17].

The primary issue addressed in this study pertains to various types of banking transactions conducted within the electronic banking platform, including activities such as remittance transfers and bill payments within the banking system. The study focuses on the dynamic prediction of electronic financial transactions using AI, as it plays a pivotal role in today's customer-centric business environment.

In this research, we will delve into the challenges and performance of the banking sector, influenced by several issues identified by [18]. These issues encompass the following: determining the transaction flow based on performance metrics by analyzing electronic payment data and validating the accuracy of the analysis through the application of multiple algorithms, comparing their levels of accuracy using the rapid miner program on the designated dataset, and identifying the most effective algorithm. We will further compare the results with those of previous studies conducted on the same dataset.

This study aims to shed light on the complexities surrounding electronic banking services and contribute to the enhancement of their efficiency and effectiveness within the banking sector. This research holds significant importance as it delves into one of the most critical financial, economic, and developmental issues of our time. Its significance is underscored by the following key elements:

- i) Global significance of electronic banking: the research addresses the pivotal role played by the electronic banking platform on an international scale. This platform is recognized for its effectiveness in fostering economic growth and contributing to comprehensive development. The study aims to underscore the importance of this role within the context of national economies [18].
- ii) Enhancing electronic banking transactions: the study explores the role of machine learning in enhancing transactions within the electronic banking platform. This enhancement is achieved through the optimization of system layers, irrespective of the specific bank branches or financial institutions involved [19].

The remainder of this paper is structured as follows: in section 2, a work related to this paper is illustrated. Then in section 3, we address our proposed model for enhancing electronic bank transactions using machine learning. Section 4 shows the experimental results and discussions. Finally, conclusions and future work.

## 2. RELATED WORK

The research endeavors to ensure the reliability of transactions across various dimensions of the system and emphasizes the need for a robust and secure dataset layer. By doing so, it strives to contribute valuable insights that can advance the field of electronic banking and its broader implications for financial and economic development. Both Arab and international studies have underscored the significance of loans in the context of enhancing electronic banking transactions through the implementation of machine learning.

Many of these studies have focused on examining the influence of electronic transactions offered by banks to cater to the needs of both institutions and individuals. Additionally, they have explored various metrics and frameworks for assessing the quality of banking services and the extent of their adoption within banking institutions. These studies offer diverse perspectives, encompassing the following areas enhancing electronic transactions, quality assessment in banking services, adoption, and application in banks. Collectively, these Arab and international studies contribute to a comprehensive understanding of the role of loans in enhancing electronic banking transactions, encompassing a wide array of facets, from the mechanisms of service provision to the assessment of service quality and the practical application within banking institutions.

### **2.1. The effect of COVID-19 on e-payment modes in Pakistan**

This study investigates the impact of e-payment modes on both providers and consumers in Pakistan during the coronavirus disease 2019 (COVID-19) pandemic. Pakistan, with its cash-based economy and limited trust in electronic commerce security, presents unique challenges for transitioning to cashless transactions. The research employs focus groups and interviews to qualitatively examine consumer and supplier behavior. The findings reveal customer dissatisfaction with the existing e-commerce system and a preference for prepayments among producers, although they are forced to offer cash on delivery. However, the pandemic has prompted an increase in cashless transactions among customers in adherence to safety protocols. The study recommends collaborative efforts between the government, private sector, and financial institutions, along with extensive marketing initiatives, to sustain this trend. While the study acknowledges limitations related to online meetings, it provides valuable insights for e-commerce stakeholders, helping them understand the drivers of consumer behavior and thereby develop effective marketing strategies, ultimately contributing to the growth and competitiveness of Pakistan's e-commerce sector [20].

### **2.2. Factors influencing gamification adoption in e-banking**

This paper explores the factors that impact the adoption of gamification in the context of e-banking. It employs the technology acceptance model (TAM) to survey 193 managers, bankers, and customers. The study identifies four key factors ease of use, usefulness, enjoyment, and convenience that positively influence the acceptance of gamification in e-banking. Furthermore, it reveals that clients engaging in gamified e-banking are more likely to manage their investments and purchase more mutual funds, resulting in increased sales. The findings provide commercial banks with insights to attract customers by expanding their network, investing in infrastructure, developing new products, enhancing safety, and building support channels. These measures aim to facilitate easier and more frequent access to e-banking services, thereby increasing the banks' competitiveness in the industry. The study also emphasizes the importance of customer responsibility in safeguarding personal information during e-banking transactions [21].

### **2.3. Core banking-redefining banking operations**

In the contemporary context, core banking represents the highest level of service a bank can offer to its customers, facilitated by cutting-edge technology. Core banking solutions (CBS) plays a central role in integrating a wide range of services that can be offered by all the bank's branches from centralized data centers [22], [23]. Beyond improving customer service, CBS enables banks to generate management information system (MIS) reports for senior management and submit various reports to regulators and government bodies. In the current banking landscape, CBS is not just an option but a necessity. Therefore, the swift adoption of CBS is not only in the best interest of banks but also benefits their customers. The ultimate beneficiaries of these consistently efficient utilities are not only financial institutions but also their customer base, collectively fostering the growth of the entire financial industry [24].

## **3. PROPOSED MODEL**

The proposed model is entitled "proposed model to enhance electronic bank transactions" using machine learning based on datasets published on the Internet and available for analysis and study. The researcher conducted several processing operations to test the results of the mechanism for enhancing electronic bank transactions using machine learning. Provides an overview of the proposed model that is designed to explore and analyze the selected dataset, to improve e-banking transactions through machine learning using rule-based machine learning [25].

The proposed model is a sequential flow that consists of several components as shown in Figure 1, The first component is the input use case diagram, the second component is identifying theories, understanding the issue, and exploring information, the third component is designating the study population and sample that represents the proposed dataset for analysis, the fourth component is cleaning the dataset and starting the information analysis, the fifth component is building the model and the last component work as

training and testing for it, so in this component is the input case for the dataset, which represented the scope of this study, and that simple data analysis method to using it in the rule-based machine learning techniques is also a part of this component.

The second component involves defining theories, understanding the issue, and exploring the available information. This component involves understanding the accessed data for analysis and extracting it to create a table or dataset. The third component is the extraction of the study population and sample that represents the proposed dataset for analysis, grouping, and sorting. In the extraction scheme, the dataset is converted into clear graphs that express the collection and sorting of data, along with their relationships. The fourth component is once the extraction scheme is completed for datasets, which is initiated by cleaning the dataset and more deeply analyzing the information by collecting and communicating it. The objective is to classify the actor, and relationship through machine learning algorithms available through and models suggested previously, which was the output of this component is structured data can be used in that models.

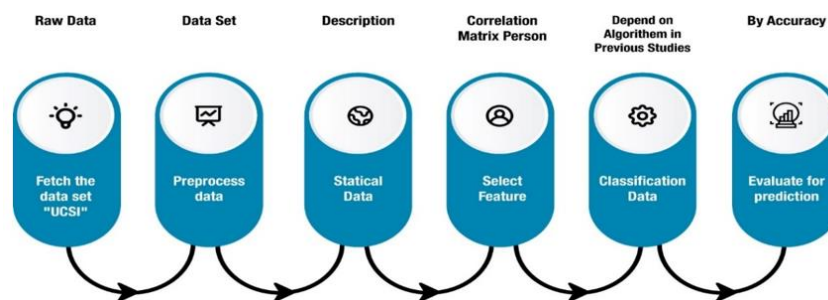


Figure 1. Proposed model

The fifth component is building the model, which involves using available classification algorithms, by browsing results to clarify the proposed model in more detail, the researcher explains each component of the model, starting from the input use case diagram, data extraction and analysis, grouping and sorting, the mechanism of building the model and accessing the process of analyzing and testing the validity of the data. All components of the proposed model used the RapidMiner software with Excel files of the CSV extension type to define the representative, and the relationship, to analyze and train with testing data.

### 3.1. Fetch the dataset from UCI component

The machine learning repository (UCI) offers a rich repository of datasets suitable for a wide range of machine learning and data science projects. It contains all of the transactions that took place for a UK-based, registered, non-store online retail business between January 12, 2009 and December 9, 2011. The business primarily sells unique giftware for all occasions. Wholesalers make up the majority of the company's clients.

### 3.2. Preprocess data component

Data cleaning is a meticulous procedure that encompasses a thorough examination of a dataset. It entails the identification and rectification of missing values, errors, and inconsistencies using suitable techniques like mean imputation or deletion. Once this process is complete, the cleaned dataset is securely stored in a designated location for subsequent use. This meticulous attention to data quality and accuracy substantially influences the dependability and trustworthiness of data-driven conclusions and insights. The result is a dataset that is free of errors and discrepancies, assuring precise, consistent, and reliable data for analysis.

### 3.3. Statistical data component

Statistical data comprises a collection of numerical or categorical information obtained through diverse methods like surveys, experiments, or observations [26]. This data is subject to analysis and interpretation, facilitating conclusions, predictions, and informed decision-making. Various methods and techniques exist for statistical data analysis, including descriptive statistics (e.g., mean, median, mode), inferential statistics (e.g., hypothesis testing, confidence intervals), and regression analysis, which identifies relationships between variables. Statistical data can be presented in formats like tables, graphs, and charts for easier comprehension. Additionally, assessing data quality measures like accuracy, completeness, and

consistency ensures the reliability and validity of statistical data [27]. In the realm of common statistical methods, the equations are applied:

- Mean: the arithmetic mean of a set of values is calculated by adding all values together and then dividing by the total count [28].

$$\text{mean} = \frac{(\text{sum of values})}{\text{number of values}} \quad (1)$$

- Median: in an ordered set of values, the median is the middle value, found after arranging values from smallest to largest [28].

$$\text{Median} = \left(\frac{n+1}{2}\right)\text{th} \quad (2)$$

Where 'n' is the number of items in the data set, and 'th' signifies the (n)th number.

- Mode: the mode is the most frequently occurring value in a set of data.
- Standard deviation: this measures the spread or variation within a set of values and is derived from the square root of the variance [28].

$$\text{standard deviation} = \text{sqrt}(\text{variance}) \quad (3)$$

- Variance: it quantifies the extent to which values in a set deviate from the mean. Variance is computed by summing the squared differences between each value and the mean, then dividing by the count of values minus one [29].

$$\text{variance} = \frac{\text{sum}((\text{value} - \text{mean})^2)}{\text{number of values} - 1} \quad (4)$$

- Regression analysis: it's a statistical technique demonstrating relationships between two or more factors, commonly depicted by simple linear regression [30].

$$y = a + bx \quad (5)$$

where 'a' is the y-intercept, 'b' is the slope, 'y' is the dependent variable, and 'x' is the independent variable. The goal is to estimate the best-fitting values for 'a' and 'b' based on the data.

### 3.4. Select feature component

To identify the most pertinent features based on correlation results, the following procedural steps should be undertaken:

- Understanding correlation: correlation, a statistical metric, quantifies the degree of interrelation between two or more variables. Within the correlation matrix of the UCI dataset, it calculates significant positive or negative associations between attributes. It serves to gauge the strength and direction of connections among variables. Correlation values range from -1 to +1, where -1 signifies a perfect negative correlation (inverse relationship), +1 indicates a perfect positive correlation (direct relationship), and 0 denotes no correlation at all. This definition is based on [31].
- Visualization of correlation: a scatter plot, where one variable's values are plotted on the x-axis and the others on the y-axis, can be employed to visualize the correlation. An upward-sloping line from left to right on the scatter plot denotes a positive correlation, while a downward-sloping line signifies a negative correlation. Scattered points with no discernible pattern indicate no correlation [32].
- Data analysis utility: correlation is commonly applied in data analysis to unveil patterns and relationships among variables. It aids in recognizing potential predictors for an outcome variable, assessing the strength of relationships, or pinpointing confounding variables. It is crucial to emphasize that correlation does not infer causation, and other influencing factors must be considered before drawing causal inferences.
- Selecting highly correlated variables: variables exhibiting the highest correlation coefficients are identified as either positively or negatively related. These variables are considered the most interrelated.
- Correlation heatmap: a heatmap of the correlation matrix is employed to visually identify highly correlated variables. This facilitates the identification of redundant variables that can be prudently excluded from the dataset.

- vi) Feature selection methods: implement feature selection techniques such as recursive feature elimination or principal component analysis to discern the most critical features. These methods aim to reduce the dataset's feature count while preserving its predictive capacity.
- vii) Model performance assessment: evaluate the model's performance using the selected features and compare it to the original model employing all features. If the performance remains comparable or improves, adopting the reduced feature set is advisable.

These comprehensive steps collectively enable the identification of salient features and the refinement of datasets for effective machine-learning analysis.

### 3.5. Classification data component

Classification techniques encompass various approaches to solving classification tasks. Linear models are valuable tools when the relationship between predictors and outcomes is presumed to be linear. However, they may not be the ideal choice for all datasets, particularly those exhibiting intricate, non-linear relationships between features and outcomes. In such cases, deep learning stands out as a potent resource, particularly when confronted with complex and high-dimensional datasets. For classification tasks necessitating high accuracy, resilience, and swiftness, random forest emerges as a robust option. Lastly, SVMs represent a versatile and formidable algorithm suited for a broad spectrum of classification tasks, excelling especially in the management of high-dimensional datasets. In our analysis, we gauged the accuracy of each model individually and employed it as a benchmark for comparing the four algorithms.

### 3.6. Evaluate prediction

The realm of data mining and machine learning acknowledges that an imbalanced class distribution within a dataset can detrimentally affect the performance of trained classifiers. While various techniques have been proposed to address such imbalances, most of them lack a solid theoretical foundation. Curiously, conventional theoretical analyses of machine learning don't account for the impact of class imbalance on classifier performance. The foundational principles of statistical learning establish the requisite number of examples to estimate a classifier's accuracy based on its complexity (VC dimension) and the desired level of confidence. Strikingly, these formulas do not consider class imbalance.

In this study, we concentrate on assessing classifier performance by utilizing precision and recall measures, which are frequently recommended for handling imbalanced data classification.

$$Accuracy = Precision / Recall \quad (6)$$

Upon close examination, it becomes evident that when precision is reasonably high, the discrepancy between precision and recall is only a minor constant factor away from the accuracy, which factors into the class imbalance. This observation leads to the corollary that a greater number of examples is both necessary and sufficient to address class imbalance effectively. This conclusion is further substantiated by empirical evidence.

## 4. EXPERIMENTAL RESULTS

A proposed model aimed at enhancing e-bank transactions was tested by researchers using four distinct models on the "Online Retail II" dataset from the UCI library e-store. This dataset was considered suitable after thorough preprocessing. The models employed for classification were the linear model, deep learning, random forest, and SVM, implemented through the RapidMiner software. The outcome of this analysis revealed a 90% correct classification rate, with 10% of cases being classified incorrectly.

To explain the proposed model in more detail, and conduct experiments and the result of each component, the researcher explains each component by step and result, and after that summarizes its result in for each component. Where is organized as follows: start with the input use case diagram, extract diagram, grouping and sorting, then follow with support vector machine SVM, linear model, deep learning, random forest and models the predication and lastly a summary of experiments.

### 4.1. Fetch the data set

The input data for the Online Retail II dataset is a collection of transactional data from an online retailer in the United Kingdom. The data includes details of purchases made by customers over a period of two years, from December 2009 to December 2011. The output data for the Online Retail II dataset can be various insights and information that can be derived from the input data through data analysis and processing. The process of analyzing the Online Retail II dataset can involve various steps such as data cleaning, data transformation, data exploration, statistical modeling, and visualization.

#### 4.2. Preprocess data

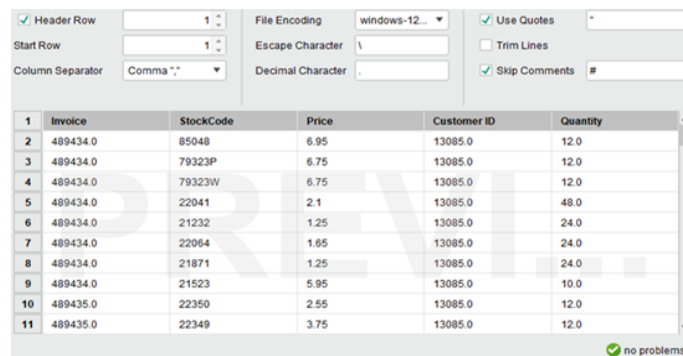
The initial phase of this procedure involves utilizing the data retrieved during the first stage of the Online Retail II dataset. The ensuing output comprises the preprocessed iteration of the Online Retail II dataset, intended for subsequent analytical endeavors. The refined dataset, poised for statistical analysis, is delineated in Figure 2, encompassing the following attributes:

- i) Invoice No: This pertains to the numerical identifier featured on the invoice, denoting a nominal value. Each transaction is assigned a unique six-digit integral number, with instances of cancellations indicated by the prefix "c."
- ii) StockCode: A nominal identifier representing the code allocated to each distinct product or item, characterized by a five-digit numerical code.
- iii) Price: The nominal unit cost associated with each product.
- iv) Customer ID: A nominal identifier comprising a unique five-digit integral number assigned to individual customers.
- v) Quantity: A numeric indicator denoting the quantity of each item per transaction.
- vi) Country: The nominal designation of the country of residence for each customer.

These attributes form the basis for subsequent statistical analyses and modeling within the context of the Online Retail II dataset.

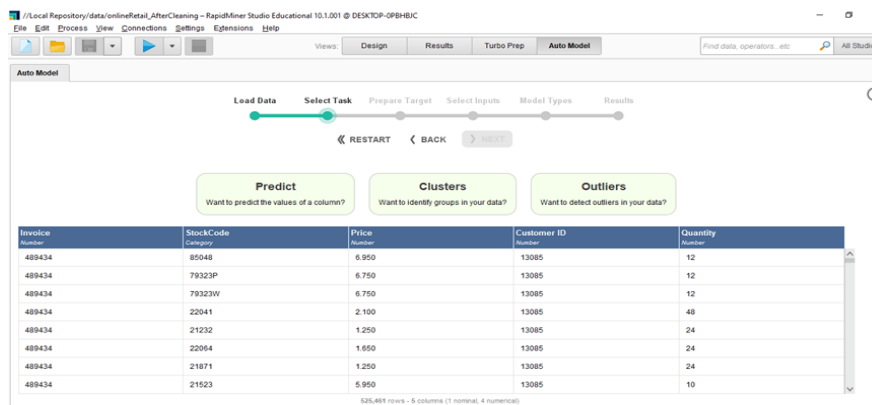
#### 4.3. Preprocess data

In the step of analyzing the data that was listed and clarified in the second stage, after cleaning the data well, we notice the process of selecting auto model, through which we can choose the dataset to be analyzed and then choose the type of analysis, whether it is for prediction or collection, as the researcher's choice here is the class classification which stands for expectation since the dataset has been labeled and the quality column has been placed as the main column for the analysis as shown in Figure 3. The output data will be the statistical analysis report, which will include various measures such as mean, median, mode, standard deviation, correlation coefficient, and more.



|    | Invoice  | StockCode | Price | Customer ID | Quantity |
|----|----------|-----------|-------|-------------|----------|
| 1  |          |           |       |             |          |
| 2  | 489434.0 | 85048     | 6.95  | 13085.0     | 12.0     |
| 3  | 489434.0 | 79323P    | 6.75  | 13085.0     | 12.0     |
| 4  | 489434.0 | 79323W    | 6.75  | 13085.0     | 12.0     |
| 5  | 489434.0 | 22041     | 2.1   | 13085.0     | 48.0     |
| 6  | 489434.0 | 21232     | 1.25  | 13085.0     | 24.0     |
| 7  | 489434.0 | 22064     | 1.65  | 13085.0     | 24.0     |
| 8  | 489434.0 | 21871     | 1.25  | 13085.0     | 24.0     |
| 9  | 489434.0 | 21523     | 5.95  | 13085.0     | 10.0     |
| 10 | 489435.0 | 22350     | 2.55  | 13085.0     | 12.0     |
| 11 | 489435.0 | 22349     | 3.75  | 13085.0     | 12.0     |

Figure 2. Specify data set with format columns



Auto Model

Load Data → **Select Task** → Prepare Target → Select Inputs → Model Types → Results

RESTART ← BACK → NEXT

**Predict**  
Want to predict the values of a column?

**Clusters**  
Want to identify groups in your data?

**Outliers**  
Want to detect outliers in your data?

| Invoice Number | StockCode | Price | Customer ID | Quantity |
|----------------|-----------|-------|-------------|----------|
| 489434         | 85048     | 6.950 | 13085       | 12       |
| 489434         | 79323P    | 6.750 | 13085       | 12       |
| 489434         | 79323W    | 6.750 | 13085       | 12       |
| 489434         | 22041     | 2.100 | 13085       | 48       |
| 489434         | 21232     | 1.250 | 13085       | 24       |
| 489434         | 22064     | 1.650 | 13085       | 24       |
| 489434         | 21871     | 1.250 | 13085       | 24       |
| 489434         | 21523     | 5.950 | 13085       | 10       |

326,465 rows - 5 columns (1 nominal, 4 numerical)

Figure 3. Select labeled column to begin analysis to data set

**4.4. Select feature**

At this step, the characteristics that the researcher would like to work on were chosen, as the quality column was chosen for the process of starting the analysis to carry out the classification and expectation of the correct results and the percentage that should be collected based on a specific column. Quality was selected for the initiation process of the analysis. The input data is the preprocessed dataset with the selected labeled column as shown in Figure 4, where the output data is a subset of the original dataset that contains only the relevant features for the analysis. The process involves selecting the relevant columns/features from the dataset based on their significance and relevance to the analysis objectives. For instance, in the given scenario, the "Quality" column was selected for the initiation process of the analysis as shown in Figure 4.

As illustrated in Figure 5, the data of the quality column selected to be analyzed is displayed, and this column is selected as the input for the chosen models to initiate the analysis. There are four columns that the researcher intends to analyze concerning the quality column. It is observed that the customer's column has the highest correlation coefficient.

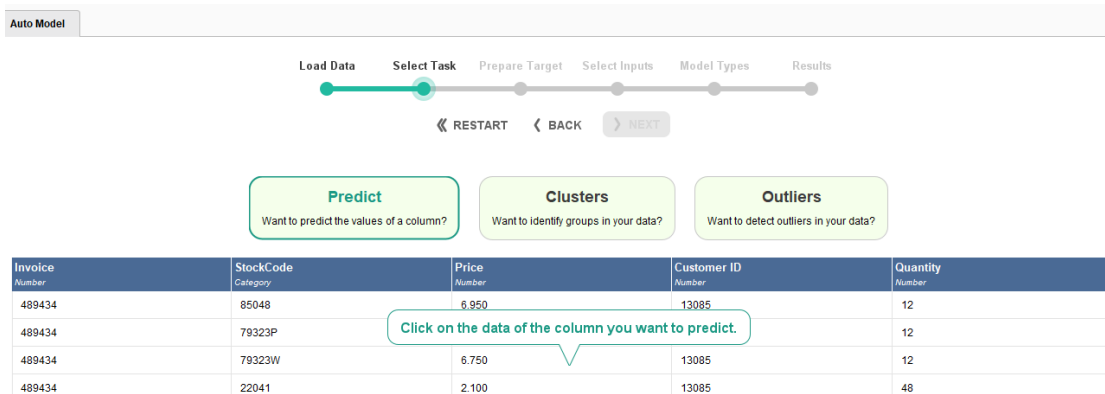


Figure 4. "Quality" column was selected for the initiation process

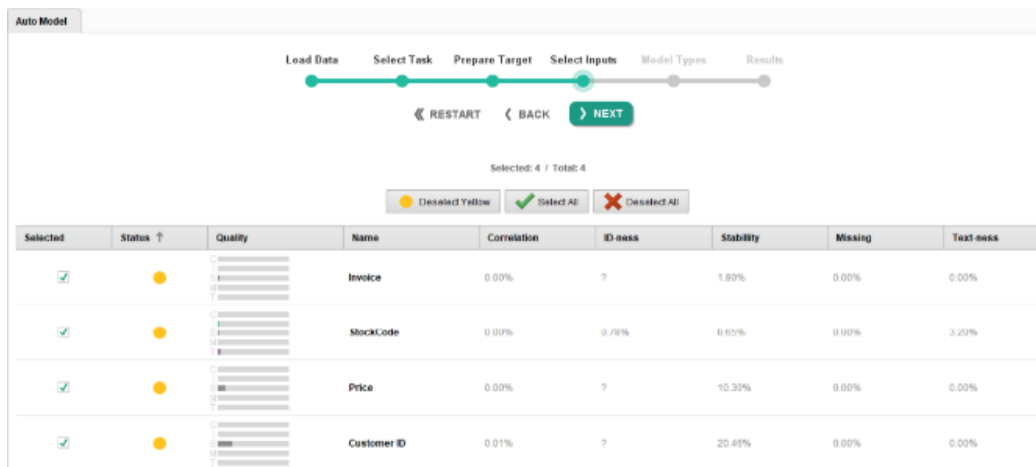


Figure 5. Input data to model type stage

**4.5. Classification data**

In this step, the input is the preprocessed dataset and the output is the predicted classification label for each instance as shown in Figure 5. The models are evaluated based as output on their accuracy and other performance metrics, and the best-performing model is selected as the final classification model. Once the model is trained, it can be used to predict the classification label of new instances of the same dataset or similar datasets by selecting models.

The process involves training and testing several classification models, namely SVM, linear model, deep learning, and random forest, using the selected features from the previous step as shown in Figure 6.



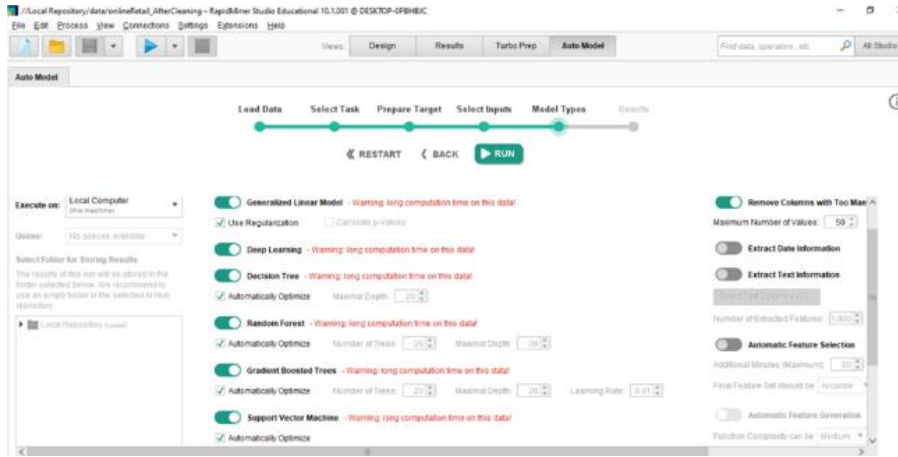


Figure 6. Models and algorithms that represent workable models on the chosen dataset to begin analyst

**4.6. Evaluate for predicate**

The researcher notes the emergence of results after working with these algorithms and shows several special results from standard regression, maximum values, minimum values, and line coefficients for that data. It also shows the total time spent for each algorithm in the training and testing work of that algorithm. There are also other details of the data, including showing the correlation coefficient in each of the algorithms that have been worked on as shown in Figure 7.

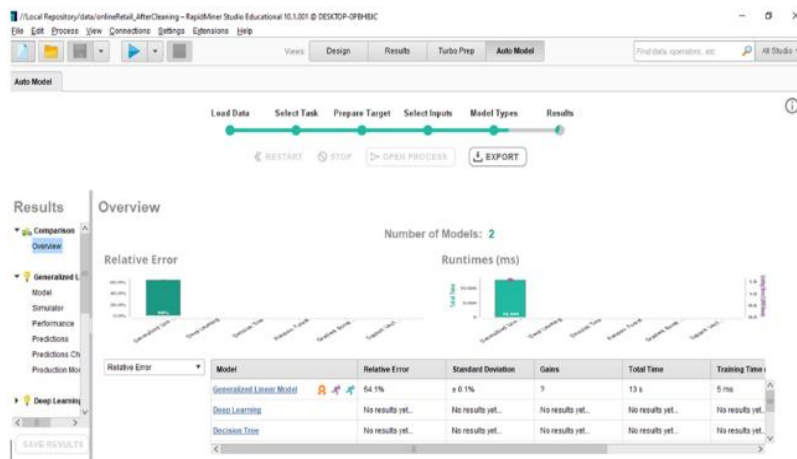


Figure 7. Final result to an analyst for models (ready for training and testing)

**4.7. Prediction**

In the last component of our proposed model, the researcher compares the output of the SVM, linear model, deep learning, and random forest as shown in Figure 8. Where the (7) is used to determine the model's accuracy:

$$Accuracy = \left( \frac{True\ Positives + True\ Negatives}{True\ Positives + False\ Positives + True\ Negatives + False\ Negatives} \right) \tag{7}$$

where:

- True positives: the number of correct positive predictions.
- True negatives: the number of correct negative predictions.
- False positives: the number of incorrect positive predictions.
- False negatives: the number of incorrect negative predictions.

For instance, suppose we have a double characterization issue where we need to anticipate regardless of whether a client will purchase an item. Have a dataset of 100 clients, and we fabricate a model that predicts the buy conduct of every client. The model correctly predicts 80 customers who will purchase and 10 customers who will not; however, it incorrectly predicts 5 customers who will purchase and 5 customers who will not purchase. Using the equation above, we can calculate the accuracy of the model as (8). This means that the model correctly predicts the purchase behavior of 90% of the customers in the dataset.

$$Accuracy = \left( \frac{80 + 10}{80 + 5 + 10 + 5} \right) = 90\% \tag{8}$$

Figure 8 illustrates the preprocessing of columns derived from a comprehensive store database, encompassing information about purchases and invoice numbers. These columns encompass data about the product number, invoice price, customer details, and the number of items purchased following the extraction of data from the Excel file. SVM, in particular, yielded impressive results, achieving a precision of 90.3%, a recall of 88.6%, and an F-measure of 87.7%. In contrast, other models (linear model, deep learning, random forest) exhibited lower performance, with an accuracy of 76.2%, a recall of 76%, and an F-measure of 72.1%.

Precision, a critical metric, was notably higher in the proposed model, at 94.7%, compared to the related work's 76.2%. Additionally, the recall rate for the proposed model stood at 94.3%, surpassing the related work's 76%. The precise nature of the collected data contributed to this enhanced recall. These results remained constant in all six iterations, with each iteration employing different machine learning algorithms, and the dataset consistently sourced from the open-source UCI library datasets.

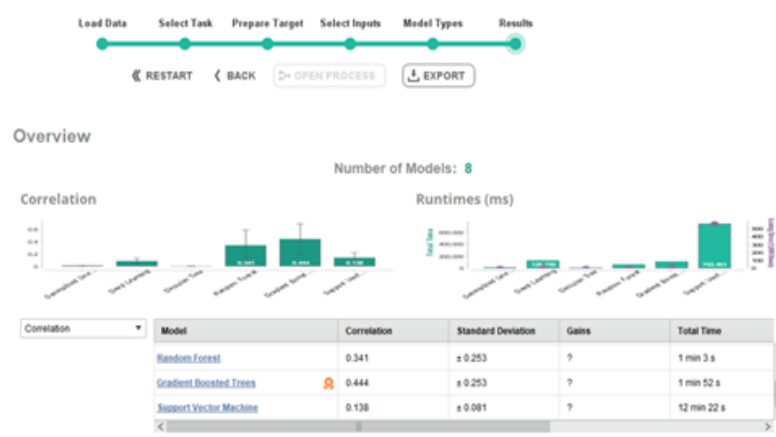


Figure 8. Final result

### 5. CONCLUSION





The study successfully achieved its intended objectives, which encompassed employing machine learning algorithms to investigate a select dataset related to a store's sales, specifically focusing on electronic sales and payments. The study sought to explore the impact of various factors such as product quality, price, and their association with electronic purchase promotions. The SVM model exhibited notable precision, achieving a rate of 90.3%, along with a recall rate of 88.6% and an F-measure of 87.7%. In contrast, the related work showed lower performance, with a precision of 76.2%, recall of 76%, and an F-measure of 72.1% for the other models (linear model, deep learning, random forest). Furthermore, the study employed precision as a key metric, yielding a significantly higher precision rate of 94.7% in comparison to the related work's 76.2%. The recall rate for the study also stood out, achieving a rate of 94.3%, surpassing the related work's 76%. This was attributed to the high precision of the collected data, resulting in an elevated recall. These findings were consistent across six iterations, each utilizing different machine learning algorithms, with the dataset originating from the open-source UCI library datasets.

### REFERENCES





[1] A. Tang *et al.*, "Canadian association of radiologists white paper on artificial intelligence in radiology," *Canadian Association of Radiologists Journal*, vol. 69, no. 2, pp. 120–135, May 2018, doi: 10.1016/j.carj.2018.02.002.  
 [2] R. Alazaidah *et al.*, "Website phishing detection using machine learning techniques," *Journal of Statistics Applications and Probability*, vol. 13, no. 1, pp. 119–129, Jan. 2024, doi: 10.18576/jsap/130108.

- [3] S. Kapoor, A. Sharma, A. Verma, V. Dhull, and C. Goyal, "A comparative study on deep learning and machine learning models for human action recognition in aerial videos," *The International Arab Journal of Information Technology*, vol. 20, no. 4, 2023, doi: 10.34028/iajit/20/4/2.
- [4] B. Ljungberg *et al.*, "European association of urology guidelines on renal cell carcinoma: the 2019 update," *European Urology*, vol. 75, no. 5, pp. 799–810, May 2019, doi: 10.1016/j.eururo.2019.02.011.
- [5] H. A. Hamzah and M. Sadry Abu Seman, "The role of web engineering in e-learning application development: a review study," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 27, no. 3, p. 1576, Sep. 2022, doi: 10.11591/ijeecs.v27.i3.pp1576-1588.
- [6] J. Dai and M. A. Vasarhelyi, "Imagineering audit 4.0," *Journal of Emerging Technologies in Accounting*, vol. 13, no. 1, pp. 1–15, Dec. 2016, doi: 10.2308/jeta-10494.
- [7] B. Edwin *et al.*, "Smart agriculture monitoring system for outdoor and hydroponic environments," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 25, no. 3, p. 1679, Mar. 2022, doi: 10.11591/ijeecs.v25.i3.pp1679-1687.
- [8] S. Berhil, H. Benlahmar, and N. Labani, "A review paper on artificial intelligence at the service of human resources management," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 18, no. 1, p. 32, Apr. 2020, doi: 10.11591/ijeecs.v18.i1.pp32-40.
- [9] A. Shankar and C. Jebarajakirthy, "The influence of e-banking service quality on customer loyalty," *International Journal of Bank Marketing*, vol. 37, no. 5, pp. 1119–1142, Jul. 2019, doi: 10.1108/IJBM-03-2018-0063.
- [10] A. Al-Fayoumi and M. Hassan, "Interactive visual search system based on machine learning algorithm," *The International Arab Journal of Information Technology*, vol. 16, no. 1A, 2016.
- [11] A. S. Al-Sherideh *et al.*, "Development of a secure model for mobile government applications in Jordan," *Journal of Statistics Applications and Probability*, vol. 13, no. 1, pp. 145–155, Jan. 2024, doi: 10.18576/jsap/130110.
- [12] O. Raphaeli, A. Goldstein, and L. Fink, "Analyzing online consumer behavior in mobile and PC devices: a novel web usage mining approach," *Electronic Commerce Research and Applications*, vol. 26, pp. 1–12, Nov. 2017, doi: 10.1016/j.elerap.2017.09.003.
- [13] T. S. Szopiński, "Factors affecting the adoption of online banking in Poland," *Journal of Business Research*, vol. 69, no. 11, pp. 4763–4768, Nov. 2016, doi: 10.1016/j.jbusres.2016.04.027.
- [14] M. Alghaili, A. J. AlBdairi, Z. Li, and M. Katiyalath, "Generating embedding features using deep learning for ethnics recognition," *The International Arab Journal of Information Technology*, vol. 20, no. 4, 2023, doi: 10.34028/iajit/20/4/13.
- [15] M. Geogy and A. Dharani, "A scrutiny of the software requirement engineering process," *Procedia Technology*, vol. 25, pp. 405–410, 2016, doi: 10.1016/j.protcy.2016.08.125.
- [16] G. Elumalai and M. Ganesan, "Deep learning based hand wrist segmentation using mask R-CNN," *The International Arab Journal of Information Technology*, vol. 19, no. 5, 2022, doi: 10.34028/iajit/19/5/10.
- [17] M. M. Mohd Sharif, R. Maskat, Z. Baharum, and K. Maskat, "A scoping review of topic modelling on online data," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 31, no. 3, p. 1633, Sep. 2023, doi: 10.11591/ijeecs.v31.i3.pp1633-1641.
- [18] A. A. Chahadah, G. A. E. Refae, and A. Qasim, "The use of data mining techniques in accounting and finance as a corporate strategic tool: an empirical investigation on banks operating in emerging economies," in *International Journal of Economics and Business Research*, 2018, pp. 442–452, doi: 10.1504/IJEBR.2018.092149.
- [19] B. Chaimaa, E. Najib, and H. Rachid, "E-banking overview: concepts, challenges and solutions," *Wireless Personal Communications*, vol. 117, no. 2, pp. 1059–1078, Mar. 2021, doi: 10.1007/s11277-020-07911-0.
- [20] F. Khan, S. Ateeq, M. Ali, and N. Butt, "Impact of COVID-19 on the drivers of cash-based online transactions and consumer behaviour: evidence from a Muslim market," *Journal of Islamic Marketing*, vol. 14, no. 3, pp. 714–734, Feb. 2023, doi: 10.1108/JIMA-09-2020-0265.
- [21] T.-L. Anh, V.-P. G. Anh, N.-M. T. Quynh, N.-N. Le, T.-N. T. Huyen, and H.-N. Trung, "Factors affecting the acceptance of gamification application in e-banking," *International Journal of Data and Network Science*, vol. 7, no. 2, pp. 601–608, 2023, doi: 10.5267/j.ijdns.2023.3.011.
- [22] G. Samara, "Certificate revocation management in VANET," *International Journal of Cyber-Security and Digital Forensics (IJCSDF)*, vol. 1, no. 2, pp. 115–121, 2012.
- [23] M. A. A. Rahman, B. Ismail, K. Naidu, and M. K. Rahmat, "Review on population-based metaheuristic search techniques for optimal power flow," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 15, no. 1, p. 373, Jul. 2019, doi: 10.11591/ijeecs.v15.i1.pp373-381.
- [24] A. Sarkar and J. Singh, "Financing energy efficiency in developing countries-lessons learned and remaining challenges," *Energy Policy*, vol. 38, no. 10, pp. 5560–5571, Oct. 2010, doi: 10.1016/j.enpol.2010.05.001.
- [25] N. R. Al-Milli, "Hybrid genetic algorithms with simulating annealing for university course timetabling problems," *Journal of Theoretical and Applied Information Technology*, vol. 29, no. 2, pp. 100–106, 2011.
- [26] A. Al-Shaikh, A. Shaheen, M. Rasmi Al-Mousa, K. Alqawasmii, A. S. Al Sherideh, and H. Khattab, "A comparative study on the performance of 64-bit ARM processors," *International Journal of Interactive Mobile Technologies (IJIM)*, vol. 17, no. 13, pp. 94–113, Jul. 2023, doi: 10.3991/ijim.v17i13.39395.
- [27] R. Mumtaz, V. Samawi, A. Alhroob, W. Alzyadat, and I. Almukahel, "PDIS: a service layer for privacy and detecting intrusions in cloud computing," *International Journal of Advances in Soft Computing and its Applications*, vol. 14, no. 2, pp. 15–35, Aug. 2022, doi: 10.15849/IJASCA.220720.02.
- [28] L. Lista, *Statistical methods for data analysis in particle physics*, vol. 941. in *Lecture Notes in Physics*, vol. 941. Cham: Springer International Publishing, 2017. doi: 10.1007/978-3-319-62840-0.
- [29] E. A. Curtis, C. Comiskey, and O. Dempsey, "Importance and use of correlational research," *Nurse Researcher*, vol. 23, no. 6, pp. 20–25, Jul. 2016, doi: 10.7748/nr.2016.e1382.
- [30] G. James, D. Witten, T. Hastie, and R. Tibshirani, "An introduction to statistical learning," in *Springer Texts in Statistics*, vol. 103. New York, NY: Springer New York, 2013. doi: 10.1007/978-1-4614-7138-7.
- [31] A. Sciacchitano and B. Wieneke, "PIV uncertainty propagation," *Measurement Science and Technology*, vol. 27, no. 8, p. 084006, Aug. 2016, doi: 10.1088/0957-0233/27/8/084006.
- [32] A. Shaheen, R. Al-Sayyed, and A. Sleit, "Improving visual analyses and communications of ontology by dynamic tree (case study: computer system)," *International Journal of Advanced and Applied Sciences*, vol. 4, no. 5, pp. 62–66, 2017, doi: 10.21833/ijaas.2017.05.011.





**BIOGRAPHIES OF AUTHORS**

**Wael Alzyadat**     is Associate Professor who works at the Software Engineering Department, Al-Zaytoonah University of Jordan. The ongoing projects are the Development of a New Big Data Model for Decision Making: Case Study -Drugs Chemical Structures and Big Data coherence with KDD. He can be contacted at email: wael.alzyadat@zu.edu.jo.







**Ameen Shaheen**     received his B.Sc. in Computer Information System in 2010, and MS in Computer Information System in 2012, and Ph.D. in Computer Science in 2018. He currently works as Assistant Professor at the Software Engineering Department, at Al-Zaytoonah University of Jordan. His expertise is in database, algorithms, and systems analysis. He can be contacted at email: a.shaheen@zu.edu.jo.







**Ala'a Al-Shaikh**     is an Assistant Professor of Computer Science at the Department of Cybersecurity, the Faculty of Information Technology, Zarqa University, Jordan. He received his Ph.D. in Computer Science from the University of Jordan in 2020. He served in different academic institutions in Jordan and Saudi Arabia. His greatest-of-all-time achievement that he is always proud of is the computerization of the General Associate Degree Examination in Jordan, which is the second national examination in Jordan that grants students who pass it the associate degree as per the Law of Higher Education of Jordan. His research interests are in the fields of computational intelligence, machine learning, artificial intelligence, parallel computing, distributed systems, cloud computing, algorithms, e-learning, and cybersecurity. He can be contacted at email: ashaikh@zu.edu.jo.



**Aysh Alhroob**     currently works as a Dean, the Faculty of Science and Information Technology and a full Professor at the Department of Software Engineering, Al-Zaytoonah University, Jordan. He has done his Ph.D. at the University of Bradford, UK. His expertise is in data and text mining, big data analysis, software testing, and software requirements. He can be contacted at email: aysh@zu.edu.jo.



**Ziyad Al-Khasawneh**     received his B.Sc. in Software Engineering in 2010, and MS in Software Engineering in 2023. He worked as Senior Dotnet Developer at Access to Arabia company. He can be contacted at email: ziyad.khasawneh@zu.edu.jo.