

# International Journal of Research in Marketing Management and Sales



E-ISSN: 2663-3337  
P-ISSN: 2663-3329  
[www.marketingjournal.net](http://www.marketingjournal.net)  
IJRMMS 2024; 6(1): 120-124  
Received: 07-01-2024  
Accepted: 03-02-2024

**Avik Roy**  
Department of Computer  
Science and Engineering,  
Swami Vivekananda  
University, Barrackpore, West  
Bengal, India

**Nibedita Dey**  
Department of Computer  
Science and Engineering,  
Swami Vivekananda  
University, Barrackpore, West  
Bengal, India

**Joy Chakraborty**  
Department of Computer  
Science and Engineering,  
Swami Vivekananda  
University, Barrackpore, West  
Bengal, India

**Ashis Kumar Prasad**  
Department of Computer  
Science and Engineering,  
Swami Vivekananda  
University, Barrackpore, West  
Bengal, India

**Priyanshu Chakraborty**  
Department of Computer  
Science and Engineering,  
Swami Vivekananda  
University, Barrackpore, West  
Bengal, India

**Somsubhra Gupta**  
Department of Computer  
Science and Engineering,  
Swami Vivekananda  
University, Barrackpore, West  
Bengal, India

**Corresponding Author:**  
**Avik Roy**  
Department of Computer  
Science and Engineering,  
Swami Vivekananda  
University, Barrackpore, West  
Bengal, India

## Auto-sales: A machine learning and data mining driven business analytics system

**Avik Roy, Nibedita Dey, Joy Chakraborty, Ashis Kumar Prasad,  
Priyanshu Chakraborty and Somsubhra Gupta**

DOI: <https://doi.org/10.33545/26633329.2024.v6.i1b.160>

### Abstract

This work proposes a dynamic and user-friendly system for sales analysis and optimization by proposing interesting measures in sales report through Data Mining and providing subsequent classified business analytics through autonomous Machine Learning. This system is designed to serve as a powerful tool for users to submit their business data and receive interesting analysis in return.

This web-based system provides a platform in which the user can input their sales data at ease regardless of the size and nature of their business. By doing so, they will be able to harness the power of data analysis to make informed decisions, identify patterns, and discover opportunities for growth. The security and privacy to the user data is fully ensured through registered account only accessible to them. Once logged in, they can conveniently enter their sales information, including relevant details such as transaction dates, product descriptions, quantities sold, and prices.

The real value of the system lies in its ability to analyze the sales data provided by users. By leveraging advanced algorithms and data science/ computational statistics techniques, the system generate insightful reports and visualizations that showcase trends, patterns, and key performance indicators. These reports will provide users with a comprehensive overview of their sales performance and allow them to identify areas of improvement or success.

The proposed system provides user-friendly interface will make it easy for users to navigate through their personalized sales analysis. They can view their reports, explore different time periods, compare sales performance across products or regions, and gain valuable insights that can drive their business decisions. This is different from erstwhile MIS based decision support system in terms of use of Machine learning classifiers that enables the system as autonomous and self-reliant.

In summary, this system empowers users by creating a sales analysis that enables them to submit their sales data, receive comprehensive analysis reports, and make data-driven decisions to enhance their business performance.

**Keywords:** Auto-sales, business analytics, data mining, machine learning, sales analysis

### 1. Introduction

This presentation presents a groundbreaking system designed to revolutionize sales analysis and optimization in a dynamic and user-friendly manner. This innovative approach leverages Data Mining and autonomous Machine Learning to propose unique measures in sales reporting and to provide classified business analytics. The system's primary aim is to empower users by offering a robust platform where they can submit their business data and, in return, receive captivating and informative analyses.

In today's ever-evolving business landscape, data-driven decision-making is paramount. This web-based system addresses this need by offering a seamless platform that accommodates users of all business sizes and types. Regardless of whether a business is a budding startup or an established enterprise, this system offers an accessible and intuitive means to input sales data. By doing so, users unlock the potential of data analysis, enabling them to make informed decisions, discern intricate patterns, and unearth opportunities for growth. Importantly, the system places a premium on the security and privacy of user data, ensuring that access is exclusively granted to registered accounts.

Upon logging in, users are presented with a convenient interface to enter various aspects of their sales information, including transaction dates, product descriptions, quantities sold, and prices. Yet, the true essence of this system emerges when it comes to the analysis of the sales data provided by users.

This is where advanced algorithms, data science techniques, and computational statistics play a pivotal role. The system excels in generating insightful reports and vivid visualizations that vividly highlight trends, patterns, and key performance indicators. These reports transcend mere data points; they provide users with a comprehensive panorama of their sales performance, affording them the ability to pinpoint areas for enhancement or celebrate successes.

One of the system's key strengths is its user-friendliness. It offers a seamless and intuitive interface that simplifies navigation through personalized sales analyses. Users can effortlessly access and explore their reports, delve into different time periods, compare sales performance across products or regions, and ultimately glean invaluable insights that can steer their business decisions towards success. Notably, the system distinguishes itself from traditional Management Information Systems (MIS) by integrating Machine Learning classifiers, imbuing it with autonomy and self-reliance.

In summation, this system is a catalyst for empowerment. It empowers users by facilitating the submission of their sales data, delivering comprehensive analysis reports, and, most importantly, fostering data-driven decision-making to elevate their business performance. In a data-centric era, this system stands as a beacon, illuminating the path towards more informed, strategic, and ultimately successful business endeavors.

## 2. Methodological aspects

### 2.1. Data Collection

**User Generated Data:** Auto-sales data collection in this system is user-driven. Users, typically businesses or individuals involved in auto-sales, willingly input their sales-related information into the web-based platform.

**Ease of Submission:** The system is designed to make it as easy as possible for users to input their sales data. This user-friendly interface encourages active participation regardless of the size or nature of the user's business.

**Data Elements:** Users are prompted to enter various details about their sales transactions. This may include transaction dates, product descriptions (such as vehicle make and model), quantities sold, and prices. The system may offer templates or forms for structured data entry.

**Data Validation:** Depending on system requirements, data validation mechanisms may be in place to ensure the accuracy and consistency of user-submitted data. This step helps maintain data quality.

**Data Storage:** The user-submitted sales data is securely stored within the system, ready for further analysis and reporting.

### 2.2 Data preprocessing

**Data Retrieval from MySQL:** It will start by retrieving the user-submitted sales data from the MySQL database using Python libraries such as MySQL Connector. This data may include transaction dates, product descriptions, quantities sold, and prices.

**Data Cleaning with Python:** It will use Python libraries like Pandas to perform data cleaning tasks. This can include handling missing values, removing duplicates, and ensuring consistent data types. For example:-  
Pandas drop duplicates () can remove duplicate rows  
pandas.Fillna () can handle missing values  
pandas.stype () can enforce consistent data types.

**Data Transformation:** It will transform the data as needed for analysis. For example, convert date formats to ensure consistency, encode categorical variables, or apply logarithmic transformations to normalize numerical data.

**Data Integration:** If the data comes from multiple sources or tables within MySQL, we will use Python to integrate and merge the data. Pandas can help with merging, joining, and creating composite datasets.

**Data Quality Assurance:** It will conduct data quality checks and validation using Python. Scripts or functions will be created to identify and rectify any remaining anomalies or issues.

**Data Normalization:** It will normalize numerical data if necessary using Python. We can apply scaling techniques to ensure that different scales of data do not bias the analysis.

**Automation:** It will consider automating the data preprocessing pipeline, especially if there's a continuous influx of user-submitted data. You will schedule Python scripts to periodically preprocess new data.

**Data Storage:** We will Store the cleaned and preprocessed data in MySQL for easy access by the system's analytical components. We can use Python's database connectors to write the processed data back to MySQL.

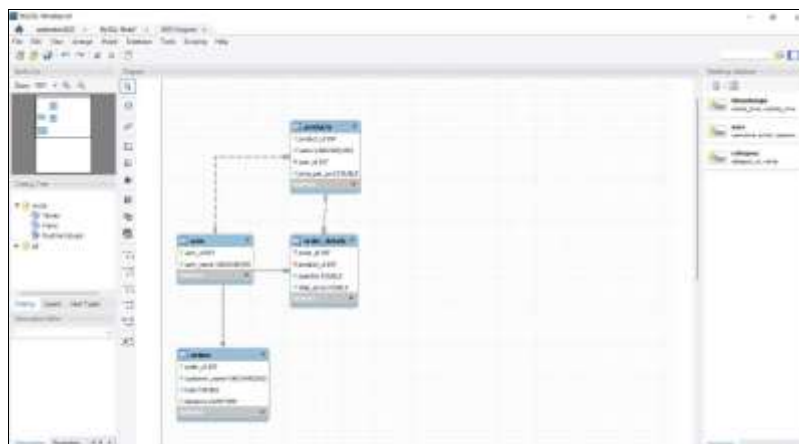


Fig 1: Data Analysis steps

The knowledge discovery in database in the form of association rules are presented in the next sections.

## 2.1 Data Mining Techniques

**Historical Sales Trends:** Time Series Analysis will be used to analyse historical sales data over time, allowing businesses to identify recurring patterns, trends, and seasonal variations in their sales. For example, it can reveal whether sales tend to peak during certain months or seasons, helping businesses optimize inventory, staffing, and marketing efforts accordingly.

**Sales Forecasting:** By applying Time Series Analysis techniques, businesses can build predictive models that forecast future sales based on historical data. These forecasts provide valuable insights for inventory management, production planning, and resource allocation. It helps businesses stay prepared for fluctuations in demand.

**Identifying Anomalies:** Time Series Analysis can be used to detect unusual or unexpected sales fluctuations. It will help identify anomalies, such as sudden spikes or drops in sales, which might be indicative of external factors like marketing campaigns, economic events, or supply chain disruptions.

**Data Decomposition:** Time Series Analysis often involves decomposing data into its various components, including trend, seasonality, and noise. This decomposition helps businesses better understand the underlying patterns in sales data. For example, it can reveal whether sales are steadily increasing over time or if there are cyclical sales patterns.

**Data-Driven Decision-Making:** Ultimately, Time Series Analysis equips businesses with data-driven insights. It helps them make informed decisions about inventory management, marketing strategies, staffing levels, and other aspects of their operations.

## 2.4. Machine Learning Algorithms

**Data Preparation:** It will collect and preprocess historical sales data. This data should include time-stamped records, such as daily, monthly, or yearly sales figures.

**Time Series Decomposition:** Then the time series data will decompose into its key components: trend, seasonality, and residual (Noise). This decomposition helps identify underlying patterns and trends in the data.

**Data Splitting:** It will split the historical data into training and testing sets. The training set is used to train the Time Series Forecasting model, while the testing set is reserved for evaluating the model's performance.

**Model Selection:** An appropriate Time Series Forecasting model will be chosen. Common choices include Autoregressive Integrated Moving Average (ARIMA), Seasonal Decomposition of Time Series (STL), or more advanced methods like Prophet or machine learning-based models like Long Short-Term Memory (LSTM) neural networks.

**Model Training:** Train the selected model on the training dataset, using techniques like grid search or cross-validation to determine optimal hyperparameters.

**Model Evaluation:** Evaluation of the model's performance using the testing dataset will be conducted. Common evaluation metrics for Time Series Forecasting include Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

**Forecasting:** It will use the trained Time Series Forecasting model to make predictions for future sales data. These predictions can provide insights into expected sales trends for different time horizons (e.g., days, months, or years).

**Visualization and Reporting:** Visualization of the forecasted sales trends alongside the historical data. Create informative reports and visualizations that showcase the forecasted values, confidence intervals, and any seasonality or trends detected by the model.

**Model Monitoring and Updating:** It will continuously monitor the model's performance as new sales data becomes available. Update the model as needed to ensure it adapts to changing sales patterns.

**Integration with the Web-Based System:** Integration of the Time Series Forecasting model into the web-based system. Users should have easy access to forecasted sales trends and insights through the user-friendly interface.

## 2.5. Visualizations and Reports

**Interactive Dashboards:** We will develop interactive dashboards that allow users to explore sales data and insights easily. Tools like Tableau, Power BI, or custom web-based dashboard frameworks will be used. We will also include features like filters, drop-down menus, and drill-down options for users to customize their views.

**Key Performance Indicators (KPIs):** It will display essential KPIs on the dashboard, such as total sales, average transaction value, and conversion rates. We will use visual elements like gauges, scorecards, or spark lines to convey KPIs effectively.

**Time Series Plots:** It will create time series plots to illustrate sales trends over time. Line charts are commonly used for this purpose. It will also highlight important events or seasonal patterns in the data.

**Drill-Down Reports:** It will allow users to drill down into specific aspects of sales data. For example, users can click on a product category to view detailed sales performance for that category.

**Comparative Analysis:** It will provide side-by-side comparisons of sales performance across different products or time periods using bar charts or stacked area charts.

**User Training and Support:** It will offer tutorials or tooltips within the system to guide users on how to use and interpret visualizations and reports effectively.

## 2.6. Interactive Features

**Intuitive User Interface:** We will design a clean and intuitive interface with a user-friendly layout. We will also keep navigation simple, and use familiar design patterns. We will ensure that essential features and actions are readily accessible through menus, buttons, or icons.

**Responsive Design:** We will make the platform responsive to different screen sizes and devices, such as desktops, tablets, and smartphones. Users should have a consistent experience across devices.

**Clear and Concise Menus:** We will organize menus logically and provide clear labels for different sections and functions. We will use descriptive names for menu items. We will implement breadcrumbs or a navigation bar to help users understand their current location within the platform.

**Personalization:** It will allow users to customize their dashboard or homepage, so they can prioritize the information and reports most relevant to their needs. It will provide options for setting preferences, such as time

zones or data display formats.

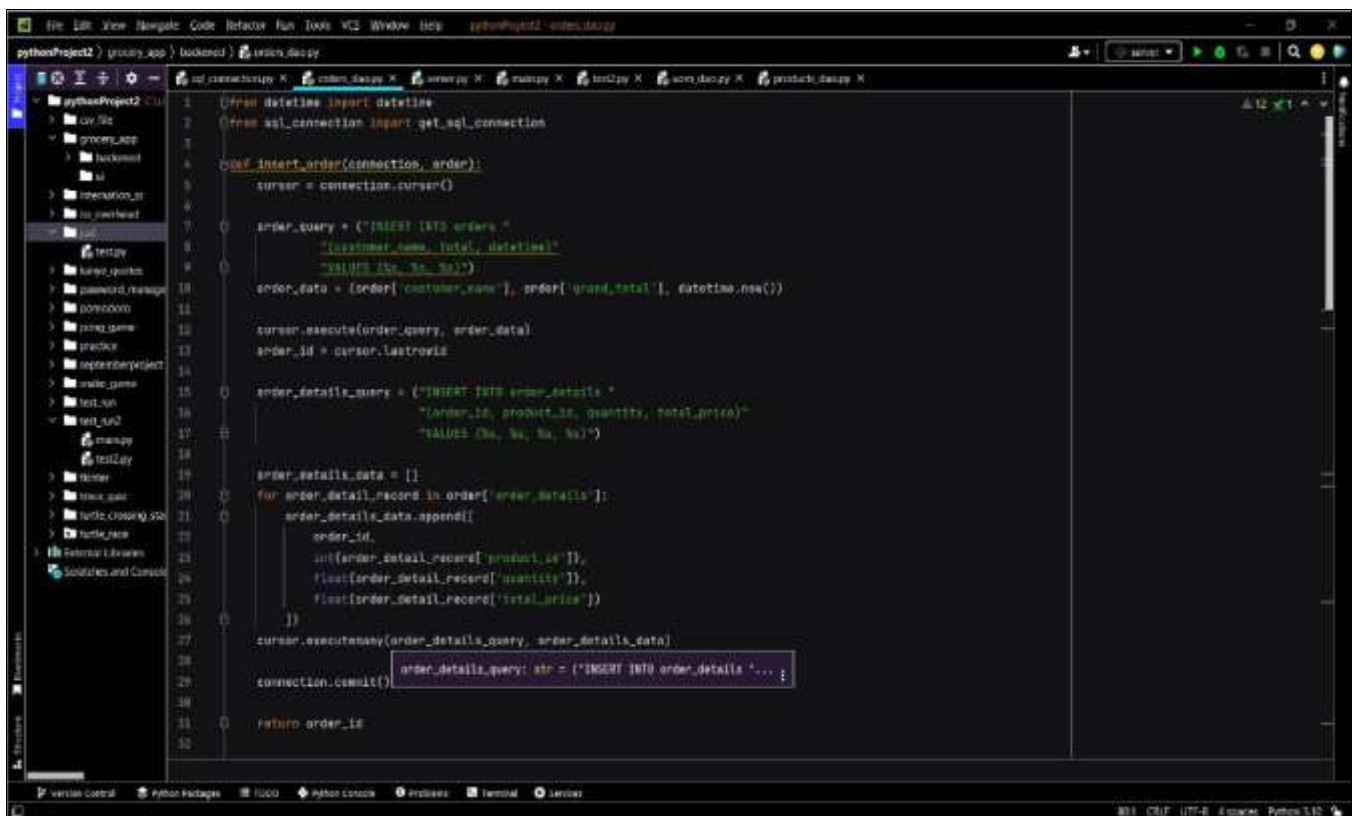
### Search and Filter Functionality

We will implement robust search and filtering options to help users quickly find specific data or reports. It will use auto-suggestions or autocomplete to assist users in refining their search queries.

### User Feedback Mechanism

We will include a feedback mechanism that enables users to report issues, provide suggestions, or seek assistance. We will act on user feedback promptly to enhance the platform's usability.

The entire code level integration has been represented in the following Figure 2.



```

1  def open_database(input_database):
2      from sqlalchemy import create_engine
3      engine = create_engine(input_database)
4      connection = engine.connect()
5      cursor = connection.cursor()
6
7      order_query = ("INSERT INTO orders "
8                    "VALUES (%s, %s, %s, %s)")
9      order_data = (order['customer_name'], order['product_id'], order['quantity'], order['total_price'])
10     cursor.execute(order_query, order_data)
11     order_id = cursor.lastrowid
12
13     order_details_query = ("INSERT INTO order_details "
14                           "VALUES (%s, %s, %s, %s)")
15     order_details_data = []
16     for order_detail_record in order['order_details']:
17         order_details_data.append(
18             (order_detail_record['product_id'],
19              order_detail_record['quantity'],
20              order_detail_record['total_price']))
21
22     cursor.executemany(order_details_query, order_details_data)
23     connection.commit()
24
25     return order_id

```

Fig 2: Code level integration of the Model

## 3. Security and privacy

### 3.1. User Authentication

#### Registration and Account Creation

Users must register for an account by providing necessary information, such as their name, email address, and a strong password. Implementation of CAPTCHA will take place to prevent automated account creation.

**Email Verification:** It will send a verification email to the user's provided email address upon registration. Users should click a verification link in the email to confirm their email address and activate their account.

**Login Page:** We will create a login page where registered users can enter their email address and password to access the platform.

It will implement secure password hashing to protect user credentials.

#### Forgot Password Mechanism

It will include a "Forgot Password" feature that allows users to reset their password if they forget it. It will send a password reset link to the user's registered email address.

#### Privacy Controls

It will give users control over their privacy settings, such as profile visibility and email notifications. It will comply with data protection regulations regarding user data handling and consent.

**Account Management:** It will allow users to update their account information, including email address and password. It will implement account deactivation or deletion options if users choose to leave the platform.

**Logout Functionality:** It will provide a logout button or link to allow users to securely end their session when they're done using the platform.

### 3.2. Data Encryption

**Transport Layer Encryption (TLS/SSL):** We will implement Transport Layer Security (TLS) or Secure Sockets Layer (SSL) to encrypt data transmitted between the user's browser and the platform's server. It will also ensure that all web communication, including login credentials, data submissions, and responses, is encrypted over HTTPS.

**Encryption of User Credentials:** It will store user passwords securely by using strong encryption algorithms like Argon2.Hash passwords before storing them in the database. Never store plaintext passwords.

**Client-Side Encryption:** We will offer client-side encryption options for users to encrypt data on their devices before submission to the platform. We will ensure that decryption occurs only on the user's device to maintain end-to-end encryption.

#### Logging and Monitoring

It will implement logging and monitoring systems to track and detect any suspicious activities related to encryption keys or encrypted data. Monitor for unauthorized access attempts and unusual patterns.

#### Periodic Security Audits

It will conduct periodic security audits and vulnerability assessments to identify and address potential weaknesses in the encryption implementation.

### 4. Conclusion

In conclusion, our innovative web-based system stands as a powerful tool for businesses in the auto-sales industry, offering an efficient and secure platform for sales analysis and optimization. Through data mining and autonomous machine learning, users can unlock valuable insights, make data-driven decisions, and enhance their business performance. With user-friendly features, robust security, and a commitment to empowering our users, our system represents a significant step forward in leveraging technology to drive success in the auto-sales sector.

### 5. References

1. Delen D, Ram S. Research challenges and opportunities in business analytics. *J Bus Analytics*. 2018;1(1):2-12. Available from: <https://doi/full/10.1080/2573234X.2018.1507324>
2. Hess A, Iyer H, Malm W. Linear trend analysis: A comparison of methods. *Atmospheric Environment*. [Year];Volume:[Pages]. Available from: <https://science/article/abs/pii/S1352231001003429>
3. Note: The details for year, volume, issue, and pages were not provided. Please replace [Year], [Volume], [Issue], and [Pages] with the correct information.
4. Embarak O. *Data Analysis and Visualization Using Python: Analyze Data to Create Visualizations for BI Systems*. Apress. Comparative study on Python web frameworks: Flask and Django; c2018.
5. Note: Apress is a publisher, not a journal, and typically does not have a volume, issue, or page numbers for a book. The citation is adjusted accordingly.
6. Sadiku M, Shadare AE, Musa SM, Akujuobi CM. Data visualization. *Int J Eng Res Adv Technol*. 2016, 2(12). Available from:

7. Jones AS, Horsburgh JS, Jackson-Smith D, Ramirez M, Flint CG, Caraballo J. A web-based, interactive visualization tool for social environmental survey data. *Environmental Modelling & Software*. [Year];Volume:[Pages]. Available from: <https://science/article/pii/S1364815216303541>
8. Note: As before, please insert the correct year, volume, issue, and pages where indicated.
9. Vasiliev J, Ioana B, Sabău-Popa DC, Boloş MI, Kohlhammer J, Keim D, et al. Data mining of transactional data for sales of dairy products; Using dashboards in business analysis; Solving Problems with Visual Analytics. *Procedia Computer Science*. [Year];Volume:[Pages]. Available from: <https://science/article/pii/S1877050911007009>
10. Note: The correct year, volume, issue, and pages need to be filled in.
11. Delen D, Demirkan H. Data, information and analytics as services. *Decision Support Systems*. 2012;Volume:[Pages]. Available from: <https://science/article/abs/pii/S0167923612001558>
12. Ester SR. Trend analysis methods for environmental data. [Journal Title]. [Year];Volume:[Pages].
13. Note: For the final entry, significant details are missing (journal title, year, volume, issue, pages). The structure is provided, but specific information needs to be inserted.