EJTAS
ISSN 2786-7447
EUROPEAN JOURNAL OF THEORETICAL AND APPLIED SCIENCES
OPEN ACCESS
PEER REVIEWED

Intelligent Sales Forecasting Technique Application

Stella A. Agomah
Department of Computer Science Abia State University, Utruru, Nigeria

Ikenna K. Ukabuiro
Department of Computer Science Abia State University, Utruru, Nigeria

Suggested Citation

Agomah, S.A., & Ukabuiro, I.K. (2023). Intelligent Sales Forecasting Technique Application. *European Journal of Theoretical and Applied Sciences, 1*(6), 641-653.
DOI: 10.59324/ejtas.2023.1(6).64

Abstract:

The primary objective of this study was to design and implement a machine learning-based sales forecasting system to enhance the production capacity and sales trajectory. The deployed model features a web-based interface that allows users to input parameters to generate predictions. The application of an intelligent forecasting technique, namely a machine learning model, significantly contributed to determining the optimal manufacturing output for a specific product in this study. The data analysis was conducted utilizing statistical software known as Tableau. The machine learning algorithm employed for constructing the model was the multiple linear regression model, which is particularly well-suited for trend analysis. The supplied dataset was utilized to train and test a supervised machine learning model, which was subsequently deployed on a local web server. Furthermore, a database system was effectively implemented to facilitate data storage, retrieval, and manipulation. The model was evaluated using two commonly employed metrics, the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE), within the Jupyter notebook environment. The resulting evaluation scores were 2.364858669808942 for RMSE and 1.7610409547966064 for MAE. These metrics were deemed effective in accurately predicting outcomes and efficiently presenting results.

Keywords: *Model, algorithm, web server, CRISP-DM, database, machine learning, MySQL, supervised machine learning, RMSE, MAE.*

Introduction

Daily, individuals engage in decision-making processes that can yield favorable or unfavorable consequences for their forthcoming circumstances. Just like in the physical realm, managers within the business domain are responsible for making decisions that significantly impact the future trajectory of their respective organizations. The options mentioned above may yield either a positive or negative result. Hence, the fundamental essence of forecasting lies in the ability to generate precise and reliable predictions regarding forthcoming occurrences.

Hence, forecasting is the interdisciplinary practice of formulating forecasts regarding future events, encompassing both the creative and empirical aspects of the process. The process involves utilizing mathematical models to project future outcomes based on past data. Occasionally, the determination of a matter relies on subjective judgment or proactive anticipation of forthcoming events. In essence, forecasting refers to the act of estimating or predicting an open state of events. It is impossible to achieve complete accuracy in forecasting due to the inherent uncertainty of the future. All human behaviors inherently entail a certain degree of predicting, yet formal forecasting approaches are increasingly gaining traction and being widely adopted within commerce.

The present study project employs a linear regression model to estimate the sales of products within a corporation. Developing a projection of forthcoming values for significant metrics, such as product demand or economic indicators, can yield substantial benefits for enterprises. Multiple forecasting techniques exist, which can be categorized into two distinct groups: causal and time-series (Gellert, 2013). Using linear regression, a time-series technique, involves the application of fundamental statistical principles to forecast the forthcoming values of a target variable (Gellert, 2013). According to Gupta and Hiza (2008), firms of varying sizes possess resources such as staff, equipment, cash, and materials, yet supply availability may be limited. In the event of an inexhaustible abundance of those resources, the necessity for a management tool such as linear programming would be rendered obsolete. Given the constraints imposed by restricted resource availability, management must make informed decisions regarding resource allocation to optimize profitability, minimize losses, and effectively utilize production capacity. Nevertheless, there are specific concerns regarding this matter that utilizing quantitative methodologies, including linear programming, can effectively address.

This study focuses on the manufacturing capacity of a theoretical brewery company, referred to as Company X. The determination of optimal production outcomes for a specific product would be highly advantageous. Given the corporation's reputation for manufacturing diverse products, using the linear regression technique would effectively address the problem at hand. Linear regression is the most suitable adaption in this particular context. Moreover, linear regression is employed to tackle an organization's sales estimates.

The Hypothetical Brewery Company X is recognized as one of the largest brewers in Nigeria. In addition to catering to the domestic Nigerian market, this product also engages in export activities to various regions within West Africa. The establishment of the Company took place in Lagos, with subsequent expansions leading to the establishment of breweries in Kaduna, Ibadan, Enugu, and Aba as well (source: https://www.drinks.ng/history-breweries/).

The merger between Hypothetical Brewery Company X and Consolidated Breweries Plc was finalized in December 2014, including three additional breweries in Ijebu-ode, Awo-Omama, and Makurdi. This merger also led to an expansion of the Company's portfolio of brands. According to a source (https://www.drinks.ng/history-breweries/), Brewery Company X is presently operating 10 breweries, facilitating the distribution of its products across Nigeria.

Review Related Works

In their study, Nashirah and Sofian (2017) employed a statistical methodology to predict cryptocurrency's exchange rate under a context characterized by significant volatility. The study employed the Autoregressive Integrated Moving Average (ARIMA) model as its statistical approach. The researchers examined the autocorrelation and partial autocorrelation functions to estimate the ARIMA model's parameters. The findings derived from their research indicate that the first difference of the Bitcoin exchange rate exhibits characteristics of stationarity. The presented model demonstrates an R-squared value of 0.444432, indicating that it accounts for about 44.44% of the total variability in the response data relative to its mean. When doing an error analysis comparing anticipated values to actual data, it was found that the mean absolute percentage error for ex post forecasting was 5.36%.

In their study, Aras et al. (2017) conducted a sales forecast for a multinational furniture retailer in Turkey. They employed various statistical methods, including state space models, ARIMA and ARFIMA models, neural networks, and the Adaptive Network-based Fuzzy Inference System (ANFIS). Additionally, they explored the effectiveness of combining some of these models. A comparison evaluation was conducted to assess the forecasting performances of different approaches using weekly sales data for ten products. The performance evaluation findings indicate that no individual models exhibited superior performance compared to all other models when considering the entire time series data. The combined approaches' estimates yielded a statistically significant improvement in forecasting accuracy compared to the separate models.

In their study, Adeniran et al. (2018) sought to anticipate domestic aviation passenger demand in Nigeria by employing a single-moving average and exponential smoothing techniques. The authors also conducted a comparison evaluation of both models using the Mean Squared Deviations (MSD) in order to ascertain the strategy that would yield the most accurate forecast for the year 2018. The dataset about domestic aviation passenger traffic from 2010 to 2017 was utilized. In the comparative analysis of the single-moving average approach and exponential smoothing, it was observed that the former yielded superior forecast results. The authors elucidated the significance of their research in facilitating the strategic decision making process of various entities within the aviation industry, such as airlines and airports.

Martinus and Arinanda (2020) studied sales forecasting of motorcycle parts in Indonesia using the auto-regressive integrated moving average model. The researchers utilized sales data of 62 motorcycle parts from January 2017 to February 2019, obtained from the Central Statistics Agency in Indonesia. The model's performance was assessed using the Mean Squared Error (MSE). A Mean Squared Error (MSE) value of 61.737 was achieved. In their study, Sahu and Kumar (2014) evaluated various forecasting techniques to predict sales of sterilized flavoured milk in the region of Chhattisgarh. The researchers utilized one year's weekly data to analyze the sales of sterilized flavoured milk. The researchers analyzed several forecasting models, including the naïve model, moving average, double moving average, basic exponential smoothing, and the semi-average approach. The evaluation of the forecasting techniques was conducted by assessing their accuracy through the utilization of statistical metrics such as Mean Forecast Error (MFE), Mean Absolute Deviation (MAD), Mean Square Error (MSE), and Root Mean Squared Error (RMSE).

In their study, Mohit et al. (2017) employed various machine learning algorithms to predict sales within the context of Rossmann, a pharmaceutical retail company. The dataset utilized by the organization comprises historical sales data and supplementary information about medication retailers. Initially, a statistical method known as ARIMA was employed. However, this model could have effectively captured the nonlinear patterns present in the dataset of the company. Consequently, alternative nonlinear models, including Neural Network, XGBoost, and Support Vector Machine models, were subsequently employed. The nonlinear models demonstrated superior performance compared to the statistical method, as evidenced by the lower Root Mean Squared Value. Composite models were constructed to enhance the system's performance by combining ARIMA with ANN, ARIMA with XGBoost, and ARIMA with SVM. The empirical findings demonstrate that the hybrid models outperformed each component model. Based on the acquired results, it was concluded that the composite models exhibited superior performance in estimating sales for the drug company.

In their study, Yucesan et al. (2017) devised a forecasting model utilizing an artificial neural network (ANN) to predict the sales of a furniture manufacturing company in Turkey. The researchers employed MATLAB software to analyze the aggregated monthly sales data of the company situated in the Black Sea region of Turkey. Using Bayesian regularisation is a fundamental component of Artificial Neural Networks (ANNs). The findings, as assessed by performance metrics, indicate that the utilization of the Artificial Neural Network (ANN) model trained using Bayesian rules exhibits a high level of accuracy in predicting the monthly sales of the observed furniture firm.

More data must be collected to accurately predict the sales of recently introduced products. In this nature, human expert systems emerge as the subsequent viable alternatives for predicting sales. The human specialists depend on their domain expertise on past sales data of comparable products, which they utilize to project future sales figures for novel products. In their study, Karb et al. (2020) put up an analytical methodology for transferring knowledge from existing stock products to newly developed products. A network-based transfer learning approach was devised to examine the efficacy of transfer learning in the field of food sales forecasting within deep neural networks. The researchers conducted a study in which they employed an Australian food retailing company as a case study to evaluate the effectiveness of this particular method. The application of this approach has been found to enhance the predictive accuracy of deep neural networks in the context of food sales forecasting, as evidenced by empirical findings.

Implementing intelligent sales forecasting methodologies necessitates feature engineering in specific business contexts, which is typically challenging, time-intensive, and reliant on specialist expertise. In their study, Zhao and Wang (2017) introduced a methodology to address the abovementioned obstacles. This strategy involves utilizing a convolutional neural network (CNN) to automatically extract meaningful features from structured data, enhancing the learning process's effectiveness. When provided with unprocessed log data, the proposed methodology autonomously identifies significant elements and employs them to predict sales outcomes. The proposed methodology's efficacy is demonstrated using a substantial real-world dataset in experimental trials.

In their study, Venishetty et al. (2020) employed machine learning algorithms to predict sales patterns for truck components. The initial step involved in the analysis was normalizing the time series data to mitigate and eliminate any redundant information. Subsequently, feature extraction and selection algorithms were employed. Various machine learning models analyzed the normalized time series data, including Support Vector Machine Regression, Ridge Regression, Gradient Boosting Regression, and Random Forest Regression. The empirical findings indicate that the ridge regression technique exhibits superior accuracy in sales forecasting when compared to alternative machine learning models.

Deep learning algorithms have been well recognized for their superior forecasting capabilities compared to machine learning algorithms. This is mainly attributed to their remarkable capacity to retain knowledge over time and effectively discern temporal correlations. In their study, Suleka et al. (2019) utilized a specialized iteration of the Long Short Term Memory (LSTM) network known as the LSTM model with peephole connections to do sales forecasting. The model was constructed by utilizing characteristics derived from historical data. The researchers conducted a comparative analysis of the performance outcomes of their constructed model with several machine learning methods, such as the Extreme Gradient Boosting (XGB) model and the Random Forest Regressor model (RFR). The constructed LSTM model exhibited superior performance to the machine learning models, showing a notable improvement of 12% to 14%.

In their study conducted in 2017, Chen and Lu proposed a strategy for projecting computer retailing sales that integrates clustering and machine learning methodologies. The model they initially suggested divided training data into groups using the clustering technique, grouping data with comparable properties or patterns. The machine learning approaches are employed to train the forecasting model for each group. The forecasting model trained on the chosen cluster was employed to predict sales using the set that exhibited data patterns most similar to the test data. The study utilized three clustering algorithms: self-organizing maps (SOM), growing hierarchical self-organizing maps (GHSOM), and K-means. Additionally, two machine learning techniques, support vector regression (SVR) and extreme learning machines (ELM), were employed. The empirical outcomes of their study revealed that the combination of GHSOM and ELM resulted in improved forecasting performance of the model.

In order to enable the most effective people scheduling for managing crew load, Schimdt et al. (2022) provided a case study on machine learning models for precise sales forecasting of a mid-sized restaurant. A number of approaches are directly compared to popular recurrent neural network (RNN) models. To facilitate forecasting, the data's features were manipulated to assist in selecting an ideal subset of strongly correlated features. Using a carefully selected test dataset, the models were evaluated by comparing their effectiveness in forecasting time steps of one day and one week. Three distinct models were tested, namely RNN, LSTM, and TFT. The RNN model outperformed the others, exhibiting a notable SMAPE score of 19.5% in the optimal outcome.

In 2020, Yiyang proposed a sale prediction model using the XGBoost algorithm and a thorough feature engineering procedure to address the issue of anticipating Walmart's sales.The analysis utilizes the sales data of Walmart supermarkets obtained via Kaggle.

They successfully utilized all the characteristics from many dimensions to make accurate forecasts. Empirical findings demonstrate the successful development of a model outperforming alternative machine learning methodologies.

The RMSE measure of the models was 0.141 and 0.113 lower than that of the logistic regression and ridge algorithms, respectively.

Abdellatief *et al.* (2021) conducted sales forecasting for Egypt's leading automotive manufacturing firm. They examined the influential elements that affect sales volume, including unit selling price, inflation rate, per capita income, fuel price, and past sales.The researchers conducted training on a Neural Network (ANN), Adaptive Neuro-Fuzzy inference system (ANFIS), and Multiple Linear Regressions (MLR) to assess and choose the most effective model. The empirical findings from their study demonstrated that the artificial neural network (ANN) technique had more remarkable performance than the other strategies.

To create a real-time online system for a specific supermarket, Kishana (2013) used an SPSS application and the k-means clustering technique. The purpose of this system is to forecast revenues across different seasonal cycles. The implemented model served as an intelligent tool that extracted inputs directly from sales data records and autonomously refreshed segmentation statistics at the conclusion of each work day. The proposal underwent a successful testing and implementation process lasting three months. After receiving observational tests, 2138 clients were categorized into four (4) groups indistinguishable.

The items were classified using the nearest mean technique. An analysis of variance (ANOVA) was conducted to assess the clusters' stability.

The results were reported to have exceptional precision.

Huang et al. (2015) presented a new trigger system that can effectively correlate specific types of goods with a predictive model, leading to improved sales forecasts across various product categories. Several correlated criteria were identified for classification.

They incorporated multiple classical prediction models as fundamental models for classification. Their proposed trigger model was compared with various individual models, and the results demonstrated that the trigger model exhibited superior accuracy compared to the single models. The authors found that sellers can employ the proposed system to forecast various goods sales accurately.

The sales data underwent analysis utilizing clustering techniques such as K-Means and EM. These algorithms unveiled intriguing patterns that can be used to enhance sales revenue and attain greater sales volume. In their study, Sastry et al. (2013) validated that partition approaches such as K-Means and EM algorithms are more appropriate for analyzing sales data than density based methods like DUBSCAN and OPTICS or hierarchical methods like COBWEB.

Rohaan and colleagues (2022) introduced a technique for incorporating advanced demand intelligence (ADI) into B2B sales forecasting. The researchers utilized data from request for quotation (RFQ) requests to conduct an analysis and acquire knowledge by applying machine learning and natural language processing methodologies. They implemented their method in a substantial after-sales care and repair provider as part of a case study. Upon investigation, it was determined that their approach achieves a 70% accurate recall of actual sales. This represents a performance improvement of 2.5 times compared to the repair and maintenance provider's current labor intensive manual process.

Furthermore, their research demonstrates the potential improvement in performance that may be expected when implementing supervised machine learning for B2B sales forecasting.

Every firm must be aware of its competitors, including factors such as client demand, purchasing patterns, and sales performance. In their study, Isa et al. (2019) introduced a sales analysis approach that utilizes prediction and association techniques on sales data. They implemented their suggested methodology on the sales data of Chan Furniture and Electrical Appliances Company. It was discovered that the company experiences varying levels of sales throughout the year, and the sales achieved are below the anticipated amount. They utilized their model to forecast the products that will be marketed monthly. The association model was employed to ascertain the products the customer purchased together. The discovered result was utilized to build a business strategy for Chan Furniture.

The advancements in information technologies have significantly impacted our lives in various ways. Traditional media organizations have experienced a growing necessity to forecast the sales of print publications, such as newspapers and magazines. The primary focus of forecasting newspaper/magazine sales has been constructing regression models using sample datasets.

However, these models are plagued by overfitting issues. Recent theoretical investigations indicate that support vector regression (SVR) can mitigate the problem of overfitting. The study by Yu et al. (2013) utilized support vector regression (SVR) to predict sales of newspapers/magazines. The findings demonstrated that SVR is a more practical approach for this type of forecasting.

Data Collection

Data collection is the systematic process of gathering, quantifying, and examining data for research purposes, employing established and validated methodologies. Data collecting is an essential and crucial initial stage of research, whether in the sciences or the arts, regardless of the research topic. The data collection must be comprehensive and dependable to ensure the research findings' accuracy, reliability, and usability.

Data collection methods can be broadly categorized into two groups: primary and secondary methods. Primary data collection refers to the researcher's direct collection of data from original or firsthand sources. Examples of approaches employed in this method include surveys, observations, experiments, questionnaires, and personal interviews. The Secondary method utilizes strategies to acquire pre existing data collected by another individual. Secondary data-gathering systems encompass sources such as books, journals, websites, and journal articles.

This study employed the following strategies to collect data and obtain necessary information.

i. Interviews: An interview is a data-collecting technique where two or more individuals engage in a structured exchange of information by posing and responding to a series of questions. The researcher formulates the questions to elicit information from the interview participants regarding a particular subject or series of topics. The study conducts interviews with the beer company in Enugu to ascertain their distribution methods for delivering products to various towns and cities and the quantity of items distributed in response to demand.

ii. Document sampling: This refers to gathering factual information, data, and knowledge from existing documents rather than from human sources.

By scrutinizing pertinent literature, the researcher uses this methodology to acquire empirical information regarding the current system. The data utilized in constructing the linear regression model for sales prediction was obtained from the brewery in Enugu as a pre existing documented file.

Data Analysis Results

Four (4) columns — month, location, product, and number [crates/cartons] — along with about 500 rows of data were used to create the sales forecasting system as seen in Table 1 below:

Table 1. Table Showing the Different Columns of the Data

Column	Description
Month	The month of the year product was sold
Location	Location in Enugu where brewery product was sold
Product	Brewery product sold
Quantity	Quantity of the product sold given in either craters or cartons

General data analysis

The sales trend analysis, depicted in Figure 1, demonstrates the significant deviation of the 2020 sales data and the reason for its exclusion from the software system's data.

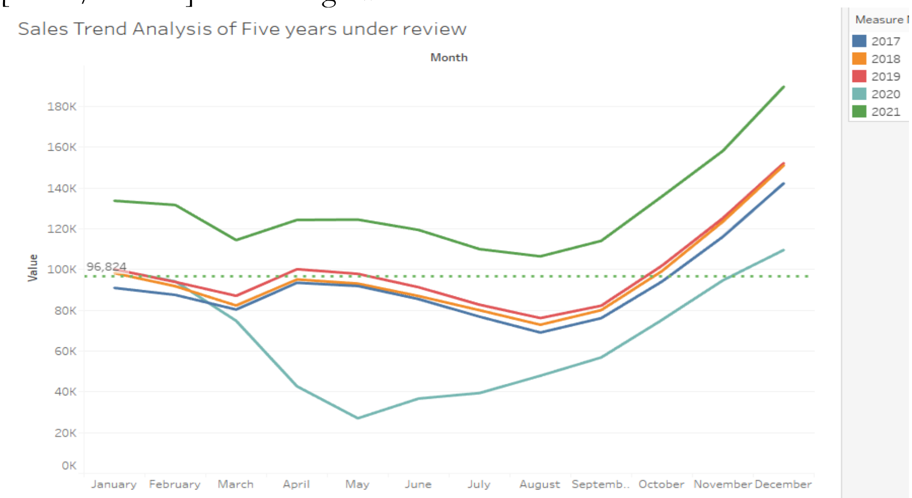

Figure 1. Line Chart Showing the Trend of Sales Over Five (5) Years Period

The Comparison Between Month and Quantity

The plot depicted in Figure 2 illustrates the cumulative sales of brewery products during the designated months within various years under examination. There is a discernible trend of increasing monthly sales as the years progress. For instance, the sales figures for January 2019 surpassed those for January 2018. This observation suggests a persistent and ongoing need for the products offered by the company.

The Comparison Between Location and Quantity

As shown in Figure 3 below, the plot demonstrates that the cumulative sales suggest

New Haven has the most significant demand for items across all months evaluated. This indicates that further promotional efforts and advertising campaigns could be directed towards this locality due to its notable product consumption rate.

Figure 2. Quantity of Products Sold in 5 Specific Months Across the Years Considered

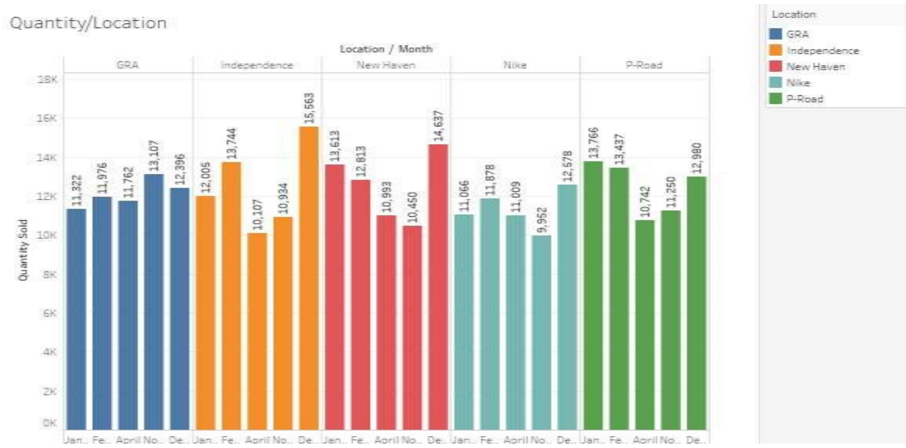

Figure 3. Sales of Products in the Different Locations in Enugu

The Comparison Between Product and Quantity

The diagram depicted in Figure 4 illustrates that the various products exhibit their peak sales throughout December. This implies that the current period is the most feasible time of the year to enhance the manufacturing of those products. Upon examining the data, it is evident that there is a minimal disparity in the sales of a specific product throughout the months under consideration. This suggests that these months may present plausible opportunities for augmenting production.

Product vs Quantity:

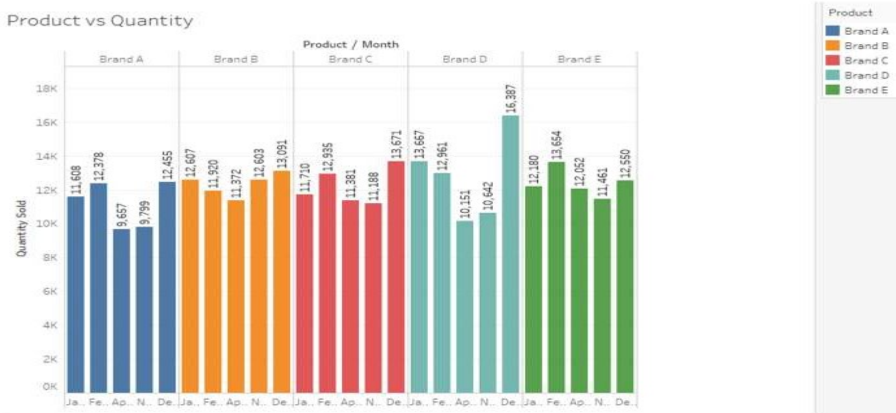

Figure 4. Sales of the Different Products

Implementation of the Linear Regression Model

A linear regression model was trained to forecast future sales using the characteristics of Month, Year, Location, Product, and Number of Products Sold. The following is a presentation of the formula used for the trained model:

$$y = \beta0 + \beta1.x1 + \beta2.x2 + \beta3.x3 + \beta4.x4 \quad (1)$$

Where:

y = Quantity of products sold (predicted variable or dependent variable)

x1 to x4 represents the independent variables;

x1= Month

x2= Year

x3= Location

x4= Product

β0 = The value of the dependent variable (y) when all independent variables (x1 to x4) are equal to zero is referred to as the intercept, also known as the y-intercept.

The regression coefficients (β1 to β4) quantify the impact on the dependent variable (y) resulting from a one-unit change in each respective independent variable (x1 to x4). The coefficient β1 represents the impact on the dependent variable y resulting from a one-unit change in the independent variable x1 while keeping all other independent variables constant.

The model underwent testing utilizing two assessment metrics: the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE).

The Root Mean Squared Error (RMSE)

The Root Mean Squared Error (RMSE) is a commonly employed statistic for assessing the performance of a regression model. Root Mean Square Error (RMSE) is a quantitative measure utilized to evaluate the average discrepancy between the projected values generated by a model and the actual values present within a given dataset. A lower Root Mean Square Error (RMSE) score indicates a higher level of accuracy in the model's ability to fit a given dataset. The root mean square error (RMSE) value obtained for the Sales forecasting model is 2.3648586698089442.

The formula for calculating RMSE is presented as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(P_i - O_i)^2}{N}} \quad (2)$$

Where:

i - equivalent to variable i

N - stands for the sample size

Pi - represents the predicted value for the i-th observation

Oi - represents the observed value for the i-th observation

The Mean Absolute Error (MAE)

The Mean Absolute Error (MAE) is an alternative statistic that can be employed to assess the performance of a regression model. The Mean Absolute Error (MAE) quantifies the extent of the discrepancy between the predicted value of an observation and its actual value. Like the root mean squared error (RMSE), a lower mean absolute error (MAE) number indicates a higher model performance. The Mean Absolute Error (MAE) value obtained for the Sales Forecasting model is 1.7610409547966064.

The formula for calculating the Mean Absolute Error (MAE) is presented as follows:

$$\text{MAE} = \frac{\sum_{i}^{N} abs(P_i - O_i)}{N} \quad (3)$$

The abbreviation MAE stands for Mean Absolute Error.

The variable "i" represents a numerical value, whereas "N" denotes the size of the sample.

The abbreviation "abs" stands for the mathematical concept of absolute value.

The symbol "Pi" represents the expected value for the ith observation.

Oi represents the observed value for the ith observation.

The Implementation of a Novel System

A predictive model was developed with the existing dataset to facilitate sales forecasting. Additionally, this system can generate reports based on factors such as year, month, product, and location. The high-level architecture of the new system is depicted in Figure 5.

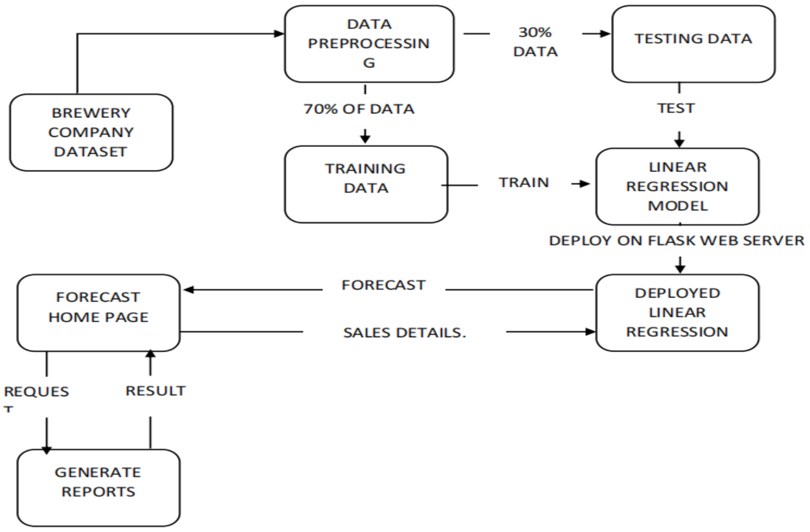

Figure 5. A High Level Architecture of the New System Showing the Training and Testing of the Sales Forecasting Model and Report Generation

The Following is a Description of the System

The system that has been developed consists of eight essential modules, which are afterward organized into the following categories modules:

- The module for model training and testing
- The module for forecasting
- The module for report creation

Model Training Module

The Model Training Module encompasses the underlying processes in constructing an appropriate machine learning model capable of performing accurate predictions. The linear regression model was trained using 70% of the available dataset and evaluated on the remaining 30% afterward to assess its performance. After training and testing, the model was deployed on a local web server using the FLASK Python module. Utilizing this Python module facilitated the creation of a web application that offered an interface capable of direct interaction with said module for sales forecasting.

The module for forecasting

This constitutes the primary component of the novel system. After the model's training, testing, and deployment, an interface was developed with a unique capability to interact with the model. This module offers a graphical user interface that establishes a connection to the linear regression module that has been constructed. The module provides an HTML page, including a form that allows users to input certain information or parameters. This information is then transmitted to the machine learning model for forecasting.

The report generation module

This is a significant component of the sales forecasting and reporting system. The present module allows the user to access and examine reports about sales transactions conducted in previous periods. The sales data can be observed or generated based on various parameters, including year, product, month, and location.

Discussion

The Sales Forecasting and Report generating system underwent testing to validate its adherence to the system specifications and to ensure the accurate flow and exchange of data inside the newly implemented system. The various parts of the software underwent testing and were determined to function accurately and efficiently. The testing methodology for evaluating the new system consists of unit, integrated, and system testing. Unit testing is a crucial process in software development that focuses on assessing the functionality of individual components or modules at the code level, ensuring that they perform as intended according to their design specifications. Integrated testing is a comprehensive process that combines all individual parts or modules to assess their functionality and effective communication. System testing is conducted to verify that the fully assembled and integrated system, as a cohesive entity, satisfies the predetermined criteria.

The successful performance of the new system is demonstrated by:

- Its accurate sales forecasting,
- The right generation of reports,
- Proper handling of sales records in the database and
- Efficient exchange of data between various modules.

Conclusion

Our work concludes that the aforementioned points substantiate the idea that:

1. Upon analyzing the dataset, it is evident that the year 2020 did not provide a suitable basis for drawing inferences in the realm of business. Consequently, it was deliberately omitted during the construction of the model. The research accomplished all of the stated objectives outlined for the developed system. The model underwent evaluation and showed proficiency in

making accurate predictions and quickly presenting the outcomes.

2. A Linear regression model, which falls under supervised machine learning, was utilized to train and evaluate its performance in predicting income and sales for a hypothetical brewery company, Company X.

3. A module has been successfully designed that facilitates the generation of sales reports by considering various characteristics. The system helps with effective record keeping and has a module that allows staff members to enter new records.

4. Additionally, the trained model was implemented on a local server to facilitate utilization by the personnel of the brewery.

Conflict of Interests

No conflict of interest.

References

Adeniran, A. O., Kanyio, O.A., & Owoeye, A.S. (2018). Forecasting methodsfor domestic Air Passenger Demand in Nigeria. *Journal of Applied Researchon Industrial Engineering, 5*(2), 146-155.

Adeniran, A. O., Kanyio, & Olufunto, A. (2018). Long Term Forecasting ofInternational Air Travel Demand in Nigeria. *American International Journal of Multidisciplinary Scientific Research, 1*(2), 16-24.

Bob, A. (2012). Why is accurate sales forecasting such a challenge? Retrieved from https://www.inflexion-point.com/Blog/bid/90329/Why-is-accurate-salesforecasting-such- a-challenge.

Chen, I.F. & Lu, C.J. (2017) Sales forecasting by combining clustering and machine-learning techniques for computer retailing. *Neural Computing & Applications, 28*, 2633–2647. https://doi.org/10.1007/s00521-016-2215-x

Etiene, T. (2017). *Six Strategies for Application Deployment.* Retrieved from https://thenewstack.io/deploymentstrategies/#:~:text=this%20git%20repository.,Recreate,boot%20duration%20of%20the%20application

Hillier, W. (2021). *What is the difference between Regression and classifications?* Retrieved from https://careerfoundry.com/en/blog/dataanalytics/regression-vs-classification

Huang, W., Zhang, Q., Xu, W., Fu, H., Wang, M. & Liang, X. (2015). A NovelTrigger Model 1213 for Sales Prediction with Data Mining Techniques. *Data Science Journal, 14*, 15.

Isa, N., Yusof, N. S. M., & Ramlan, M. A. (2019). The Implementation of DataMining Techniques for Sales Analysis using Daily Sales Data. *International Journal of Advanced Trends in Computer Science and Engineering, 8*(1.5), 74-80.

Sastry, S. H., Babu, P., & Prasada, M. S. (2013). Analysis & Prediction of Sales Datain SAP ERP System using Clustering Algorithms.

Karb, T., Kuhl, N. & Hirt, R. (2020). A Network-Based Transfer Learning Approachto Improve Sales Forecasting of New Products. *Paper presented at the Twenty-Eight European Conference on Information Systems.* Marrakesh, Morocco.

Karri, L. (2018). *What is system integration?* Retrieved from https://www.youredi.com/blog/what-is-system-integration

Kashwan, K.R. & Velu, C.M (2013). Customer segmentation using clustering anddata mining techniques. *International Journal of Computer Theory and Engineering, 5*, 856- 861.

Kotsiantis, S. (2007). Supervised Machine Learning: A Review of ClassificationTechniques. *Informatica Journal, 31*(2007), 249-268.

Kui, Z. & Can, W. (2017). *Sales Forecast in E-commerce using Convolutional Neural Network.*

Martinus, M., Ernawati, & Komang, A. (2020). Motorcycle Parts Sales ForecastingUsing Auto-regressive Integrated Moving Average Model. *International Journal of Computer Theory and Engineering, 12*(1), 28 -31

Martinovic, J., & Damnjanovic, V. (2006). The sales forecasting techniques. *International Scientific Days.*

Abdellatief, M., Shaaban, E.M. & Abu-Raya, K.A. (2019). Egyptian Case Study Sales forecasting model for automotive section. *International Conference on Smart Applications, Communications and Networking (SmartNets)*, 1-6. https://doi.org/10.1109/SmartNets48225.2019.9069751

Mohit, G., Yogesh, K., Prachi, S., Sandeep, U., Vijay, S., & Sunil. (2017). Forecasting of sales by using fusion of machine learning techniques. *Paper presentedat the International Conference on Data Management, Analytics and Innovation.* Pune, India

Nashirah, A.B. & Sofian R. (2017). Autoregressive Integrated Moving Average (ARIMA) Model for forecasting cryptocurrency Exchange rate in High Volatility Environment: A New insight of Bitcoin Transaction. *International Journal of Advanced Engineering Research and Science, 4*(11), 130- 137

Niu, Y. (2020). Walmart Sales Forecasting using XGBoost algorithm and Featureengineering. *International Conference on Big Data & Artificial Intelligence&Software Engineering (ICBASE)*, 458-461. https://doi.org/10.1109/ICBASE51474.2020.00103

Rohaan, D., Topan, E., & Groothuis Oudshoorn, C.G.M. (2022). Using supervise machine learning for B2B sales forecasting: A case study of spare parts sales forecasting at an after-sales service provider. *Expert Systems with Applications, 188*.

Serkan, A., Ipek, D.K., & Cigdem, P. (2017). Comparative Study on retail sales forecasting between single and combination methods. *Journal of Business Economics and Management, 18*(5), 803-832.

Schmidt, A, Md Kabir, W.U. & Md Hoque, T. (2022). "Machine Learning Based Restaurant Sales Forecasting. *Machine Learning and Knowledge Extraction, 4*(1), 105-130. https://doi.org/10.3390/make4010006

Venishetty, S.V., Huseyin, K., & Alain, B. (2020). Forecasting Sales of Truck Components: A Machine Learning Approach. Proceedings of *IEEE 10th International Conference on Intelligent Systems.*

Yucesan, M., Gul, M., & Celik, E. (2017). Application of Artificial Neural NetworksUsing Bayesian Training Rule in Sales Forecasting for Furniture Industry.

Yu, X., Qi, Z. & Zhao, Y. (2013). Support Vector Regression for Newspaper/Magazines Sales Forecasting. *Information Technology and Quantitative Management*, 17(2013), 1055-1062.