Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4, 1795-1807
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i4.6383

Intelligent inventory prediction: A machine learning framework using random forest for inventory forecasting

Nagarajan Deivanayagampillai[1,2], Thangavel Bhuvaneswari[3*], Yasothei Suppiah[4]
[1]Department of Mathematics, Rajalakshmi Institute of Technology, Chennai, India; dnrmsu2002@yahoo.com (N.D.).
[2]Postdoctoral Researcher, Faculty of Engineering and Technology, Multimedia University, Jalan Ayer Keroh Lama,75450 Melaka. Malaysia.
[3,4]Faculty of Engineering and Technology, Multimedia University, Jalan Ayer Keroh Lama,75450 Melaka. Malaysia; t.bhuvaneswari@mmu.edu.my (T.B.). yasothei.suppiah@mmu.edu.my (Y.S.).

Abstract: This study proposes a machine learning-based inventory forecasting strategy using random forest classifiers to improve demand prediction and anomaly detection in retail sales. It aims to address the limitations of traditional inventory management by capturing nonlinear demand dynamics and interdependencies. The model is developed and validated using real-world inventory data, with the United Kingdom as the primary market. Data preprocessing includes handling 135,080 missing CustomerID fields, 1,454 missing Description fields, and correcting anomalies in Quantity and Unit Price. The random forest classifier is employed to identify complex sales patterns and enhance demand forecasting accuracy. Experimental findings demonstrate a significant improvement in prediction accuracy and inventory optimization over traditional methods. The model effectively captures regional sales variations and adapts to changing demand trends, enabling more precise inventory decisions. The proposed framework contributes to the development of intelligent and adaptive inventory management systems, allowing businesses to make data-driven decisions, optimize stock levels, and reduce inventory risks, and provides valuable insights for businesses aiming to enhance inventory forecasting. It highlights the importance of machine learning in improving demand prediction, minimizing errors, and adapting to evolving market conditions.
Keywords: *Anomaly detection, Inventory, Random Forest classifier, Sales prediction.*

1. Introduction

High accuracies of prediction have been achieved with an ensemble of trees and majority vote among them. Random vectors are normally utilized to train the ensemble of trees and direct the individual tree development within the ensemble. In the last few years, many papers addressed the issue of finding the best inventory management policies. Most of them concerned finding the optimal purchase lots of a particular facility facing a known demand behaviour [1]. Deep reinforcement learning has been an adaptable and strong method of useful inventory replenishment policy design They demonstrate that the use of transfer learning of learned heuristics enhances training stability. This approach not only enhances inventory cost-effectiveness and computational efficiency but also training stability and thus enhances the confidence in the inventory management policies that the reinforcement learning algorithm comes up with De Moor, et al. [2]. A contribution to a literature that has classically been disaggregated, though it has been considered important that the forecasting and forecasting constituencies be consolidated to address an aggregation of similar issues. The end to a call for a research agenda underscores the necessity to move forward not only in theory development and empirical analysis but to further strengthen the knowledge base within this domain [3]. An overarching

History: Received: 10 February 2025; Revised: 8 April 2025; Accepted: 11 April 2025; Published: 19 April 2025
* Correspondence: t.bhuvaneswari@mmu.edu.my

framework that is sufficiently general to be applicable within every inventory model, every demand distribution, and every parameter estimator is presented. Inventory need is highly dependent upon forecasts of demand, but it has classically been assumed within the literature that demand distributions are known [4]. The suggested policy is proven to be less costly than an identical system where the lost demand is backordered rather than being lost, and to be asymptotically optimal under an infinite-cost lost sale under mild distribution assumptions about the demand. Numerical results prove the performance is better than under prevailing policies [5]. The policy has been designed to blend promotional and trend information with periodic ordering decisions irrespective of assumptions about the demand distribution in an efficient manner. Up till now, supervised learning has been used mainly in the application of inventory control with a crisp definition of the optimal policy and an easy-to-apply decision rule such as the problem of a newsvendor [6]. This paper proposes a random forest classifier (RFC) technique with an emphasis on feature selection for diagnosing lymph disease. The dimensionality of the lymph disease set of features is reduced at the initial stage with the use of multiple feature selection methods. The identified subset of features at the second stage is utilized in training the RFC with a view of providing an adequate classification. Experimental results reveal that the highest classification of 92.2% was provided by the GA-RFC model. The GA helped diminish the dimension of the input space of features from eighteen to six features with a boost in efficiency of the classification [7]. Random Forest employs bagging, also termed as bootstrap aggregating, to build a combination of CART-like classifiers. Random subset selection of variables at each CART node prevents correlation among classifiers. The method resists noise and overtraining since resampling does not occur based on weights. The method also happens to be computation-efficient relative to the application of boosters and slightly lighter relative to standard bagging [8]. These tests were carried out with a multisource geographic and remote sensing database. The performance of Random Forest has been compared and contrasted with results that utilize bagging and boosting methods [9]. Random Forests possess a few advantages relative to the remaining image classification methods. This paper will detail the Random Forest algorithm and how it operates with strengths and limitations [10]. This study attempts to quantify the results demonstrate that Random has fewer parameters that the users need to set compared with SVMs and that it has a simpler setup compared with SVMs [11]. A domain-based Random Forest of decision trees approach is introduced to predict the interaction of proteins. The entire set of pairs of domains are considered, and the predictions are made based on the entire set of the protein domains. Better sensitivity (79.78%) and specificity (64.38%) are demonstrated using the Saccharomyces cerevisiae [12]. The Random Forest algorithm has been adopted as the classifier due to the algorithmic efficiency with large datasets and the availability of individual feature scores per category. The measurable outcomes highlight the added benefit through the fusion of the multispectral imagery and LiDAR imagery optically and the importance of the employment of the full-waveform LiDAR features to capture the vegetated cover and the built structures [13]. This paper introduces new results of the experiments show that the new Random Forest classifier performs better than the conventional Random Forest algorithm [14]. Effective inventory management rests with correct demand prediction, yet the standard methods fall short of capturing the nonlinear patterns and demand interdependencies. These methods have been shown to significantly enhance the accuracy of classification. Among the most well-known ensemble methods are bagging and boosting [8]. Machine learning algorithms are used in the business decision-making process to forecast backorders for products while providing flexibility for decision makers, increasing process transparency, and keeping the precision high. A range approach is used to specify different levels of predictive features in order to cater to the varying nature of real-time data, which can be influenced by machine or user error. The adjustable range offers the required flexibility to decision managers [15]. This study addressed the backorder forecasting problem by comparing various binary classification machine learning algorithms. Model calibration was performed, with subsequent post-hoc explain ability by the SHAP model for the choice of the most significant features contributing to material backorders [16]. In this work, we present a backorder forecasting model using a deep neural network that effectively balances between backorders and orders fulfilled.

Various balancing methods The balanced data are then used in the fully connected deep neural network presented for the development of the predictive model [17]. In this paper, a new explanatory predictive model using convolutional neural networks (CNN) for predicting backorders in inventory management is proposed Backorders [18]. In this paper, several deep learning (DL) methodologies are presented to leverage the advantages of DL in predicting likely delays in the supply of products in intricate supply chains [19]. There is a proposed machine learning algorithm with an under-sampling process [20]. In the current study, various demand forecasting models were applied to predict demand for grocery products using the application of machine learning algorithms [21]. This paper analysed machine learning algorithm of inventory [22]. The performance of the aforementioned models is ultimately compared and analysed [23]. This paper assumes that the distributions of both demand and supply are unknown. A computationally efficient online learning algorithm is developed to address this uncertainty [24]. They examine stochastic periodic-review inventory systems with lost sales in which the decision maker does not know the actual distribution of the demand and bases decisions on only past sales history [25]. A classical inventory control problem is resolved here: the multiproduct periodic-review lost-sales inventory model with warehouse capacity constraint. This deeply researched problem is approached from the perspective of learning the demand for censored data [26]. Both the lost-sales model with zero lead times and the multiproduct backlogging model with positive lead times are discussed in this paper under unknown cyclic demands, fixed joint-order cost, and order constraints [27]. A traditional periodic-review lost-sales inventory model with lead times is investigated in this paper that is not only complex to optimize but also widely applied in numerous practical applications where efficient inventory management is necessary despite stochastic demands and supplies. The complexity is caused by the interaction between lost sales, replenishments, and lead times that makes traditional optimization inefficient. In this study, contributions are made to develop more efficient inventory management policies [28]. Inventory planning forms the basis for precise demand forecasting, yet the traditional approaches are unable to capture the demand relationships as well as the nonlinear behavior of the demand. The paper introduces a Machine learning inventory forecasting framework with the use of Prediction for the sales and Anomaly detection in retail transaction of Random Forest classifier networks which learn complex patterns of the past. Our proposed system is developed with real inventory data sets and attains higher predictive accuracies with the use of machine learning techniques.

2. Random Forest Classifier

2.1. Bagging (Bootstrap Aggregating)

Random Forest is based on the bagging technique, which involves training multiple decision trees on randomly sampled subsets of the training data.

$D = \{(x_i, y_i)\}_{i=1}^{N}$ be the training dataset with N samples.

θ_t be the random variable representing a specific bootstrap sample for the t th decision tree

$h_t(x; \theta_t)$ be the prediction function trained on θ_t

Each tree is trained on a bootstrap sample Dt drawn randomly with replacement from D, meaning some instances may appear multiple times while others may be excluded.

2.2. Decision Tree Model

Each decision tree $h_t(x)$ splits the data based on a criterion such as Gini impurity or entropy

or a node containing mmm samples, the Gini impurity is defined as:

$G = 1 - \sum_{j=1}^{C} \quad p_j^2$

where: C is the number of classes,

pj is the proportion of samples belonging to class j.

A split is chosen that minimizes the weighted Gini impurity.

ISSN: 2576-8484
Vol. 9, No. 4: 1795-1807, 2025
DOI: 10.55214/25768484.v9i4.6383

2.3. Entropy (Information Gain)
Entropy measures the uncertainty of a node:
$$H = -\sum_{j=1}^{C} P_j log_2 P_j$$

The information gain from splitting on feature x is
$$IG(x) = H_{parent} - \sum_k \frac{N_k}{N} H_{child_k}$$
where:

N_k is the number of samples in child node k,

$H_{child_k}$is the entropy of the child node. The split with the highest Information Gain is selected.

H_{parent} represents the entropy of the parent node before the dataset is split based on a given feature x.

2.4. Random Feature Selection

Each tree randomly selects a subset of features at each node to split the data. If there are MMM total features, the subset size is typically:
$$m = log_2 M$$
where mmm is the number of randomly chosen features. This reduces correlation among trees and enhances model generalization.

2.5. Majority Voting for Classification

Each decision tree in the forest makes an independent prediction, and the final classification is determined using majority voting.
For an input sample x, let the predictions from T trees be:
$$h_1(x), h_2(x), \dots, h_T(x)$$
The final prediction H(x) is given by
$$H(x) = argmax \sum_{t=1}^{T} 1. h_t(x)$$
where:
1(.) is the indicator function, which is 1 if $h_t(x)$=y and 0 otherwise.
The class "y" with the most votes is selected.

2.6. Probability Estimation
Random Forest can estimate class probabilities by averaging the probabilities predicted by all trees
$$p\left(\frac{y}{x}\right) = \frac{1}{T}\sum_{t=1}^{T} P_t(\frac{y}{x})$$
where $P_t(\frac{y}{x})$is the probability assigned to class y by tree t.

2.7. Expected Error Reduction
The generalization error of a Random Forest can be approximated as:
$$E = \rho. E_T + (1-\rho). E_B$$
where:

E_T is the error of an individual tree.

E_B is the error from bagging.

ρ is the correlation between trees.

Since bagging reduces variance, and random feature selection reduces correlation ρ, Random Forest achieves lower overall error than individual decision trees.

ISSN: 2576-8484
Vol. 9, No. 4: 1795-1807, 2025
DOI: 10.55214/25768484.v9i4.6383

3. Database Information

The Online Retail Data Set in the UCI Machine Learning Repository is a large transaction database containing sales data from an online store operating in the UK during the period 1st December 2010 to 9th December 2011. The database contains 541,909 instances and contains vital attributes such as InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country. With a 22.6 MB Excel file size, it provides a large source of data to study the buying behavior of customers, product demand, and selling trends. The dataset exhibits characteristics typical of real-world transactional data, including missing values (especially in the CustomerID and Description fields) and potential data anomalies such as duplicate or canceled transactions, which can be identified through negative quantity values. Researchers and data analysts can leverage this dataset for various applications, including market basket analysis, customer segmentation, sales forecasting, anomaly detection, customer lifetime value estimation, and product recommendation systems. Given its detailed transactional nature, the dataset is particularly useful for understanding consumer purchasing patterns and optimizing retail strategies through machine learning and data mining techniques.

There is transactional data regarding inventory management with the notable attributes that help in analyzing customer behavior and sale patterns. An Invoice No uniquely identifies each transaction with the transaction prefix starting with a 'C' indicating cancellation orders. The Stock Code and Description identify the specific item transacted with the Quantity field indicating the unit count purchased per transaction and the date and time of purchase with the InvoiceDate field. The Unit Price field indicates the price per unit in pounds (£) and will help in the analysis of the revenue. The customer has a unique identification with the field of CustomerID and will help in customer-centric analysis. The field of Country indicates the geographical area of the customer and will help in analyzing the sale of the respective area. The combination of the attributes will help in a detailed analysis of inventory patterns, customer behavior, and demand analysis.

Table 1.
Retail description

	Quantity	InvoiceDate	UnitPrice	CustomerID
Count	541909.000000	541909	541909.000000	406829.000000
Mean	9.552250	2011-07-04 13:34:57.156386048	4.611114	15287.690570
Min.	-80995.000000	2010-12-01 08:26:00	-11062.060000	12346.000000
25%	1.000000	2011-03-28 11:34:00	1.250000	13953.000000
50%	3.000000	2011-07-19 17:17:00	2.080000	15152.000000
75%	10.000000	2011-10-19 11:27:00	4.130000	16791.000000
Max.	80995.000000	2011-12-09 12:50:00	38970.000000	18287.000000

In Table 1, the dataset comprises 541,909 transactions with eight attributes related to inventory management, offering valuable insights into sales patterns and customer behavior. However, a preliminary analysis reveals certain data quality issues that need to be addressed. The Description column has 1,454 missing values, which may indicate incomplete product details, while the CustomerID column has 135,080 missing entries, suggesting that some transactions are not linked to registered customers, possibly due to guest checkouts or incomplete records. The data types are mostly appropriate, with InvoiceNo, StockCode, and Country stored as categorical variables, while InvoiceDate is correctly formatted as a datetime type, allowing for time-based analysis. The Quantity is represented as an integer, indicating whole units of products, while UnitPrice and CustomerID are stored as float values, though CustomerID might require conversion to an integer. Potential issues include invoices starting with 'C', which denote cancellations that should be excluded when analyzing sales trends, as well as negative or zero values in Quantity and UnitPrice, which may reflect returns, refunds, or data entry errors. Given these aspects, data cleaning is essential to manage missing values, remove anomalies, and refine the dataset before applying predictive models or conducting inventory forecasting to ensure accurate and meaningful analysis.

Table 2.
Null retail details

InvoiceNo	0
StockCode	0
Description	1454
Quantity	0
InvoiceDate	0
UnitPrice	0
CustomerID	135080
Country	0

In Table 2, the description and CustomerID. The Description field has 1,454 missing values that might be indicative of incomplete product details and errors during the data entry process. Even more critical, the field of CustomerID has 135,080 missing values that imply that a large chunk of the transaction does not relate to registered customers. This might be the result of guest purchases and incomplete customer data that might influence customer segmentation and repeated purchase analysis. Other fields such as InvoiceNo, StockCode, Quantity, InvoiceDate, UnitPrice, and Country are filled with no missing values. Considering such missing values, suitable preprocessed techniques such as imputation and exclusion need to be applied before further analysis to support the integrity and consistency of the data.

There are 22,190 transactions with 3,684 unique goods purchased by 4,372 individual buyers in 37 countries. The heterogeneity of the database emphasizes the scope of analyzing worldwide purchase behavior, customer preferences and tastes, and inventory patterns. The numerous counts of unique goods signifies a diversified inventory that may call for good inventory management and demand forecasting techniques. The geographic coverage of the database in multiple countries also poses an opportunity to investigate cross-country variation of purchase behavior. The identification of such vital metrics will optimize inventory levels, refine customer segmentation, and maximize selling strategies toward making good business decisions.

Table 3.
Retails country data

United Kingdom	356728
Germany	9480
France	8475
EIRE	7475
Spain	2528
Netherlands	2371

In Table 3 The database has 37 countries' transactions with most of them belonging to the United Kingdom with 356,728 transactions making it the top market. Other countries that follow include 9,480 in Germany, 8,475 of France, and 7,475 of EIRE(Ireland) showing good sales within the continent of Europe. Other countries that also play a large role include 2,528 in Spain, 2,371 in the Netherlands, 2,069 in Belgium, and 1,877 in Switzerland. Out of Europe, 1,258 from Australia, 1,086 from Norway, 358 from Japan, and 291 from the USA also play a large role in the database.

Also, 241 of the transactions are labeled as "Unspecified," and it might be that they need further analysis to identify the source of the transactions. The occurrence of countries with lower transaction figures, including Saudi Arabia (10) and Bahrain (17), indicates minimal penetration within the respective countries. The figures overall imply a geographically dispersed customer base and the need for localized sales analysis and customized inventory management strategies. The set of checks has passed, the database is now set for further analysis such that demand prediction, customer segregation, and inventory optimization will be feasible.

In Fig 1, the dataset provides a comprehensive view of inventory transactions across 37 countries, with the United Kingdom being the dominant market. The presence of 22,190 transactions, 3,684 unique products, and 4,372 distinct bar charts visualizes the top 10 countries with the highest number of transactions in the dataset. The United Kingdom overwhelmingly dominates the transaction count, far surpassing all other countries. Germany, France, and EIRE (Ireland) follow, but their transaction volumes are significantly lower. This suggests that the dataset is primarily influenced by UK-based transactions, making it the focal point for sales and inventory analysis. The skewed distribution may indicate a UK-centric business operation, requiring tailored strategies for international expansion.

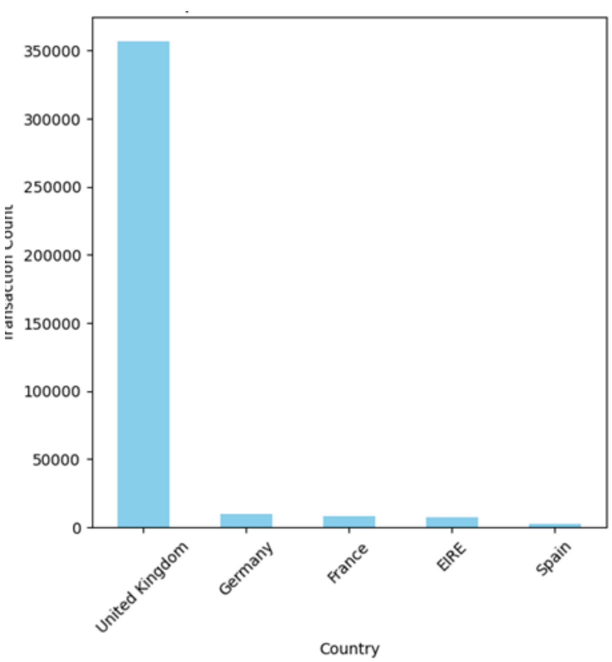

Figure 1.
Country with most number of transaction

In Figure 2, the monthly patterns of 2011 and 2010 sales during the period of the time are reflected in the line graph. The constant growth of the sales during 2011 with a sudden jump during November may be attributed to holiday and seasonal demand. The sudden fall during December may be attributed to lower holiday aftermath sales or lost data towards the latter half of the month. The overall direction of the line reflects a strong growth of the sales during the year with clear seasonality affecting the patterns of the revenue.

ISSN: 2576-8484
Vol. 9, No. 4: 1795-1807, 2025
DOI: 10.55214/25768484.v9i4.6383

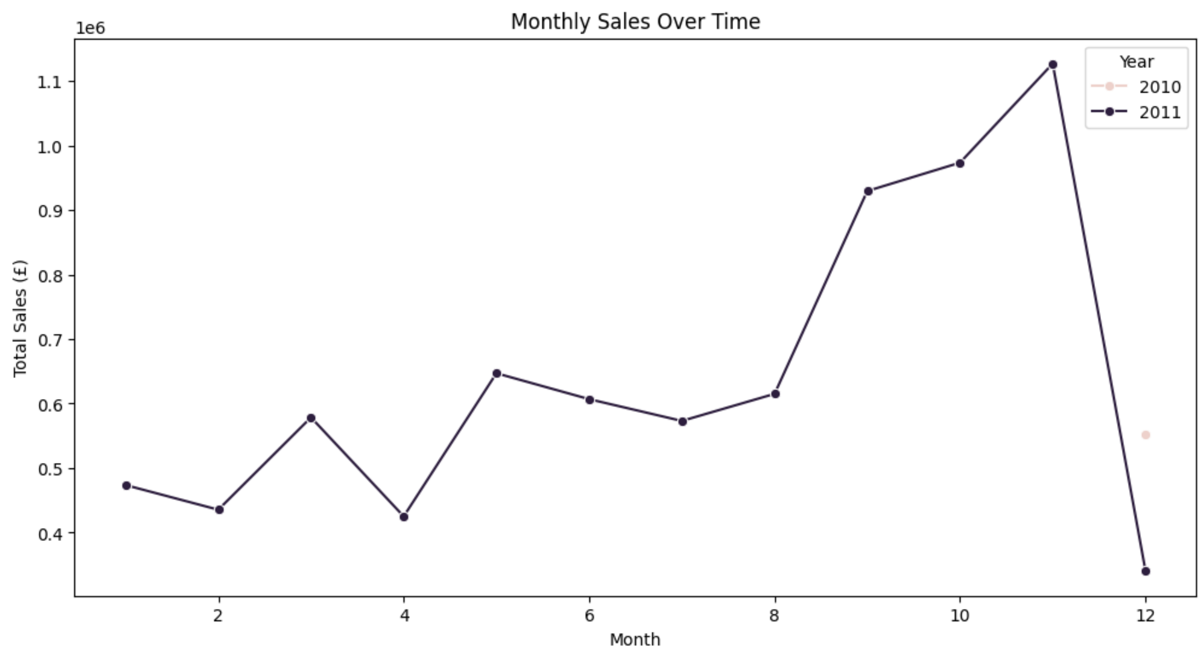

Figure 2.
Monthly sales over time

In Table 4 the precision, Recall, and the F1-score are 1.00 each, indicating that the model accurately classifies with no errors whatsoever.

Accuracy: 100% (1.00) on 80, 321 samples.

Macro Average & Weighted Average: They are 1.00 each, which signifies that the classes performed equally and consistently throughout.

This means that the model has either been highly trained or possibly overfit if real-world data will be different. This is the ideal situation if this is an experiment; otherwise testing against unseen or varied data will be valuable in verifying its strength.

Table 4.
Classification Report

	Precision	Recall	f1-score	Support
0	1	1	1	78635
1	1	1	1	1686
Accuracy			1	80321
Maco average	1	1	1	80321
Weighted average	1	1	1	80321

4. Application I Prediction for the next 30 days' sales

4.1. Algorithm

Loads the Online Retail Dataset.
Converts InvoiceDate to a datetime format.
Aggregates daily sales revenue.
Fits an ARIMA (5,1,0) model to the sales data.
Forecasts the next 30 days of sales.
Plots both the historical sales and forecasted sales.

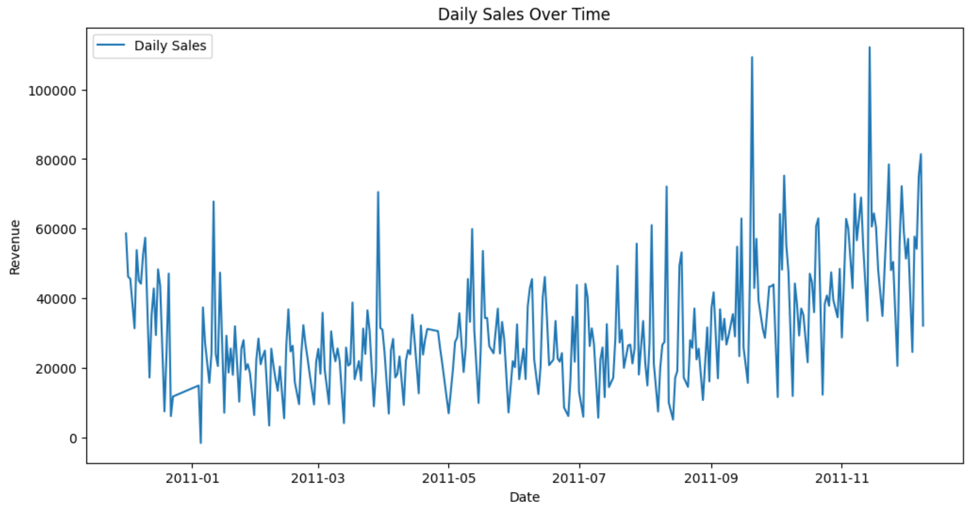

Figure 3.
Daily sales over time

In Figure 3 this dataset is valuable for various analytical applications in retail and e-commerce. Market Basket Analysis helps identify frequently bought-together items using association rule mining techniques like the Apriori algorithm, enabling businesses to optimize cross-selling strategies. The time series plot of daily sales revenue from the Online Retail Dataset reveals significant fluctuations in revenue over time, indicating variability in customer purchases. Several sharp peaks suggest high sales days, possibly due to promotions, seasonal demand, or bulk purchases by certain customers. Despite these fluctuations, an overall upward trend is noticeable toward the end of the dataset, which could indicate growing customer engagement, an expanding product range, or increased market reach. Understanding these trends can help businesses optimize inventory, forecast demand, and implement targeted marketing strategies to enhance revenue.

In Fig 4 This set of data has multiple analytical applications within the area of business and commerce. Market Basket Analysis helps in the identification of the often purchased-together item pairs based on the rules supplied the supplied time series plot plots the observed sale revenue against a 30-day prediction. The blue plot of the sale figures has a lot of variation with multiple maxima and indicates the variation of the sale activities with the passage of time. The red dashed line indicates the predicted sale and displays a reasonably stable trend with small variances. The prediction follows the patterns of the previous few days and captures the underlying behavior of the data. This prediction will be useful in inventory planning, demand prediction, and strategic business decision-making to maximize business activities.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4: 1795-1807, 2025
DOI: 10.55214/25768484.v9i4.6383

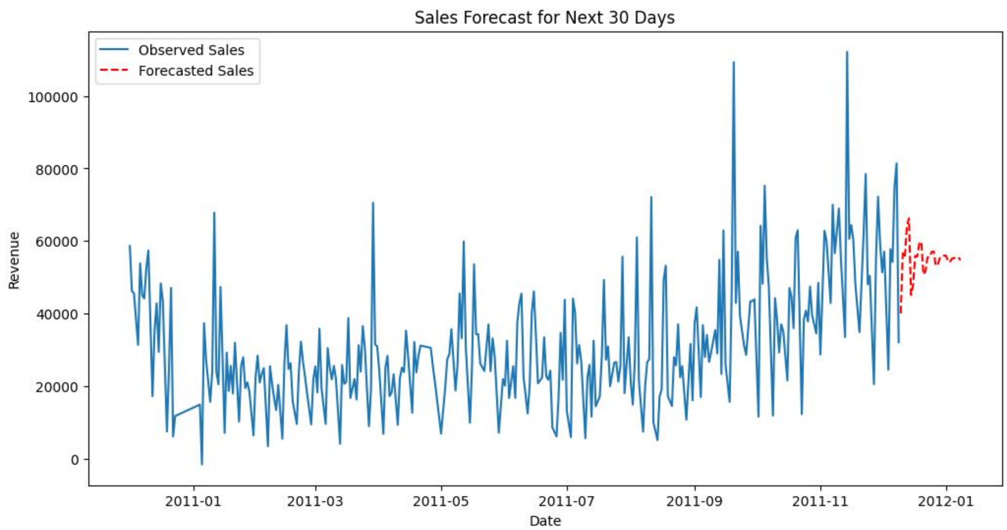

Figure 4.
sales forecast for the next 30 days

5. Application II Anomaly Detection in Retail Transactions

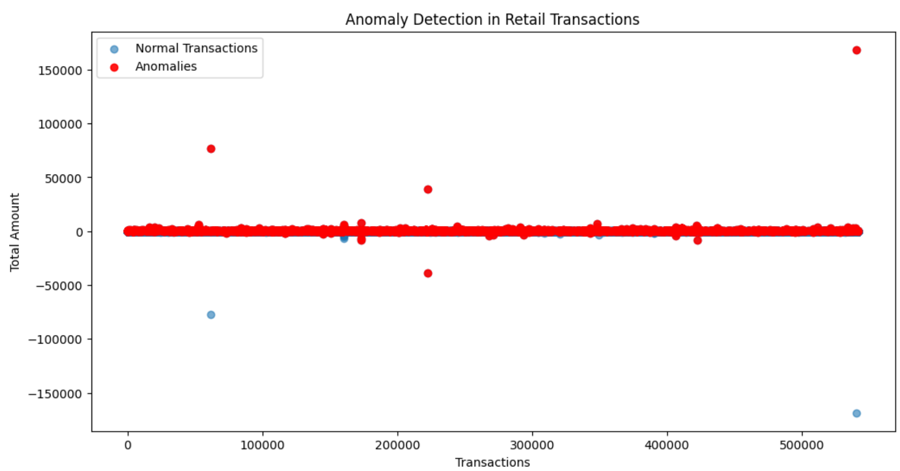

Figure 5.
Anomaly detection in a retail transaction

In Fig 5 The scatter plot illustrates anomaly detection in retail transactions, where normal transactions are depicted in blue, and anomalies are marked in red. The x-axis represents the number of transactions, while the y-axis shows the total transaction amount. Most transactions are concentrated around zero, indicating a high volume of small or balanced transactions. However, several extreme outliers are noticeable, with some transactions exhibiting significantly high or low amounts. These anomalies could indicate fraudulent activities, unusual purchasing behavior, or data entry errors. Identifying such irregularities is crucial for fraud detection, financial risk assessment, and ensuring the integrity of retail transaction data.

Table 5 shows That the detected anomalies in the retail transactions exhibit abnormal patterns of purchase that deviate widely from typical transactions. They comprise large-quantity orders such as an 80,995-unit purchase of "Paper Craft, Little Birdie" with a transaction value of 168,469.60 and the occurrence of a negative unit quantity such as a manual transaction with the unit quantity of -1. They may be indicative of errors in entering the data, attempts at fraud, bulk purchase orders placed by wholesalers, and abnormal returns. The occurrence of abnormally large unit prices and transaction values also signifies the possibility of price errors and price manipulation. The detection and analysis of the anomalies identify the fraud ensure the correctness of the data and optimize the inventory management.

Table 5.
Detected Anomalies

	InvoiceNo	StockCode	Description	Quantity	UnitPrice	TotalAmount	CustomerID
65	536374	21258	Victorian sewing box large	32	10.95	350.40	15100.0
178	536387	79321	Chilli Lights	192	3.82	733.44	16029.0
179	536387	22780	Light garland butterflies pink	192	3.37	647.04	16029.0
180	536387	22779	Wooden owls light garland	192	3.37	647.04	16029.0
181	536387	22466	Fairy tale cottage nightlight	432	1.45	626.40	16029.0
540421	581483	23843	Paper craft , little birdie	80995	2.08	168469.60	16446.0
540425	581485	20749	Assorted colour mini cases	84	6.35	533.40	17389.0
540442	581487	21137	Black record cover frame	120	3.39	406.80	15694.0
541541	C581499	M	Manual	-1	224.69	-224.69	15498.0
541702	581566	23404	Home sweet home blackboard	144	3.26	469.44	18102.0

6. Conclusion

In summary, the dataset captures a complete picture of inventory transactions in 37 countries with the leading market in the United Kingdom. The 22,190 transactions, 3,684 distinct products, and 4,372 individual customers demonstrate the promise of the dataset in providing meaningful analysis of demand prediction, inventory optimization, and customer behavior analysis. The 135,080 and 1,454 missing values in the CustomerID and Description columns need careful treatment to preserve the integrity of the data. The anomalies of the Quantity and Unit Price also need closer attention to correct errors and cancel orders. The geographic dispersion of the transactions opens the door to examining the sale patterns of the different geographic areas and the formulation of inventory management strategies per area. This dataset has the promise of making a major contribution toward the enhancement of prediction analytics in inventory management and decision-making with the use of the correct pre-processed and modeled techniques.

This set of data has numerous analytical use cases in the field of commerce and retailing. Market Basket Analysis identifies often-purchased-together goods with association rule mining methods such as the Apriori algorithm, allowing companies to maximize cross-selling efforts. Customer Segmentation involves the use of techniques such as K-Means or DBSCAN to categorize buyers based on purchase behavior and allow them to target marketing efforts. Sales Forecasting employs the use of time-series models such as ARIMA and Prophet to estimate future sales patterns and aid inventory management and demand planning efforts. Anomaly Detection aids in the detection of fraudulent transactions and abnormal purchase behavior, making fraud detection and prevention efforts better. Lastly, the use of Customer Lifetime Value (CLV) Estimation allows companies to estimate future customer value and inform better customer retention and targeting of marketing efforts. These analytical techniques aid in making better business decisions, boost sales, and enhance business performance overall.

Transparency:
The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

Reference
[1] A. J. Clark and H. Scarf, "Optimal policies for a multi-echelon inventory problem," *Management Science*, vol. 6, no. 4, pp. 475-490, 1960. https://doi.org/10.1287/mnsc.1040.026
[2] B. J. De Moor, J. Gijsbrechts, and R. N. Boute, "Reward shaping to improve the performance of deep reinforcement learning in perishable inventory management," *European Journal of Operational Research*, vol. 301, no. 2, pp. 535-545, 2022. https://doi.org/10.1016/j.ejor.2021.10.045
[3] T. E. Goltsos, A. A. Syntetos, C. H. Glock, and G. Ioannou, "Inventory–forecasting: Mind the gap," *European Journal of Operational Research*, vol. 299, no. 2, pp. 397-419, 2022. https://doi.org/10.1016/j.ejor.2021.07.040
[4] D. Prak and R. Teunter, "A general method for addressing forecasting uncertainty in inventory models," *International Journal of Forecasting*, vol. 35, no. 1, pp. 224-238, 2019. https://doi.org/10.1016/j.ijforecast.2017.11.004
[5] W. Van Jaarsveld and J. Arts, "Projected inventory-level policies for lost sales inventory systems: Asymptotic optimality in two regimes," *Operations Research*, vol. 72, no. 5, pp. 1790-1805, 2024. https://doi.org/10.1287/opre.2021.0032
[6] J. F. Van Der Haar, A. P. Wellens, R. N. Boute, and R. J. Basten, "Supervised learning for integrated forecasting and inventory control," *European Journal of Operational Research*, vol. 319, no. 2, pp. 573-586, 2024. https://doi.org/10.1016/j.ejor.2024.07.004
[7] A. T. Azar, H. I. Elshazly, A. E. Hassanien, and A. M. Elkorany, "A random forest classifier for lymph diseases," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 2, pp. 465-473, 2014. https://doi.org/10.1016/j.cmpb.2013.11.004
[8] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognition Letters*, vol. 27, no. 4, pp. 294-300, 2006. https://doi.org/10.1016/j.patrec.2005.08.011
[9] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forest classification of multisource remote sensing and geographic data," presented at the IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium, IEEE, 2004.
[10] N. Horning, "Random Forests: An algorithm for image classification and generation of continuous fields data sets," in *Proceedings of the International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences, Osaka, Japan*, 2010, vol. 911, pp. 1-6.
[11] M. Pal, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217-222, 2005. https://doi.org/10.1080/01431160412331269698
[12] X.-W. Chen and M. Liu, "Prediction of protein–protein interactions using random decision forest framework," *Bioinformatics*, vol. 21, no. 24, pp. 4394-4400, 2005. https://doi.org/10.1093/bioinformatics/bti721
[13] L. Guo, N. Chehata, C. Mallet, and S. Boukir, "Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 1, pp. 56-66, 2011. https://doi.org/10.1016/j.isprsjprs.2010.08.007
[14] A. Chaudhary, S. Kolhe, and R. Kamal, "An improved random forest classifier for multi-class classification," *Information Processing in Agriculture*, vol. 3, no. 4, pp. 215-222, 2016. https://doi.org/10.1016/j.inpa.2016.08.002
[15] S. Islam and S. H. Amin, "Prediction of probable backorder scenarios in the supply chain using Distributed Random Forest and Gradient Boosting Machine learning techniques," *Journal of Big Data*, vol. 7, no. 1, p. 65, 2020. https://doi.org/10.1186/s40537-020-00345-2
[16] C. Ntakolia, C. Kokkotis, P. Karlsson, and S. Moustakidis, "An explainable machine learning model for material backorder prediction in inventory management," *Sensors*, vol. 21, no. 23, p. 7926, 2021. https://doi.org/10.3390/s21237926
[17] M. Shajalal, P. Hajek, and M. Z. Abedin, "Product backorder prediction using deep neural network on imbalanced data," *International Journal of Production Research*, vol. 61, no. 1, pp. 302-319, 2023. https://doi.org/10.1080/00207543.2021.1901153
[18] M. Shajalal, A. Boden, and G. Stevens, "Explainable product backorder prediction exploiting CNN: Introducing explainable models in businesses," *Electronic Markets*, vol. 32, no. 4, pp. 2107-2122, 2022. https://doi.org/10.1007/s12525-022-00599-z

[19] M. M. Bassiouni, R. K. Chakrabortty, K. M. Sallam, and O. K. Hussain, "Deep learning approaches to identify order status in a complex supply chain," *Expert Systems with Applications*, vol. 250, p. 123947, 2024. https://doi.org/10.1016/j.eswa.2024.123947

[20] P. Hajek and M. Z. Abedin, "A profit function-maximizing inventory backorder prediction system using big data analytics," *IEEE Access*, vol. 8, pp. 58982-58994, 2020. https://doi.org/10.1109/ACCESS.2020.2983118

[21] N. Vairagade, D. Logofatu, F. Leon, and F. Muharemi, "Demand forecasting using random forest and artificial neural network for supply chain management," presented at the Computational Collective Intelligence: 11th International Conference, ICCCI 2019, Hendaye, France, September 4–6, 2019, Proceedings, Part I 11, Springer International Publishing, 2019.

[22] H. Bousqaoui, I. Slimani, and S. Achchab, "Comparative analysis of short-term demand predicting models using ARIMA and deep learning," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 4, p. 3319, 2021. https://doi.org/10.11591/ijece.v11i4.pp3319-3328

[23] A. Husna, S. H. Amin, and B. Shah, *Demand forecasting in supply chain management using different deep learning methods. In Demand forecasting and order planning in supply chains and humanitarian logistics*, 140-170 ed. United States: IGI Global, 2021.

[24] B. Chen, J. Jiang, J. Zhang, and Z. Zhou, "Learning to order for inventory systems with lost sales and uncertain supplies," *Management Science*, vol. 70, no. 12, pp. 8631-8646, 2024. https://doi.org/10.1287/mnsc.2022.02476

[25] J. Ding, W. T. Huh, and Y. Rong, "Feature-based inventory control with censored demand," *Manufacturing & Service Operations Management*, vol. 26, no. 3, pp. 1157-1172, 2024. https://doi.org/10.1287/msom.2021.0135

[26] S. Guo, C. Shi, C. Yang, and C. Zacharias, "An online mirror descent learning algorithm for multiproduct inventory systems," University of Miami Business School Research Paper No. 4806687, 2024.

[27] X.-Y. Gong and D. Simchi-Levi, "Bandits atop reinforcement learning: Tackling online inventory models with cyclic demands," *Management Science*, vol. 70, no. 9, pp. 6139-6157, 2024. https://doi.org/10.1287/mnsc.2023.4947

[28] C. Lyu, H. Zhang, and L. Xin, "Ucb-type learning algorithms with kaplan–meier estimator for lost-sales inventory models with lead times," *Operations Research*, vol. 72, no. 4, pp. 1317-1332, 2024. https://doi.org/10.1287/opre.2022.0273

DOI: 10.55214/25768484.v9i4.6383