© March 2025| IJIRT | Volume 11 Issue 10 | ISSN: 2349-6002

Data Exploration and Demand Prediction in Medical Supply: A Machine Learning Approach

Vibinsuryaa M[1], Solomon A[2], Praveenkumar A K[3], Srikrishnaprasanth N[4], Dr. A. S. Muthanantha Murugavel

[1,2,3,4]*Computer Science and Engineering, karpagam College of Engineering, India*

[5]*Co-Author, Computer Science and Engineering, Karpagam College of Engineering, Coimbatore*

Abstract-*The healthcare industry relies heavily on the well-organized management of medical supplies to ensure timely and quality patient care. Accurate demand prediction for medical supplies is critical to prevent shortages, optimize stock levels, and reduce wastage. This study explores data-driven approaches to medical supply management by conducting an in-depth analysis of historical demand data and smearing machine learning techniques to foresee future demand. The data exploration phase involves cleaning and preprocessing data from various sources, shadowed by analyzing trends, seasonality, and anomalies. Insights gained from data exploration help identify key factors influencing demand, such as seasonal patterns, economic conditions, and population health metrics. The study creates and tests various machine learning models for demand prediction, such as time series models, regression techniques, and advanced deep learning methods. Prediction accuracy and the capacity to identify intricate correlations in the data are the two main criteria used to evaluate models. To identify the best-fit perfect for medicinal supply-demand forecasting, the study evaluates the effectiveness of more modern machine learning techniques against more conventional statistical methods.*

Keywords: Demand Prediction, Seasonality Analysis, Supply Chain Optimization.

1. INTRODUCTION

In the healthcare business, reliable demand forecasting for vital drugs is critical to avoiding shortages that jeopardize patient care as well as costly surpluses. It is difficult to forecast the demand for medications because of variables that affect consumption patterns in complex ways, such as seasonal fluctuations, local healthcare requirements, and unforeseen outbreaks. These swings pose serious difficulties for legislators and healthcare providers attempting to ensure steady and effective supply levels. By examining past consumption statistics to find significant plans and tendencies, this research seeks to usage machine-learning techniques to increase the precision of demand estimates for human medications. The project aims to assist healthcare companies in improving inventory management, cutting waste, and guaranteeing timely access to necessary pharmaceuticals by creating a strong predictive model. Data from human medicine is the primary focus at first, with intentions to apply learned lessons to other fields, such as veterinary medicine, in later stages. By utilizing machine learning techniques, this research tackles the need for a more accurate and flexible method of forecasting medical demand. The project intends to help healthcare providers maintain optimal inventory levels, reduce shortages, and improve the availability of essential medications by creating a predictive model that can precisely predict demand trends based on past data. The primary goal of this study is to increase the precision of request predicting for diabetic medications, which is crucial for maximizing output, cutting left over, and guaranteeing patients' appropriate access to these medications. By guaranteeing a steady supply of medications and lowering the possibility of stockouts, precise demand foretelling can also enhance persistent care. To guarantee accurate forecasts, this study will concentrate on determining the main elements impacting demand, choosing suitable machine learning algorithms, and assessing their effectiveness.

2. LITERATURE REVIEW

Anupriya Jain, Vikram Karthikeyan, Sahana B., Shambhavi B.R., Sindhu K., and Balaji S. [1] Focusing on demand forecasting for e-commerce using SARIMA and LSTM methods. Their dataset is from a superstore but includes only a partial number of product categories, which restricts their analysis.

Sara Jebbor et al. [2] select AI-based forecasting models for hospital requests during disasters, using ANN, RNN, ANFIS, and learning-FIS. Their learning is limited to one hospital.

Muhtasim Hafiz, Md Sabir Ibna Sazzad, Khalid Ibne Hasan, Jamil Hasnat, and Mahbub C. Mishu [3] aim to predict medicine requests in Bangladesh with an AI-based LSTM model. They fold data from local pharmaceutical companies. A key limitation is that the model is not well-trained on larger datasets.

Ismail El Kafazi, Rachid Bannari, and Abdellah Abouabdellah [4] compare forecasting methods, precisely linear regression and polynomial curve fitting, for renewable energy production. They composed data from Morocco's energy distribution department, but the accuracy of their results is interrogated.

Gökçe Candan et al. [5] foresee healthcare demand using Neuro-Fuzzy methods and ANN, with data from a pharmaceutical company. However, their model emphasizes on a single drug.

Houria Laaroussi, Fatima Guerouate, and Mohamed Sbihi [6] present a context for forecasting tourism demand using LSTM and SVR, based on data from Morocco's tourism laboratory. Their study highlights a need for enhanced accuracy.

Lei Hao et al. [7] tackle optimization in medication storage and distribution using various procedures and data from multiple pharmacies. Their research faces challenges in balancing replacement and distribution.

Shweta and Dinesh Kumar [8] examine pharmaceutical supply chain subjects in India using AHP and Fuzzy AHP, based on data from the BPPI. Their conclusions may not generalize to other contexts.

Anish Palkar, Mitali Deshpande, Shweta Kalekar, and Shree Jaswal [9] discover demand forecasting for liquor feasting using LSTM networks. They utilize the Iowa Liquor Sales Dataset from Kaggle and discuss several techniques, including SVM, SVR, and artificial neural networks. The restricted range of machine learning techniques practically restricts their learning.

Francois Mbonyinshuti et al. [10] forecast trends for vital medicines using systems like linear regression, ANN, and Random Forest. They obtained data from the eLMIS system but faced contests in using all the data due to scope restrictions.

3. METHODOLOGY

The methodology for this study was developed to provide a comprehensive approach to demand forecast in medical stores using advanced data analytics and machine learning approaches. The approach started with data collecting from medical stores, which included six months of historical sales records with attributes including Medicine_name, Quantity_sold, Seasonal_regular categorization, and Total_sales. The dataset was divided into regular and seasonal drugs to allow for targeted research. Data quality was ensured through preprocessing methods such as managing lost values, encoding definite variables, and scaling numerical structures. Temporal information such as Month and Year were converted into cyclical formats to preserve seasonality. Data integrity was ensured by outlier identification and treatment, and model performance was evaluated using an 80%-20% traintest split. Seasonal trends, regional variations, and consumption patterns were found using exploratory data analysis, or EDA. While visualizations like histograms, heatmaps, and line charts showed sales trends over time and between districts, aggregates by Medicine_name revealed the most often used medications. For predictive modeling, machine learning models like Random Forest, ARIMA, and LSTM were used. For long-term forecasting, Random Forest captured nonlinear interactions and feature relevance, ARIMA handled time-series trends, and LSTM handled sequential dependencies. Metrics like RMSE and R2 were used to train and validate the models, and visuals were used to evaluate their performance. The findings were compared to existing literature, demonstrating the strength of the suggested approach while addressing limitations such as data size and feature scope for future enhancement.

Time series forecasting includes a sequence of steps that help to generate a model that predicts the future values of a time series. Here are the general steps complicated in time series prediction:

3.1 Data Collection

In order to provide reliable machine learning analysis for demand forecast in medical stores, the dataset utilized in this work was meticulously gathered. It includes sales data from seven different medical stores during the last six months. With 8,795 data arranged in a table format, the collection contains important information regarding past sales and product performance. Store_ID, Medicine_name, Quantity_sold, Date, Medicine_category, and Seasonal_regularity are some of the essential components. By capturing crucial elements such as the kind and quantity of medications sold, these variables enable a more thorough examination of regular demand patterns and seasonal fluctuations. To enable more focused prediction modeling, the information further separates drugs into regular (such as cardiovascular medications) and seasonal (such as antivirals) groups. The identification of regional and temporal demand fluctuations is enhanced by the combination of geographical data (district) and temporal factors (month and year). A comprehensive analysis of sales trends is made possible by this large feature set, which enables predictive models to offer useful insights. Medical stores may guarantee accurate demand forecasting and effective inventory management by using the processed dataset as a base for sophisticated machine learning techniques.

3.2 Data Exploration

The exploratory data analysis (EDA) on the medical store datasets used a variety of statistical and analytical techniques to identify patterns and trends.

Regular Medicine Summary:		
Medicine_name	**Quantity_sold**	**Frequency**
321 Domperidone (50mg)	**721**	**13450**
198 Aceclofenac (200mg)	**815**	**9420**
154 Azithromycin (300mg)	**532**	**1567**
276 Cefixime (100mg)	**654**	**1980**
429 Ofloxacin (500mg)	**498**	**3120**
Seasonal Medicine Summary:		
Medicine_name	**Quantity_sold**	**Frequency**
267 Cefixime (250mg)	**5252**	**45210**
189 Azithromycin (400mg)	**4100**	**51890**
492 Domperidone (20mg)	**2890**	**40120**
97 Aceclofenac (150mg)	**2675**	**38960**
329 Ofloxacin (300mg)	**2301**	**20540**

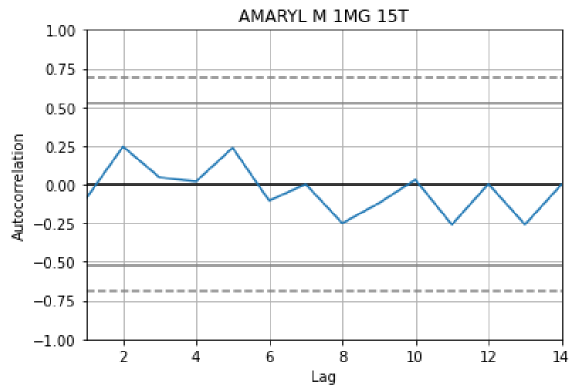

Fig 2: Seasonality based Autocorrelation for AMARYL 1MG 15T

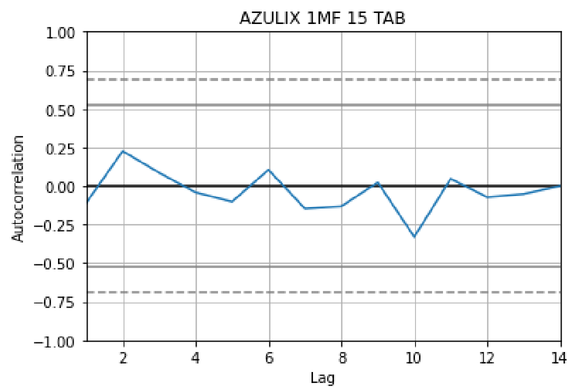

Fig 3: Seasonality based Autocorrelation for AZULIX 1MF 15TAB

In conclusion, the Autocorrelation function (ACF) can be used to detect seasonality in time series data by plotting the ACF for the seasonal difference of the time series. Significant correlation at seasonal lags indicates the presence of seasonality in the data.

4. MODEL IMPLEMENTATION

ARIMA (Autoregressive Integrated Moving Average) and LSTM (Long Short-Term Memory) are among the most widely used methods for time series forecasting, each employing distinct techniques to capture and model patterns in temporal data.

LSTM, a type of recurrent neural network, is precisely designed to handle long-term dependencies in sequential information. It can selectively retain or discard past data, allowing it to identify complex patterns and non-linear dynamics. This capability makes LSTM particularly effective for forecasting time series data such as weather conditions and stock market trends. However, it does have higher data and computational requirements.

On the other hand, ARIMA models focus on the autocorrelation structure within a time series to predict future values. They consist of components that address autoregression, differencing, and moving averages, making ARIMA a popular choice for forecasting in finance and economics. Although ARIMA models are generally easier to interpret and implement, they require a stationary time series for optimal performance and may struggle with non-linear or non-stationary patterns.

4.1 ARIMA Model Implementation

The ARIMA model is a powerful tool for accurately forecasting medicine demand based on historical sales data. By following these clear steps, you can effectively implement the ARIMA model:

1. Begin by determining the ARIMA model parameters (p, d, and q) using ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots. This foundational analysis will guide your model selection.
2. Next, fit the ARIMA model to your sales data with user-friendly software like Python's statsmodels or R's forecast package. These tools make the fitting process seamless and efficient.
3. To ensure the reliability of your model, evaluate its accuracy with metrics such as Mean Squared Error (MSE) or Root Mean Squared Error (RMSE), along with visual inspections and statistical tests like the Ljung-Box test. This thorough validation will bolster your confidence in the results.
4. Finally, leverage the fitted ARIMA model to generate insightful forecasts for future demand. By inputting the latest sales data and applying the model’s coefficients, you can make informed decisions for your inventory and supply chain management.

4.2 LSTM Model Implementation

Here are the key steps to implement an effective LSTM model for forecasting medicine demand:

1. Split the Data: Divide the preprocessed data into training, validation, and testing sets for thorough evaluation.
2. Normalize the Data: Adjust the data to have a zero mean and unit variance, enhancing model learning.
3. Define the Model: Create the LSTM architecture by specifying cells, layers, activation functions, and hyperparameters.
4. Compile the Model: Set the loss function, optimizer, and essential training parameters for successful training.
5. Train the Model: Train on the training data while validating against the validation set to ensure performance.
6. Evaluate the Model: Use the testing data to assess the model with metrics like MSE, MAE, and R^2.
7. Make Predictions: Leverage the trained model to forecast future demand by inputting the latest sales data.

These streamlined steps will help you build a powerful forecasting tool efficiently.

5. RESULT EVALUVATION

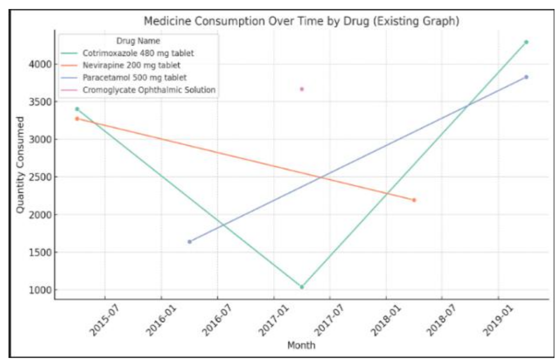

Fig 4: ARIMA Model for Seasonality Medicines

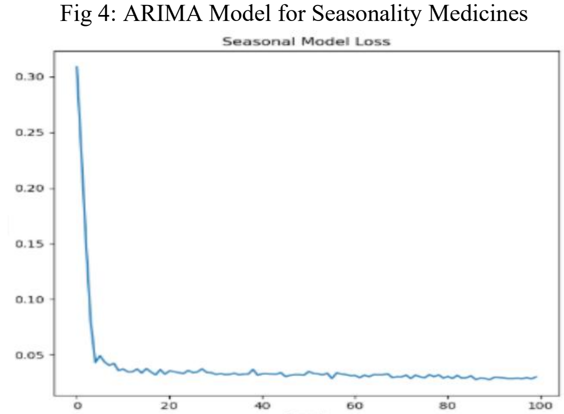

Fig 5: LSTM Model for Seasonality Medicines

In the above Figs, the prediction values for sales of medicine is calculated. The attributes ‘week’ and ‘name of medicines’ are passed to the model. This prediction helps retailers to understand the behavior of the selling pattern in the whole state.

6. CONCLUSION

This research paper demonstrates that the use of ARIMA for demand forecasting in the medical sector can lead to precise predictions and significant cost savings. Key steps in this process include the collection and preprocessing of data, as well as the analysis of stationarity and seasonality, which are essential for preparing the data for prediction. The final stages involve implementing the model and making predictions. Overall, this methodology provides a valuable solution to the challenges associated with forecasting demand for medical supplies.

7. REFERENCES

[1] Palkar, Anish, et al. "Demand Forecasting in Retail Industry for Liquor Consumption using LSTM."*2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*. IEEE, 2020.

[2] Hafiz, Muhtasim, et al. "Predicting the Demand of Prescribed Medicines in Bangladesh using Artificial Intelligent (AI) based Long Short-Term Memory (LSTM) Model." *Proceedings of the International Conference on Computing Advancements*. 2020.

[3] Jain, Anupriya, et al. "Demand forecasting for e-commerce platforms." *2020 IEEE International Conference for Innovation in Technology (INOCON)*. IEEE, 2020.

[4] Laaroussi, Houria, and Fatima Guerouate. "Deep learning framework for forecasting tourism demand." *2020 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)*. IEEE, 2020.

[5] El Kafazi, Ismail, et al. "Energy production: a comparison of forecasting methods using the polynomial curve fitting and linear regression." *2017 International Renewable and Sustainable Energy Conference (IRSEC)*. IEEE, 2017.

[6] Mbonyinshuti, Francois, et al. "The prediction of essential medicines demand: a machine learning approach using consumption data in Rwanda." *Processes* 10.1 (2021): 26.

[7] Kumar, Dinesh. "Analysis of issues of generic medicine supply chain using fuzzy AHP: a Pilot study of Indian public drug distribution scheme." *International Journal of Pharmaceutical and Healthcare Marketing* 15.1 (2021): 18-42.

[8] Jebbor, Sara, Chiheb Raddouane, and Abdellatif El Afia. "A preliminary study for selecting the appropriate AI-based forecasting model for hospital assets demand under disasters." *Journal of Humanitarian Logistics and Supply Chain Management* 12.1 (2022):1-29.

[9] Candan, Gökçe, M. fatih TASKİN, and Harun YAZGAN. "Demand forecasting in pharmaceutical industry using neuro-fuzzy approach." *Journal of Management and Information Science* 2.2 (2014): 41-49.

[10] Hao, Lei, Hongfeng Wang, and Qi Yan. "A Data-Driven Storage Assignment Strategy for Automated Pharmacy." *2021 11th International Conference on Information Science and Technology (ICIST)*. IEEE, 2021.

[11] Elmasdotter, Ajla, and Carl Nyströmer. "A comparative study between LSTM and ARIMA for sales forecasting in retail." (2018).Battineni, Gopi, et al. "Applications of machine learning predictive models in the chronic disease diagnosis." *Journal of personalized medicine* 10.2 (2020): 21.

[12] Fernandes, Bruno, et al. "Long short-term memory networks for traffic flow forecasting: exploring input variables, time frames, and multi-step approaches." *Informatica* 31.4 (2020): 723-749.

[13] Hong, Jun-Ki. "LSTM-based Sales Forecasting Model." *KSII Transactions on Internet & Information Systems* 15.4 (2021).

[14] Arun Kumar, K. E., et al. "Forecasting of COVID-19 using deep layer recurrent neural networks (RNNs) with gated recurrent units (GRUs) and long short-term memory (LSTM) cells." *Chaos, Solitons & Fractals* 146 (2021): 11086.