DOI: 10.6977/IJoSI.202409_8(3).0006
N. Qassrawi, M. Azzeh /Int. J. Systematic Innovation, 8(3), 63-83 (2024)

Drug sales forecasting in the pharmaceutical market using deep neural network algorithms

Noura Qassrawi[1], Mohammad Azzeh[2*], Mohammad Hijjawi[1]
[1] Department of Computer Science, Applied Science Private University, Amman, Jordan
[2] Department of Data Science, Princess Sumaya University for Technology, Amman, Jordan
* Corresponding author E-mail: m.azzeh@psut.edu.jo
(Received 21 September 2023; Final version received 08 January 2024; Accepted 29 February 2024)

Abstract

Drug sales and price forecasting have become an attractive investigation topic due to their important role in the pharmaceutical industry, A sales forecast helps every business to make better business decisions in overall business planning, budgeting, marketing, and risk management. The traditional forecasting method focuses on a conventional statistical model, which highly depends on the availability of historical sales data. However, for new drug entities, where not enough historical data is available, new methods of Machine Learning are applied. The aim of this paper is to identify an efficient Deep Neural Network algorithm suitable to forecast drug sales and pricing by applying Deep Neural Network Algorithms such as Multilayer Perceptron, Convolutional Neural Network, and Long Short-Term Memory, which are expected to perform well on this issue. The results are carried out to determine the efficiency of these algorithms by evaluating the performances of the models using MAE and RMSE performance metrics to identify the best algorithm for Drug Sales and Price Forecasting. The accepted accuracy should be more than 80\% of the actual value for quantity which is less than three thousand by unit and less than two dollars (USD) for price, Based on the results of the experiments Long Short Term Memory performed better than MLP and CNN for generating predictions with average Root Mean Square Error of for sales is 1.28(k) and Mean Absolute Error of about 0.85(k), and with average Root Mean Square Error for USD Prices is about 0.75, and Mean Absolute Error is about 0.44. The forecasts are then used to adjust stock levels according to the predictions.

Keywords: Drug sales Forecasting, Multilayer Perceptron, Convolutional Neural Network, Long Short-Term Memory.

1. Introduction

The pharmaceutical industry involves the discovery, research, development, manufacturing, distribution and marketing of pharmaceutical products, which are drugs or medications used for the treatment or prevention of diseases (Taylor, 2015). It is one of the most resilient, fastest growing, and research-intense industries (Lakner et al., 2019). The industry is dominated by Multinational companies mainly in the United States (US) and European Union (EU). These companies invest highly in the research and development of innovative drugs which are protected by IP law once approved by the relevant regulatory bodies (such as the USFDA). This patent protection (for about 20 years) enables these companies to set high prices for their products to recover the large costs incurred in the discovery phase. Upon the expiry of an innovative drug's patent, a reformulation of the same active ingredient can be developed by other companies. This is called a generic drug. The process of developing a generic drug is much less expensive and less time consuming. For these reasons, in addition to the presence of a number of competitors producing the same active ingredient, the prices of generic drugs are significantly lower than those of innovative (originator's) drugs.

Most local and regional pharmaceutical manufacturing companies based in the MENA region are generic companies that produce copies of innovative drugs after the expiry of their patents. Drug pricing policies are country-dependent1. For example, in the Kingdom of Saudi Arabia and Jordan, drug pricing is regulated by the Saudi FDA and the Jordanian FDA, respectively. The price depends on many factors, including whether the product is originator or generic, the patent is valid or expired, and the number of generic products in the market. There are also continuous governmental initiatives to control drug prices. On the other hand, the Pharmaceutical market is a highly fragmented market that includes international, regional, and local companies. There are no regulations to control drug prices. For the same product, the price may differ between pharmacies, warehouses, or regions. The presence of illegal and parallel imports only complicates the issue.

Pharma manufacturers strive to enhance their efficiencies through accurate planning and sales forecasting. Sales forecasting is vital across various functions in the industry; it plays a major role in formulating effective business plans and gaining valuable and powerful insights that empower companies to make important decisions related to their costs, performance, and profitability (Tiriveedhi, 2018). Time is a very important variable in the analysis of sales data using the Time-series method of forecasting. Data is collected at equally spaced intervals of time. This data from current and past periods is used to predict the values in future periods as well as show trends and repeating patterns (seasonality). Sales forecasting is a complex process as it involves different internal (company-related) and external factors (political, economic…etc.). In this paper, we aim to identify the critical factors that influence drug products' sales, and then Artificial Neural Network algorithms are applied to forecast the sales of drugs, these algorithms will be studied and compared against each other to identify the most efficient algorithm suitable to forecast sales of drugs based on the results obtained for a given dataset to provide the most accurate predictions.

This paper aims to identify the critical features that affect drug sales and also to identify the machine learning algorithms suitable for resolving the drug sales forecasting problem and find the most efficient algorithm among the chosen algorithms based on the models' performances for the given dataset. Particularly, we studied three of Deep Artificial Neural Network algorithms for sales and pricing forecasting: Long Short-Term Memory (LSTM), Multi-Layer-Perceptron (MLP), and Convolutional Neural Network (CNN). The current research is driven by the following research questions:

- **RQ1**: What are the critical features that influence drug sales?

Motivation: This research question's motivation is to find the critical features that will influence the sales of drugs in the private pharmaceutical market. This will help us to improve the quality of the results.

- **RQ2**: How can the Deep Artificial Neural Network algorithms be chosen to resolve the sales forecasting problem?

Motivation: The motivation for this research question is to examine suitable Deep Artificial Neural Network models for sales forecasting.

- **RQ3**: Which Deep Neural Network model is efficient for forecasting the sales of drugs in the private pharmaceutical market?

Motivation: The motivation of this research question is to identify efficient Deep Neural Network algorithms among the selected algorithms for forecasting the sales of the drugs in the pharmaceutical market based on the obtained results.

The critical features identified from RQ1 are used to develop the Machine Learning model using different Deep Neural Network algorithms that were selected in RQ2. An efficient machine learning model is then identified by comparing the performances of these models using various metrics such as Mean Absolute Error and Root Mean Square Error Metrics for the dataset to address RQ3.

2. Background

Sales prediction is an important part of modern business intelligence (C. H. Wang & Yun, 2020). It can be a difficult problem, especially in the case of lack of data, missing data, and outliers' behavior. Sales can be analyzed as a time series. At present, different time series models have been developed, for example, by Holt-Winters, ARIMA, GARCH, SARIMAX, SARIMA.. ect. (Andrawis et al., 2011; Doganis et al., 2006; Lim & Zohren, 2021). Time-series approaches have some

[1] https://www.sfda.gov.sa/en

DOI: 10.6977/IJoSI.202409_8(3).0006
N. Qassrawi, M. Azzeh /Int. J. Systematic Innovation, 8(3), 63-83 (2024)

limitations in sales forecasting, below are some of them: 1) Historical data for a long time period is required to capture seasonality. However, we often do not have historical data for a target variable, for example, when a new product is launched. Simultaneously, we have a sales time series for similar products, so we could expect that our newly launched product will have the same sales pattern (Ensafi et al., 2022). 2) Drug sales data could have missing data and outliers, which should be cleaned before using a time series approach (Ensafi et al., 2022) , and 3) We should study all critical factors that affect drug sales to predict the more accrue value.

The drug sales forecast in the pharmaceutical market is a regression problem. Practice shows that by Machine-learning models, we can find patterns in the time series also complicated patterns in the dynamics of sales using supervised machine-learning methods. There are several methods to forecast the future sales for the products in different business areas. Forecasts are used for planning production and other business activities such as purchasing materials, inventory management and often more across most industries. Traditional forecasting approaches were primarily focused on experienced employee opinions or statistical analysis of previous data such as time series and linear regression, but in recent years Machine Learning and Artificial Neural Networks techniques have been implemented with great success in this field (Boyapati & Mummidi, 2020) . The next subsections describe the main ideas of the above forecasting methods.

2.1 Time-Series Forecasting

Time series is a time-dependent sequence of observations of a variable (Benidis et al., 2022). Based on the rate at which the data is collected, time series is categorized into two types: Continuous time series and Discrete-time series. A continuous-time series is a sequence of observations made continuously through time. A time series is a discrete time series when the observations are collected at fixed or equal intervals of time such as Daily closing price of Google stock, Monthly sales of cars, Yearly rate of change of global temperature and so on (Yeasmin et al., 2022). A continuous time series can be sampled at equal intervals of time to form a discrete time series (Lim & Zohren, 2021). Time series data can be analyzed for several purposes such as to describe the (seasonal or trend) variations of time series data, to use variations of one time series to gain insights into another time series (Lim & Zohren, 2021). Time series forecasting is useful to develop a model by analyzing the past observations of a time series to describe the relationships in the time series. This model is then used to predict future values of the time series (Ji et al., 2016).

2.2 Artificial Neural Networks Algorithms

Artificial Neural Networks (ANN) comprises of multiple nodes that initiate biological neurons of the human brain. The neuron is connected by links, interact with each other using these links. The node takes the input data through the input layer and operates on the hidden layer's data. These operations result is passed to the other neurons. After computation, the result is passed to the output layer (Hossen et al., 2017). when the hidden layers are more than two in any neural network, it is known as a deep neural network, which uses a cascade of multiple layers of non-linear processing units for feature extraction. The output of the current layer is fetched to the next layer as an input.

CNN is a feed-forward neural network. Although based on the traditional architecture of a neural network, CNN includes input layers, hidden layers, and output layers. CNN was designed to automatically and adaptively learn spatial hierarchies of features, from low- to high-level patterns. CNN is a mathematical construct typically composed of three types of layers (or building blocks): convolution, pooling, and fully connected layers. The first two, convolution and pooling layers, perform feature extraction, whereas the third, a fully connected layer, maps the extracted features into final output, such as classification. A convolution layer plays a key role in CNN, which is composed of a stack of mathematical operations, such as convolution, a specialized type of linear operation (Kim et al., 2016). CNN's convolutions are popularly known to work on spatial or 2D data. What is less popular is that there are also convolutions for 1D data. This allows CNN to be used in more general data type including texts and other time series data. Instead of extracting spatial information, you use 1D convolutions to extract information along the time dimension as shown in Figure 1.

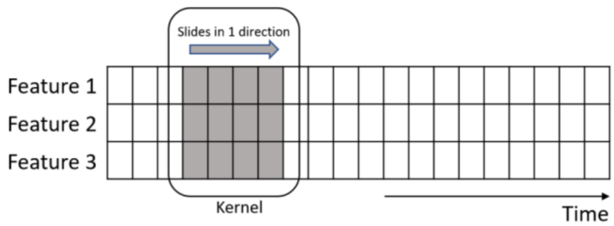


Fig 1. Conv1D Convolving on time dimension

2.3.2 Multi-layer Perceptron (MLP):

Multilayer Perceptron is a type of feed-forward neural network where the information flows from the input layer towards the output layer through the hidden layer as shown in Figure 2. Multilayer Perceptron, or MLPs for short, can be used to model univariate, multivariate, and multi-step time series forecasting problems (Dai et al., 2008). Rectified Linear Unit is used as activation function for Multilayer Perceptron algorithm. MLP makes use of a supervised learning algorithm called backpropagation for training the network (Lim & Zohren, 2021). In backpropagation, the error is propagated backward throughout the network. The error is calculated by taking the difference between the network output and the actual output. The network parameters called weights are modified to minimize this error based on this method. This process is repeated several times until a stopping condition is reached.

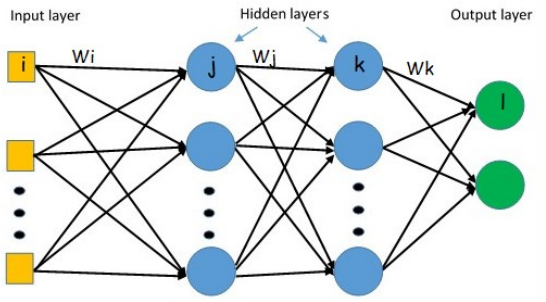


Fig 2. Multilayer Perceptron

$$y = f(x_1 w_1 + x_2 w_2 + \cdots + x_n w_n) \qquad (1)$$

where 'x' and 'y' represents input and output of the network. 'W_i' represents the connection weights between two input and hidden layers, W_j represents the connection weights between 1st and 2nd hidden layer, W_k denotes connection weights between hidden layer and output layer, 'f' represents the activation function, 'i','j','k','l' represents the nodes in the network (Yeasmin et al., 2022).

Long Short-Term Memory networks, or LSTMs for short, can be applied to time series forecasting. There are many types of LSTM models that can be used for each specific type of time series forecasting problem (Boyapati & Mummidi, 2020). Long Short-Term Memory (LSTM) is a type of recurrent neural network. It consists of regulators or LSTM units called gates. LSTM can learn long-term dependencies because of the usage of the gating mechanism and a memory cell. LSTM can overcome the vanishing gradient and exploding gradient problems faced by RNN (Dai et al., 2008; Yeasmin et al., 2022). A typical LSTM consists of a memory cell, input gate, output gate, and forget gate. The memory cell can remember information over arbitrary time periods. The gating mechanism regulates the flow of information to and from the cell. A typical LSTM consists of a memory cell, input gate, output gate, and forget gate. The memory cell can remember information over arbitrary time periods, and the gating mechanism regulates the flow of information to and from the cell (Z. Wang & Lou, 2019).

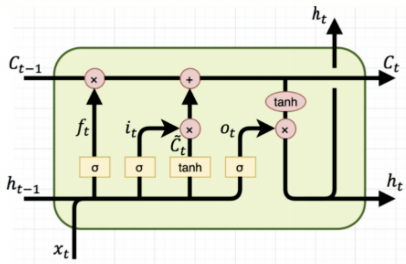


Fig 3. Long Short-Term Memory

In Figure 3, c_t is the cell state which is used to carry information throughout the sequence chain which acts as memory. Forget gates represented by ft are used to determine which information should be eliminated from the cell state. For this purpose, they subject the input vector to a sigmoid function and then perform a pointwise multiplication operation with cell state C_{t-1}. The input gate it is used to determine the values (or input vector ($[h_{t-1}, x_t]$) we are going to update. tanh function is used to generate new values from the input vector. The result from the input gate and tanh are combined by making use of a pointwise multiplication operation which is added to the cell state by making use of a pointwise addition operation. Finally, the output gate represented by o_t determines which values of the input vector, we need output by applying a sigmoid function to the input vector. The value of hidden state h_t is calculated by subjecting the cell state to a *tanh* function and then multiplying it with the output gate using pointwise multiplication. This information is then passed along the chain in a sequence and the above process is repeated (Z. Wang & Lou, 2019).

3. Related Work

Previously, several comparative studies between traditional models and neural networks have been carried out, researchers have extensively worked and examined alternative methods to find out the most efficient sales prediction methodology, and they have identified important features of time series data to enable the sales prediction methodology to forecast sales efficiently. This has been done by using statistically-based methods like Linear Autoregressive models (AR) which are flexible to model many stationary processes (Hossen et al., 2017). The ARMA (Autoregressive Moving Average) model is used for short-term time series forecasting. The ARMA model failed because it only gave a linear relationship between features and could not accurately predict the evolution of non-linear and non-stationary data. Whenever there is highly fluctuating time series data due to seasonal factors or time trends it shows a degraded performance (Hossen et al., 2017). Therefore, most statistical methods are limited to non-linear and stationary time series forecasting assuming an AR type structure. To overcome the challenge of linear statistical time-series models, many non-linear machine-learning models like artificial neural networks (ANNs) have been proposed in the literature (Aras & Kocakoç, 2016; Kamruzzaman et al., 1 C.E.). ANNs belong to the data-driven approach, where training depends on the available data with little prior rationalization regarding relationships between variables (Aras & Kocakoç, 2016). ANNs do not make any assumptions about the statistical distributions of the underlying time series and they can naturally perform non-linear modeling (Aras & Kocakoç, 2016). As a result, ANNs are self-adaptive by nature. Recently we have seen the higher performance of Artificial Neural Networks (ANN) in classification and regression problems and have received focused attention in the time series forecasting methods. When we compare ANN with the normal statistical techniques, we find that ANN has many unique features such as: 1) non-linear and data-driven, 2) Not having a requirement for an explicit underlying model and 3) it is applicable to complicated models.

A literature review was performed in this study to identify a suitable neural network model for drug sales forecasting. To handle this work, some of the methods such as Machine Learning models, hybrid models, and statistical models will be helpful. (Dai et al., 2008) Have implemented an Artificial Neural Network for time series forecasting. They tried to study the use of artificial Neural Network in time series forecasting. They proposed that ARNN gives the best results to predict consumer goods' sales compared to SVM and Arima. ANNs are also known for their ability to map non-linear functions, making them suitable for sales forecasting. (Bing et al., 2014) proposed an algorithm to predict the stock price movement with an accuracy up to 76.12% by investigating public social media information represented in tweets data. Bing used a model to analyze public tweets and hourly stock price trends. NLP techniques have been used along with data mining techniques to identify correlation patterns between public sentiment and numeric stock prices. This study examines whether there is an internal association in the multilayer hierarchical structures, and found a relation between internal layers and the top layer of unstructured data. This study considers only daily closing values for historical stock prices. (Islek & Oguducu, 2015)studied with the use of bipartisan graphic clusters that clustered different warehouses according to sales behavior. They discussed the application by applying the Bayesian network algorithm in which they managed to produce the enhanced forecasting experience. (Siami-Namini & Namin, 2018) compared the accuracy of ARIMA and LSTM when it was forecasting time series data as representative techniques. These two techniques have been implemented and applied to a set of financial data, and the results have shown that LSTM is superior to ARIMA. (Kraus & Feuerriegel, 2017) used LSTM with transfer learning using text mining through financial news and the stock market data, Similarly, (Ding et al., n.d.) implemented Deep Stock Ranker, an LSTM based model for stock ranking using 11 technical indicators. (Mehdiyev et al., 2017) proposed a new multi-stage approach to deep learning for multivariate time series classification issues. They used the stacked LSTM autoencoders after extracting the features from the time series data in an unsupervised manner. The objective of the case study is to predict post-processing activities depending on the detected steel surface defects using the time series data obtained from the sensors installed in various positions of the steel casting process facility and the steel's chemical properties.

Meanwhile, in some of the papers, CNN models were preferred. (Kim et al., 2016) use CNNs in a bank telemarketing case study, whereby the aim is to predict whether a customer will take up a particular marketing campaign based on a number of numeric and nominal features per customer. The results for this study yield an impressive 76.70% accuracy, which yields the highest accuracy amongst 7 classifiers. In order to incorporate external features in the forecasting model. In (Sci & 2014, n.d.) the author uses CNNs to predict stock price changes based on the image of the time series plot. The author also attempts to color code the time series, however, the results of this approach were not positive. (Ding et al., n.d.) designed a model for stock market prediction driven by events. First, events are extracted from financial news and represented by word embedding as dense vectors. They trained a deep CNN to model on stock price events both short term and long-term influences. Their proposed model in S&P 500 index prediction and individual stock prediction gave better performance than SVM. They also use a deep convolutional neural network to model short- and long-term influences of events of stock price movements. Results from this study show that CNNs can capture longer-term influence of news events than standard feed-forward networks. The authors of (Processing & 2017, 2017)used 250 features for the prediction of the private brokerage company's real data of risky transactions. They used CNN and LSTM for stock price forecasting. (Zhang et al., 2003) use convolutional neural networks for recognition of human activity (HAR). Their methodology capitalizes on the fact that a combination of unsupervised learning and supervised classification of features can increase the discriminative power of features. (Hernández et al., 2016) proposed an Auto-encoder and MLP based deep learning model to predict the daily accumulated rainfall. The authors used Auto-encoder to select features and MLP to predict. The results showed a better performance of their proposed model than other approaches. To make predictions, they used the deep feed-forward neural network. From the above literature review LSTM, CNN, and MLP are recommended as a deep learning models used for sales forecasting. Previously, most of the studies focused on considering the metrics as a mean absolute error, mean squared error, root mean squared error and k-fold cross-validation is used for training and testing data. Metrics like mean absolute error and root mean squared error are considered in this research. In this study, a stratified K-fold cross-validation technique is used for training and testing to increase the results' efficiency. In this study, LSTM, CNN, and MLP are chosen for sales forecasting.

4. Methodology

An experiment is chosen as a research method to answer the research questions because the experiment is considered a suitable research method for dealing with quantitative data as experiments would give greater control over variables. The experiments aim to evaluate the performance of Deep neural networks Multilayer Perceptron, Convolutional Neural Network, and Long Short-Term Memory on quarterly sales data extracted from the order management database of the private pharmaceutical sales in the Iraqi market. The efficient algorithm among the chosen algorithms is identified by analyzing and comparing the experiment results on the given dataset. We can describe the procedure followed in this experiment as follows: 1) Extracting the required data for the sales. 2) Applying Multilayer Perceptron, CNN, and LSTM algorithms. 3) The performance of the output can be enhanced by comparing metrics such as Mean Absolute Error and Root Mean Square Error, and 4) Based on assessment tests, the best suitable algorithm can be selected. We used Python, Pandas Sklearn and seaborn libraries to accomplish this work.

4.1 Study area and Dataset

The importance of data in the Pharmaceutical Industry is growing rapidly. Market data is one of the key factors that help companies monitor their performance and make strategic decisions to improve their efficiency. In this study, the private pharmaceutical market in Iraq (which is one of the most challenging markets in MENA) is selected to be the study area. Dataset is provided by Advanced Marketing Statistics company AMS. The data is collected quarterly at the SKU level (Stock Keeping Units) in volume and price (price is at pharmacy purchase level; i.e.: from drug store to pharmacy). The data covers the period from Q3-2010 To Q1-2021. AMS data reflects the quantitative sales of pharmaceutical products in the private sector of the Iraqi Market. It doesn't include tenders or governmental sales data. This dataset covered medicines that have had sales in the Iraqi market from at least 2013 which consists of about 80k records:

Table 1. Data set description

No.	Features	Description
1	ATC2	Anatomical Therapeutic Area at level 2
2	Company Type	Gx / Rx
3	Product	Brand Name
4	Molecule	Active Ingredient of the product
5	Product Source	Legal/ Smuggled
6	Launch Date	The quarter at which a product has been first marketed
7	No. of Competitors	No. of other brands of the same molecule that are available in the market
8	Quantity	No. of standard units sold in a specific period of time
9	Price	Standard unit price in USD
10	Year	2010 to 2021
11	Quarter	Q1(Jan, Feb, March) Q2(April, May, June) Q3(July, Aug, Sep) Q4(Oct, Nov, Dec)

The features in the previous dataset represent drug description, and sales are recorded on a quarterly basis. This data set might contain discrepancies in the names or codes, and also might contain missed data or outliers or errors; So, we should prepare data as in data preprocessing. The dataset is then transformed into a supervised learning dataset using the sliding window approach.

4.2 Data Preprocessing

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Data preprocessing is used in database-driven applications such as customer relationship management and rule-based applications (like neural networks). In Machine Learning (ML) processes, data preprocessing is critical to encoding the dataset in a form that could be interpreted and parsed by the algorithm (Zhang et al., 2003). Data goes through a series of steps during preprocessing: 1) Data Cleaning: Data is cleansed through processes such as filling in missing values or deleting rows with missing data, smoothing the noisy data, or resolving the inconsistencies in the data (García et al., 2016). 2) Data Integration: Data with different representations are put together and conflicts within the data are resolved (García et al., 2016).

4.3 Sliding Window

Time Series data can be transformed into a supervised learning problem by making use of Sliding Window Method. This transformation will enable us to use standard linear and nonlinear machine learning algorithms. A time series dataset can be transformed by making use of previous time steps as inputs and new time steps as output variables (Benidis et al., 2022).

4.4 Kwiatkowski, Phillips, Schmidt and Shin (KPSS) Test

Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) Test is a statistical test that is used to test whether a time series is stationary or not. It is used in this study to determine whether a time series is trend stationery or it consists of a unit root. The null hypothesis of the test is that the time series is stationary or trend stationary towards and the alternative hypothesis of a unit root series.

Null Hypothesis: If null hypothesis is accepted, the time series is considered to be trend stationary. Alternative Hypothesis: If the null hypothesis is rejected, the time series consists of unit root, meaning it is non-stationary. The results from the tests can be interpreted as follows: If p-value > 0.05 the null hypothesis is not rejected; the time series is trend stationary. Otherwise, the null hypothesis is rejected, and the time series consists of a unit root i.e., the series is non-stationary.

4.5 Feature Selection:

Feature selection refers to a class of methods for assigning values to input features to a predictive model which determines the relative significance of each factor while forecasting. Feature selection scores provide an overview of the model. Most significant scores are determined using a prediction approach that was fitted to the dataset. Inspecting the score of importance gives insight into that particular model and what features are the most essential and least important to the model while making a prediction. Feature Importance can be used to enhance a predictive model. This can be accomplished by selecting those features to remove (lowest scores) or those features to retain, using the importance scores. This is a type of selection of features that can simplify the modeling problem, accelerate the modeling process, and in certain cases, improve model performance.

4.6 Performance Metrics:

Performance Metrics should be selected depending on the regression problem and the dataset used for the experiment. Several metrics can be used while evaluating how well a model is performing. It is necessary to understand how each metric is measured to select the evaluation metric to better assess the model. This thesis's main objective is to compare the performance of neural network algorithms by evaluating all of these performance metrics such as Root Mean Square Error, and Mean Absolute Error. Root Mean Square Error as shown in equation 2 is the square root of the difference between the values predicted by the model and the real or observed values. The value of RMSE indicates the fit of the models on a particular dataset. Values close to zero implies a better fit thereby reducing the impact of outliers

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - yi^{\wedge})} \qquad (2)$$

Where 'n' represents the feature variables, 'y_i' represents actual values and '$y_i^{\wedge}$' represents the predicted or forecasted value.

Mean Absolute Error as shown in equation 3 is calculated by taking the average of absolute difference between values predicted by the model and the real or actual values. Similarly, the accuracy of the model is higher when MAE values are close to zero.

$$MEA(x, y) = \frac{1}{\text{Nsamples}} \sum_{i=0}^{\text{Nsamples}-1} |yi - xi| \quad (3)$$

4.7 Walk Forward Validation:

In time series modeling, the predictions over time become less and less accurate and hence it is a more realistic approach to re-train the model with actual data as it becomes available for further predictions. Since training of statistical models are not time consuming, walk-forward validation with sliding windows is the most preferred solution to get the most accurate results as shown in Figure 4. Walk Forward Validation by sliding windows is a re-sampling technique used to evaluate the machine learning model because it keeps the temporal order in the dataset while splitting time series data. The data is divided into training and test splits as fixed window size (*n*) training set (*1-n-1*) Test set (*n*) for each time then window slide one step or time stamp and predict (*n*) value. each window evaluated by RMSE & MAE, these results are stored to get the average of our performance metrics (*RMSE & MAE*).

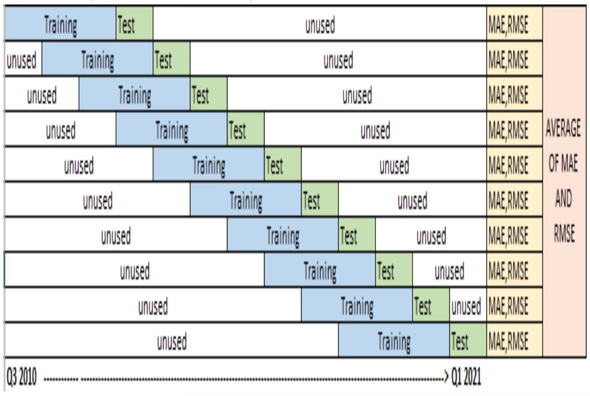
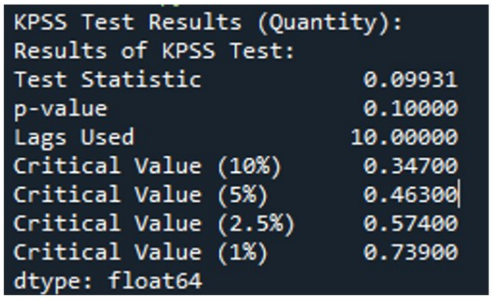


Fig 4. Walk Forward Validation

5. Results

5.1 Stationarity Test

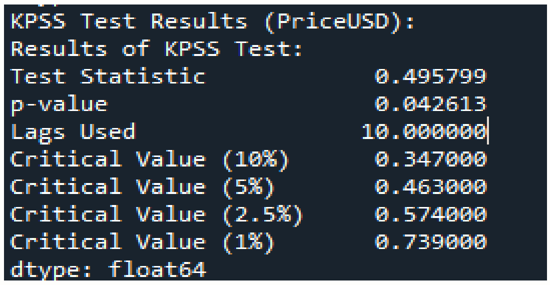
Kwiatkowski, Phillips, Schmidt, and Shin Test was used to test the stationarity of the dataset and the results obtained are shown in Figure 5 and Figure 6.



```
KPSS Test Results (Quantity):
Results of KPSS Test:
Test Statistic            0.09931
p-value                   0.10000
Lags Used                10.00000
Critical Value (10%)      0.34700
Critical Value (5%)       0.46300
Critical Value (2.5%)     0.57400
Critical Value (1%)       0.73900
dtype: float64
```

Fig 5. Kwiatkowski, Phillips, Schmidt and Shin Test (Quantity)

Based on Figure 5, the p-value is greater than 0.05 So, the null hypothesis cannot be rejected. This indicates the data is stationary which can further be utilized for performing time series analysis.



```
KPSS Test Results (PriceUSD):
Results of KPSS Test:
Test Statistic            0.495799
p-value                   0.042613
Lags Used                10.000000
Critical Value (10%)      0.347000
Critical Value (5%)       0.463000
Critical Value (2.5%)     0.574000
Critical Value (1%)       0.739000
dtype: float64
```

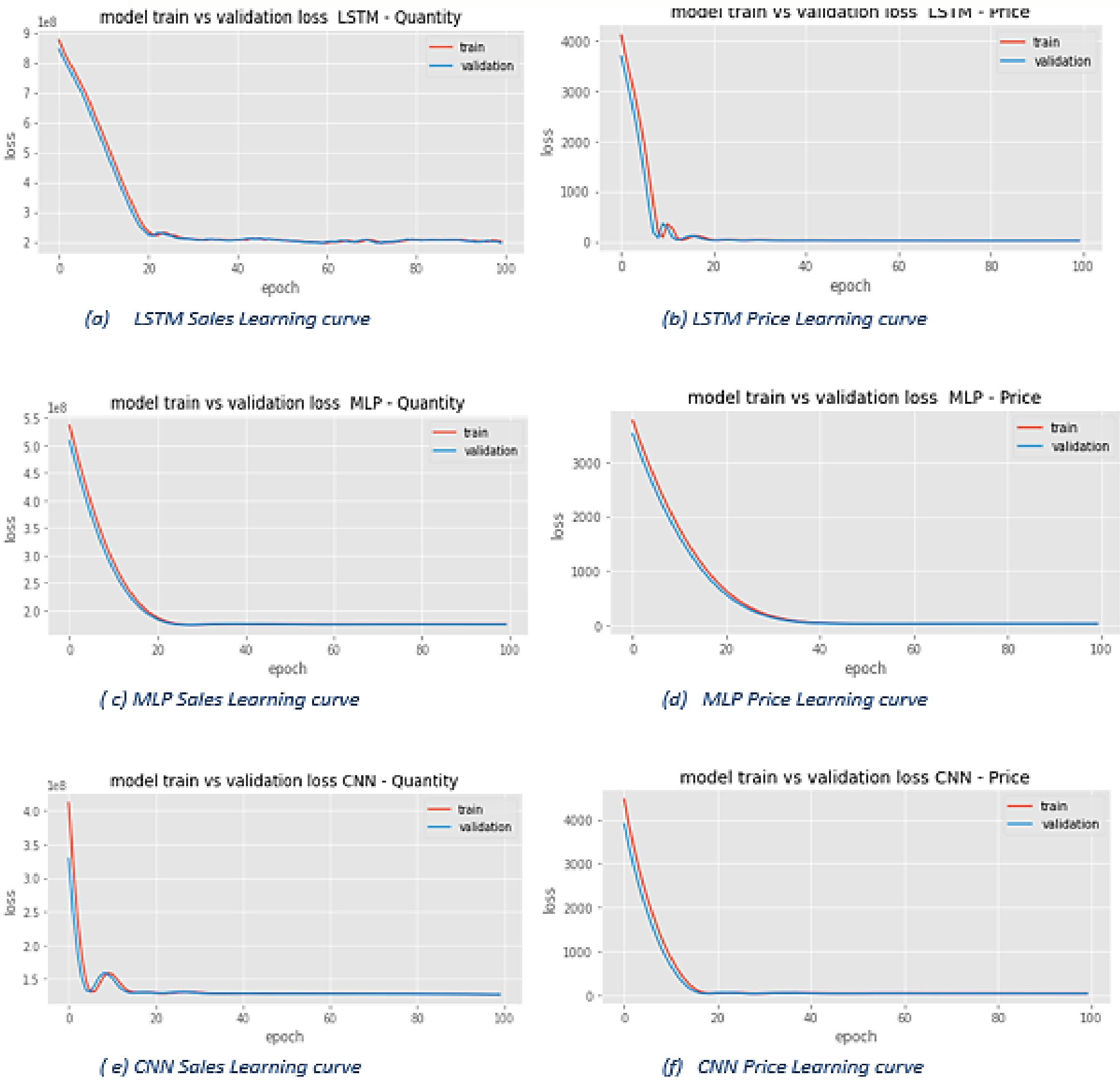
Fig 5. Kwiatkowski, Phillips, Schmidt, and Shin Test (USD PRICE)

Based on Figure 6, the p-value is less than 0.05 So, the null hypothesis is rejected. This indicates that the data is non-stationary. However, we want to test Deep learning models to perform the analysis of this type of time series. LSTM, MLP, and CNN have been used for forecasting sales and prices of medicines in the Iraqi pharmaceutical market, and the following results depict how LSTM, MLP, and CNN have performed based on various performance metrics used in this thesis.

5.2 Learning curve

Learning curves are used to evaluate the performance of a model to diagnose whether the model underfits, overfits, or is a good fit on the chosen dataset. They can also be used to determine if the statistical properties of the training dataset are relative to the properties of the validation dataset. Learning curves are used to identify learning performance changes by taking a plot of loss or error over time. It can also be used to identify how well a model generalizes to unseen data by using a validation dataset. Based on the structure and dynamics of the learning curve, the configuration of the model can be changed to enhance the learning and performance of the model.

DOI: 10.6977/IJoSI.202409_8(3).0006
N. Qassrawi, M. Azzeh /Int. J. Systematic Innovation, 8(3), 63-83 (2024)



(a) LSTM Sales Learning curve

(b) LSTM Price Learning curve

(c) MLP Sales Learning curve

(d) MLP Price Learning curve

(e) CNN Sales Learning curve

(f) CNN Price Learning curve

Fig 6. (LSTM, MLP, CNN) Learning curves

Figure 7 (a - f) represents the Learning curves of LSTM, MLP, and CNN for Sales and Prices. We can see that the training and validation errors decrease to an optimal point with a minimal difference in loss values between them, indicating that the used ANN models are a good fit.

5.3 Forecast results

LSTM, MLP & CNN are trained using walk forward validation approach and the data is split in 80% to 20% ratio where 80% was used for the training set and 20% was used as the test set. The training set is further split into training and validation sets for in 5 steps as the used approach. The performance of these models is estimated using the performance metrics RMSE and MAE and the results obtained are shown below for each type of data sets, the used dataset in this thesis was classified based on the source of medicines in the market as below:

1- Actual (To Market) dataset. (40% OF DATASET)
2 -Projected medicines (In Market). (40% OF DATASET)
3- Smuggled medicines. (20% OF DATASET)

DOI: 10.6977/IJoSI.202409_8(3).0006
N. Qassrawi, M. Azzeh /Int. J. Systematic Innovation, 8(3), 63-83 (2024)

Table 2. Performance metrics for Actual medicines (To Market) dataset.

ID	Medicine	LSTM				MLP				CNN			
		Sales (k)		Price USD		Sales (k)		Price USD		Sales (k)		Price USD	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	ADOL-TAB 500MG 96`S	1.9	2.5	0.07	0.14	55.2	86.0	0.12	0.22	18.4	29.1	0.06	0.13
2	AMARYL-TAB 2MG 30'S	0.30	0.38	0.38	0.66	17.6	21.8	0.70	0.97	1.63	2.83	0.41	0.66
3	SUPRAX-CAP 200MG 8`S	0.33	0.61	0.35	0.47	1.80	2.95	0.50	0.66	0.39	0.61	0.31	0.43
4	CONCOR-TAB 10MG 30`S	0.005	0.006	0.21	0.48	10.3	12.6	0.52	0.87	2.46	3.69	0.47	0.79
5	NOVATEN-TAB 100MG 28`S	1.23	1.54	0.005	0.009	34.2	49.0	0.006	0.01	16.8	20.9	0.006	0.009
6	CRESTOR-TAB 10 MG 28'S	0.63	0.99	0.85	1.95	5.85	7.28	2.52	4.35	1.96	3.19	1.39	2.47
7	ZESTRIL-TAB 5MG 28'S	1.81	2.97	0.26	0.56	3.88	4.90	0.81	1.11	1.08	1.93	0.22	0.51
8	APROVEL-TAB 150MG 28`S	0.12	0.22	1.60	2.34	1.22	1.51	2.68	3.74	0.17	0.72	1.65	2.79
9	PLAGIN-TAB 75MG 30'S	0.49	0.60	0.12	0.14	1.57	2.11	0.12	0.16	0.60	0.87	0.13	0.14
10	ATACAND-TAB 8MG 28`S	1.29	2.09	0.83	1.62	10.99	13.94	1.46	2.42	4.41	6.05	1.00	1.82
11	SUPRAX-CAP 200MG 8`S	0.16	0.26	0.35	0.47	1.17	1.61	0.51	0.66	0.29	0.54	0.31	0.43
12	PLATIL-TAB 75MG 30'S	0.024	0.034	0.73	1.05	14.35	17.91	0.70	1.31	4.53	6.02	0.52	1.02
13	MYOGESIC-TAB 450MG 20`S	0.26	0.34	0.09	0.17	62.02	80.85	0.19	0.27	27.93	37.96	0.10	0.17
14	JOINTACE-CAP 30'S	0.075	0.124	0.20	0.29	3.96	5.42	0.26	0.42	1.99	3.22	0.194	0.289
15	NEXIUM-TAB 40MG 14`S	2.41	3.07	0.31	0.89	34.39	42.02	1.85	2.74	11.36	18.27	0.47	1.02
Average Performance Metrics		0.74	1.05	0.42	0.75	17.23	23.33	0.86	1.33	6.27	9.06	0.48	0.85

This table obtained the MAE and RMSE for Sales & USD Prices for 15 products Selected randomly from the Actual dataset. Outliers in Actual (To-Market) Data have been overcome by dividing the sales data by 2. Suppliers normally ship large quantities of drugs to secure enough stock for at least six months. **Figure 8** below represents the actual sales and the forecasted sales of NEXIUM-TAB 40MG 14`S using LSTM, MLP, and CNN algorithms, where the green line indicates the predicted value of the target variant and the blue line represents the actual values.

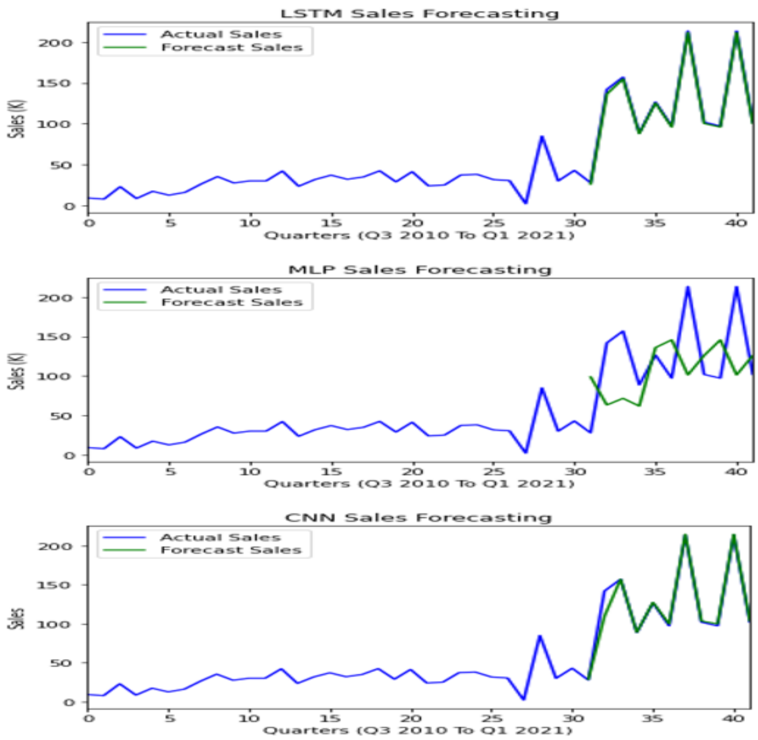


Fig 7. NEXIUM-TAB 40MG 14`S (Actual Sales (k))

Figure 8 represents the actual prices and the forecasted prices of **NEXIUM-TAB 40MG 14`S** using LSTM, MLP, and CNN algorithms, where the green line indicates the predicted price of the target variant and the blue line represents the actual price.

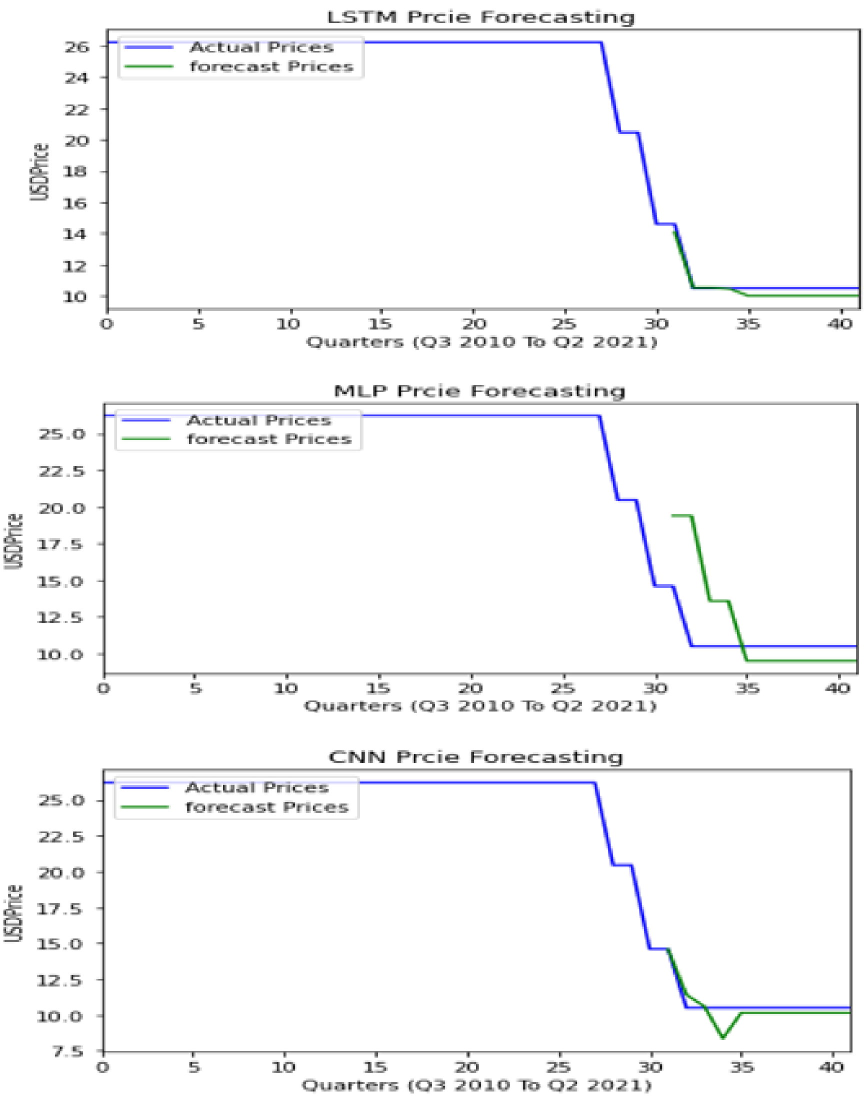


Fig 8. NEXIUM-TAB 40MG 14`S (Actual USD Price)

Table 3 obtained the MAE and RMSE for Sales & USD Prices for 15 products Selected randomly from the Projected dataset. Outliers replaced by "Mean" value.

Table 3. Performance metrics for Projected medicines

ID	Medicine	LSTM				MLP				CNN			
		Sales (k)		Price USD		Sales (k)		Price USD		Sales (k)		Price USD	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	ACTIFED-TAB 30`S	0.037	0.045	0.036	0.044	3.641	4.830	0.06	0.14	0.196	0.442	0.038	0.044
2	ALDACTON-TAB 25MG 20 'S	0.027	0.036	0.20	0.23	3.66	4.55	0.32	0.53	0.420	0.727	0.223	0.335
3	MOTILIOSYR-TAB 10MG 30`S	1.483	3.250	0.047	0.100	16.31	20.32	0.079	0.158	5.015	7.64	0.047	0.100
4	TOFRANIL-TAB 10 MG 50`S	0.015	0.036	0.315	0.530	2.68	3.86	0.59	0.80	0.317	0.484	0.312	0.529
5	LIBROXIDE-TAB 10MG 10`S	6.19	8.78	0.007	0.008	34.25	42.66	0.007	0.009	8.62	13.89	0.007	0.008
6	XENICAL-CAP 120MG 84'S	0.011	0.022	1.439	2.965	1.238	1.666	12.80	18.66	0.093	0.195	4.769	6.866
7	CO-AMOXI-1000 TAB 12`S	1.62	4.64	0.240	0.378	25.64	32.74	0.355	0.593	13.59	16.92	0.285	0.439
8	LOFRAL-5 OP 30'S	0.050	0.078	0.120	0.159	15.03	18.35	0.145	0.208	6.975	9.375	0.138	0.169
9	BUTADIN-TAB 2MG 10`S	5.209	8.49	0.002	0.003	103.7	145.4	0.002	0.003	52.29	80.26	0.001	0.002
10	IMURAN-TAB 50MG 100`S	0.062	0.094	1.579	2.439	2.16	2.83	2.38	3.43	0.605	0.654	1.69	2.53
11	DAKTACORT-CREAM 15G	0.282	0.360	0.346	0.739	10.07	11.56	0.639	0.99	4.35	6.72	0.434	0.758
12	CATAFLAM-TAB 50MG 20`S	0.261	0.665	0.260	0.410	13.72	19.08	0.319	0.527	5.76	9.43	0.146	0.349
13	ADVIL SINUS-CAP 20`S	0.71	0.826	0.242	0.372	12.40	14.54	0.46	0.635	2.298	3.748	0.239	0.365
14	RABEZOLE-TAB 20MG 28`S	0.106	0.129	0.269	0.455	6.29	11.76	0.357	0.718	2.83	6.173	0.236	0.453
15	DILATREND-TAB 25MG 30`S	0.510	0.810	1.16	1.95	2.18	2.74	1.73	2.55	0.605	0.905	1.255	2.065
Average Performance Metrics		1.10	1.88	0.42	0.72	16.86	22.46	1.35	2.00	6.93	10.5	0.65	1.00

Figure 9 represents the actual sales and the forecasted sales of CO-AMOXI-1000 TAB 12`S using LSTM, MLP, and CNN algorithms, where the green line indicates the predicted value of the target variant and the blue line represents the actual values.

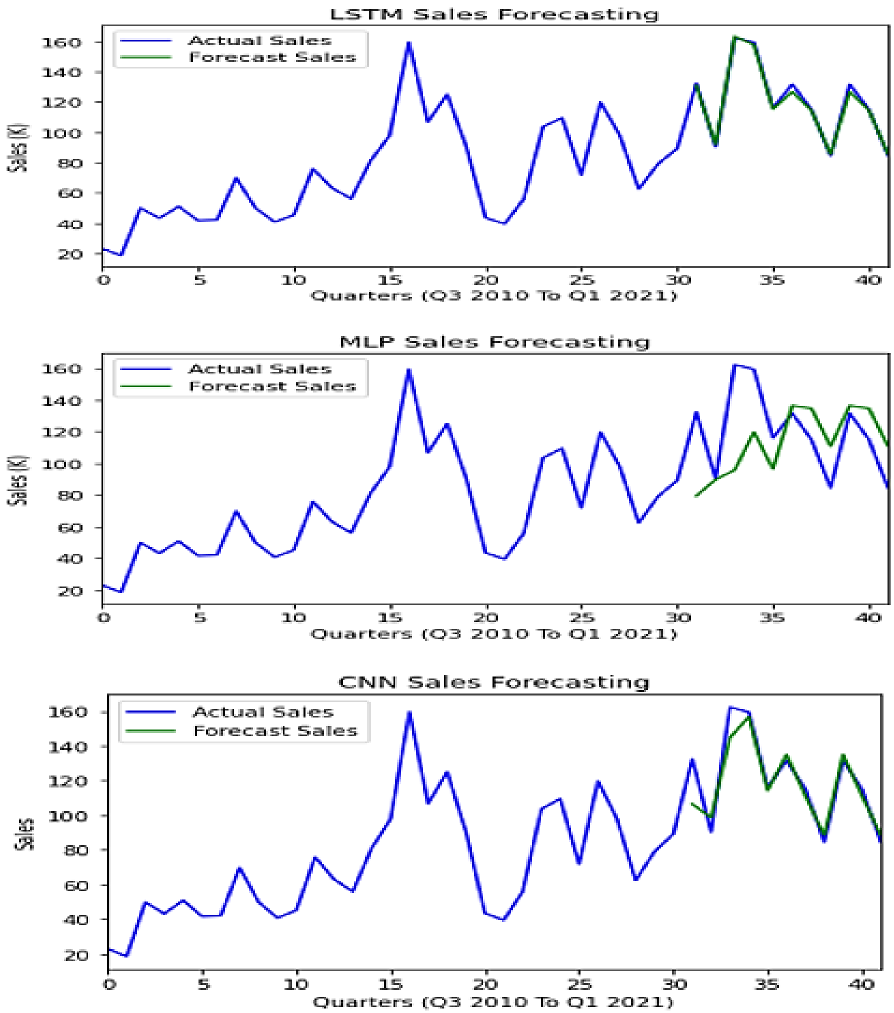


Fig 9. CO-AMOXI-1000 TAB 12`S (Projected Sales(k))

Figure 10 represents the actual prices and the forecasted prices of CO-AMOXI-1000 TAB 12`S using LSTM, MLP, and CNN algorithms, where the green line indicates the predicted price of the target variant and the blue line represents the actual price.

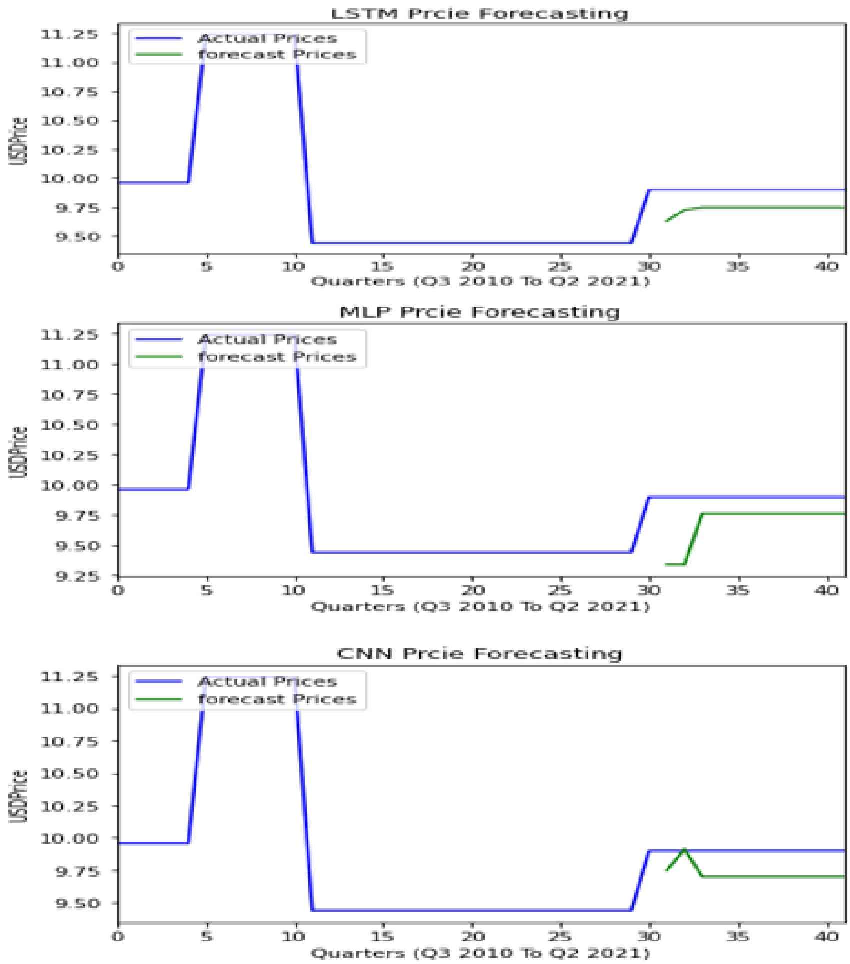


Fig 10. CO-AMOXI-1000 TAB 12`S (USD Price)

Table 4 obtained the MAE and RMSE for Sales & USD Prices for 15 products Selected randomly from the Smuggled dataset. Figure 11 represents the actual sales and the forecasted sales of NEXIUM-TAB 40MG 14`S using LSTM, MLP, and CNN algorithms, where the green line indicates the predicted value of the target variant and the blue line represents the actual values.

Table 4. Performance metrics for Smuggled medicines

ID	Medicine	LSTM				MLP				CNN			
		Sales (k)		Price USD		Sales (k)		Price USD		Sales (k)		Price USD	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	TOFRANIL-TAB 25MG 50'S	0.571	0.842	0.222	0.459	9.05	12.70	0.470	0.684	1.514	2.195	0.227	0.460
2	ALDOMET-TAB 250MG 30`S	0.550	0.757	0.201	0.363	5.68	6.689	0.257	0.570	2.46	3.76	0.198	0.372
3	AMARYL-TAB 4MG 30'S	1.324	1.461	0.516	0.794	12.15	17.54	0.926	1.176	2.449	3.307	0.535	0.800
4	ACTIFED-TAB 12`S	1.934	2.422	0.066	0.086	32.68	42.68	0.067	0.092	12.48	20.6	0.055	0.075
5	ISOPTIN-TAB 40MG 30'S	0.155	0.345	0.261	0.523	1.60	2.10	0.74	0.89	0.169	0.308	0.263	0.530
6	NEXIUM-TAB 40MG 14`S	0.022	0.031	0.703	1.13	3.46	4.01	1.3	1.95	0.850	1.42	0.77	1.19
7	PLAVIX-TAB 75MG 28`S	0.44	0.52	1.317	2.179	3.75	5.064	3.574	4.356	0.28	0.52	1.90	2.87
Average Performance Metrics		0.71	0.91	0.47	0.79	9.77	12.97	1.05	1.39	2.89	4.59	0.59	0.90

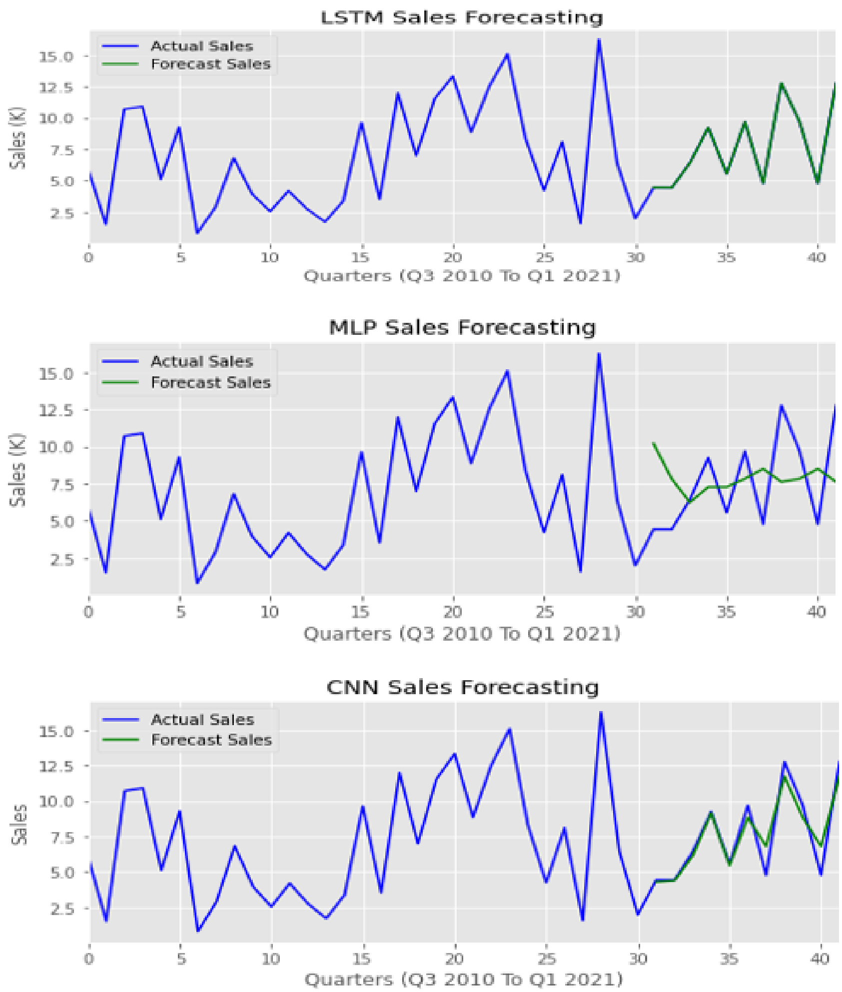


Fig 11. NEXIUM-TAB 40MG 14`S (Smuggled Sales forecasting)

Figure 12 represents the actual prices and the forecasted prices of NEXIUM-TAB 40MG 14`S using LSTM, MLP, and CNN algorithms, where the green line indicates the predicted price of the target variant and the blue line represents the actual price.

DOI: 10.6977/IJoSI.202409_8(3).0006
N. Qassrawi, M. Azzeh /Int. J. Systematic Innovation, 8(3), 63-83 (2024)

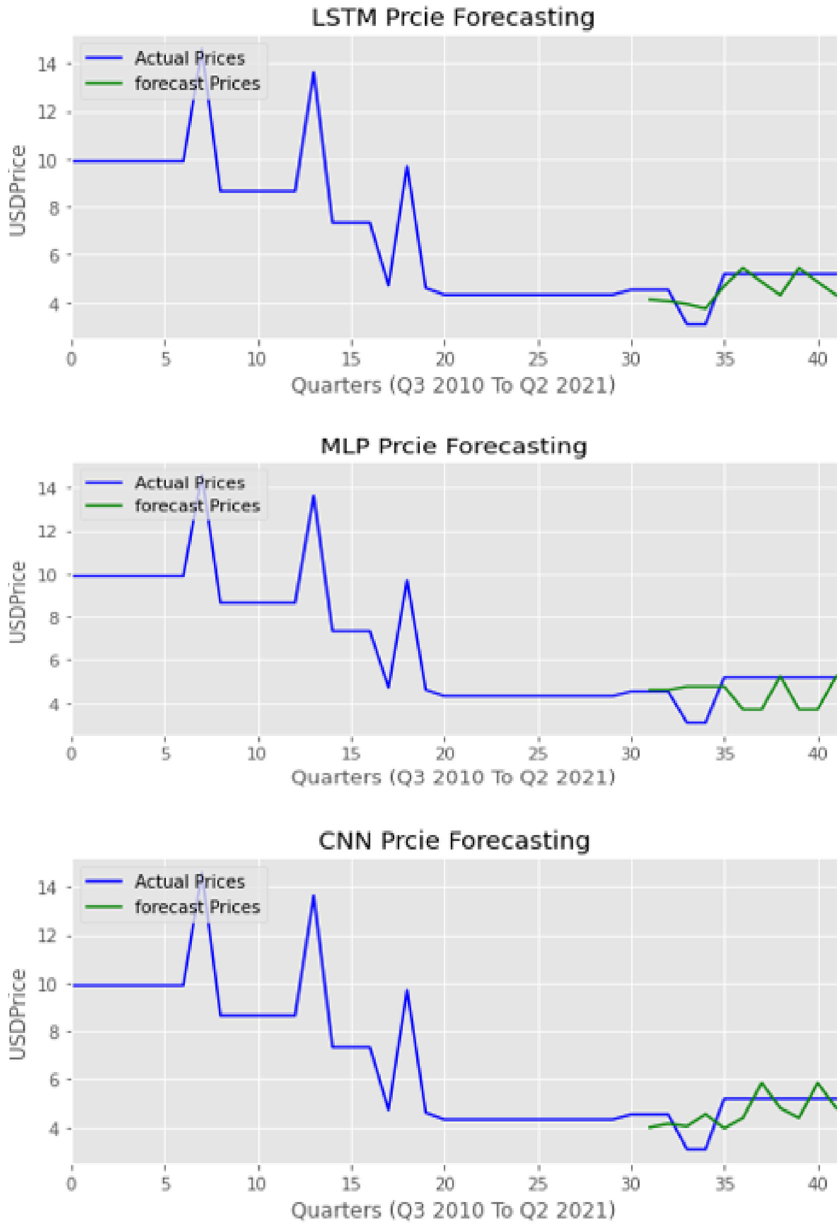


Fig 12. NEXIUM-TAB 40MG 14`S (Smuggled USD Price)

5.3 Analysis of experiment results

Figure 14 represents the average mean absolute error of Long Short-Term Memory, Multilayer Perceptron, and Convolutional Neural Network algorithms for Sales forecasting for the three types of data. Long Short-Term memory performed better in the three types of data when compared to Multilayer Perceptron and Convolutional Neural networks. Multilayer Perceptron has the highest average error across the three types of data.

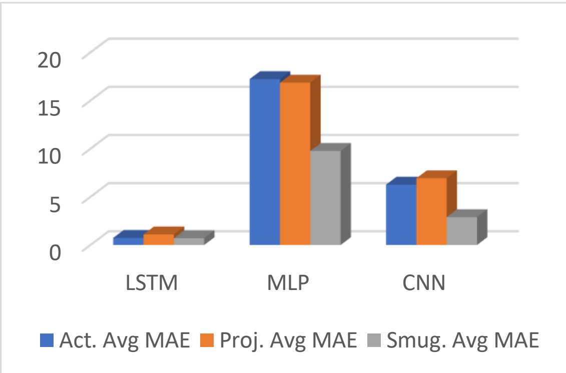


Fig 13. Mean Absolute Error for Sales Forecasting

Figure 15 represents the average mean absolute error of Long Short-Term Memory, Multilayer Perceptron, and Convolutional Neural Network algorithms for USD Price forecasting for the three types of data. Also, we can see that Long Short-Term memory performed better in the three types of data when compared to Multilayer Perceptron and Convolutional Neural networks. Multilayer Perceptron has the highest average error across the three types of data.

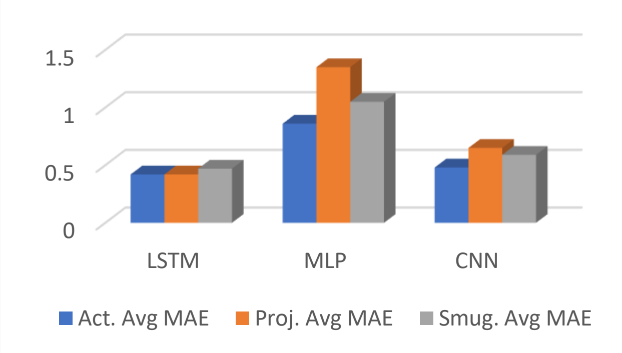


Fig 14. Mean Absolute Error for USD Price Forecasting

Figure 16 represents the Root Mean Square error of Long Short-Term Memory, Multilayer Perceptron, and Convolutional Neural Network algorithms for Sales forecasting for the three types of data. Long Short-Term memory performed better in the three types of data compared to Multilayer Perceptron and Convolutional Neural networks. However, Multilayer Perceptron has the highest average error across the three types of data.

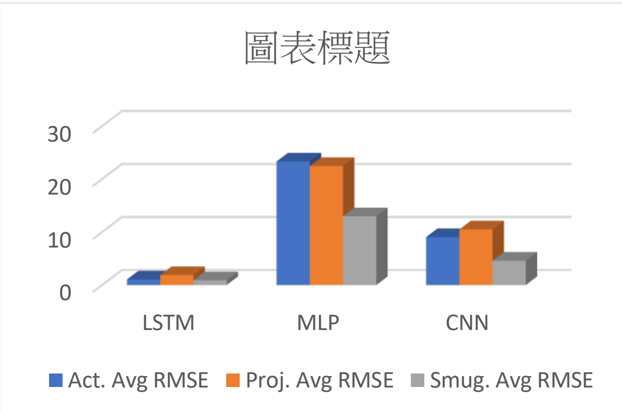


Fig 15. Root Mean Square Error for Sales Forecasting

Figure 17 represents the Root Mean Square error of Long Short-Term Memory, Multilayer Perceptron, and Convolutional Neural Network algorithms for Price (USD) forecasting for the three types of data. Long Short-Term memory performed better in the three types of data compared to Multilayer Perceptron and Convolutional Neural networks. However, Multilayer Perceptron has the highest average error across the three types of data.

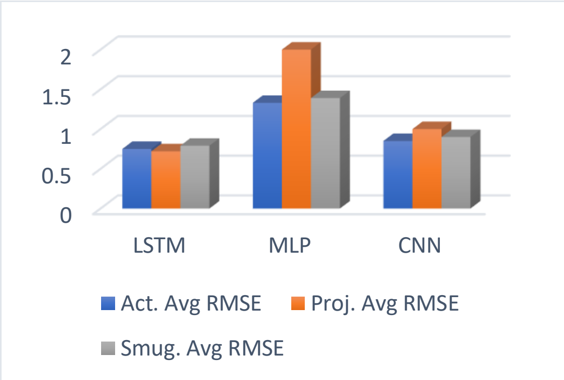


Fig 16. Root Mean Square Error for Price (USD) Forecasting

5.4 Performance Evaluation

The main objective of this thesis is to compare MLP, CNN, and LSTM forecasting accuracy in medicine sales and Prices in different data sources.

Table 5. Comparison of performance evaluation results for sales forecasting

Algorithm	MAE (k)	RMSE (k)
LSTM	0.85	1.28
MLP	14.62	19.58
CNN	5.363	8.05

The average error as shown in Table 5 indicates that LSTM based forecasting performed better than the other two algorithms with a Root Mean Square Error of about 1.28(k) and a Mean Absolute Error of about 0.85(k).

Table 6. Comparison of performance evaluation results for Price (USD) forecasting

Algorithm	MAE	RMSE
LSTM	0.44	0.75
MLP	1.09	1.57
CNN	0.57	0.92

The average error as shown in Table 6 indicates that LSTM based forecasting performed better than the other two algorithms with a Root Mean Square Error of about 0.75 and a Mean Absolute Error of about 0.44.

6. Discussion

RQ1. What are the critical features that influence drug sales?

Our objective was to forecast future sales and prices. Year & Quarter are the main features required to build date for the time series, Sales (Quantity), and USD Prices are our targets. Data sources also affect sales volumes and the change of the prices and each source has different behavior in the market. based on that we split our dataset into three other datasets to study them separately.

RQ2. How can the Deep Artificial Neural Network algorithms be chosen to resolve the sales forecasting problem?

Deep neural network algorithms had been selected based on the literature review and what is recommended for time series, we have chosen LSTM, MLP, and CNN and checked if these models have a good fit for the data by reviewing the learning curve.

RQ3. Which Deep Neural Network model is efficient for forecasting the sales of drugs in the private pharmaceutical market?

Long Short-Term Memory proved to be an efficient Deep learning algorithm. From the experiment, Long Short-Term memory performed better when compared to Multilayer Perceptron and Convolutional Neural networks. This is because LSTM makes use of temporal information from the data. The average Root Mean Square Error of Long Short-Term Memory for sales is 1.28(k) and Mean Absolute Error of about 0.85(k). The average Root Mean Square Error of Long Short-Term Memory for USD Prices is about 0.75, and the Mean Absolute Error is about 0.44. Which is less when compared to that of Multilayer Perceptron and Convolutional Neural networks.

Final remarks Different sources of data, moving stock between agents or pharmacies, the change of sample size & universe size in the different periods between (2010 to 2021) and the smuggled sales for agents and pharmacies coming from neighboring countries are the primary reasons that make data tracking very complex. The proposed model was implemented using Deep Artificial Neural Network algorithms, and the main objective was to obtain a suitable algorithm for sales and price forecasting from selected algorithms (LSTM, MLP, CNN).

7. Conclusion

Sales forecasting plays an important role in the business sector in every field. With the sales forecasts improved, sales revenue analysis will help decision-makers get the required details to estimate both the revenue and the income. Artificial Neural Networks have been selected among several machine learning models because they can effectively handle non-linear data. In this thesis, three Artificial Neural Network algorithms: Multilayer perceptron, Convolutional Neural Network, and Long Short-Term Memory are identified as fitting machine learning algorithms for drug sales and price predicting. The three algorithms are evaluated using RMSE and MAE performance metrics. The model with the lowest value is considered to be an efficient model for generating forecasts. Based on the results from the experiment, Long Short-Term Memory performed better than MLP and CNN for generating predictions with an average Root Mean Square Error of for sales is 1.28(k) and a Mean Absolute Error of about 0.85(k), and average Root Mean Square Error for USD Prices is about 0.75, and Mean Absolute Error is about 0.44. The forecasts are then used to adjust stock levels according to the predictions. As part of future work, the predictions can be enhanced by including a set of influential factors as feature variables in the dataset, such as the effect of smuggled sales on legal sales and pricing in the market, bonuses, and discounts offered on specific variants. These are a sample of the factors that could be used to understand the variations in time series data, allowing us to further improve the model performance to generate reliable forecasts. The potential future work is to build a recommended system by finding a relation between sales and the price change from historical data to recommend the best price for a new product in a specific ATC to make the best sales.

Acknowledgment

The authors thank Applied Science Private University and Princess Sumaya University for Technology for their support to carry out this research project.

References

Andrawis, R. R., Atiya, A. F., & El-Shishiny, H. (2011). Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition. International Journal of Forecasting, 27(3), 672–688. https://doi.org/10.1016/J.IJFORECAST.2010.09.005

Aras, S., & Kocakoç, T. D. (2016). A new model selection strategy in time series forecasting with artificial neural networks: IHTS. Neurocomputing, 174, 974–987. https://doi.org/10.1016/J.NEUCOM.2015.10.036

Benidis, K., Sundar Rangapuram, S., Flunkert, V., Research, A., Yuyang Wang, G., Maddix, D., Caner Turkmen, U., Gasthaus, J., Bohlke-schneider, M., Salinas, D., Stella, L., Aubet, F., Callot, L., Wang, Y., Turkmen, C., Bohlke-Schneider, M., Aubet, F.-X., Turkmen, C., Gasthaus, J., … Maddix, D. (2022). Deep Learning for Time Series Forecasting: Tutorial and Literature Survey. ACM Computing Surveys, 55(6). https://doi.org/10.1145/3533382

Bing, L., Chan, K. C. C., & Ou, C. (2014). Public sentiment analysis in twitter data for prediction of a company's stock price movements. Proceedings - 11th IEEE International Conference on E-Business Engineering, ICEBE 2014 - Including 10th Workshop on Service-Oriented Applications, Integration and Collaboration, SOAIC 2014 and 1st Workshop on E-Commerce Engineering, ECE 2014, 232–239. https://doi.org/10.1109/ICEBE.2014.47

Boyapati, S. N., & Mummidi, R. (2020). Predicting sales using Machine Learning Techniques. https://www.diva-portal.org/smash/get/diva2:1455353/FULLTEXT02

Dai, J., Zhang, P., Mazumdar, J., Harley, R. G., & Venayagamoorthy, G. K. (2008). A comparison of MLP, RNN and ESN in determining harmonic contributions from nonlinear loads. IECON Proceedings (Industrial Electronics Conference), 3025–3032. https://doi.org/10.1109/IECON.2008.4758443

Ding, X., Zhang, Y., Liu, T., & Duan, J. (n.d.). Deep Learning for Event-Driven Stock Prediction.

Doganis, P., Alexandridis, A., Patrinos, P., & Sarimveis, H. (2006). Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. Journal of Food Engineering, 75(2), 196–204. https://doi.org/10.1016/J.JFOODENG.2005.03.056

Ensafi, Y., Amin, S. H., Zhang, G., & Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. International Journal of Information Management Data Insights, 2(1), 100058. https://doi.org/10.1016/J.JJIMEI.2022.100058

García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., & Herrera, F. (2016). Big data preprocessing: methods and prospects. Big Data Analytics 2016 1:1, 1(1), 1–22. https://doi.org/10.1186/S41044-016-0014-0

Hernández, E., Sanchez-Anguix, V., Julian, V., Palanca, J., & Duque, N. (2016). Rainfall prediction: A deep learning approach. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9648, 151–162. https://doi.org/10.1007/978-3-319-32034-2_13/COVER

Hossen, T., Plathottam, S. J., Angamuthu, R. K., Ranganathan, P., & Salehfar, H. (2017). Short-term load forecasting using deep neural networks (DNN). 2017 North American Power Symposium, NAPS 2017. https://doi.org/10.1109/NAPS.2017.8107271

Islek, I., & Oguducu, S. G. (2015). A retail demand forecasting model based on data mining techniques. IEEE International Symposium on Industrial Electronics, 2015-September, 55–60. https://doi.org/10.1109/ISIE.2015.7281443

Ji, S., Yu, H., Guo, Y., & Zhang, Z. (2016). Research on sales forecasting based on ARIMA and BP neural network combined model. ACM International Conference Proceeding Series. https://doi.org/10.1145/3028842.3028883

Kamruzzaman, J., Sarker, R. A., & Begg, R. (1 C.E.). Artificial Neural Networks: Applications in Finance and Manufacturing. Https://Services.Igi-Global.Com/Resolvedoi/Resolve.Aspx?Doi=10.4018/978-1-59140-670-9.Ch001, 1–27. https://doi.org/10.4018/978-1-59140-670-9.CH001

Kim, K. H., Lee, C. S., Jo, S. M., & Cho, S. B. (2016). Predicting the success of bank telemarketing using deep convolutional neural network. Proceedings of the 2015 7th International Conference of Soft Computing and Pattern Recognition, SoCPaR 2015, 314–

317. https://doi.org/10.1109/SOCPAR.2015.7492828

Kraus, M., & Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. Decision Support Systems, 104, 38–48. https://doi.org/10.1016/J.DSS.2017.10.001

Lakner, Z., Kiss, A., Popp, J., Zéman, Z., Máté, D., & Oláh, J. (2019). From Basic Research to Competitiveness: An Econometric Analysis of the Global Pharmaceutical Sector. Sustainability 2019, Vol. 11, Page 3125, 11(11), 3125. https://doi.org/10.3390/SU11113125

Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: a survey. Philosophical Transactions of the Royal Society A, 379(2194). https://doi.org/10.1098/RSTA.2020.0209

Mehdiyev, N., Lahann, J., Emrich, A., Enke, D., Fettke, P., & Loos, P. (2017). Time Series Classification using Deep Learning for Process Planning: A Case from the Process Industry. Procedia Computer Science, 114, 242–249. https://doi.org/10.1016/J.PROCS.2017.09.066

Processing, N. A.-F. in S., & 2017, undefined. (2017). Neural networks for financial market risk classification. Isaac-Scientific.ComN AbroyanFrontiers in Signal Processing, 2017•isaac-Scientific.Com, 1(2). https://doi.org/10.22606/fsp.2017.12002

Sci, A. S.-S. U. D. C., & 2014, undefined. (n.d.). Convolutional networks for stock trading. Cs231n.Stanford.EduA SiripurapuStanford Univ Dep Comput Sci, 2014•cs231n.Stanford.Edu. Retrieved September 21, 2023, from http://cs231n.stanford.edu/reports/2015/pdfs/ashwin_final_paper.pdf

Siami-Namini, S., & Namin, A. S. (2018). Forecasting Economics and Financial Time Series: ARIMA vs. LSTM. https://arxiv.org/abs/1803.06386v1

Taylor, D. (2015). The pharmaceutical industry and the future of drug development. Pharmaceutical in the Environment. https://pubs.rsc.org/en/content/chapterhtml/2015/9781782622345-00001?isbn=978-1-

Tiriveedhi, V. (2018). Impact of Precision Medicine on Drug Repositioning and Pricing: A Too Small to Thrive Crisis. Journal of Personalized Medicine 2018, Vol. 8, Page 36, 8(4), 36. https://doi.org/10.3390/JPM8040036

Wang, C. H., & Yun, Y. (2020). Demand planning and sales forecasting for motherboard manufacturers considering dynamic interactions of computer products. Computers & Industrial Engineering, 149, 106788. https://doi.org/10.1016/J.CIE.2020.106788

Wang, Z., & Lou, Y. (2019). Hydrological time series forecast model based on wavelet de-noising and ARIMA-LSTM. Proceedings of 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2019, 1697–1701. https://doi.org/10.1109/ITNEC.2019.8729441

Yeasmin, N., Amin, S. H., & Tosarkani, B. M. (2022). Machine Learning Techniques for Grocery Sales Forecasting by Analyzing Historical Data. 21–36. https://doi.org/10.1007/978-3-030-85383-9_2

Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. Applied Artificial Intelligence, 17(5–6), 375–381. https://doi.org/10.1080/713827180

AUTHOR BIOGRAPHIES

Noura Qassrawi is Master student pursuing MSc in computer Science at Applied Science Private University. His research interests include Machine learning and data mining.

Mohammad Azzeh is a full professor at the Department of Data Science at Princess Sumaya University for Technology. He earned his PhD in Computing from University of Bradford in 2010, Bradford, UK. M.S.C in Software Engineering from University of the West of England, Bristol, UK. Dr. Mohammad has published over 50 research articles in reputable journals and conferences such as IET Software, Software: Evolution & Process, Empirical Software Engineering and Systems & Software. His research interests focus on Software Cost Estimation, Empirical Software Engineering, Data Science, Mining Software Repositories, Machine Learning for Software Engineering Problems. Dr. Mohammad was Conference chair of CSIT2016 and CSIT2018, and he is co-chair of many IT-related workshops.

Mohammad Hijjawi is an associate professor at the Department of Computer Science in the Faculty of Information Technology at Applied Science University. He has published many research articles in reputable journals and conferences. His research interests focus on Natural Language Processing and Health informatics.