Industrial Financial Forecasting using Long Short-Term Memory Recurrent Neural Networks

Muhammad Mohsin Ali[1], Muhammad Imran Babar[2], Muhammad Hamza[3], Muhammad Jehanzeb[4], Saad Habib[5], Muhammad Sajid Khan[6]

APCOMS, Rawalpindi, Pakistan.UET, Taxila, Pakistan

Abstract—**This research deals with the industrial financial forecasting in order to calculate the yearly expenditure of the organization. Forecasting helps in estimation of the future trends and provides a valuable information to make the industrial decisions. With growing economies, the financial world spends billions in terms of expenses. These expenditures are also defined as budgets or operational resources for a functional workplace. These expenses carry a fluctuating property as opposed to a linear or constant growth and this information if extracted can reshape the future in terms of effective spending of finances and will give an insight for the future budgeting reforms. It is a challenge to grasp over the changing trends with an effective accuracy and for this purpose machine learning approaches can be utilized. In this study Long Short-Term Memory (LSTM), which is a variant of Recurrent Neural Network (RNN) from the family of Artificial Neural Networks (ANN), is used for forecasting purposes along with a statistical tool IBM SPSS for comparative analysis. In this study, the experiments are performed on the data set of Pakistan GDP by type of expenditure at current prices - national currency (1970-2016) produced by Economic Statistics Branch of the United Nations Statistics Division (UNSD). Results of this study demonstrate that the proposed model predicted the expenses with better accuracy than that of the classical statistical tools.**

Keywords—Financial forecasting; prediction; long-short term memory; recurrent neural networks; artificial neural networks; IBM SPSS

I. INTRODUCTION

Forecasting is concerned with the process of estimating the future trends by doing the analysis on the basis of historical data. Financial forecasting determines the trends by utilizing the previous data and provides valuable information to make future decisions and define strategies for financial management. Financial forecasting is highly vital as different companies and firms are closed due to bankruptcy and the obvious reason behind it was their strategies that were not well-defined and they were unable to compete their rivals. They were not able to see what is coming in the future and when it comes, they were not prepared with the consequences. This looks like a troublesome issue to comprehend however, there is a solution to every problem. As the time goes on, organizations spare their vital data in a very much characterized and appropriate way since it may be useful for them in later phases of the business to analyze the trends. This is where the solution lies.

Whatever, that is occurring in the organization is being spared either it is related to their products or manufacturing or their finances. Forecasting the financial conditions, in terms of expenses, will be vital for a company to survive and that is the idea of this research. Much of the work is done in the domain of stock market [1-5], power and load [6-9], building energy consumption [10-12], electric price [13-16], weather forecasting [17, 18] and so on. However, the focus of this research is on expense forecasting of the companies to save them from bankruptcy.

In this research a technique is proposed to predict the financial expenses of an organization. The data of the company are analyzed and helpful data are isolated. The legitimacy of the data is highly vital to ensure the results of the proposed intelligent technique for financial forecasting. Different machine learning (ML) techniques are used for financial forecasting. The understanding of the given outcomes is highly vital with the goal that they can be utilized later for various purposes. There are couple of imperative things to consider like the information being utilized is from dependable sources and not fabricated. Secondly, system being designed needs not to be error prone. Forecasting is never 100 % yet it should be as close as it is conceivable. Salaries, business charges, office lease, telephone charges or any other, it appears as though there's no conclusion to the costs related with maintaining a business or to a specific individual. In any case, your capacity to get a firm handle on these money outpourings can assume an imperative part in your definitive achievement or disappointment. Regardless of whether you're contemplating a startup venture or you've been doing business for some time, precisely determining your costs can profit your venture in a number of potential ways.

To improve the procedure of financial forecasting there are different strategies which can be utilized to interface each sort of cost to different factors or cost drivers, for example, income or headcount, which have just been conjecture in the budgetary projections. Obviously, there will be expenses which are settled down and cannot be connected to different factors and should be assessed in total financial terms. There is a need of system that should keep all implying factors in to account and forecast expenses with logical and dependable results. Forecasting helps to create simulation in understanding problems of many sorts for example the weather, a simple analysis can lead us to predict the chances of rain in a particular way that the farmer can sow the seeds accordingly and predicting has led us to many possibilities same can be said in financial sectors and we can consider prediction of the bankruptcies on a corporate level and have counter measure ready at the time of need.

II. RESEARCH BACKGROUND

Forecasting has ever been a hot topic in research horizon. The topic for corporate firm bankruptcies prediction has also been in considerable discussions for the discipline of finance and risk management and various bankruptcies' prediction models are available in the form of light statistical models and artificial intelligence (AI) techniques [19]. The existing techniques involve decision tree, logistics regression and artificial neural networks (ANNs) and among all of them ANN became one of the most popular technique in accurate prediction. The ANN application has also been proven to determine the mortgage applicant solvency, rating of corporate bonds and fraud prevention. The basic intent of all these studies is to compare and contrast the performance and accuracy of the ANN and to justify its usage in the research.

In [19] a comparative analysis of two packages, one is statistical named as SYSTAT [20] and the other is software package named as BRAINMAKER [21], is performed. SYSTAT is a personal statistical package used in discriminant analysis. The test concluded with all variables included in each discriminant analysis. BRAINMAKER software package is based on neural network. After experimentation it was found that neural network outperformed the discriminant analysis resulting in neural network's learning more efficient than the classical methods.

Another study carried out in [22] focuses on the parallelization of the back propagation algorithm over a network that forecasts S&P500 Index in terms of prices of financial instruments, a series prediction problem, the ANN has been used extensively. Stock market forecasting with demands reaching at its peak as people getting more connected to the stock business. The challenge remained for predicting a strong forecast although a few data mining and machine learning techniques are significant enough to do the job. The use of ANN with a windowing operator for prediction is efficiently working along with the time series data [23].

Different techniques are used till now and ANN is the most prominent among them for prediction. Two theories commonly used for stock market prediction are "Efficient market hypothesis (EMH)" and "The random walk theory." Two commonly used approaches for stock market prediction are "Technical analysis" and "Fundamental analysis." All the stated theories and approaches are not of much use to us now but they still provide the foundation for new research. ANN consists of three layers which are helpful in case of nonlinear relationship in between inputs and output. There are several guidelines available that can aid in forecasting [2].

Feedforward NN was first used because of its simple approach and stock price data of IBM for 5000 days was used. Data of 1000 days was used for training and rest for validation purpose. Though it did not provide good results but it gave valuable insight for future experiments that were conducted very later and resulted in an accuracy of nearly 90% by using a three-layered back propagation ANN. When compared with the available statistical models such as GARCH, EGARCH, GJR GARCH, and IGARCH, the ANN performed better than all.

Many experiments were also made where ANN was used along with support of other algorithms like genetic algorithm, fuzzy logic etc. which further improved the results up to 74%. Now a day hybrid methods are being used like back propagation algorithm with Feedforward ANN or with Genetic Algorithm and these hybrid methods resulted up to 98% of accuracy [2].

Without using historical data, load and power forecasting has been proven by using ANN and the electrical energy produced with wind plays a crucial role but the unpredictable nature requires an approach to manage the resources, however, prediction requires a historical dataset to train the neural network for prediction. The drawback exists when there are no historical datasets of the newly created wind farms so the proposed approach requires no extensive historical data set [24]. The model purposed is the self-adaptive artificial network and such prediction model exist using the hidden layer complexity such as Multi-layer Perceptron, Support Vector Machine and Radial Basis Function all of which require high computation power. To overcome the computation complexity the authors used single layered model Functional Link ANN (FLANN) that results in trigonometric base output vector, Chebyshev Neural Network (ChNN) that results in polynomials base output vector and Legendre Neural Network (LeNN) that results in Chebyshev polynomials base output vector having the lower load to process.

Price Forecasting is a useful technique and is constantly being developed because of its benefits. The given research throws some light on the price forecasting of electricity for the day ahead. Different models are present for this purpose, but ANN is preferred because of its dependable results [25]. Support Vector Machine (SVM) Model is developed for price forecasting in case of rapidly changing environment and is used because of its speed and minimizing risks. ANN is used mainly to tackle with the nonlinear relationship in between inputs and output. One thing to be noted is that results were forecasted for a more stable environment with less changes.

For buildings the prediction of demands for its energy consumption holds a great importance. Energy if properly used and managed can be helpful in lowering the operational cost of building. Certain factors are involved in energy demands of a building which includes climate, structure of building and people inside the building. There are few methods available to predict energy demands [26]. Physical method is based on engineering methods for calculating building's energy demands but is less popular because of difficulty in gathering required information like physical parameters of building. With the passage of time different new models based on AI like ANN and SVM were used to obtain better results. However, ANN is used in this research because of its better performance. Data of a building located in Italy was used for experimentation with energy consumption in KWH and an ANN model known as Nonlinear Auto Regressive Neural Network (NAR) was chosen. It has a feedback arrangement and is suitable for forecasting in such cases where only single series of information is present like in the problem of building energy consumption. Results concluded that NAR is an effective model and can give dependable results in case of building energy consumption [26].

ANNs are also widely applicable in weather forecasting. For the past few decades numerous efforts are made to forecast weather. ANNs have played a vital role for prediction purposes and many forecasting models are based on it. The researchers proposed a new approach based on ensemble of neural networks for forecasting [27]. The proposed model intends to solve the redundancy issue by a mutual information sharing process. As a whole the process consists of four basic stages. In the first stage the features were selected from the normalized data. In the second stage an ensemble of four neural networks, which include Multi-layered Perceptron (MLP), Radial Basis Function (RBF), General regression neural network (GRNN) and Time delay neural network (TDNN) was developed for forecasting. The proposed model was applied and compared with existing techniques of SOM and Voting, the proposed model predicts successful results [27].

A research was presented for weather forecasting that includes the technique Back-Propagation Neural Network (BPNN). BPNN has the ability to learn on itself adjusting its weights to achieve the desired results. Different numerical as well as other techniques are present for the purpose of forecasting but still there is a margin of improvement. ANN is brilliant concept when it comes to forecasting [28]. The results conclude that the proposed system is performing very well under the given attributes including Temp. (°C), Dew Point (°C), Humidity (%), Sea Level Pressure (hPa), Visibility (km), Wind (km/h), Gust Speed (km/h) and Precip (cm). Using the BPPN for forecasting on the given attributes produced satisfying results and can result in replicating the existing methods used for forecasting all around the world.

Since weather prediction continues to be a day to day need as it effects over the agriculture and industrial sectors both and is dependent on the predicted data in order to function but the demand is not only limited to that warning for natural disaster can turn out to be a big life saver. The prediction varies for location to location and time as well, to start with to be accurate in regard of time differences for when the data is being recoded and for when it is being used for prediction. A research was presented to predict the temperate weather that included the collection assimilation & analysis of data and the methodology comprises of multi-layer perceptron neural network [29].

The data collected is of historic past of 5 years having the observation or the selected input attributes of temperature, wind speed, wind direction and atmospheric pressure. The analytical data is given to the ANN for the prediction purposes through usage of well-known forecasting models. These models have been known for their effectiveness over predictions, the training is conducted using the back-propagation algorithm and the first set of generated output was parsed through ID3 to generate rules. The ID3 tree had the outcome from the neural network resulting in to the final output which were the generated rules. The experiment resulted in outcome from the neural network which then given to the ID3 decision tree leading to the conversion of the output to the desired one and through this implementation it is clear demonstration of how effective is the integration of the neural network with the intelligent system as compared to the traditional meteorological approaches [29].

A weather forecast model presented in [30] is based on Multilayer Feedforward Artificial Neural Network (MLFANN) that is based on patterns that are process of conversion from input sets into the outputs. The training of the network is through Resilient Propagation algorithm or RPROP Algorithm. It has proven that it produced the most accurate results during the test periods which were conducted to compare the Conjugate Gradient in terms of the performance [30]. The data considered was for the weather forecasting on a daily basis over in Tiwi, San Rafael, Albay and Philippines. The datasets had the data of between 2012 up to 2015 and the gathered data were provided by the Advanced Science and Technology Institute (DOST – ASTI). With division of the data set resulted into two data set one imputed and the other one with the removed missing values. Three models were created for both the data sets with around 40000 iterations, the networks number of hidden neurons were based through a theorem (the Kolmogorov's Theorem) the weight training went through first set a random value and then the second set by IWI methodology. The study resulted in a successful 98.96743% accurate prediction although the optimal model contained 10 neurons in the hidden layer and the dataset comprising of removed missing values is utilized.

III. COMPARATIVE ANALYSIS

A detailed comparative analysis of multiple techniques based on the concept of Neural Network is presented in Table I. Table I shows that many techniques like Hybrid Feed Forward and Back Propagation as well as Chebyshev Neural Network (ChNN) show results near to 90 % when it comes to accuracy. Another added benefit is that they also have a very high learning rate among all. For our case expense forecasting which is related to financial forecasting, Hybrid Feed Forward and Back Propagation technique shows perfect results and has been proved. For the purpose of comparative analysis four characteristics are considered like learning rate, effective usage, efficiency and resource requirements. Learning rate describes that how much the proposed technique is efficient in terms of learning, effective usage explains the domain in which the technique is applied, efficiency is based on performance as compared to the contemporary techniques and resource requirements depicts the computational needs of the proposed technique.

TABLE I.   COMPARATIVE ANALYSIS OF EXISTING TECHNIQUES

	Learning Rate	Effective usage	Efficiency	Resource Requirements
SYSTAT (DA) [20]	Mid	Financial	88.25% (Ratio With competitor)	High
BRAINMAKER (BP) [21]	High	Financial	97.5% (Ratio With competitor)	High
Parallelization (BP) [22]	High	Generic	61% faster	Mid
Hybrid Feed Forward and Back Propagation [2]	High	Financial	Up 97% more accurate	High
Functional Link ANN (FLANN) [24]	Very High	Power/load	Faster as compared ChNN & LeNN	Low
Chebyshev Neural Network (ChNN) [24]	Very High	Power/load	90% overall performance	Low
Legendre Neural Network (LeNN) [24]	High	Power/load	N/A	Low
Long Short-Term Memory (LSTM) [31]	High	Financial/Pattern Prediction	59% (Average Accuracy)	High

IV. PROPOSED SOLUTION

After going through the detailed literature review, we stumbled upon many methods to setup a neural network. Among those were the classical methods including Back Propagation and more complex mathematical models such as LeNN. As we understand from the study among these stood RNN type method LSTM with a memory to redirect its training under short terms as well as longer terms of time. The proposed model is shown in Fig. 1.

A. LSTM

Conventional Neural Networks were of incredible use to us but they had one issue. Every time they had to start from scratch until the perfect combination was found but what if there is a neural network which can save the output and utilizing it again by feeding it to input. Recurrent Neural Network (RNN) is a type of neural networks that is based on the concept of memory. Once the feed forward network reaches to output the learning rate and error correction are performed during the back propagation to get the answer right. They have been utilized effectively in taking care of numerous issues like speech recognition, financial solutions and so forth.

Recurrent Neural Networks (RNNs) have been used in and in addition with blend of different models. An example of this is calculating Financial Volatility. Long Short-Term Memory Networks (LSTMs) are a variant of RNN that are now most popular in the field. There are countless examples which can prove its working and usability. LSTMs have been used in multiple fields for prediction purposes and have given noteworthy outcomes.

In transportation, calculating the flow of traffic has been a difficult task because of being highly non-linear in nature. An approach was made to solve this problem using LSTMs. Simply using traffic flow information gave acceptable results but with few other variables like occupancy and speed along with neighboring traffic information showed better results [32].
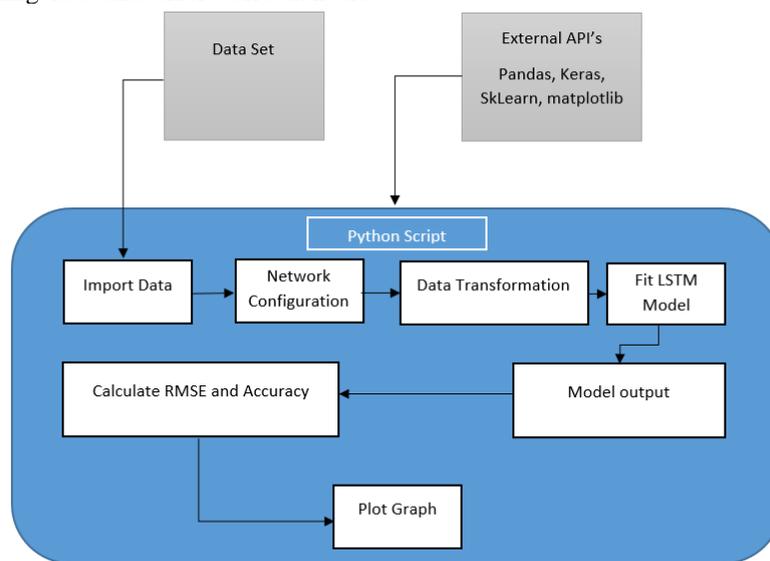

Fig. 1.   Proposed Model for Industrial Financial Prediction.

LSTMs were also used in stock market prediction and they came out successful with an average accuracy of 55.9%. It told whether the price of the specific stock is going up or not in the coming 15 minutes and for this purpose the data from 2008 to 2015 were used. Google Tensor Flow was used for performing LSTMs and with such large dimension of input the results were obtained without reducing or handpicking the attributes [31].

RNN was also used for prediction in the field of energy consumption. Short term load forecasting was done on 69 customers of New South Wales. The important part to consider was that the research focused on forecasting on single customer and not as a whole and several experiments resulted that approaches that are successful in grid forecasting may not be very successful in this area and LSTMs produce better results [33].

For calculating the remaining life of Lithium-Ion batteries LSTMs were used. As interesting it may seem, it is of immense use for minimizing battery risks. Different cells were used for this purpose on two different temperatures (25 °C and 40 °C). Different current rates were used to get maximum information and diversity. It is a long-term prediction as it calculates the time of failure for the batteries and the results show that LSTM's give better results than existing techniques like Support Vector Machine (SVM) [34].

LSTMs are a kind of recurrent neural network which can hold the memory of previous prediction and are capable enough for learning long term dependencies. It is widely been used in speech prediction, uses a default squashing function (tanh) which keeps the range of outputs to be in 1 and -1 depending upon the calculated voted outcome. LSTM initially predicts a value and then stores it to be used along with the next value of input. LSTM loops the previous predicted values. It does that by using 4 stages initial prediction, ignoring, forgetting and selection. All 4 states are separate mini networks with the squashing functions.

1) Initial Input: The first gate where the new information is processed initially.

2) Ignoring: Should a new information be stored or the information are to be ignored. The gate has a sigmoid function that decides what new information be updated.

3) Forgetting: Choses what previous output needs to be forget and drop the information of the old subject votes for what not.

4) Selection: This gate has another sigmoid squashing function which choses the most valid information to be output. Visualization of the above process can be seen in Fig. 2.

Mathematical representation

Input Gate: $\qquad\qquad f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$

Update Gate: $\qquad\quad i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$

$$\tilde{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

Forgetting Gate: $\qquad C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$

Output Gate: $\qquad\quad o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$

$$h_t = o_t * \tanh(C_t)$$

B. Tuning the Model

LSTM uses an activation function tanh and a basic RMSprop optimizer. Although the network can function over these basic models but we tend to change the activation functions to sigmoid, Softmax and SELU to better tune the prediction. We also tuned the optimizer to an Adamax Stochastic optimizer with a variable learning rate and shuffle of the population set to be restricted for the efficiency and better performance. In addition to that Equation 1 is used to measure accuracy to better visualize the results.

$$Accuracy = \frac{\sum_{n=0}^{n} 100 - ((\frac{P_n - E_n}{P_n}) * 100)}{n} \qquad (1)$$

The detail of the activation functions is given in Table II.

C. Execution

The model proposed in Fig. 1 is implemented by using Python with Keras API. The script first parses the data through Pandas API. Further the data are transformed from time series to supervised and stationary data. The data are pre-processed. Thus, the training begins based on the configured model of the LSTM. The proposed model forecast financial values and visualizes the results accordingly. Fig. 3 depicts the overall flow chart of the proposed LSTM model.
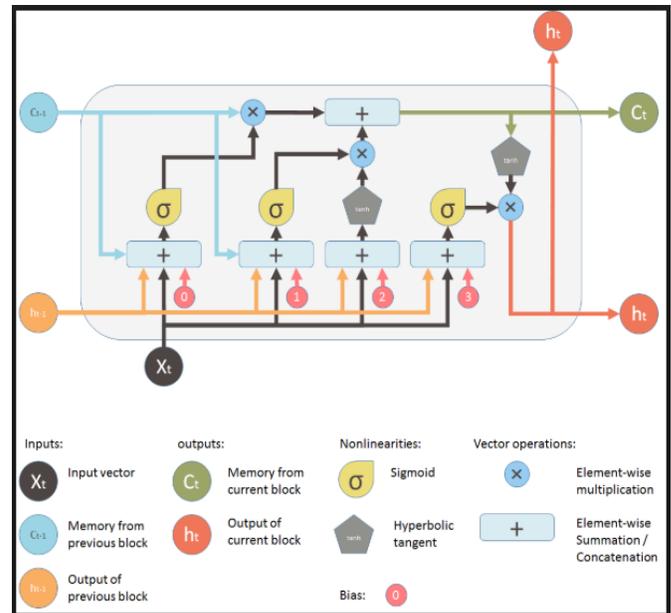

Fig. 2. Architecture of LSTM.

TABLE II.   DETAILS OF ACTIVATION FUNCTIONS

Name	Formula	Range	Plot
Tanh	$f(x) = \tanh(x) = \dfrac{(e^x - e^x)}{(e^x + e^x)}$	(-1,1)	
Sigmoid	$f(x) = \sigma(x) = \dfrac{1}{1 + e^{-x}}$	(0,1)	
Softmax	$f_i(\vec{x}) = \dfrac{e^{x_i}}{\sum_{j=1}^{j} e^{x_j}}$	(0,1)	
Scaled Exponential Linear Unit (SELU)	$f(\alpha, x) = \lambda \begin{cases} \alpha(e^x - 1) & for \ x < 0 \\ x & for \ x \geq 0 \end{cases}$	$(-\lambda\alpha, \infty)$	
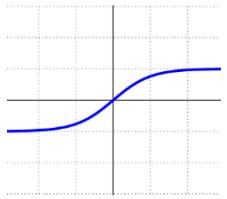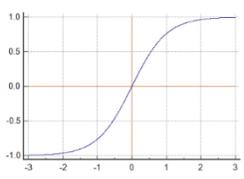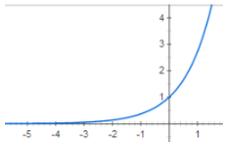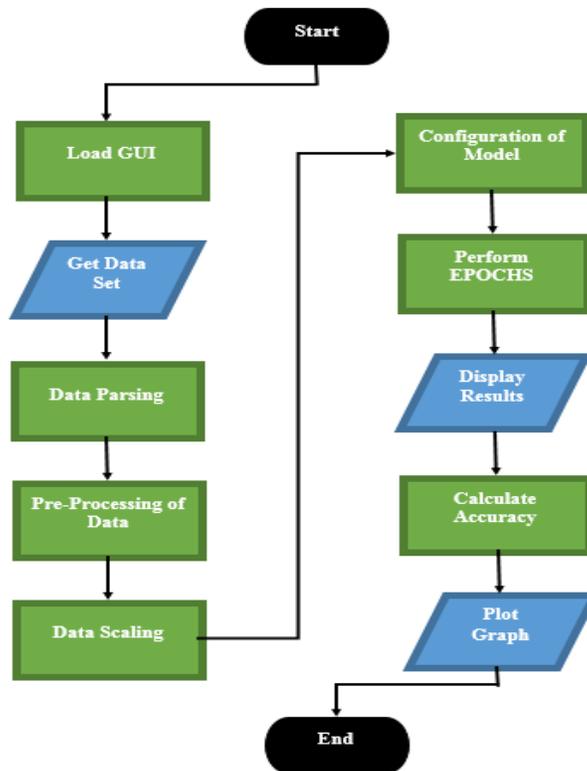
Fig. 3.   Flow Chart of the Proposed Model's Execution.

V. RESULTS AND ANALYSIS

For the past years the statistical methods and tools have been used for predication. The financial world has relied upon such techniques as they delivered good results. But what we believe is that the use of machine learning would result in better time series prediction as compared to the classical methods. There are many renowned tools like IBM SPSS. These methods help in recognition of the attributes to be valid and statistically correct. Fig. 4 shows the attribute processing report.

Time series analysis is a sequence of points in a given time. For example, a consumption of electricity at end of every month recorded over a period of time let's say years. We will end up with a time series of monthly based stretching over a number of years. Time series possesses a trend in it, a trend is said to be a change in general direction. Classical methods such as used in SPSS Time Series Analysis failed to fully grasp the change of trend as shown in Fig. 5.

Table V and Fig. 6 shows the results from the system which uses the default activation function (Tanh) for LSTM. In order to maximize the performance, we can use different available activation functions along with making tweaks in Learning Rate, EPOCHS, Number of Neurons in hidden layer and optimizers. This would be done using hit and trial method until we get the best combination to be used for getting the most accurate results. Comparison of Activation Functions (Network Configurations and Plots)

The above graph visualization in Fig. 6 is of the Tanh activation function in terms of expected and predicted values generated by the proposed model resulting in an average accuracy of 90.95%. Table III shows the network configuration used to generate the plot are Number of Neurons 10, Epochs 3000, Learning rate 0.002 and optimizer used is Adamax which are also mentioned in the table above. The graph has two lines of blue and orange color where each color represents expected and predicted values, respectively.

Fig. 7 is visualization of the Sigmoid activation function in terms of expected and predicted values generated by the proposed model resulting in an average accuracy of 89.46%. Table IV shows the network configuration used to generate the plot are Number of Neurons 10, Epochs 3000, Learning rate 0.002 and optimizer used is Adamax which are also mentioned in the Table VI. Fig. 7 represents the expected and predicted values respectively.

Fig. 8 is visualization of the Softmax activation function in terms of expected and predicted values generated by the proposed model resulting in an average accuracy of 91.64%. Table V shows the network configuration used to generate the plot are Number of Neurons 10, Epochs 3000, Learning rate 0.002 and optimizer used is Adamax which are also mentioned in the table above. The graph has two lines of blue and orange color where each color represents expected and predicted values respectively.

Fig. 9 is visualization of the SELU activation function in terms of expected and predicted values generated by the proposed model resulting in an average accuracy of 92.36%. Table VI shows the network configuration used to generate the plot are Number of Neurons 10, Epochs 3000, Learning rate 0.002 and optimizer used is Adamax which are also mentioned in the table above. The blue and orange lines represent expected and predicted values respectively.

A. Comparison of Activation Functions (Plot)

Fig. 10 and Table VII are helping in understanding the accuracy of all functions used in the experimentation in order to measure the function with highest accuracy. After performing the experiment several time using various activation functions, it is observed that the LSTM model using SELU function outperforms tanh, sigmoid and softmax functions having an average accuracy of 92.34%.

TABLE III. NETWORK CONFIGURATION AND RESULTS OF TANH FUNCTION

Activation Function: Tanh	
Neurons	10
Epochs	3000
Learning Rate	0.002
Optimizer	Adamax
Accuracy	90.945718 %

TABLE IV. NETWORK CONFIGURATION AND RESULTS OF SIGMOID FUNCTION

Activation Function: Sigmoid	
Neurons	10
Epochs	3000
Learning Rate	0.002
Optimizer	Adamax
Accuracy	89.461604%

TABLE V. NETWORK CONFIGURATION AND RESULTS OF SOFTMAX FUNCTION

Activation Function: Softmax	
Neurons	10
Epochs	3000
Learning Rate	0.002
Optimizer	Adamax
Accuracy	91.643963%

TABLE VI. NETWORK CONFIGURATION AND RESULTS OF SELU FUNCTION

Activation Function: Scaled Exponential Linear Unit (SELU)	
Neurons	10
Epochs	3000
Learning Rate	0.002
Optimizer	Adamax
Accuracy	92.356115%

Case Processing Summary

| | Cases | | | | | |
| | Included | | Excluded | | Total | |
	N	Percent	N	Percent	N	Percent
Finalconsumptionexpenditure	47	100.0%	0	0.0%	47	100.0%

Report

Finalconsumptionexpenditure

Mean	N	Std. Deviation
5.013E+12	47	7.3352E+12

Fig. 4.   Attribute Processing Report.
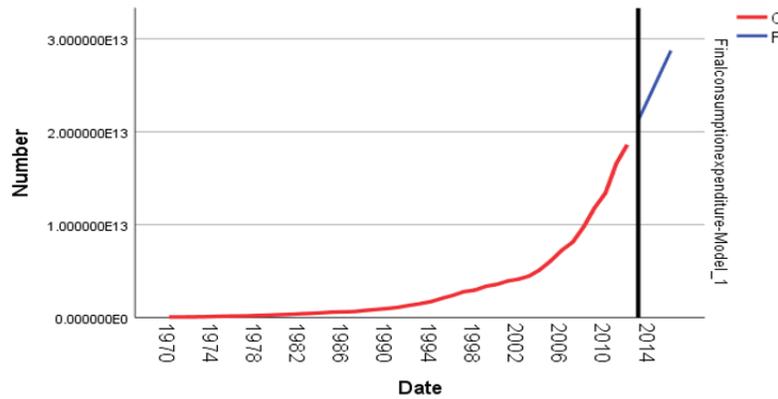
Fig. 5.   Statistical Forecasting on Data Set.
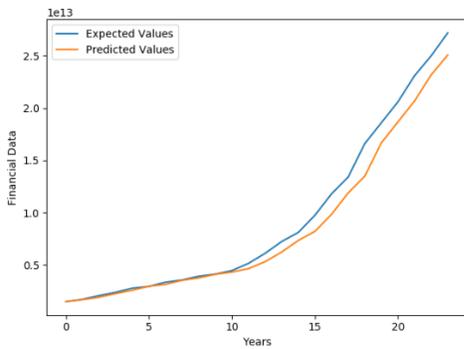

Fig. 6.   Graph of Tanh Function Results.
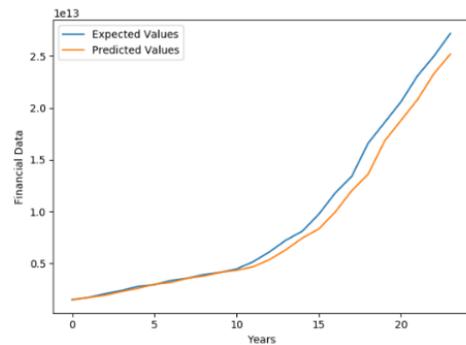
Fig. 8.   Graph of Softmax Function Results.
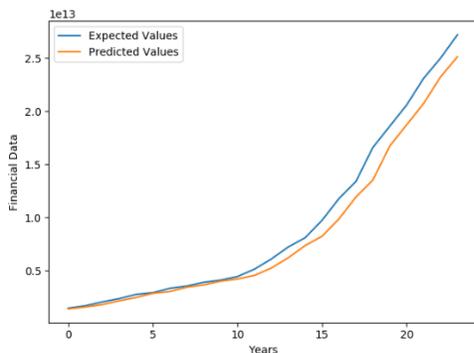
Fig. 7.   Graph of Sigmoid Function Results.
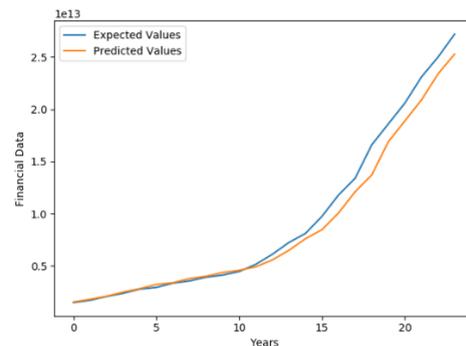
Fig. 9.   Graph of Softmax Function Results.

TABLE VII.   EXPECTED AND PREDICTED VALUES FROM ALL ACTIVATION FUNCTIONS

Year	Expected "Final consumption expenditure"	Tanh Function	Sigmoid Function	Softmax Function	Scaled Exponential Linear Unit (SELU) Function
1994	1490000000000	1492147941446	1416830417788	1520000000000	1546896795749
1995	1710000000000	1681736572456	1596830417788	1700000000000	1801833265018
1996	2070000000000	1905435251712	1816830417788	1920000000000	2097350911903
-	-	-	-	-	-
2014	23100000000000	20681566936051	20734847291469	20810000000000	20896554042625
2015	25000000000000	23177904562711	23234165990781	23310000000000	23391085687923
2016	27200000000000	25082044218373	25134989737868	25210000000000	25296475122737
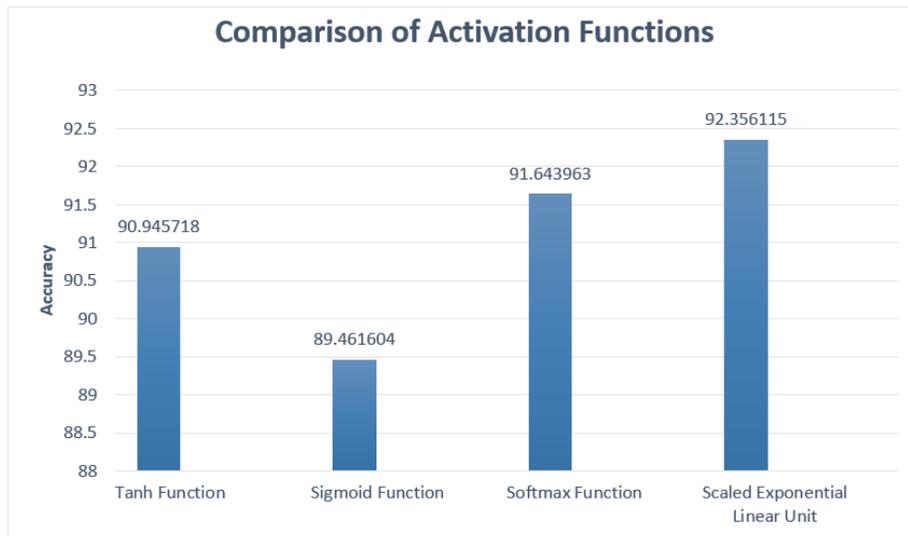
Fig. 10. Comparison of Activation Functions.

VI.   EXPERIMENT

The statistical method when tested under the condition was SPSS and expert model of its own best finding. Half of the data was used for it to analyze the trend and make a forecast for the rest of the values so we can compare the expected and predicted values. Fig. 11 show the actual curve and trend of data. Fig. 12 shows the forecasted trend of data.

As it is observed that the analysis failed to grasp upon the curve and didn't follow the trend rather it follows a linear growth path in the positive direction. Table VIII shows the output for the respective forecast in terms of UCL upper limit and LCL lower limit respectively.

A. Experimental Data Set

The experimental data set is used to train and validate the proposed system. The data set needs not to be manufactured rather it should have trends for forecasting purposes. The data set used was "Pakistan GDP by Type of Expenditure at current prices - national currency". It is produced and maintained by the Economic Statistics Branch of the United Nations Statistics Division (UNSD). Getting into the particulars of the data, it expands over 48 years beginning from 1970 up to 2016 [35]. It consists of following nine attributes.

- Final consumption expenditure
- Household consumption expenditure
- General government final consumption expenditure
- Gross capital formation
- Gross fixed capital formation
- Changes in inventories
- Exports of goods and services
- Imports of goods and services and
- Gross domestic product (GDP)
- Fig. 13 shows the partial dataset for experimental purposes.
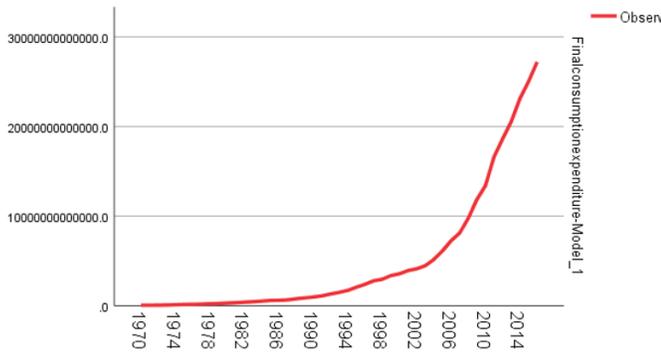
Fig. 11. Actual Curve and Trend of Data.
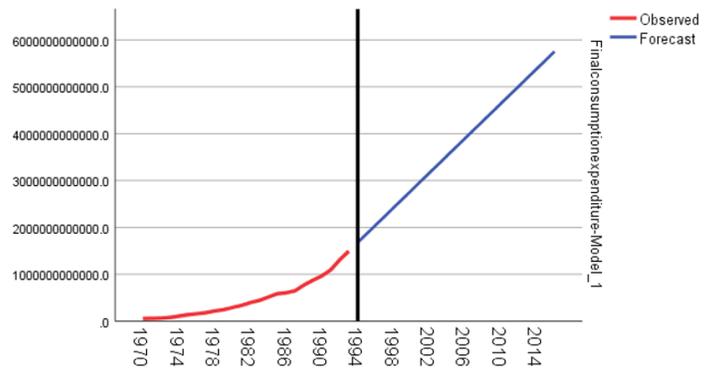
Fig. 12. Forecasted Trend of Data.

TABLE VIII. FORECASTED RESULTS OF 4 YEARS

Forecast					
Model		**2008**	**2009**	**2010**	**2011**
Finalconsumptionexpenditure-Model_1	Forecast	4270986450158.1	4456215747031.0	4641445043903.8	4826674340776.7
	UCL	6171717092146.6	6541483486414.0	6916855821347.0	7297673656143.6
	LCL	2370255808169.5	2370948007648.0	2366034266460.7	2355675025409.9

Fig. 13. Partial Data Set.

Subsequent to acquiring the data set another critical thing is to do the validation of data set and in addition finding the right attributes for the system to process. To do this we used a tool named "WEKA". This tool gave all the needed information regarding the data by executing the algorithm of attribute selection and provides us with facts like how much of our data is unique, how many missing values are there and are all the entries in data distinct or not. This is a vital step to execute as our outcomes are subject to it and in case the data is forged our results would not be accurate. Fig. 14 depicts the results that "WEKA" produced after running the attribute selection algorithm on the given data set.

Fig. 14. WEKA Attribute Selection Details.

B. Experimental Setup

Data and pre-processing: The experiment is executed by reading the data in .csv format by Pandas that reads and parses it. Then the dataset is splitted into two parts one for the training and another one for the testing. Now before we fit a machine learning algorithm or the LSTM model over it. In order to achieve this first we need to transform the data into a supervised learning problem with having an input x and output y in our time series case t-1 and t respectively. Then the stationary transformation is required to make model result with more skillful forecast. Lastly, we scale it as the model processes the input and output in range of (-1, 1) range that normalizes the data by using minmax scalar.

1) Model: After data pre-processing the optimizer provided by the Keras is configured by hit and trial method and agreed to keep the Adamax optimizer with a flexible learning rate. By default, LSTM has a tanh activation function but we tend override it with various different functions like sigmoid, SELU and Softmax. The model is tested based on various epochs, neuron populations and learning rates.

2) Results: The predicted values are displayed in comparison with the expected values. The results from the model are fed to the graph plotting function the matplotlib. In the process the accuracy of the model and the RMSE are also calculated to measure the accuracy of the proposed model.

C. Final Outcome

Based on selected dataset attributes the experiment is performed and the results are detailed in Table IX. The results produced by IBM SPSS and ANN (LSTM) are compared in terms of accuracy. The results predicted by Long Short-Term Recurrent Neural Network are more accurate as compared to the IBM SPSS.

The RMSE value calculated during experimentation is1221190424306.566.

The results in Table IX reflect the six years of testing data (Initial 3 years, Final 3 years). It is observed that for the year 1994 there was an unexpected change in the trend but still the proposed LSTM model predicted closer to the following trend. Another observation from the experiment is that SPSS only produced one promising result for the year 1996 out of six years. In contrast to the proposed LSTM it is observed that in Table IX LSTM model presented 5 promising results out of total six entries.

The conventional LSTM model uses Tanh as a default activation function, these settings did show promising results but the model used with SELU through research had better training curve and performed better as shown in Table X.

Clearly RMSE value of LSTM with default function is much higher than that of LSTM with SELU function. The function with lesser RMSE value is the one to be preferred. The difference between RMSE of both the functions is:

Precise difference = 38071328299.673

The difference shows how much closer results to actual value are produced by the SELU function when compared with the default function.

TABLE IX.   TABLE OF EXPERIMENT RESULTS

Year	Expected "Final consumption expenditure"	IBM SPSS	ANN(LSTM)
1994	1490000000000	1677776293937	1619476432991
1995	1710000000000	1863005590810	1780559280565
1996	2070000000000	2048234887683	1993555488109
-	-	-	-
2014	23100000000000	5382362231395	20837874049091
2015	25000000000000	5567591528268	23337874049091
2016	27200000000000	5752820825141	25237874049091

TABLE X.   ACCURACY OF TANH AND SELU

Functions of LSTM	Historic Data Requirement	RMSE	Average Accuracy
Tanh (Default)	Yes	1241633196117.760	90.94%
SELU	Yes	1203561867818.087	92.34%

VII. CONCLUSION

The impetus of the study was to construct and develop an artificially intelligent expense forecasting system that could grasp over a trend inside the data using artificial neural networks and machine learning techniques. After thorough study of the available algorithms and mathematical models, Long Short-Term Memory (LSTM) was used. Root Mean Square Error (RMSE) was used to calculate the error and accuracy was obtained by the difference between actual and predicted values. The proposed model had produced accurate result up to 91% average accuracy. The base model was further tweaked with variation of different activation function for testing the efficiency. The system can be used to anticipate the future trend of numeric financial data and that it can help in more refined budget definitions, spending criteria and crises avoidance.

REFERENCES

[1] G. Chan, "Forecasting the S&P 500 index using time series analysis and simulation methods," Massachusetts Institute of Technology, 2009.

[2] C. S. Vui, et al., "A review of stock market prediction with Artificial neural network (ANN)," in Control System, Computing and Engineering (ICCSCE), 2013 IEEE International Conference on, 2013, pp. 477-482.

[3] N. Nazir and M. Dmutto, "Stock Market Prediction Using Artificial Neural Network," International Journal of Advanced Engineering, Management and Science, vol. 2, 2006.

[4] T. Kimoto, et al., "Stock market prediction system with modular neural networks," in Neural Networks, 1990., 1990 IJCNN International Joint Conference on, 1990, pp. 1-6.

[5] Y. Zhang and L. Wu, "Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network," Expert Systems with Applications, vol. 36, pp. 8849-8854, 2009.

[6] H. S. Hippert, et al., "Neural networks for short-term load forecasting: A review and evaluation," IEEE Transactions on power systems, vol. 16, pp. 44-55, 2001.

[7] D. Wang and Z. Sun, "Big data analysis and parallel load forecasting of electric power user side," Proceedings of the CSEE, vol. 35, pp. 527-537, 2015.

[8] R. Hu, et al., "A short-term power load forecasting model based on the generalized regression neural network with decreasing step fruit fly optimization algorithm," Neurocomputing, vol. 221, pp. 24-31, 2017.

[9] K. G. Boroojeni, et al., "A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon," Electric Power Systems Research, vol. 142, pp. 58-73, 2017.

[10] K. Amasyali and N. M. El-Gohary, "A review of data-driven building energy consumption prediction studies," Renewable and Sustainable Energy Reviews, vol. 81, pp. 1192-1205, 2018.

[11] U. Berardi, "A cross-country comparison of the building energy consumptions and their trends," Resources, Conservation and Recycling, vol. 123, pp. 230-241, 2017.

[12] S. Singh and A. Yassine, "Big data mining of energy time series for behavioral analytics and energy consumption forecasting," Energies, vol. 11, p. 452, 2018.

[13] A. R. Gollou and N. Ghadimi, "A new feature selection and hybrid forecast engine for day-ahead price forecasting of electricity markets," Journal of Intelligent & Fuzzy Systems, vol. 32, pp. 4031-4045, 2017.

[14] J. Nowotarski and R. Weron, "Recent advances in electricity price forecasting: A review of probabilistic forecasting," Renewable and Sustainable Energy Reviews, 2017.

[15] J. Saez-Gallego and J. M. Morales, "Short-term forecasting of price-responsive loads using inverse optimization," IEEE Transactions on Smart Grid, 2017.

[16] A. Mirakyan, et al., "Composite forecasting approach, application for next-day electricity price forecasting," Energy Economics, vol. 66, pp. 228-237, 2017.

[17] W. Y. Cheng, et al., "Short-term wind forecast of a data assimilation/weather forecasting system with wind turbine anemometer measurement assimilation," Renewable Energy, vol. 107, pp. 340-351, 2017.

[18] C. Feng, et al., "A data-driven multi-model methodology with deep feature selection for short-term wind forecasting," Applied Energy, vol. 190, pp. 1245-1257, 2017.

[19] R. L. Wilson and R. Sharda, "Bankruptcy prediction using neural networks," Decision support systems, vol. 11, pp. 545-557, 1994.

[20] L. Wilkenson, "SYSTAT: the system for statistics. SYSTAT," Inc., Evanston. 638p, 1989.

[21] J. Stanley, Introduction to neural networks: California Scientific Software, 1990.

[22] C. A. Casas, "Parallelization of artificial neural network training algorithms: A financial forecasting application," in Computational Intelligence for Financial Engineering & Economics (CIFEr), 2012 IEEE Conference on, 2012, pp. 1-6.

[23] R. I. Rasel, et al., "Financial instability analysis using ANN and feature selection technique: application to stock market price prediction," in Innovations in Science, Engineering and Technology (ICISET), International Conference on, 2016, pp. 1-4.

[24] A. D. G. Reddy and L. T. Varma, "Wind power forecasting without using historical data," in Proceedings of the International Conference on Advances in Electrical Engineering (ICAEE), 2014, pp. 1-3.

[25] K. K. Nargale and S. Patil, "Day Ahead Price Forecasting in deregulated electricity market using artificial neural network," in Energy Efficient Technologies for Sustainability (ICEETS), 2016 International Conference on, 2016, pp. 527-532.

[26] S. Ferlito, et al., "Predictive models for building's energy consumption: An Artificial Neural Network (ANN) approach," in AISEM Annual Conference, 2015 XVIII, 2015, pp. 1-4.

[27] A. Ahmadi, et al., "Hybrid model for weather forecasting using ensemble of neural networks and mutual information," in Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International, 2014, pp. 3774-3777.

[28] S. S. Baboo and I. K. Shereef, "An efficient weather forecasting system using artificial neural network," International journal of environmental science and development, vol. 1, p. 321, 2010.

[29] R. Nayak, et al., "An artificial neural network model for weather forecasting in Bhopal," in Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on, 2012, pp. 747-749.

[30] K. L. M. D. Sobrevilla, et al., "Daily weather forecast in Tiwi, Albay, Philippines using Artificial Neural Network with missing values Imputation," in Region 10 Conference (TENCON), 2016 IEEE, 2016, pp. 2981-2985.

[31] D. M. Nelson, et al., "Stock market's price movement prediction with LSTM neural networks," in Neural Networks (IJCNN), 2017 International Joint Conference on, 2017, pp. 1419-1426.

[32] D. Kang, et al., "Short-term traffic flow prediction with LSTM recurrent neural network," in Intelligent Transportation Systems (ITSC), 2017 IEEE 20th International Conference on, 2017, pp. 1-6.

[33] W. Kong, et al., "Short-term residential load forecasting based on LSTM recurrent neural network," IEEE Transactions on Smart Grid, 2017.

[34] Y. Zhang, et al., "Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries," IEEE Transactions on Vehicular Technology, 2018.

[35] Quandl. (2016, 04/06/2018). Pakistan GDP by Type of Expenditure at current prices-national currency. Available: https://www.quandl.com/data/UNAE/GDPCN_PAK-GDP-Current-Prices-National-Currency-Pakistan