Comparing Machine Learning Models for Predicting the Global Internet Usage

MSc Research Project
Data Analytics

Tijo Sebastian
Student ID: x21139237

School of Computing
National College of Ireland

Supervisor: Mr. Hicham Rifai

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Tijo Sebastian
Student ID:	x21139237
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Mr. Hicham Rifai
Submission Due Date:	01/02/2023
Project Title:	Comparing Machine Learning Models for Predicting the Global Internet Usage
Word Count:	6001
Page Count:	17

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	1st February 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Comparing Machine Learning Models for Predicting the Global Internet Usage

Tijo Sebastian
x21139237

Abstract

The internet is increasingly important today. Because every aspect of life is in some manner closely tied to the internet, it is impossible for us to think about a tomorrow without it. The goal of this forecast is to predict network usage using past observations. The internet usage forecast can be beneficial for multiple reasons. The internet service providers are more benefited from this research so that they can plan for their future and implement necessary steps wherever required. The Long Short-term Memory (LSTM) model and the Simple Exponential Smoothing (SES) model are implemented in this research. Both LSTM and SES models are widely used for time series forecasts. The required data is gathered from an open-source platform. The analysis confirmed that the SES model outperformed the LSTM model in accuracy. The model accuracy is measured with appropriate metrics.

1 Introduction

The Internet is becoming a vital tool for our modern world. We cannot think about a world without the Internet because each sector in life is directly connected to the Internet one way or another. Internet usage prediction aims at predicting network traffic by using the previous network traffic data. This can serve as a proactive approach to network management and planning tasks. In modern society, the Internet and its applications have become a primary communication tool for all types of users to carry out daily activities. Internet usage has been increasing throughout the years. Especially with the appearance of smartphones and the Internet of things, a lot of people and devices are becoming connected. This has introduced an overload on the network, which makes the study of Internet usage prediction important. Even though a lot of people can benefit from this research, the majority will be Internet service providers. Understanding network usage allows a service provider to identify different factors contributing to the traffic and therefore make decisions accordingly. If they found a drop in the usage for a particular country, they can plan according to that. The most populated countries like China and India will be a better chance to invest in them. In addition, efficient methods of resource management, such as bandwidth, can be used to gain performance, reduce costs, and improve the quality of service (QoS).

Numerous works have been implemented with LSTM and the SES models. Each work has been done in different areas which proves that these models apply to almost every situation. Those works are explained in the related work in the coming section. Siami-Namini et al. (2019) analyzed different forecasting models with LSTM models and proved that the LSTM model gave better results compared to the rest. This paper also suggests the Bi-directional Long-short term memory (Bi-LSTM) model surpasses the standard LSTM model with text parsing and word prediction in the input. However, this research is not making use of that.Smyl (2020) explains the SES model implementation as a better practice when it comes to time series data. But the problem with the research was the absence of time data. However, this problem will be solved here.

The data required for this research has been gathered from a public data source which has data over the years 1970 to 2019. But this work investigates global internet usage from 2010 till 2019. With those data in hand, this work will try to predict future internet usage. A Time series model has been used to predict internet traffic volumes. In this research, a Long short-term memory (LSTM) model, as well as the Simple Exponential Smoothing (SES) models, are used to perform forecasting. The popularity of the newest deep learning methods has been increasing in several areas, but there is a lack of studies concerning time series prediction related to global internet usage. Several works have been developed using recurrent neural networks (RNN) and they have shown that RNN is a competitive model. Network traffic is a time series, which is a sequence of data regularly measured at uniform time intervals. A time series can be a stochastic process or a deterministic one. To predict a time series is necessary to use mathematical models that truly represent the statistical characteristic of the sampled traffic. For adaptive applications that require real-time processing, the choice of the prediction method must consider the prediction horizon, computational cost, prediction error, and response time.

The research question is to what extent Long Short-term Memory model can outperform a Simple Exponential Smoothing model in forecasting internet usage? The main aim of this project is to predict future global internet usage with the time series models. After the analysis, the performance is evaluated using appropriate metrics. The research question raised in this research is as follows.

“To what extend the prediction can be done for global internet usage and which model or machine learning method works perfectly with time series data?”

The objectives withstand the research questions are as follows:

- Analysing the existing work to assist this research.
- Implementing the machine learning models such as LSTM and SES.
- Evaluating the models with appropriate metrics.

This section is followed by related work which explains the existing work in Section 2. The methodology and the design specification are explained in Section 3 and Section 4. The implementation is explained in Section 5 which explains the machine learning models. Section 6 explains the implementation and ends with a conclusion and future work in Section 7.

2 Related Work

Dong et al. (2013) forecasts the time series of high-resolution solar irradiance using the exponential smoothing state space (ESSS) model. compared to that of existing widely used trend models utilizing residual analysis and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) stationarity test. The simulation results says that this model has high performance compared with forecasting models. Initially ¡author name¿ did a comparison of the Fourier trend model and other trend models is presented. The forecasting equations for exponential smoothing are then developed. The exponential smoothing technique’s state

space models are then introduced to produce a 95 percent confidence interval. The procedure for selecting the best model for each step of the forecast is then presented. The comparative application is based on Colorado's solar irradiance time series data, whereas the first application is based on Singapore's solar irradiance time series data. This paper only looks at samples from 07:00 to 19:30 because it is useless to predict solar irradiation after sundown. For power system utilities, the 95% confidence range for the ESSS model's forecasting accuracy is an excellent forecasting interval of roughly 90%. The ESSS model has the broadest range of low forecasting errors when compared to other models' forecasting horizon plots, indicating that it performs the best when predicting up to 20 minutes in advance with less than 20 minutes worth of average data.
This research introduces a new framework for modeling innovation state space that is based on a trigonometric formulation and can handle all this seasonal complexity.De Livera et al. (2011) stated that traditional models used for exponential smoothing assume the error process value is not correlated. This cannot be the same all the time. The properties of the series that are not explicitly permitted in the state definition could be the cause of this auto-correlation. The call center data, for instance, may be affected by annual seasonal impacts, but due to the small sample size, this cannot be clearly modeled. To reduce the problems with the non-linear models, this paper would not consider the linear homoscedastic models which in return reduces the forecast only to positive time series data. The simplest straightforward modification of the conventional seasonal innovations models to accommodate several seasonal periods is the BATS model. The suggested modeling framework was used for three complex seasonal time series, and the results showed that the trigonometric models performed better outside of the sample and required significantly fewer estimates than conventional seasonal exponential smoothing methods.
This paper introduces a hybrid forecasting method which is a combination of both the exponential smoothing model as well as the LSTM model. The ES equations allow the approach to successfully capture the primary characteristics of the individual series, such as seasonality and level, while the LSTM networks provide cross-learning and non-linear trends. The hybrid model has 3 elements which are de-seasonalisation and adaptive normalization, generating the forecasts and ensembling. For the exponential smoothing model, the models of Holt and Holt and Winters with multiplicative seasonality were used. The data used for modeling do not contain any time data, so Smyl (2020) introduces a drastic method known as de-seasonalisation where the quality will likely to get worse at the end. The error criteria used in the model is a combination of MAPE and MASE. Back-testing was carried out by shortening each series' last horizon number of points—typically one, but occasionally two—and training the algorithm on a set of such truncated series. The number of epochs, learning rate, and training percentile is some parameters considered for hyperparameter tuning. The main difficulty faced by this approach is the lack of timestamps in the dataset.
The Holt-Winters smoothing method and robust variations of the exponential smoothing approach are described in this paper. However, the Holt-Winters forecasting performance significantly declines in the presence of outlier observations. This study presents a robust version of the Holt-Winters and exponential smoothing methods. Robustness regarding extreme values is achieved by de-weighting observations with surprisingly high or low one-step-ahead prediction errors. Furthermore, the simulation study demonstrated that when the error distribution is fat-tailed, the robust Holt-Winters technique also produces better forecasts. A little higher mean squared forecast error at a normal error distribu-

tion is the only tiny cost to be paid for this outlier protection. The recursive updating algorithms used in the robust exponential and Holt-Winters smoothing methods present the standard method on pre-cleaned data.Gelper et al. (2010) suggests a robust version of the Kalman filler to the state space method. The smoothing technique used in this work runs up to observation 100 before predicting observations 101 and 105 for a simulation of 1000 time series of length 105. The presented convincing Holt-Winters method produces precise and steady forecasts, is robust to outliers, and performs well with various kinds of data. As the underlying data generation process is unknown and outliers are frequent, this is crucial for realistic business forecasting.

The Levenberg-Marquardt (LM) algorithm and the hybrid exponential smoothing method are employed in this paper to propose a novel neural network training method that aims to enhance the generalization capabilities of previously employed methods for training neural networks for short-term traffic flow forecasting. By predicting short-term traffic flow conditions on the Mitchell motorway in Western Australia, the suggested strategy was assessed. Chan et al. (2011) proposes the hybrid exponential smoothing and Levenberg-Marquardt (LM) algorithm (EXP-LM), a straightforward but efficient method to train neural networks to achieve strong generalization capability in the anticipated short-term traffic flow. EXP-LM combines the LM algorithm and the exponential smoothing method's processes. Prior to using LM for training, the lumpiness in the traffic flow data has been removed using the exponential smoothing method in the suggested EXP-LM. The findings show that, on average, EXP-test LM mistakes are lower than those of the other examined algorithms. As a result, adopting EXP-LM generally results in neural networks with higher generalization capabilities for traffic flow predictions.

Bi et al. (2021) introduces a novel hybrid prediction method called SG and TCN-based ST-LSTM for network traffic prediction by using the concepts of Savitzky-Golay (SG) filter and temporal convolutional network (TCN). The results clearly point out that this method outruns every method in terms of accuracy. Most of the existing methods fail to predict the non-linear characteristics of the network. Although nonlinear models are more accurate than traditional prediction ones, they have a low ability to capture long-term dependence. In recent years, deep learning methods have become a mainstream approach for time series prediction. Recurrent neural networks (RNN) have shown their advantages in the task of sequence modeling. As its typical example, long short-term memory (LSTM) network can not only capture long-term dependence but also effectively solve the gradient disappearance issue in traditional RNN. On the other hand, by combining RNN and convolutional neural network (CNN) architectures, a recent temporal convolutional network (TCN) employs causal convolutions to possess very long effective history sizes and performs well in sequence modeling. While LSTM has been extensively used for time series prediction, in recent years, some research reveals that TCN may perform better under some circumstances. It is observed that TCN can effectively extract high- and low-frequency information from sequences, while LSTM is good at capturing long-term dependence. Experimental results say that the hybrid method showed better performance when compared with typical stand-alone methods.

For comparing the effective network traffic forecast, Oliveira et al. (2016) makes use of four different artificial neuron approaches. They are mainly, multi-layer perceptron (MLP) which uses back-propagation as a training algorithm, MLP with resilient back-propagation (Rprop), recurrent neural network (RNN), and the deep learning stacked auto-encoder (SAE). MLP makes use of supervised training whereas SAE uses a greedy algorithm for unsupervised learning. MLP is compared with both standard back-propagation and resilient back-propagation. This paper argues that deep learning methods are far better when it is compared with conventional algorithms. Conventional algorithms like back-propagation will not be able to perform better since a neural network has more than three hidden layers. The experiments conducted make use of the open-source codes of different libraries such as Deep Learn Toolbox and an Encog machine learning framework. The Encog framework was utilized for multi-layer perceptron with back-propagation (MLP-BP), multi-layer perceptron with resilient back-propagation (MLP-RP), and RNN prediction and training. The Deep Learn Toolbox for MATLAB was used to create the SAE. The SAE approach is more sophisticated than the others for time series prediction because it includes an additional unsupervised training phase that sets the neural network weights for the final stage of fine-tuning. The SAE was inferior even with the added complexity. Because of this, this method is not suggested for time series prediction. The results conclude that MLP and RNN are better than SAE. This work aims to implement the deep learning technique called DBN and the continuous restricted Boltzmann (CRBM) technique in the future.

Siami-Namini et al. (2019) analyses the Bi-directional LSTM (Bi-LSTM) and the LSTM models by checking the additional layers of the data training and compares whether the BiLSTM model predicts more accurately than that of a regular LSTM model. This work observed that the BiLSTM models offer better accuracy than the traditional ARIMA model as well as the regular LSTM model. It is proved that the Long short-term memory is far beyond when it is compared with the ARIMA model. The conventional time series methods utilize linear regression as well as the moving average models for the predictions. These models' performance cannot be trusted when subjected to long-term predictions. In a Bidirectional LSTM model, the input model is utilized twice for the training. After the analysis, the results say that the Bidirectional LSTM outperforms the normal LSTM model. For some types of data, such as text parsing and word prediction in the input sentence, the superior performance of BiLSTM over the standard unidirectional LSTM is understandable. It was unclear, nevertheless, if training numerical time series data twice and learning from both the present and the past would improve time series forecasting because some contexts might not exist, as seen in text parsing. When the loss value for both models was compared, the LSTM model achieves a loss value of 0.0256 whereas the BiLSTM shows a maximum of 0.0874 when the Epoch was set to 1. This paper concludes that the additional layer will improve accuracy of 37.78 percent more.

In this paper, a neural network model that combines deep neural networks with long short-term memory networks (LSTM) is proposed (DNN). To increase the prediction model's accuracy, the auto-correlation coefficient is incorporated. It can offer greater precision than other conventional models. Additionally, the neural network comprising LSTM and DNN has certain advantages in the accuracy of the big granularity data sets after considering the auto-correlation features. To demonstrate the effectiveness of the LSTM model and how accuracy improved when auto-correlation was considered, several experiments were conducted utilizing real-world data. Zhuo et al. (2017) offers a neural network model that may be used to merge LSTM and DNN to address the issue of network traffic with auto-correlation. The experimental findings demonstrate the effectiveness of LSTM as a timing sequence forecast model using a real data set amassed both domestically and internationally. Prediction of network traffic with high accuracy offers some assistance in addressing potential network congestion, anomalous attacks, etc. One of the many RNN variants is the LSTM. Numerous RNN models are both well-liked and well-represented in numerous problems, including GRU. The next step is to investigate

various RNN structures and combine network traffic data's properties to see if there is a way to increase prediction accuracy even more.
The prediction of multivariate time series data has numerous applications, but the complexity and the non-linearity among the inter-dependencies complicate this idea. Shih et al. (2019) proposes a set of filters to extract time-invariant temporal patterns. The dataset used in this paper has been retrieved from four different sources. There are both linear and non-linear inter-dependencies in the data-sets, which are drawn from the actual world. Furthermore, there are substantial periodic patterns in the data-sets for solar energy, traffic, and electricity that point to daily or weekly human activity. The methods used in this work are the auto-regression model, VAR model with L-2 regularization, LSVR, Gaussian process model, self-exciting auto-regression model, LSTNet-Skip, and LSTNet-Attn models. AR, LRidge, LSVR, GP, and SETAR are traditional methods, whereas LSTNet-Skip and LSTNet-Attn are new methods. The argument that the traditional attention mechanism does not perform well on these tasks is first supported by the results, which show that the proposed model and LSTM perform better under conditions with similar hyper-parameters and trainable weights. Additionally, the suggested model performs better in terms of precision, recall, and F1 score and learns more effectively than LSTM throughout the learning phase.

2.1 Summary

The literature review was done on 10 papers. Among them, 5 were done on a paper describing LSTM models and the remaining 5 were done on SES. While reviewing the papers, every paper has done a comparison with all other forecasting models or techniques, and the LSTM and ES models provided better accurate results. Most of the papers have their accuracy prediction done using respective error predictors such as RMSE, and MAPE values.Siami-Namini et al. (2019) has a unique way mentioned among the papers which analyze the BiLSTM and the LSTM models by checking the additional layers of the data training and compares whether the BiLSTM model predicts more accurately than that of a regular LSTM model. This paper confirms that the BiLSTM provided better results. Whereas Smyl (2020) explains the limitation they have faced during the forecasting. The main problem was not having enough or required data for prediction.

3 Methodology

This section discusses the methodology followed in the research. This includes data acquisition, data pre-processing, dealing with missing data, and exploratory data analysis. When it comes to time series analysis, the Long short-term memory model (LSTM) and the Exponential smoothing (SES) methods are ones that provide accurate results. This research follows the KDD methodology, and it is explained with the figure below.

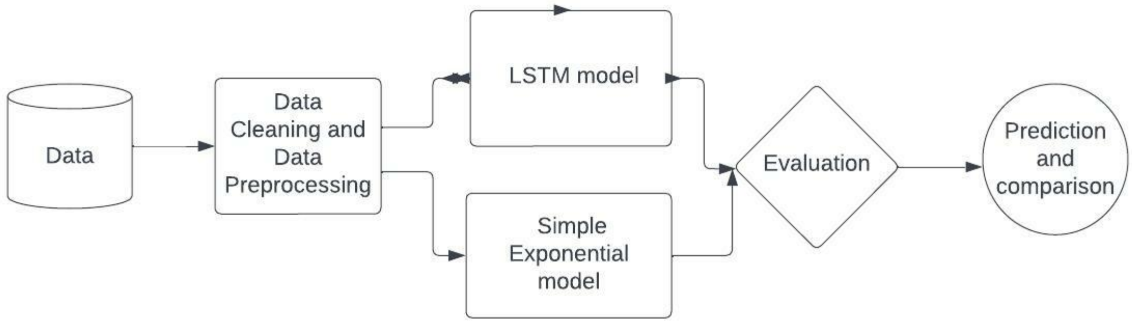

Figure 1: Research Methodology

3.1 Data Acquisition

In this research, the required data has been gathered from a single source known as the Kaggle data-sets. This data source is a combination of 4 data-sets, but this research will make use of the two among them. This dataset includes global internet usage data over the years 1970 to 2019. The columns in the dataset are country name, country code, year, mobile cellular subscriptions, and the individuals using the internet.

- Mobile cellular subscriptions

The mobile cellular subscription value is taken among 100 people in a particular country in a particular year. This has a total value of 2395, and this includes 253 unique countries. The value starts from the year 2010 till 2019.

- Individuals using the internet

The number of individuals using the internet is a percentage value where it explains the percentage of individuals using the internet over the years. This has a total value of 2395, and this also includes the data of 253 unique countries from 2010 to 2019.
All the above dataset is fetched into a Jupyter environment using pandas and the analysis is done in Python.

3.2 Data Pre-processing

This section explains the data pre-processing steps followed in this research after the data acquisition.

3.2.1 Dealing with missing data

The original dataset contains internet usage data from 1970 till 2019. This research is trying to predict the near future based on past observations. Because of that this research only considered values starting from the year 2010 till the year 2019. This cut down the original dataset of over 11000 values into 2395 values which are very recent. The initial dataset has a lot of missing values, and the new dataset has no null values present. This is confirmed with isnull.sum () function.

3.2.2 Feature Engineering

• The original dataset contains only yearly data, and no month was specified. To perform various operations on the dataset, this was not enough. Because of this, this research makes use of the month as well. the month was added to the original dataset using excel operations.
• The original dataset has values ranging over the years 1970 to 2019. This research is trying to predict the near future usage of the internet where also the original dataset has a lot of null values. This problem was cleared after cutting down the dataset into values ranging from the year 2010 to 2019.
• This feature extraction is made possible using Excel operation. The year was selected separated using the control key and removed all those entries falling out of 2010 to 2019 range.

3.2.3 Feature Scaling

The combined dataset has different features each of them measured on a different scale. This change may affect the performance when it is subjected to machine learning models. To normalize this diversity among them, this research makes use of the min-max scaling which transforms all features in a scale ranging from 0 to 1. This is made possible through the MinMaxScaler package, it is imported from the Sklearn library. This can be explained with the below formula.

$$\text{xnew} = ((\text{x} \quad \text{xmin})/(\text{xmax} \quad \text{xmin})) \quad (\text{max} \quad \text{min}) + \text{min}$$

3.2.4 Exploratory Data analysis

This section includes some initial investigations carried out on the dataset before the machine learning models were applied. Visualization techniques such as matplotlib as well as Tableau were used for this. This helps to analyze the internet usage patterns and the anomaly present in the dataset.

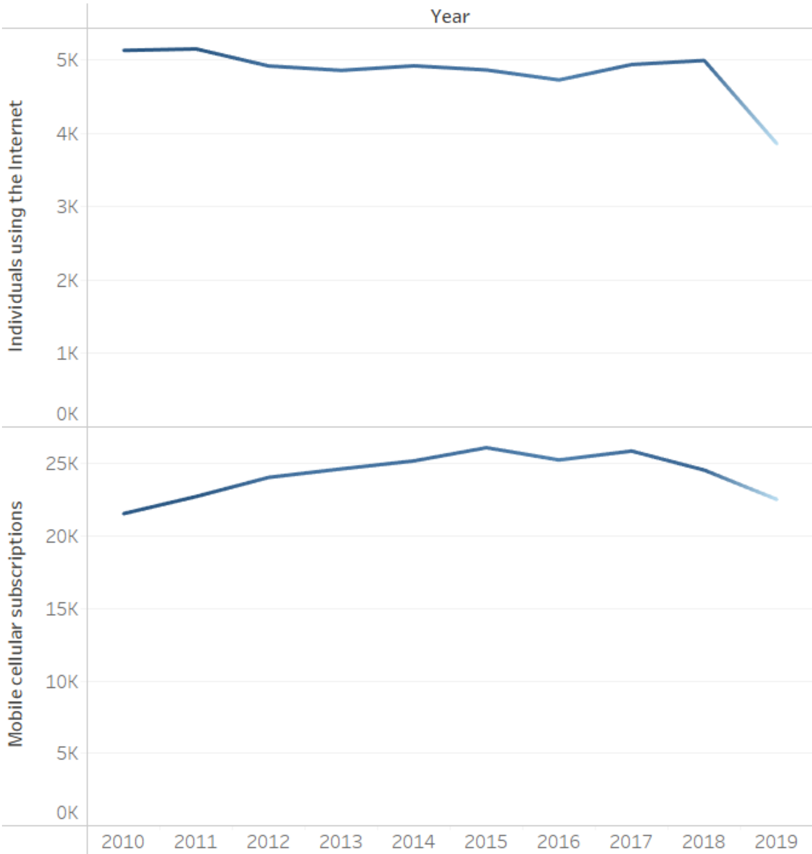

Figure 2: Database plotting

Year of Year	Individuals using the Internet	Mobile cellular subscriptions
2010	5,125.87	21,565.80
2011	5,145.61	22,747.06
2012	4,913.57	24,069.78
2013	4,853.74	24,657.07
2014	4,915.87	25,209.56
2015	4,859.41	26,122.04
2016	4,723.60	25,274.56
2017	4,933.49	25,890.91
2018	4,988.96	24,573.83
2019	3,855.95	22,548.53

Figure 3: Total values in Dataset

4 Design Specification

The overall architecture of this research can be formulated into 3 layers. They are the database layer, the application layer, and the presentation layer. This is represented using an architecture diagram below.

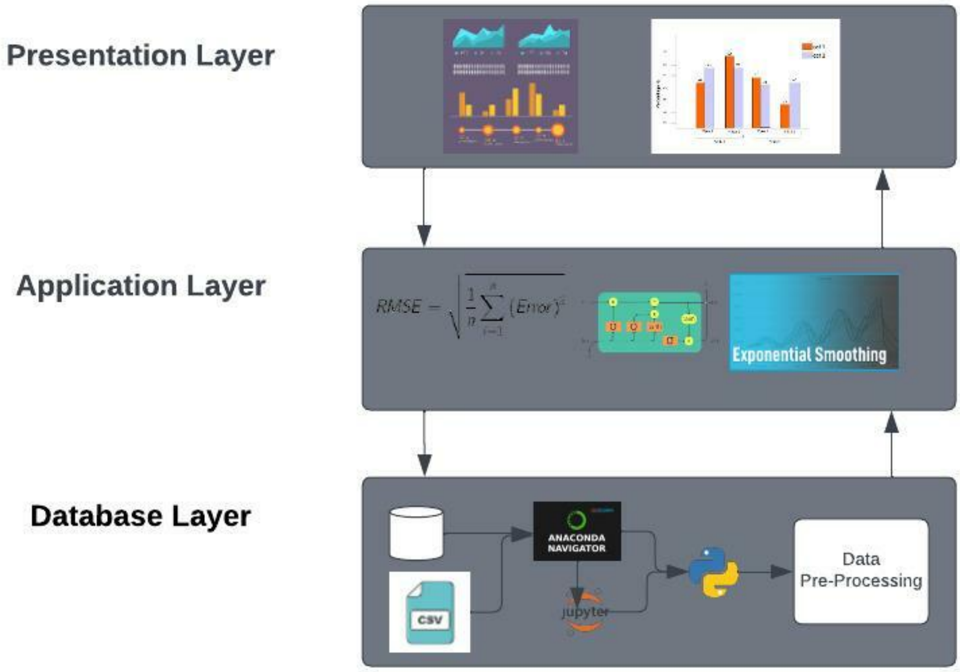

Figure 4: Design Specification

•Database layer
The data required for this research is gathered in this layer. The entire data is obtained from Kaggle datasets, and it is fetched into a Jupyter environment.
• Application layer
The machine learning models are performed on the fetched data in this layer. This research has time series data and the two machine models which are suitable for the time series data prediction are implemented. The two machine learning models used in this research are Long Short-term Memory (LSTM) model and the Simple Exponential Smoothing (SES) model.
• Presentation layer
The results obtained from the machine learning models are evaluated in this layer and the best-performing model is analyzed based on the accuracy criteria. This forecasting will help the service providers to plan according to the data and improve their market.

5 Implementation

5.1 Data Preparation

After finishing the data pre-processing and the feature engineering, all the null values are cleared, and the data is ready for modeling. The test and training split of the actual data depends upon the size of the dataset. After the pre-processing and the excel operations, the original dataset of 11000 values has been cut down to 2395 rows. Within this data, the data will be split for testing and training.
• The test and train ratio for the LSTM model is 67 and 33. That is among the 2395

rows, 67 percent of values 1604 will be used for model training whereas the remaining 33 percent which is 791 values are used for testing the model.
• Simple exponential model focus on the weighted average than the moving average for prediction. The model takes 2365 values for training the model. But for the testing of the model, it takes very recent values of size 30.

5.2 Machine Learning models

5.2.1 Long Short-Term Memory Model (LSTM)

The Long Short-Term Memory network, also known as the LSTM network, is a recurrent neural network that fixes the vanishing gradient issue. It was trained using back-propagation through time. As a result, it can be used to build substantial recurrent networks, which can then be utilized to tackle challenging sequence issues in machine learning and produce cutting-edge outcomes. Memory blocks in LSTM networks are connected by layers as opposed to neurons. A block is equipped with elements that make it more intelligent than a traditional neuron and a memory for recent sequences. Gates are components found in blocks that control their state and output. Each gate inside a block uses the sigmoid activation units to regulate whether it is triggered or not, making the change of state and addition of information flowing through the block conditional. A block functions upon an input sequence. There are 3 gates within a unit which are the Input gate, the output gate, and the forget gate.
• Input gate: This gate decides which values are used to update the memory state.
• Output gate: This gate decides the output based on the value from the input and memory of the block.
• Forget gate: This gate decides what information is not required and will be thrown away.
The LSTM model operation can also be termed a regression problem. Re-scaling the data, also known as normalizing, to the range of 0-to-1 can be a beneficial approach. After the scaling, the data will get split for training and testing. The network consists of an output layer that predicts a single value, a hidden layer with four LSTM blocks, and a visible layer with one input. For the LSTM blocks, the standard sigmoid activation function is employed. A batch size of 1 is utilized, and the network is trained over a period of 50 epochs.

5.2.2 Simple Exponential Smoothing Model (SES)

Exponential smoothing is a time series forecasting technique for uni variate data. The forecast of time series methods like the Box-Jenkins ARIMA family of methods is a weighted linear sum of recent past data or lags. Although the model explicitly utilizes an exponentially decreasing weight for past observations, exponential smoothing forecasting approaches are similar in that a prediction is a weighted sum of past observations. Past observations are specifically weighted using a geometrically diminishing ratio. The SES model is commonly used for data with no seasonality or trend. This includes only one parameter known as alpha(a).
• Alpha (a): Smoothing factor
The package from statsmodels.tsa.api import SimpleExpSmoothing is imported for using the simple exponential model. The fit settings, notably the smoothing level alpha value, are then passed to the fit () function. In the absence of this or if it is set to None, the

model will automatically choose the best value. To make a forecast, use the result object's forecast () or predict () functions. The SES model gives more importance to the weighted average rather than the moving average model. The model focus on very recent value for predicting the future. This research follows this idea and uses very recent observations for testing the model.

5.2.3 Hyperparameter Tuning

The initial LSTM model uses 3 layers LSTM layer whereas the dense layer uses 1 layer also the initial model was set up with 10 epochs only. The improve the model performance, the LSTM layer was changed to 2 and the dense layer was changed to 1. Also, the epoch value changed. No noticeable changes were observed after changing the parameters in the RMSE value. The tuning was done with the SES model also. Changes have been made to the smoothing value and the accuracy is measured on the RMSE value. Those who gave better performance have chosen for the prediction.

6 Evaluation

6.1 Root-Mean-Square Error (RMSE)

A measure that is widely used to evaluate the accuracy of the predictions made by a model is the Root-Mean-Square Error (RMSE). It calculates the discrepancies or residuals between the predicted and actual values. The statistic compares the prediction errors of several models for a specific collection of data, not across datasets. The accuracy is measured from the RMSE value.

$$\text{Accuracy}=1.96*\text{RMSE}$$

6.2 Hyperparameter Optimization

To increase the model accuracy this research has been done by increasing the epoch values, the dense and LSTM layer units. The LSTM model displayed a slight change in the RMSE after changing the epoch value. The smoothing value for the SES model was tuned for a better RMSE value. Changes have seen when this was done.

6.3 Experiment with LSTM

As mentioned in the research methodology, the models are applied on two terms they are mobile cellular subscriptions and individuals using the internet. Initially, the LSTM model is applied, and obtained RMSE value with both train and test datasets are shown in the below table. The graph obtained after the forecast of both terms is also shown below.

- Individuals using the internet

LSTM Model	RMSE value obtained
Train	16.98
Test	16.63

Table 1: RMSE values obtained for LSTM model on Individuals using the internet

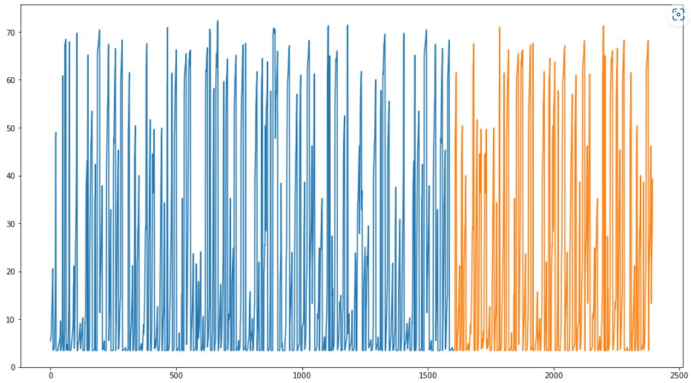

Figure 5: Prediction graph of Individuals using the Internet using LSTM

The following are the results obtained when the variable Individuals using the internet are subjected to the LSTM model. Reasonable accuracy was obtained. The graph with blue color is the actual dataset and the one in orange is the predicted graph. The RMSE value for this model is mentioned in the table below. The test got a score of 16.63 whereas the train got 16.98.

- Mobile Cellular Subscriptions

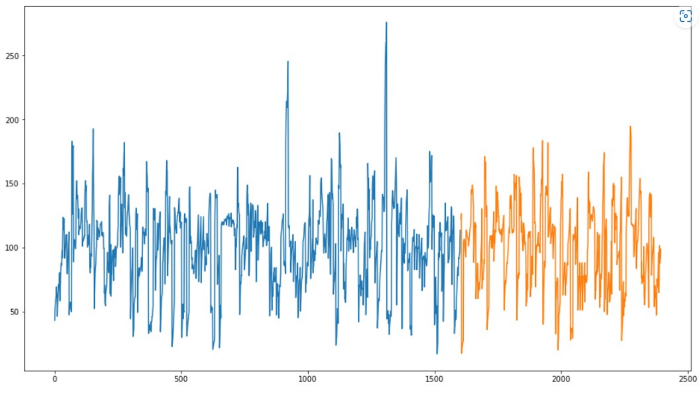

Figure 6: Prediction graph of Mobile Cellular Subscriptions using LSTM

These are the results obtained when the LSTM model was applied to the variable Mobile cellular subscriptions. The orange graph shows the predicted one and the blue is the actual dataset. This model has got an RMSE value of 20.78 for the train and 21.30 for the test dataset. Various measures were applied to the model to improve the accuracy which did not bring any major change in the RMSE value.

LSTM Model	RMSE value obtained
Train	20.78
Test	21.30

Table 2: RMSE values obtained for LSTM model on mobile cellular subscriptions

SES Model	RMSE value obtained
Train	36.44
Test	37.76

Table 3: RMSE values obtained for SES model on individuals using internet

6.4 Experiment with SES

The same data was predicted using the Simple Exponential Smoothing (SES) model. This model only considers the weighted averages which are the immediate past data for better prediction. The plots and the RMSE values obtained after the prediction is shown below.

• Individuals using the internet

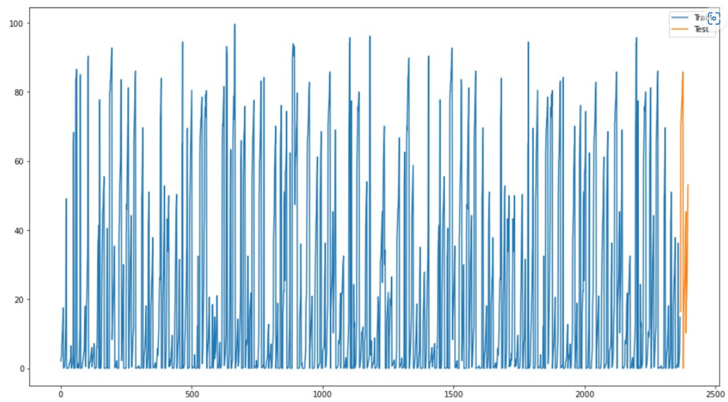

Figure 7: Prediction graph of Individuals using the Internet using SES

The prediction graph obtained after the analysis is given below. The one in orange represents the prediction graph. The model considers 30 past observations for the testing model. Tuning has been done to improve the model performance. The smoothing value with better performance has been chosen.

• Mobile cellular subscriptions

SES Model	RMSE value obtained
Train	38.55
Test	37.05

Table 4: RMSE values obtained for SES model on mobile cellular subscriptions

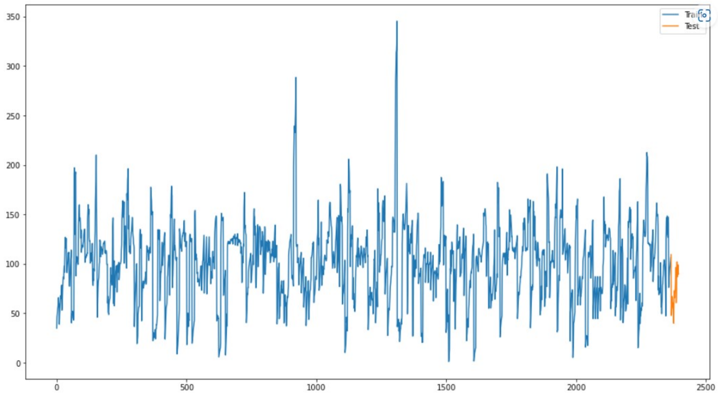

Figure 8: Prediction graph of Mobile Cellular Subscriptions using SES

The graph obtained after the prediction of mobile cellular subscription has given below. The RMSE and the R Square values obtained here are comparatively high. Tuning was done on the smoothing value and the one which gave better results has taken.

6.5 Discussion

The prediction of global internet usage and thereby comparing the machine learning model performance was not mentioned in the related works. The LSTM model as well as the SES model have been used for different areas for a long time especially when it comes to time series data. This research uses the LSTM and SES models for forecasting global internet usage. This is made possible through forecasting two variables they are individuals using the internet and mobile cellular subscriptions. The accuracy of both models is measured using the RMSE value. As mentioned before, the accuracy can be calculated from the RMSE value. The LSTM model on the variable individuals using the internet gave an accuracy of 33% on the train data and it was 32.8% on the test data whereas the mobile cellular subscriptions gave an accuracy of 41% on the test data and the accuracy of train data 40.67%. The accuracy was better for the SES model when compared with the LSTM. The SES model gave an accuracy of 71% on train accuracy and 75% on test accuracy for the variable individuals using the internet. Whereas the accuracy was 75.5% for train and 72% on the test for the mobile cellular subscriptions. When it comes to the forecasting graph, the values are more likely to fluctuate. Since the SES model takes the immediate data and considers the model with more accuracy the results can be obtained from that plot.
Overall, this research has two changes to the history they are the SES model outperforms the LSTM model when the internet usage data was predicted. The obtained prediction can also be used by internet service providers to think about their future. Since the world

advancing each second, a prediction like this also makes people think about their internet usage.

7 Conclusion and Future Work

This study was conducted on the global internet usage data to forecast the future usage rate. This forecast will help the internet service providers, as well as the individuals, think about their internet domination. Service providers can plan on their market ideas with a forecast like this. The machine learning models such as the Long Short-term Memory (LSTM) model and the Simple Exponential Smoothing (SES) model are used in this research. These two models perform better with time series data. One other aim of this research was to compare and conclude which model performs better with time series data. After the analysis, the results concluded that the SES has better accuracy compared with the LSTM model. The dataset required for this research was gathered from a public source known as the Kaggle dataset. This original dataset was split into different CSV files and has the data available over the years 1970 to 2019. This project has combined those into a single CSV file so that the data can be retrieved from a single source. Also, through the excel operations, the original data having more than 11000 values have been cut down to 2395 rows which has internet usage data over the years 2010 to 2019. Since the SES model gave a more accuracy percentage of 75%, this model can be discussed. The graph clearly shows that there is an upward trend in the individuals using the internet and the mobile cellular subscriptions showed a downward trend. The ratio of individuals using the internet will be more likely to increase in the future whereas mobile cellular subscriptions will be reduced. This would not sound perfect to the internet service providers. People will rely on internet assistance more in the future which the cellular companies cannot dominate. Their poor performance and the hidden charges can be one of the factors behind this. Each service provider's performance depends on its bandwidth and usage ratio. If this is the case, their market will be more likely to drop with more populated countries.

There were some difficulties experienced during this research. Initially, the problem was with the dataset. There is not enough data available in any sources which give to the day information about internet usage. This research proves that the SES model can perform with time series data and with more data the model accuracy will be higher. The simple exponential smoothing model is used on the data which shows no trend or seasonality. With more data in hand, this research would like to perform the double exponential smoothing as well as the triple exponential smoothing in future work. Both these models work perfectly with the linear trend and dampen the trend with seasonality considered.

8 Acknowledgement

The author would like to express sincere gratitude toward DR Hicham Rifai. His guidance and comments are the bedrock behind this research. His patience and motivation helped through the process of coding as well as report writing.

References

Bi, J., Zhang, X., Yuan, H., Zhang, J. and Zhou, M. (2021). A hybrid prediction method for realistic network traffic with temporal convolutional network and lstm, *IEEE Transactions on Automation Science and Engineering* .

Chan, K. Y., Dillon, T. S., Singh, J. and Chang, E. (2011). Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and levenberg–marquardt algorithm, *IEEE Transactions on Intelligent Transportation Systems* **13**(2): 644–654.

De Livera, A. M., Hyndman, R. J. and Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing, *Journal of the American statistical association* **106**(496): 1513–1527.

Dong, Z., Yang, D., Reindl, T. and Walsh, W. M. (2013). Short-term solar irradiance forecasting using exponential smoothing state space model, *Energy* **55**: 1104–1113.

Gelper, S., Fried, R. and Croux, C. (2010). Robust forecasting with exponential and holt–winters smoothing, *Journal of forecasting* **29**(3): 285–300.

Oliveira, T. P., Barbar, J. S. and Soares, A. S. (2016). Computer network traffic prediction: a comparison between traditional and deep learning neural networks, *International Journal of Big Data Intelligence* **3**(1): 28–37.

Shih, S.-Y., Sun, F.-K. and Lee, H.-y. (2019). Temporal pattern attention for multivariate time series forecasting, *Machine Learning* **108**(8): 1421–1441.

Siami-Namini, S., Tavakoli, N. and Namin, A. S. (2019). The performance of lstm and bilstm in forecasting time series, *2019 IEEE International Conference on Big Data (Big Data)*, IEEE, pp. 3285–3292.

Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting, *International Journal of Forecasting* **36**(1): 75–85.

Zhuo, Q., Li, Q., Yan, H. and Qi, Y. (2017). Long short-term memory neural network for network traffic prediction, *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, IEEE, pp. 1–6.