Automating Time Series Forecasting on Crime Data using RNN-LSTM

J Vimala Devi[1]

Department of CSE
Cambridge Institute of Technology
Bengaluru, India

Dr K S Kavitha[2]

Department of CSE
Dayananda Sagar College of Engineering
Bengaluru, India

Abstract—**Criminal activities, be it violent or non-violent are major threats to the safety and security of people. Frequent Crimes are the extreme hindrance to the sustainable development of a nation and thus need to be controlled. Often Police personnel seek the computational solution and tools to realize impending crimes and to perform crime analytics. The developed and developing countries experimenting their tryst with predictive policing in the recent times. With the advent of advanced machine and deep learning algorithms, Time series analysis and building a forecasting model on crime data sets has become feasible. Time series analysis is preferred on this data set as the crime events are recorded with respect to time as significant component. The objective of this paper is to mechanize and automate time series forecasting using a pure DL model. N-Beats Recurrent Neural Networks (RNN) are the proven ensemble models for time series forecasting. Herein, we had foreseen future trends with better accuracy by building a model using NBeats algorithm on Sacremento crime data set. This study applied detailed data pre-processing steps, presented an extensive set of visualizations and involved hyperparameter tuning. The current study has been compared with the other similar works and had been proved as a better forecasting model. This study varied from the other research studies in the data visualization with the enhanced accuracy.**

Keywords—*Time series analysis; deep learning; RNN; forecasting; crime data; predictive policing; machine learning*

I. INTRODUCTION

Time series analysis and forecasting[7][17] has always been crucial in the aspects of many research applications such as stock prediction, weather forecasting, supply chain management etc., So why not time series forecasting on crime data?

A time series [5] is a set of numerical values of the same entity taken at equally spaced intervals over time. A time series dataset can be collected yearly, monthly, quarterly and daily etc., any time series analysis can be explained with the help of three components such as

- Trend – Overall long-time direction of series (May be upward or downtrend trend).

- Seasonality – Repeated behaviour at fixed intervals of time.

- Cycles – Occurs when it follows up or down pattern that is not seasonal and can be of varying length.

Conventionally time series analysis has been implemented using linear methods such as AR models, ETS etc., and these methods are simple and effective in implementation for smaller datasets. Machine learning [8] and Deep learning algorithms on the other hand are able to learn the temporal dependencies among the features and do forecasting with more accuracies. Also, deep learning algorithms automatically learn features and build model whereas manual feature extraction is required. The major challenge of this research is to handle the growing volumes of crime data and to build a predictive model with improved accuracy.

This paper is intended to build a better performing forecasting model [4][6] on the crime data. The objective of the current study is to gage the forecasting capacity of the NBeats model [1] on crime data, which is a hybrid of RNN-LSTM [4]. This model will aid police personnel in optimal decision-making and resource management. This work has been compared with the previous works done in this domain and the results are tabulated.

The rest of the paper is organized into the following segments. Section II introduces the existing methodologies. Section III deliberates the proposed approach with flowchart to build a model, discusses the techniques to prepare the dataset suitable for time series analysis and also presents a wide array of data visualization. In Section IV, the outcomes of the forecasting model and measures to calculate its error percentage are presented.

II. EXISTING METHODOLOGIES

Below is the detailed discussion of the existing methodologies used for forecasting. There are additive models. auto regression models[14][16], machine learning[8][15] and deep learning models[6][12] that are used to foretell about the trend or pattern of the crimes.

A. Exponential Smoothing (ETS Models)

These ETS models[10] use weighted average of past observations. The components of the model are error, trend and seasonality. Each component can be applied either additively or multiplicatively. Additive methods are useful when the trend and seasonal variations remain constant over time whereas multiplicative methods are applicable when trend and seasonality decrease or increases in magnitude over time.

There are four ETS models:

- Simple exponential smoothing method - In simple exponential method, forecast has been measured as follows. Forecast = $weight_t y_t + weight_{t-1} y_{t-1} + weigh_{t-2} y_{t-2} + \dots (1-\alpha)_n y_n$, where t is the number of periods before the most recent period and y_t is the target value of time series in 't' and 'α'is the smoothing parameter.

- Holt's linear trend method - The simple exponential method was expanded to include forecasting data with a trend known as double exponential method or Holt's linear trend methods. This method builds off simple exponential smoothing method not only the level but also trend in its calculation. Trend in this method is always applied in a linear or additive fashion and this method is great and suitable for non-seasonal data.

- Exponential trend method - A variation of Holt's linear trend method is the exponential trend method. It uses the same components (Level and trend) but they are applied multiplicatively. This method is great for non-seasonal time series analysis.

- Holt-winters seasonal method - This method models all the three components such as level, trend and seasonality of time series analysis. It can be either implemented as additive or multiplicative model. The additive method is used in which seasonal fluctuation does not change in time whereas in the multiplicative method, seasonal fluctuation changes in time.

B. ARIMA Model

ARIMA[11] stands for auto regressive integrated moving average. It is of two types: Seasonal and Non-seasonal. Non seasonal models is built on three components: AR(p), I(d) and MA(q). (p,d,q) represents the amount of time periods to lag for in ARIMA calculation.

- p - refers to the previous periods or the number of lag observations.

- d – refers to differencing term and the number of transformations used in the process of transforming a series into non stationery one.

- q – refers to moving average.

C. Recurrent Neural Networks(RNN)

Neural networks(NN)[6] in general are simple and great for classification problems by assigning labels and for regression problems in which a continuous value can be predicted. The disadvantage with this kind of NN is that its tendency to forget what it learned or what happened in the past. When it comes to sequential data or ordered data in which data points are interdependent, these NN are a greater disadvantage. Here RNN[4]come into picture which could be imagined as they have a sense of memory to remember what happened in the past. Hence RNN is best suited for time series analysis.

III. PROPOSED APPROACH

This section introduces dataset and its attributes, data pre-processing steps, smoothing and normalization techniques, proposed architecture, an algorithm to build a model and evaluation of a model using error accuracy measures such as MAE and Smape.

A. Dataset

The data used in this paper are real time dataset which was collected from the Sacramento police Open Data portal https://data.cityofsacramento.org/search (2014-2021). The dataset contains attributes such as FID, RecordId, Offense Code, Offense _Extension, Offense Category, Description,Police District, Beat,Grid, occurrence Timestamp. Each instance of the dataset is a crime record with date and timestamp. There are totally 66 unique crime categories. The data set is a collation of the past seven years data and contains a total of 2,72,333 records approximately.

The geographical locations[9] of city of Sacremento have been divided into six districts. Each district is divided into beats and there are a total of 21 prominent beats, beats further split in to grids for better patrolling and surveillance. The dataset reports 66 unique categories of offenses such as trespass, weapon offense, petty theft, burglary, DUI alcohol, owning or possessing ammunition, conspired crime, vehicle theft, false personation etc.,

B. Proposed Architecture

In Fig. 1, a flow chart is depicted that explains the proposed methodology step by step.
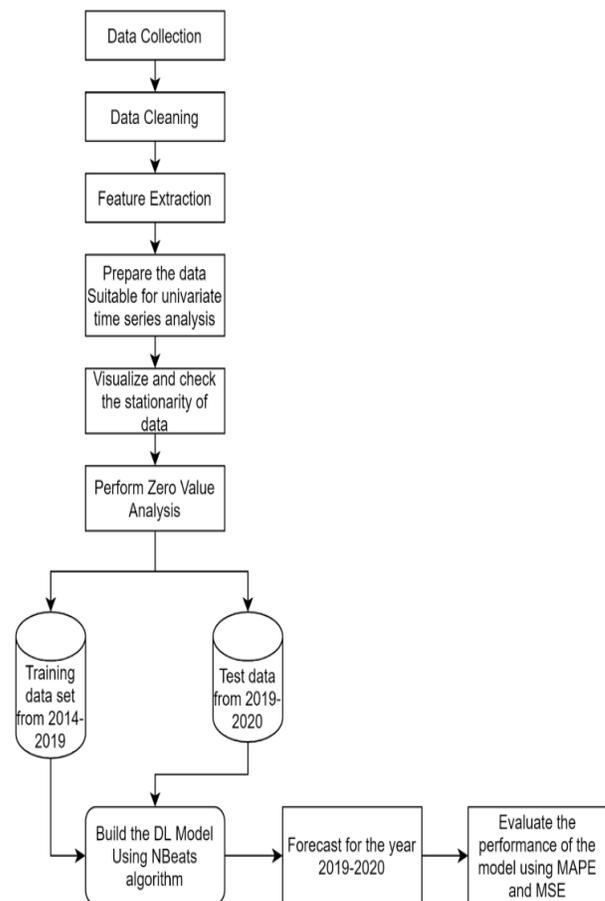
Fig. 1. Proposed Flow of this Work.

It begins with data collection, cleaning and extracting the appropriate features that suit to the requirement of time series analysis. Later, the dataset was tested to realize whether it is stationery or not. Smoothing and normalization techniques had been applied to understand their necessity. Then the dataset was divided into testing and training sets. The model had been trained on 2014 – 2020 data and was tested and forecasted on for the next year 2020-2021 data. The accuracy of the model was measured using MAE and sMAPE.

C. Data Preparation

The real time dataset has to be modified to suit Univariate time series analysis[13][17]. The dataset is cleaned by removing duplicate entries and the rows that contain null values. Less than 1% of rows are found dirty and removed. The instances in the dataset are grouped by Beat and the crime count is calculated for each day beat wise as given in Fig. 2. The zeroth row displays two crimes recorded on 01st January 2014, ten crimes recorded on 02nd January 2014 and so on in the beat '1A'.

For better performance of the model and to achieve uniformity in timeline across the Beats, the period from 2014-12-31 till 2020-12-31 have been considered and the rest of the days are ignored. To do forecasting, measurement of data should be sequential and equal with utmost one data point. For each day within the stipulated time period, the number of crimes is counted.

D. Checking the Stationerity of Series

To check the stationery of time series, a histogram had been plotted for the 500 days from 1st of January 2014 against crime count. Since the distribution of data across the plot did not follow Gaussian distribution and it looked like the distribution is squashed as given in Fig. 3, it may be concluded that the mean and variance is not the same thereby, the given time series data is not stationery dataset.

E. Zero Value Analysis

Zero value analysis is one of the smoothing techniques. With respect to our time series forecasting, zero value analysis is finding the count of days within the given period for each Beat whose crime count is nil for the day. The sample result is given in Fig. 4.

	Beat	Occurence_Date	Record_ID
0	1A	2014-01-01	2
1	1A	2014-01-02	10
2	1A	2014-01-03	8
3	1A	2014-01-04	6
4	1A	2014-01-05	11
...	...	...	...
57625	UI	2021-02-10	2
57626	UI	2021-02-11	2

Fig. 2. Snapshot that Shows Count of Crimes Day-wise for Every Beat.
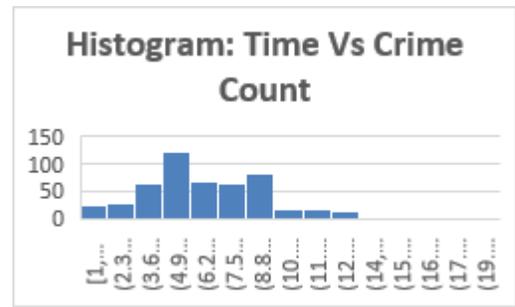
Fig. 3. A Plot to Understand the Distribution of Crime Count.

```
for the beat : ( 1A ) of 1day intervel with zero crimes:  5
-------------------------------------------------------------
for the beat : ( 1B ) of 1day intervel with zero crimes:  42
-------------------------------------------------------------
for the beat : ( 1C ) of 1day intervel with zero crimes:  77
-------------------------------------------------------------
for the beat : ( 2A ) of 1day intervel with zero crimes:  5
-------------------------------------------------------------
for the beat : ( 2B ) of 1day intervel with zero crimes:  6
-------------------------------------------------------------
for the beat : ( 2C ) of 1day intervel with zero crimes:  1
-------------------------------------------------------------
for the beat : ( 3A ) of 1day intervel with zero crimes:  12
-------------------------------------------------------------
for the beat : ( 3B ) of 1day intervel with zero crimes:  2
-------------------------------------------------------------
for the beat : ( 3M ) of 1day intervel with zero crimes:  26
-------------------------------------------------------------
for the beat : ( 4A ) of 1day intervel with zero crimes:  17
-------------------------------------------------------------
```

Fig. 4. Snapshot Displays Count of Days with Zero Crimes, Beatwise.

For a total of 2,192 days, a time series analysis[13][17] graph with the number of days on the X axis and the count of occurrences of crime on the Y-axis has been plotted for every Beat. Below are the snapshots starting from Fig. 5 till Fig. 13.

For each Beat, a time series graph has been plotted with the incrementing number of days on x-axis and count of crimes on y-axis. From the graphs, it is obvious that the given data set is not a stationary dataset as it exhibits strong seasonality and also there is no obvious upward or downward trend.
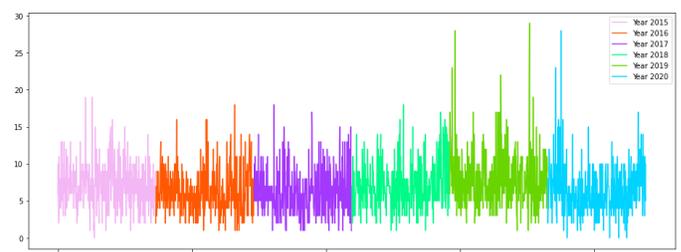
Fig. 5. Time Series Plot for Beat 1A.
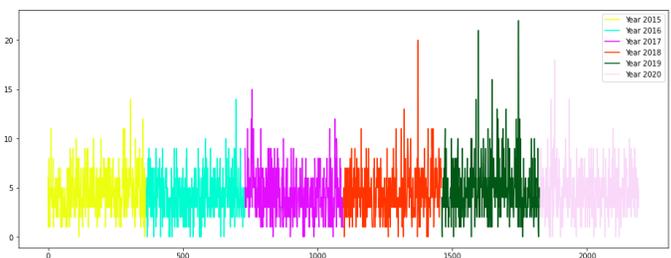
Fig. 6. Time Series Plot for Beat 1B.
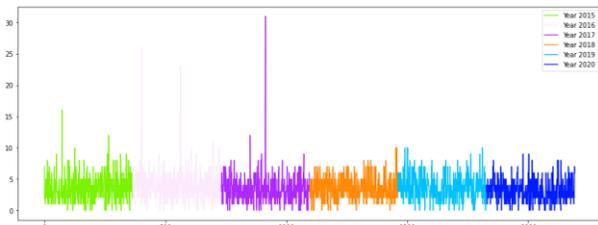
Fig. 7.   Time Series Plot for Beat 1C.
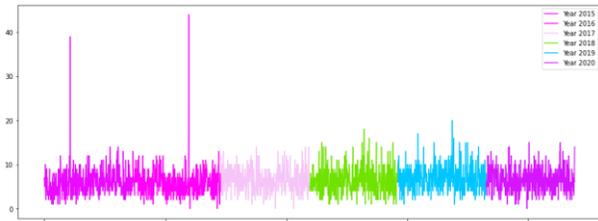
Fig. 8.   Time Series Plot for Beat 2A.
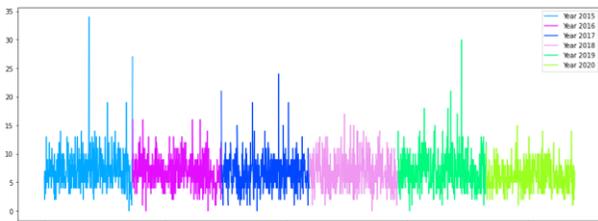
Fig. 9.   Time Series Plot for Beat 2B.
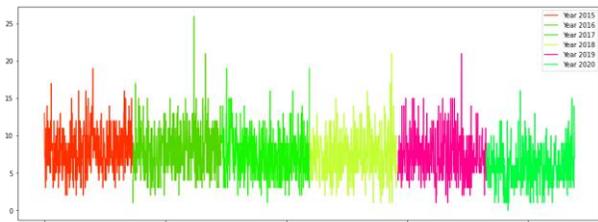
Fig. 10. Time Series Plot for Beat 2C.
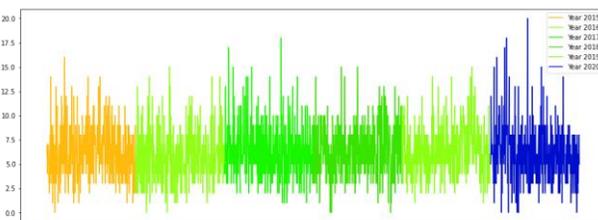
Fig. 11. Time Series Plot for Beat 3A.
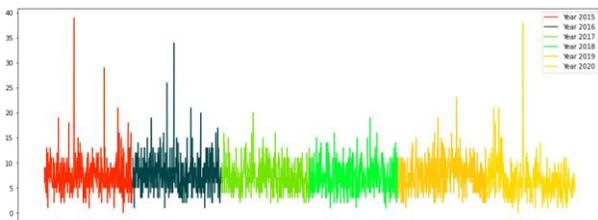
Fig. 12. Time Series Plot for Beat 3B.
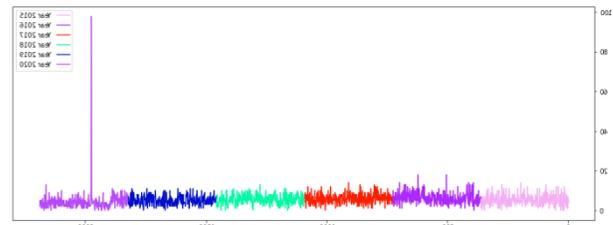
Fig. 13. Time Series Plot for Beat 3M.

Some Beats (geographical locations) recorded the highest number of crimes whereas some Beats such as Beat 3M as in Fig. 13 showed a smaller number of crimes reporting on a daily basis. Most of the Beats exhibited repeating patterns at fixed intervals of time is said to have seasonality. In most of the plots, there is a display for the presence of noise with extraordinary spikes in the growth of crimes.

A normalization technique has been applied to the feature 'count of crimes'. The count of crimes for every day has been normalized in the range of 0 to 1. The rationale behind this normalization is that in general, deep Learning algorithms that suffer from vanishing and exploding gradient descent problems. In order to overcome these problems, data normalizing have been applied.

F. Algorithm

N-Beats – Neural basis expansion analysis for interpretable time series forecasting [1][2] is a block based deep neural architecture suitable for Univariate time series analysis. It is chosen to build a pure deep learning model for forecasting based on time series that can take non stationery data with long term trends and seasonality and excels the accuracy of existing models such as ETS, ARIMA and Holt Winters, etc.

NBeats[1][2] is a hybrid model of RNN and LSTM that takes an entire window of past values and computes many forecast time point values in a single pass. For doing so, the architecture uses fully connected layers containing several blocks connected in a residual way. The first block models the past data(backcast) and predicts the future, then the second block models only the residual error from the previous block and improves the forecast values based on this error and continues to repeat.

Hence it is a residual architecture wherein multiple blocks are stacked together to avoid the risk of gradient vanishing which is common in deep learning algorithms and also has the advantages of ensembling technique. Hence the forecast value is the sum of predictions of several blocks and keeps improving based on residual errors calculated in the other blocks.

Advantages of using N-BEATS RNN over several traditional approaches:

- As all operations are parallelized, it supports quicker training of networks.

- Stacked blocks are much configurable thereby light weight networks.

- Fully Configurable Backcast and Forecast.

This model in Fig. 14 takes time series data upto Y_t, where y = datapoints upto time 't' as its input and predict future Y_{t+l} where l = length of forecast window. Let us consider t = 90 days and l = 30 days. The size of the input is always x*H where x refers to features and x = 1, also known as Lookback Period. During this period, our time series model learns the behavior in the past and tries to predict the behavior of H data points which is known as Forecast period. In our case, it is for one day.
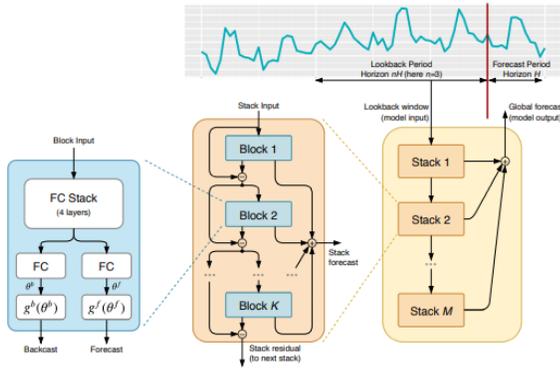
Fig. 14. NBeats Deep Layered Architecture [1].

NBeats architecture takes time series data for the lookback period of 90 days as input to stack 1, and each stack in turn is made up of multiple blocks and it is necessary to understand the structure of basic block as given below.

1) Basic blocks: A lookback period of 90 days is given as input to the stack that passes through every block and a forecast period is set for the next 30 days. The input passes through a set of 4 connected layered(FC + Relu) stack and then divided into two outputs. Each output is further passed through FC and finally we receive two outputs such as

- 90-dimension vector as backcast , (X).

- 30-dimension vector as forecast., (FC).

2) Stacking of blocks: Each stack consists of multiple basic blocks, arranged in a double residual manner. For each block,

- A vector $X(x_1, x_2, x_3, \ldots x_{90})$ is given as input to the first block and every block.

- The first block gives two vectors as outputs ie. BC_1(Backcast of block 1) and FC_1

- The input of 2nd, 3rd , and so on blocks is calculated as (X-BC_1).

- The backcast output (BC_n) of nth block is the final stack output.

- The forecast output of stack is calculated as $\sum_{k=1}^{n} FC\_k$ where n is the number of blocks in a stack.

IV. RESULT AND DISCUSSION

Any deep learning neural network is trained using stochastic gradient descent algorithm. The purpose of this algorithm is to optimize the model by updating the weights during training based on error gradient measures. The rate at which the weights are getting updated during training of a model are referred to as learning rate. A plot for tuning the hyperparameter 'Learning rate' is given in Fig. 15.

Training of any neural network involves the challenge of identifying the hyperparameter 'learning rate'. Estimating the optimal training rate is crucial to train a neural network, since learning rate describes how quickly a model is adapted to the given dataset and problem. Learning rate is a highly tunable parameter that calculates the number of weights needed to be updated so that the loss is reduced each time. The calculated learning rate is 0.000562341325190349.

The error measure of this N-Beats baseline model has been measured using mean and SMape(Symmetric Mean absolute percentage error) [3] and displayed in Table I. There are many error metrics such as MAE, MAPE, sMAPE, etc. available.

Mean absolute Error (MAE) is defined as the average of forecast errors e_t, where $e_t = F_t - A_t$.

$$MAE = 1/n\sum_{t=1}^{v} |et| \tag{1}$$

Symmetric Mean absolute Percentage error(sMAPE) is chosen over MAPE for the reason that MAPE behaves extremely undefined for actual zero values. The prediction or the future values are calculated based on the recent known values. But the resulting SMAPE measure is disappointing and the error percentage is 0.5219. Hence the optimal learning rate to train the RNN is calculated as 10-4 approximately. After training the model, the error in the model has been reduced and calculated as 0.2922. The smaller the sMAPE value, the greater is the prediction accuracy.

$$sMAPE = 1/v \sum_{i=1}^{n} |(Fi - Ai)|/(Fi + Ai)/2 \tag{2}$$

A. Comparison with Previous Works

The novelty about our work is the in-depth statistical analysis on the data set and converting it into univariate time series data set suiting to model building, thereby pruning the other dimensions. Here is the tabulation of results of this paper with recently published works given in Table II.

Different data sets have been compared in the above table that are using RNN model and their performance measures have been tabulated.
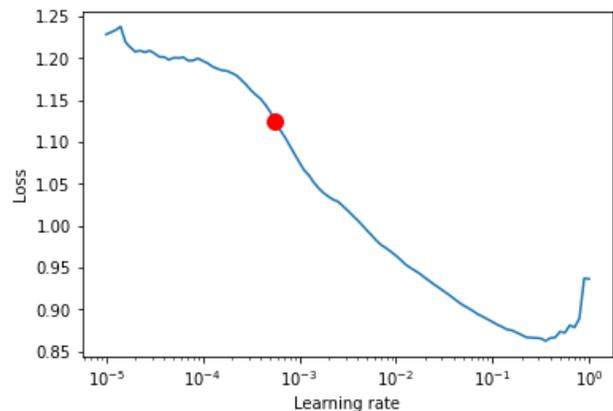
Fig. 15. Plot for Hyperparameter Tuning.

TABLE I.   ERROR ACCURACY MEASURES

Error Measures	Before training	After training with optimal learning rate	
		SMape	*MAE*
Validation set	0.4314	-	0.1932
Test set	0.5219	0.3922	0.1407

TABLE II.   COMPARITIVE ANALYSIS

References	Model	Accuracy Measures
HaseebTariq, Muhammad Kashif Hani, Muhammad Umer Sarwar, Sabeen Bari, Muhammad Shahzad Sarfraz, and Rozita Jamili Oskouei [4]	RNN-LSTM model	MAPE = 13.42
Wajiha Safat, Sohail Asghar, (Member, IEEE) and Saira Andleeb Gillani[6]	RNN	RNN-LSTM model for Chicago dataset, MAE = 11.6 RNN-LSTM model for Los Angeles dataset, MAE = 6.0
Current study	RNN-LSTM Model	MAE = 0.1407

V. CONCLUSION

It is understood that RNN-LSTM model implemented on our univariate data set displayed better performance based on MAE measure(refer in Table II) than the other works in time series forecasting. In general, error accuracy measures vary depending on the context, size of training set and the number of features. In this table, we have used MAE as performance metric. The innovation of this work lies in the systematic way of statistical analysis thereby extracting and understanding the deep insights of the data and converting the dataset suitable for univariate analysis before building a model on the data compared to the other works.

VI. FUTURE ENHANCEMENTS

In future, a hybrid model could be a better choice for building a forecasting model on this crime data and the accuracy of the mentioned model may be improved by exploring and tuning the other hyperparameters as well. This paper considered the features such as date and time and the number of occurrences of crime per day. In future, the attributes with respect to location that is geospatial coordinates can also be considered to build and improvise the model.

REFERENCES

[1] Attilio Sbrana, André Luis Debiaso Rossi, Murilo Coelho Naldi, "N-BEATS-RNN: deep learning for time series forecasting," 19th IEEE International Conference on Machine Learning and Applications (ICMLA), , 978-1-7281-8470-8/20/$31.00 ©2020 IEEE.

[2] Boris N.Oreshkin, Dmitri Carpov, Nicolahpados, Yoshua Bengio, "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting," [1905.10437] N-BEATS: Neural basis expansion analysis for interpretable time series forecasting (arxiv.org).

[3] Choosing the correct error metric: MAPE vs. sMAPE | by Eryk Lewinson | Towards Data Science.

[4] Haseeb Tariq, Muhammad Kashif Hanif , Muhammad Umer Sarwar, Sabeen Bari, Muhammad Shahzad Sarfraz, and Rozita Jamili Oskouei, "Research Article Employing Deep Learning and Time Series Analysis to Tackle the Accuracy and Robustness of the Forecasting Problem", Hindawi Security and Communication Networks Volume 2021,Article ID 5587511, 10 pages https://doi.org/10.1155/2021/5587511.

[5] Neil Shah , Nandish Bhagat and Manan Shah, Shah et al. "Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention", Visual Computing for Industry, Biomedicine, and Art (2021) 4:9pages.

[6] Wajiha Safat, Sohail Asghar , (Member, IEEE), and Saira Andleeb Gillani, "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques" in IEEE Access, Received April 24, 2021, accepted May 2, 2021, date of publication May 6, 2021, date of current version May 17, 2021. Digital Object Identifier 10.1109/ACCESS.2021.3078117.

[7] J Vimala Devi and Dr K S Kavitha, "Time Series Analysis and Forecasting on Crime data" in Press.

[8] Raza Ul Mustafa, M. Saqib Nawaz, M. Ikram Ullah Lali, Tehseen Zia, Waqar Mehmood, "Predicting The Cricket Match Outcome Using Crowd Opinions On Social Networks: A Comparative Study Of Machine learning Methods" in Malaysian Journal of Computer Science, Mar 1, 2017.

[9] S. Palaniappan, T.V.Rajinikanth and A. Govardhan "Enhancement of effective spatial data analysis using R in Indian Journal of Science and Technology, Vol 9(21),DOI:10.17485/ijst/2016/v9i21/95149, June 2016.

[10] J Vimala Devi, Pooja P and Dr K S Kavitha, "Time Series Analysis and Sacramento Crime Data Forecasting",Journal of Huazhong University of Science and Technology, ISSN-1671-451.

[11] Peng Chen, Hongyong Yuan, Xueming Shu," Forecasting Crime Using the ARIMA Model", IEEE Fifth International Conference on Fuzzy Systems and Knowledge Discovery,vol.5, 2008, pp. 627-630.

[12] Manal Almuammar, Maria Fasli,"Deep learning for Non-Stationery multivariate time series forecasting", 2019 IEEE International Conference on Big Data (Big Data).

[13] Jesia Quader Yuki, Md.Mahfil Quader Sakib, Zaisha Zamal, Khan Mohammad Habibullah, "Predicting crime using Time and Location data", ICCCM 2019, July 27–29, 2019, Bangkok, Thailand© 2019 Association for Computing Machinery. ACM ISBN 978-1-4503-7195-7/19/07.

[14] Roomika yadav, Savita Kumari Sheoran "Crime Prediction using Auto Regression Techniques for Time series data", 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering, 22-25 November 2018.

[15] Suhong Kim, Param Joshi, Parminder Singh Kalsi and Pooya Taheri, "Crime Analysis through Machine Learning" , 978-5386-7266-2/18/ ©2018 IEEE.

[16] E. Cesario, C. Catlett, and D. Talia, "Forecasting crimes using autoregressive models", in IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, Aug 2016, pp. 795– 802.

[17] Marzan, C. S., Baculo, M. J. C., de Dios Bulos, R., & Ruiz, C. ," Time Series Analysis and Crime Pattern Forecasting of City Crime Data", Proceedings of the International Conference on Algorithms, Computing and Systems - ICACS '17.

AUTHORS' PROFILE

J Vimala Devi is a B.E graduate in Computer Science and Engineering from Bharathidasan University, M.Tech in Computer Science and Engineering from Satyabama University, Chennai and currently pursuing my Ph.D in Visveswrays Technology University. She has 13 years of Experience in Teaching. She is a life member of CSI and ISTE. She published a paper titled "Fraud detection in Credit Card Transactions by using Classification Algorithms " in IEEE conference. She is doing a lot of projects on title "Descriptive analytics". She is currently working as Associate Professor, Department of CSE in Cambridge Institute of Technology, Bangalore.

Dr Kavitha K S completed BE computer science and Engineering from SIT Tumkur, MTECH from BMSCE Bangalore and PhD from Anna University, Chennai. She is having 22 years of experience in engineering colleges and has worked in various capacities. Around 40 research papers under her credit. Currently she is supervising 6 research scholars. Her areas of interest are Data mining, Machine Learning, Algorithms and Programming etc. Currently she is serving as Professor in the Department of CSE, Global Academy of Technology, Bangalore. She is a life time member of professional bodies such as CSI and IEEE.