BIGDATA ANALYTICS AND TRENDS FORECASTING CRIME DATA

Shastri VishwaShree

Assistant Professor, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad, India, shastrivishwashree@gmail.com

Karangulla Nitheesha

U.G Student, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad, India, Nitheesha16@gmail.com

Kolla Harshini

U.G Student, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad, India, harshinikolla2001@gmail.com

Vangeepuram janaki

U.G Student, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad, India, vangeepuramjanaki@gmail.com

ABSTRACT

Big Data Analytics (BDA) is a sophisticated way of advanced analytic technique against very large, diverse big data. Itis outlined as data sets whose size or sort is on the far side of the power of ancient relative databases to capture, manage, and method the information with low latency this includes tools for knowledge mining, which shifts through data sets in search of patterns and relationships prognosticative analytics, which builds models to forecast client behavior and different future actions, scenarios, and trends machine learning, which faucets varied algorithms to research large data sets deep learning, which may be a lot of advanced effect of machine learning for effective visualization of data in this paper we will apply the criminal data tothe big data analytics to know about the trends visualization and prediction of the data we also use several deep learning techniques for the accurate analysis The predictive results show that the Prophet model and Keras stateful LSTM perform better than neural network models, it is useful for police departments and law enforcement organizations to better understand crime issues whichwill enable them to track activities, predict the likelihoodof incidents, effectively deploy resources and optimize the decision-making process.

Keywords- Big data analytics, Lstm, Neural network

I.INTRODUCTION

Crimes are the threats that are happening in major cities Crime prediction and criminal identification are the key issues for the department of local government as there is an incredible quantity of crime knowledge that exists. BDA becomes increasingly crucial to organizations to address their developmental issues As one of the elemental techniques of BDA, information mining is an innovative, interdisciplinary, and growing analysis

space, which might build paradigms and techniques across numerous fields for deducing useful info and hidden patterns from data The higher downside made American state visit analysis, however, will determine whether a criminal case is formed easier Data processing is beneficial in not solely the invention of latest knowledge or phenomena however additionally for enhancing our understanding of acknowledged ones With the support of such techniques, BDA can help us simply establish crime patterns that occur during a specific area and however they're connected with time The implications of machine learning and applied mathematics techniques on crime or different big information applications like traffic accidents or time-series data, can alter the analysis, extraction, and understanding of associated patterns and trends, ultimately aiding in crime hindrance and management during this paper, progressive machine learning and massive data analyticsalgorithms are used for used for the mining of crime data from U.S. cities i.e. San Francisco. once pre- processing, as well as data filtering and normalization, Google maps primarily based on geo-mappingof the options are enforced for visualization of the statistical results. numerous approaches in machine learning, deep learning, and statistic modeling are used for future trends analysis The aim of this project is to make crime predictions using the features present in the dataset. The dataset is extracted from the official sites. With the help of a machinelearning algorithm, using python as core we can predict the type of crime that will occur in a particular area The objectivewould be to train a model for prediction. The training would be done using the training data set which will be validated using the test dataset.

II.   RELATED WORK

In this section, existing systems are discussed, and the existing systems are categorized and defined based on the technology used and application methodologies:

A.   CONVOLUTION NEURALNETWORK

Convolutional neural network (CNN), a class of artificial neural networks that have become dominant in various computer vision tasks, is attracting interest across a variety of domains, including radiology. CNN is designed to automatically and adaptively learn spatial hierarchies Of features through backpropagation by using multiple building blocks, such as convolution layers, pooling layers, and fully connected layers. Two challenges in applying CNN to radiological tasks, small dataset, and overfitting, will also be covered in this article, as well as techniques to minimize them Ithas very high accuracy in image processing and detects the feature without any human supervision and it has a lack of ability to be spatially invariant to input data and lots of training data is required and CNN do not encode the position and orientation of an object

B.   RECURRENTNEURALNETWORK

Recurrent neural networks (RNN) have feedback loops in the recurrent layer. This lets them maintain information in 'memory' over time. This model helps us to predict the outcome of the layer and this model also helps in backpropagation RNN also includes less feature compatibility compared to CNN and RNN tells about the time series prediction but, it can be difficult to train

standard RNNs to solve problems that require learning long- term temporal dependencies. This is because the gradient of the loss function decays exponentially with time (called the vanishing gradient)

C.   DESIGNING OF CRIME PREDICTION

Predicting a crime plays a pivotal part in addressing crime, violence, conflict and, instability in metropolises to promote good governance, applicable civic planning and, operation. Plentitude sweats have been made on developing crime vaticination models by using demographic data, but they failed to capture the dynamic nature of crimes in civic. lately, with the development of new ways for collecting and integrating fine- granulated crime-related datasets, there is an implicit to gaining an understanding of the dynamics of crimes and advanced crime vaticination. still, for a megacity, it is hard to make an invariant frame for all megalopolises due to the uneven distribution of data. To this end, in this paper, we exploit Spatio-temporal terms in civic data in one a mega city and also influence transfer literacy ways to support the crime prediction of other megalopolises. Specifically, we first validate the actuality of Spatio-temporal patterns in civic crime. Also, we prize the crime-related features from cross-domain datasets. Eventually, we propose a new transfer learning frame to integrate these features and model spatiotemporal patterns for crime prediction

III.   METHODOLOGY

In this design, we will make the use of concepts of machine learning and data science for crime prediction of Chicago crime data set. The crime data is uprooted from the sanctioned gate of Chicago police. It consists of crime information like position description, type of crime, date, time, latitude, longitude. Before training of the model data preprocessing will be done following this point selection and scaling will be done so that delicacy gained will be high. The K-Nearest Neighbor (KNN) classification and colorful other algorithms will be tested for crime vaticination and one with better delicacy will be used for training. Visualization of dataset will be done in terms of graphical representation of numerous cases for illustration at which time the crime rates are high or at which month the criminal activities are high.
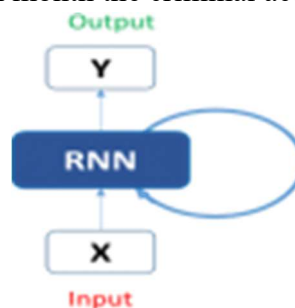

Fig no : 1. RNN RECURRENT NEURALNETWORK

The soul purpose of this design is to give a basic idea of how machine learning can be used by the law enforcement agencies to detect, predict and break crimes at an important faster rate and thus reduces the crime rate. It not confined to Chicago, this can be used in other states or

countries depending upon the availability of the dataset. This is the design which gives the most accurate results hence it is the most widely used in today technology . in the exixting model we make use of RNN algorithm. In Recurrent Neural networks, the information cycles through a circle to the middle subcaste layer.

The input subcaste 'x' takes in the input to the neural network and processes it and passes it onto the middle subcaste. The middle layer 'h' can correspond of multiple retired layers, each with its own activation functions and weights and biases. If you have a neural network where the colorful parameters of different retired layers are not affected by the previous layer, ie: the neural network does not have memory, also you can use a recurrent neural netw ork.

The Recurrent Neural Network will regularize the different activation functions and weights and impulses so that each hidden layer has the same parameters. Also rather of creating multiple hidden layers, it will produce one and cycle over it as numerous times as needed.

PROPOSED METHODOLOGY

Huge Data Analytics (BDA) is becoming an arising approach for dissecting information and separating data and their relations in a wide scope of use zones. comparable to a public arrangement in any case, there are numerous difficulties in managing a lot of accessible information. Therefore, new strategies and innovations should be contrived to examine this heterogeneous and multi-sourced information.

Big information investigation (BDA) is applied and focussed in the arenas of information science and software engineering. The origination of huge information in BDA, its examination and the related difficulties while communicating among them. on exploration holes and difficulties of wrongdoing information mining. In extra to that, this undertaking knowledge about the information digging for finding the examples and patterns in wrongdoing to be utilized fittingly and to be assistance for novices in the examination of wrongdoing information mining. As an outcome, the administration and investigation with colossal information are exceptionally troublesome and complex. To build the effectiveness of wrongdoing discovery, it is important to choose the information mining procedures reasonably. different information mining applications, particularly applications that applied to address the violations Apriori calculation to locate the viable affiliation rule and to lessen the measure of handling time.

Furthermore, there are a few procedures that have been created to break down the relationship between two itemsets allthe more adequately, for example, common data idea however the calculation was expanded the more measure of time .

DATA PREPROCESSING

Before actualizing any calculations on our datasets,a progression of pre- handling steps is performed for information molding as introduced beneath: Time is discretized a few segments to consider time arrangement estimating for the general pattern inside the information.

For some missing direction credits in Chicago and Philadelphia datasets, missing qualities, processed their mean, and afterward supplanted the missing ones. The timestamp demonstrates

the date and season of an event of every wrongdoing, we reasoned these ascribe into five highlights, Month (1-12), Day (1-31), Hour (0-23), and Minute (0-59). We likewise overlook a few highlights that are unneeded like incident Num, arranged.

We will make use of a different set of modules in thisproposed system

A.   UPLOADING CRIME DATASET:

Upload Crime Dataset Module is used to upload theSan-Francisco dataset.

B.   ISUALIZE DATASET:

Visualize Dataset module is used to view the top 10 crimes. We can see the top 10 crimes and which crime is committed g more

C.   YEARLY CRIME:

Yearly Crime module to visualize the graph, in above graph x-axis represents the year and the d y-axis represents crimes count and from above graph from 2003 to 2014 more number of crime was committed

D.   HOURLY CRIME:

Hourly Graph module to visualize the graph, in above graph x axis represents Hour and the y-axis represents crime count and in the above graph, we can see more crimes are committed at night

E.   RUN NEURAL NETWORK ALGORITHM

Run Neural Network Algorithm module is used to generate neural filter network model, its building model on different epochs to filter .

Network Algorithm module is used to generate neural network model, its building model on different epochs to filter

F.   RUN LSTM ALGORITHM:

Run LSTM Algorithm module is used to generate model using LSTM we got prediction with LSTM and LSTM got less RMSE error to compare to neural network so we can say LSTM is better than Neural Network.

G.   RMSE COMPARISON GRAPH:

RMSE Comparison Graph module is used to get an RMSE comparison graph between LSTM and neural network. In above graph x-axis represents algorithm name and y-axis represents RMSE value. From above graph we can say LSTM got less RMSE so it performance is better. If any prediction model accuracy is high then it will consider as best prediction model and its RMSE error will reduce.

H .SOFTWARE REQUIREMENTS

A Software Requirements Specification (SRS) – the specification of a condition for a software system is a complete description of the geste of a system to be developed. It includesa set of use cases that describe all the relations the druggies will have with the software. In addition to use cases, the SRS also contains non-functional conditions..

Inoperative conditions are conditions that put constraints on the design or perpetration (similar as performance engineering conditions, quality norms, or neural network model, its building model on different epochs to The analytics of fraud can work with variety of detection functions.

They work toward mitigating the impact of the attack rather than trying to proactively prevent it.

Network Algorithm module is used to generate neural network model, its building model on different epochs tofilter .

Service providers will give an access to the latest platform atall times service providers. Projects are subject to three sorts of requirements:

● Business requirements describe in business terms what must be delivered or accomplished to provide value. · Product requirements describe properties of a system or product (which could be one of several ways to accomplish a set of business requirements.)
● Process requirements describe activities performed by thedeveloping organization. For instance, process requirements
could specify a Preliminary investigation examine project feasibility, and the likelihood the system will be useful to the organization. The main objective of the feasibility study is to test the Technical, Operational and Economical feasibility for adding new modules and debugging old running system. All system is feasible if they are unlimited resources and infinite time. There are aspects in the feasibility study portion of the preliminary investigation:
design constraints).

System requirements specification: A structured collection of information that embodies the conditions of a system. A business critic, occasionally named system critic, is responsible for assaying the business requirements of their guests and stakeholders to help identify business problems and propose results. Within the systems development lifecycle sphere, the BA generally performs a liaison function between the business side of an enterprise and the information technology department or external

● **ECONOMIC FEASIBILITY**
A system can be developed technically and that will beused if installed must still be a good investment for the

from the new systems. Fiscal benefits must equal or exceed the costs. The system is economically doable. It does not beare any addition tackle or software. Since the interface for this system is developed using the being coffers and technologies available at NIC, There is nominal expenditure and provident feasibility for certain.

- **OPERATIONALFEASIBILITY**

Proposed systems are salutaryl only if they can be turned out into information system. That will meet the association's operating conditions. functional feasibility aspects of the design are to be taken as an important part of the design perpetration. This system is targeted to be in agreement with the below-mentioned issues. Beforehand, the operationt issues and stoner conditions have been taken into consideration. So there is no question of resistance from the druggies that can undermine the possible operation benefits. The well- planned design would insure the optimal application of the computer coffers and would help in the enhancement of performance status.

- **TECHNICALFEASIBILITY**

Before no system was to feed the requirements of the 'Secure structure perpetration System'. The current system developed is technically doable. It is a web-based stoner interface for inspection workflow at NIC-CSD. Therefore it provides easy access to the druggies. The database's purpose is to produce, establish and maintain a workflow among colorful realities in order to grease all concerned druggies in their colorful capacities or places. authorization to the druggies would be grounded based on the places specified. Thus, it provides the specialized guarantee of delicacy, trustability and security

IV. LONG SHORT TERM MEMORY

If in case we need to make some space for anything importantwe know which meeting could be canceled to accommodate apossible meeting.

Turns out that an RNN doesn't do so. In order to add a new information, it transforms the being information fully by applying a function. Because of this, the entire information is modified, on the total,

i. e. there is no consideration for 'important'information and 'not so important' information.

LSTMs on the other hand, make small variations to the information by proliferations and additions. With LSTMs, the information flows through a medium known as cell countries. This way, LSTMs can widely flashback or forget effects. The information at a particular cell state has three different dependences.

We'll fantasize this with an illustration. Let's take the illustration of prognosticating stock prices for a particular stock. The stock price of moment will depend upon:

The trend that the stock has been following

in the former days,perhaps a downtrend or an uptrend.

The price of the stock on the former day, because numerous dealers compare the stock's former day price before buying it. The factors that can affect the price of the stock for moment. This can be a new company policy that is being blamed extensively, or a drop in the company's profit, or perhaps an unanticipated change in the elderly These dependencies can be generalized to any problem as: The previous cell state The previous retired state The input at the current time step Another important point of LSTM is its analogy with conveyorbelts! diligence use them to move products around for different processes. LSTMs use this medium to move information around. We may have some c it flows through the different layers, just like a product may be moldered, painted or packed while it is on conveyor belt.
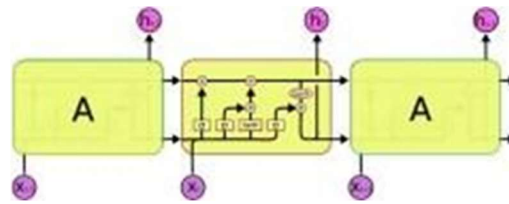

Fig2: LSTM

V.   DATA ANALYSIS
A. RESULTS

A neural community consists of a positive numbers of neurons, specifically nodes withinside the community, that are prepared in numerous layers and linked to every different pass one-of-a-kind layers. There are as a minimum 3 layers in a neural community, i.e. the enter layer of the observations, a non-observable hidden layer withinside the middle, and an output layer because the expected results. In this paper we explored the multilayer feed-ahead community, wherein every layer of nodes gets inputs from the preceding layer. The outputs of the nodes in a single layer becomes the inputs to the subsequent layer.

Fig no : 3. Hourly graph

A neural community consists of a positive numbers of neurons, specifically nodes withinside the community, that are prepared in numerous layers and linked to every different pass one-of-a-kind layers. There are as a minimum 3 layers ina neural community, i.e. the enter layer of the

observations, a non-observable hidden layer withinside the middle, and an output layer because the expected results. In this paper we explored the multilayer feed-ahead community, wherein every layer of nodes gets inputs from the preceding layer. The outputs of the nodes in a single layer becomes the inputs to the subsequent layer.

Fig 4: Range

The lstm can delete or upload facts to the cell state , Which is cautiously managed through systems known as gates. Gates are a mechanism to selectively permit facts to byskip through sigmoid neural internet layer plus a pointwise.
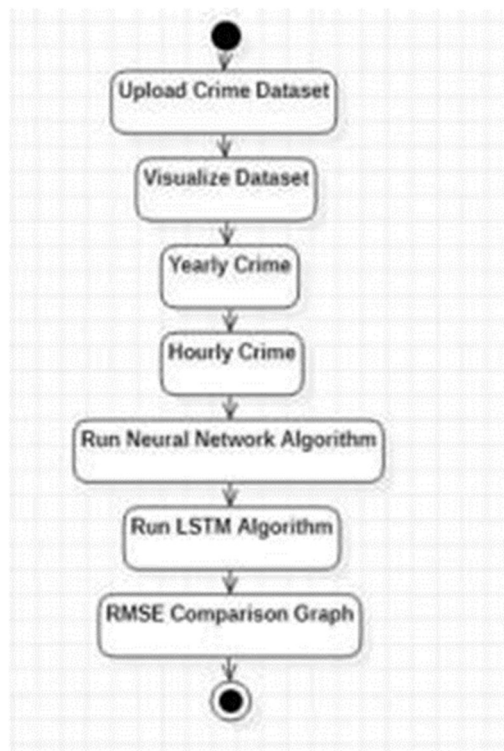
- **ANALYSIS:**

Fig no:5. Flowchart

All of the software program product necessities to be evolved at some point of the necessities definition degree waft from one or greater of those goals. The minimal data for every purpose includes a name and textual description, despite the fact that extra data and references to outside

files can be included. The outputs of the task starting stage are the configuration control plan, the nice guarantee plan, and the task plan and schedule, with an in depth list of scheduled sports for the approaching Requirements degree, and excessive stage estimates of attempt for the out stages.

VI.   CONCLUSION AND FUTURE SCOPE

With the assist of system gaining knowledge of technology, it has end up smooth to discover relation and styles amongst numerous information's. The paintings on this mission specially revolves round predicting the form of crime which may also show up if we realize the region of wherein it has occurred. Using the idea of system gaining knowledge of we've constructed a version the usage of education information set which have gone through information cleansing and information transformation. The version predicts the form of crime with accuracy of 0.789. Data visualization facilitates in evaluation of information set. The graphs encompass bar, pie, line and scatter graphs every having its personal characteristics. We generated many graphs and observed thrilling data that helped in know-how Chicago crimes datasets which could assist in taking pictures the elements which could assist in preserving society safe.

In future, we plan to complete our on- going platform for generic big data analytics which will be capable of processing various types of data for a wide range of applications. We also plan to incorporate multivariate visualization graph mining techniques.

VII.   REFERENCES

1.   Yadav, S., Timbadia, M., Yadav, A., Vishwakarma, R., & Yadav, N. (2017, April). Crime pattern detection, analysis & prediction. In Electronics, Communication and AerospaceTechnology (ICECA), 2017 International conference of(Vol. 1, pp. 225- 230). IEEE.

2.   Shamsuddin, N. H. M., Ali, N. A., & Alwee, R. (2017, May). An overview on crime prediction methods. In Student Project Conference (ICT- ISPC), 2017 6th ICT International (pp. 1-5). IEEE.

3.   Sivaranjani, S., Sivakumari, S., & Aasha, M. (2016, October). Crime prediction and forecasting in Tamilnadu using clustering approaches. In Emerging Technological Trends (ICETT), International Conference on (pp. 1-6). IEEE.

4.   Sathyadevan, S., & Gangadharan, S. (2014, August). Crime analysis and prediction using data mining. In Networks & Soft Computing (ICNSC), 2014 First International Conference on (pp.406-412). IEEE.

5.   Nath, S. V. (2006, December). Crime pattern detection using data mining. In Web intelligence and intelligent agent technology workshops, 2006. wi-iat 2006 workshops. 2006 ieee/wic/acm international conference on (pp. 41-44). IEEE.

6.   Zhao, X., & Tang, J. (2017, November). Exploring Transfer Learning for Crime Prediction. In Data Mining Workshops (ICDMW), 2017 IEEE International Conference on (pp. 1158-1159). IEEE.

7.   Al Boni, M., & Gerber, M. S. (2016, December). AreaSpecific Crime Prediction Models. In Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on (pp. 671-676). IEEE