



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 8, August 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.524

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com

Crime Type and Occurrence Prediction using Machine Learning Algorithm

K. Oviya¹, P. Malarvizhi²

PG Scholar, Department of CSE, Knowledge Institute of Technology, Salem, Tamil Nadu, India¹

Assistant Professor, Department of CSE, Knowledge Institute of Technology, Salem, Tamil Nadu, India²

ABSTRACT: Big data analytics (BDA) is a powerful method to evaluate and identify different facts, relationships, and patterns in big data. In this study, we use BDA to identify offensive words and conduct a literature review to predict the representations and patterns. Many top-notch data mining and deep learning algorithms are used. After much research and presentation, some surprising facts and examples emerged from criminal records in San Francisco, Chicago, and Philadelphia. The estimated output shows that Prophet model and Keras state LSTM outperform texture model with appropriate dimension profile determined over three years. These useful results will allow police and law enforcement agencies to better understand wrongdoings and provide information that will allow them to track training, approximate probabilities, delivery of appropriate assets, and the development of dynamic cycles.

I. INTRODUCTION

In recent years, big data analytics (BDA) has emerged as a new tool to analyze data across multiple applications and provide insights and relationships. Cities play an important role in our lives due to continuous urbanization and population growth. However, these changes have been accompanied by an increase in violent crimes and accidents. Health experts, analysts, and law enforcement agencies are trying to identify probabilities and trends to answer these questions. However, dealing with too much information is a common problem in public policy. Therefore, new methods and techniques are needed to analyze various materials and locations. The analysis of this big data allows us to effectively track events, discover event patterns, utilize resources and make timely decisions.

1.1 DATA MINING OVERVIEW

The process of extracting useful and helpful information from data is called data mining. Although data mining is still in its infancy, companies in many industries, such as retail, banking, healthcare, manufacturing, transportation, and aviation, are already using data mining tools and techniques to work with historical data. Data mining uses information technology models, statistics, and mathematics to filter home data, helping analysts identify important themes, connections, differences, patterns, inconsistencies, etc. Data mining is becoming more common in the business and public sectors. Data is widely used in industries such as banking, insurance, healthcare, and business to reduce costs, improve research, and increase sales. Data mining was originally used to detect fraud and waste in the public sector but has now evolved to be used for purposes such as evaluation and good business development.

1.2 DATA MINING CHARACTERISTICS

The problems encountered by data seekers are due to the fact that data in real-world data is not collected with the primary purpose of discovery. The main purpose of information in a database is storage, retrieval and modification. Therefore, the information provided by companies that are involved in information search usually has the following characteristics. The stored information is large and popular. Formats and data sizes. Since most of the data has been collected many times in the previous process, the data is fragmented and heterogeneous.

1.3 MODEL OF DATA MINING

Data Preparation

Real-world datasets have a number of uncertainties that must be resolved before informed discoveries can be made. This data preprocessing/cleaning can be done using visual or analytical tools. Fill in missing values, reduce noise, reduce data length and resolve differences. Interactive graphics are some of the techniques frequently used for preliminary information.

1.4 PREDICTION OF CRIME USING LSTM AND ARIMA

The short term is called LSTM. LSTM cells are used in neural networks trained to predict the future based on variable length. It is worth noting that, unlike ARIMA and Prophet, RNNs can make all types of data connections. Auto regressive integrated moving average (ARIMA) is a time series forecasting model that uses auto correlation measures to characterize temporal patterns in time series data in order to predict future prices.

1.5 ANALYSIS AND PROJECTIONS

This is a strategy that predicts the future based on past data. It requires a thorough analysis of history and current trends or tendencies to predict future developments. It uses statistics and methods. That is why it is also called statistical analysis. In other words, forecasting can help you plan how to follow your business, how to prevent disruptions and risks, how to prepare for unexpected problems, and how to improve operations to increase revenue. Predictive analytics allow you to analyze customer behavior at a more granular level.

II. LITERATURE SURVEY

2.1 NORTH CAROLINA RACE, SHELBY COUNTY, AND THE VOTER INFORMATION VERIFICATION ACT:

This is a strategy that predicts the future based on past data. It requires a thorough analysis of history and current trends or tendencies to predict future developments. It uses statistics and methods. That is why it is also called statistical analysis. In other words, forecasting can help you plan how to follow your business, how to prevent disruptions and risks, how to prepare for unexpected problems, and how to improve operations to increase revenue. Predictive analytics allow you to analyze customer behavior at a more granular level. [1]

2.2 MAINTAINING AN ONLINE BIBLIOGRAPHICAL DATABASE, THE DATA QUALITY PROBLEM:

Cite Seer and Google-Scholar are important digital libraries that provide access to (computer) research materials in this field. Both collections work like expert researchers, crawling the web and analyzing the content, categorizing all text and extracting metadata with minimal human intervention. On the other hand, traditional bibliographies such as INSPEC for engineering and PubMed for medicine are also available. The DBLP program for computer science has expanded from a small specialized database to a digital library covering most subfields of computer science. Maintaining the second layer of writing has been a major effort. This expense is only worth it in the long run when the quality of the data is stored more than the search engine type domain. In this work, we address the problems of data quality management and algorithms. We focus on specific challenges in the list. [2]

2.3 THE IGRID INDEX, REVERSING THE DIMENSIONALITY CURSE IN HIGH DIMENSIONAL SPACE SIMILARITY INDEXING:

In this work, search consistency and indexing issues are commonly addressed as problems for advanced applications. Most indexing solutions quickly become smaller in size and all queries need to access the entire database. Furthermore, new findings suggest that even a similar strategy may not be worthwhile at high altitudes. In this work, we introduce a similar indexing method, I-grid indexing, which uses a remote location whose main part is stored in a small size. Furthermore, this approach performs differently than all other cognitive training methods; Therefore, it needs to be large in size to produce similar results. [3]

2.4 LINKAGE OF BLOCKING-AWARE PRIVATE CRIME RECORDS:

This work addresses the challenge of comparing high-speed data from two independent sources without disclosing privacy to other parties (e.g., communication issues related to links to illegal files). Our main goal is to establish a security barrier that will increase the efficiency of linking data breaches while maintaining security. Although efforts have been made to link private evidence, no one has decided to use property blocking. Therefore, our proposed blocking-aware linking of private data can address large data breaches without compromising privacy. Preliminary tests have proven the feasibility of the proposal. [4]

2.5 NEAR-OPTIMAL HASHING ALGORITHMS FOR APPROXIMATE Closest NEIGHBOR IN HIGH DIMENSIONS:

In this paper, we provide an algorithm for c-approximation problems in d-dimensional Euclidean space with query time $O(dn^{1/c} + o(1))$ and space $O(dn + n^{1/c} + o(1))$. This is almost the same as the lower bound for hash-based algorithms recently discovered in [27]. We also see the quality difference of this method, which requires $dn + n \log O(1)$

n locations and has a query time $dnO(1/c2)$. Finally, we investigate the evolution of the Leech Lattice method using a fast cutting machine. [5]

III. EXISTING & PROPOSED SYSTEM

3.1 EXISTING SYSTEM

A weak learning method (WSL) is proposed to solve the translation problem of object detection in optical RSI. As the most efficient learning method possible, WSL only needs the base name of the image to indicate whether the image contains the object of interest. To this end, unlike learning models that rely on entity-labeled return boxes to provide product identification, the WSL method does not require a clear boundary and size of the product copy. The analysis paper of WSL often simultaneously mentions the boundaries of interest in each specific preparation image (annotation) and simultaneously prepares the product with interpretation (training). Disadvantages of the existing method This approach cannot successfully distinguish between normal and abnormal NVCR datasets. It may be difficult to measure the distance between samples when the material is complex. Criticism cannot be convincing. The criminal evidence that is part of the record will be close to other "normal" values and therefore will be classified as "normal". It is usually located in less populated areas that interact with the local smart community. Check their importance, such as the previous health content is not far from the main content. Due to hardware limitations, we cannot test more than 16 threads. It cannot guarantee the stability of the time, and the data will be cached when the data is restored..

3.2 PROPOSED SYSTEM

Big data analytics (BDA) is a new way to analyze data and provide information and relationships in various application areas. However, dealing with too much information is a common problem in public policy. Therefore, new methods and techniques are needed to analyze various materials and locations. Big data analytics (BDA) has been used and studied in data science and computer science disciplines for a long time. Big data, analysis and related topics related to BDA. On the detection of gaps and problems in criminal profiling. In addition, the project provides information on profiling to identify crime patterns and patterns that can be used effectively and provides new resources for criminal investigation law. Therefore, managing and analyzing data becomes very difficult and challenging. In order to be effective in crime detection, appropriate data mining techniques must be used. Many data mining applications, especially those that use a priori methods to find the most efficient shared rules and reduce processing time. Additionally, many strategies have been developed to achieve more.

3.2.1 KEYWORD INDEXES:

BFS processes the basic elements in this model to create query and response tuples using table-level granular frequency data and column-level granular time frequency data. The two rules used to maintain frequency statements for row and unit granularity levels are location granularity and laboratory granularity. BFS creates a hash table with row ordering to generate and sort candidates' questions for key questions.

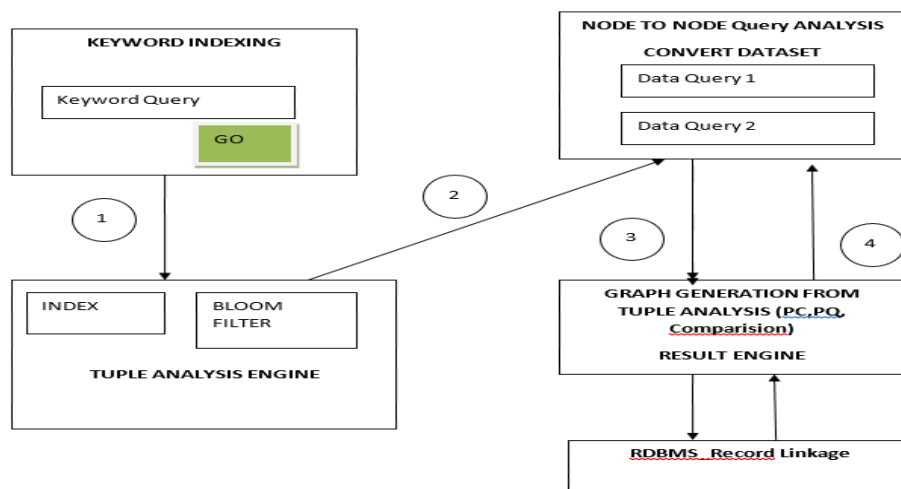


Fig no: 1 Keywords

3.3 GENERATION OF ANSWER TUPLES:

Not every attribute string in the filename will have one, as providing all the indexing requires more capacity of the module. However, the full text of the publication is not available due to the KWS effect leading to the combination of relevant tables. As a result, BFS stores cell-granular word frequency information without changing the recording information for reporting features that do not have a full record, thus avoiding the costs associated with many time uses.

3.4 NODE-NODE INDEX

In this model, we use the shortest violation data to connect the data between two nodes. Like the node keyword index, the link profile information that only comes from nodes with a violation link profile is more serious than a specific thing that can be recorded. The reason for node-to-node indexing is that in data-based data, the number of different words in a given data link field is the weight on the profile according to the number of nodes in that field.

3.5 METADATA GRAPH GENERATION

This model contains metadata for these graphs, and it is easy to imagine other important, possibly analytical models that would benefit from having more information in the form. There is no reason why this “data” should be limited to the properties of the node; the framework now simply works as a library of useful code, providing a set of interfaces and abstract classes to support implementations. An improvement.

IV. CONCLUSION

Crime poses a threat to human progress, security, and long-term development and must be stopped. Forecasting and prediction are often required by law enforcement agencies to improve crime analysis to enhance urban safety and help prevent crime. We found that most of the shortest crime link profiles are set very high. However, the traversal method ignores intermediate nodes. Statistical studies have also shown the comparative behavior of different types of networks, as well as the characteristics of the network such as integration, average shortest crime network profiles, and their averages.

V. RESULT

There are some points to consider when evaluating the results of crime prediction models Long Short-Term Memory (LSTM) and Deep Neural Networks (DNN). Powerful machine learning algorithms such as DNNs and LSTMs can capture complex patterns and dependencies in complex data such as crime analysis. They can find indirect connections and write passwords effectively. However, DNN models may have difficulty capturing the dwell time and long-term patterns of crime. LSTMs can use this memory system to identify types of violations and their consequences. By obtaining actual illegal data, the LSTM model can predict the type and probability of future illegal behavior.

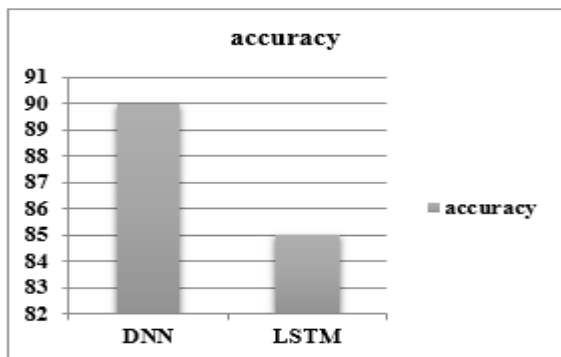


Fig no: 2 Project accuracy rate

algorithm	accuracy
DNN	90
LSTM	85

Fig no: 3 Algorithm name

VI. FUTURE WORK

Improvements to the LSTM model architecture are possible. Variations such as bidirectional LSTM, stacked LSTM layers, and hybrid models that combine LSTM with other types of recurrent or convolutional layers are all options for research. You can find great resources for capturing physical and structural effects in criminal records by examining various real estate options. Future work could focus on developing modern methods for designs to deal with predictable crimes. Researchers could use specific features to examine spatiotemporal relationships, correlations, and even online sources such as News Feeds. Additionally, the priority selection process can be used to separate the most common issues from error behavior, dimensionality reduction, and modeling needs.

REFERENCES

1. (2022). [Online] North Carolina voter registration database. Available:ftp://www.app.sboe.state.nc.us/data.
2. (2022). ([Online] The dblp computer science bibliography. Available:http://dblp.uni-trier.de/xml/
- 3.C. Aggarwal and P. Yu, "The igrid index: Reversing the dimensionality curse for similarity indexing in high dimensional space," in Proc. 6th ACM Int. Conf. Knowledge Discovery and Data Mining, 2022, pp. 119-129.
- 4.A. Al-Lawati, D. Lee, and P. McDaniel, "Blocking-aware privateCrime record linking," in Proc. 2nd International Workshop on Information Quality and Information Systems, 2021, pp. 59-68.
5. "Near-optimal hashing techniques for approximate closest neighbour in large dimensions," A. Andoni and P. Indyk. ACM Communications, vol. 51, no. 1, pp. 117-122, 2022.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details