International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

TRAFFIC ACCIDENT DATA ANALYSIS USING PANDAS

[1].K.Vijay,[2]C.Naga kowshik reddy,[3]M.Abhinay kumar,[4]M.Aravind

[1]Assistant Professor, [234]Students

Department of CSE(Data Science)

Siddhartha institute of technology & sciences,narapally

vijaykoraveni.cse@siddhartha.co.in, 23TQ1A6763@siddhartha.co.in, 23TQ1A6732@siddhartha.co.in, 23TQ1A6754@siddhartha.co.in

ABSTRACT

Road accidents are a major public safety concern worldwide. According to global transportation statistics, millions of people are injured or killed every year due to traffic accidents. The increasing number of vehicles, poor road conditions, driver negligence, and adverse weather conditions contribute significantly to road accidents. Therefore, analysing accident data is essential for understanding accident patterns and implementing preventive measures. This project focuses on Traffic Accident Data Analysis using Python. The main objective is to analyse a dataset containing information about traffic accidents and identify trends and patterns that lead to accidents. The dataset includes variables such as accident date, time, location, weather conditions, road conditions, number of vehicles involved, and casualties. Using Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn, the data is cleaned, processed, and analysed through Exploratory Data Analysis (EDA). Visualization techniques such as bar charts, line graphs, heatmaps, and pie charts are used to represent accident patterns. The findings of the analysis help identify accident-prone areas, peak accident times, and the most common causes of accidents. These insights can assist traffic authorities and policymakers in implementing strategies to reduce accidents and improve road safety

I INTRODUCTION

Road transportation is one of the most widely used modes of transportation in the world. However, with the rapid growth of vehicles and urbanization, road accidents have become a major issue affecting human life and economic development. Traffic accidents occur due to various factors such as:

• Driver behavior (overspeeding, drunk driving)

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

• Poor road infrastructure

• Weather conditions

• Vehicle mechanical failures

• Traffic congestion

Data analysis plays an important role in identifying accident trends and helping authorities take preventive measures. With the advancement of data science and machine learning, large volumes of traffic accident data can be analyzed to uncover hidden patterns. Python is widely used in data analysis due to its powerful libraries and tools. In this project, Python is used to analyze traffic accident datasets and visualize accident patterns.

II LITERATURE SURVEY

The literature survey reviews previous research and studies related to traffic accident analysis and the use of data analytics techniques to understand accident patterns and improve road safety. Many researchers and organizations have analyzed road accident data to identify the major factors that contribute to accidents. Several studies have shown that driver behavior is one of the leading causes of road accidents. Factors such as overspeeding, distracted driving, drunk driving, and violation of traffic rules significantly increase the risk of accidents. Researchers have emphasized the importance of monitoring driver behavior and implementing strict traffic regulations to reduce accident rates. Another important factor identified in previous studies is road infrastructure and road conditions. Poor road design, damaged roads, lack of proper traffic signals, and inadequate lighting can increase the likelihood of accidents. Researchers have suggested improving road infrastructure and implementing better traffic management systems to enhance road safety. Weather conditions also play a significant role in traffic accidents. Studies have shown that accidents tend to increase during rainy, foggy, or snowy weather conditions due to reduced visibility and slippery road surfaces. Therefore, weather-related data analysis helps in identifying high-risk situations and planning preventive measures. With the advancement of technology, many researchers have started using data science and machine learning techniques to analyze traffic accident data. Tools such as Python, R, and various statistical models are used to perform exploratory data analysis, visualize accident patterns, and identify correlations between different factors.

III SYSTEM ANALYSIS

The system analysis of the Traffic Accident Data Analysis project focuses on understanding the limitations of the existing manual and semi-digital methods and proposing an efficient

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

data-driven solution. In the current system, traffic accident data is often maintained in spreadsheets or physical records, making it difficult to manage, analyze, and extract meaningful insights due to the large volume of data. This leads to time-consuming processes, higher chances of human error, and lack of proper visualization or pattern identification. To overcome these challenges, the proposed system utilizes Python with the Pandas library to efficiently handle and analyze large datasets. The system performs data cleaning, preprocessing, and transformation to ensure accuracy and consistency. It enables quick analysis of accident trends based on factors such as location, time, and causes, helping in identifying high-risk areas and peak accident periods. By automating the analysis process and providing clear insights, the system improves decision-making for traffic management authorities and contributes to better road safety planning.

Existing system

The existing system for traffic accident data analysis mainly relies on manual methods and basic tools such as spreadsheets for storing and managing data. This approach makes it difficult to handle large volumes of accident records efficiently. Data analysis is often time-consuming and prone to human errors, as there is no proper automation involved. Additionally, it becomes challenging to identify important patterns such as accident-prone areas, peak accident times, and common causes of accidents. The lack of advanced analytical tools and visualization techniques further limits the ability to generate meaningful insights from the data. As a result, decision-making for traffic management and safety improvements is less effective and slower in the existing system.

Disadvantages of existing system

- Time-consuming data analysis process
- High chances of human errors
- Difficulty in handling large volumes of data
- Lack of automation
- No real-time data processing

Proposed system

The proposed system for Traffic Accident Data Analysis utilizes Python and the Pandas library to efficiently process and analyze large volumes of accident data. Unlike the existing manual methods, this system automates data handling tasks such as data cleaning, preprocessing, and transformation, ensuring accuracy and consistency. It enables quick and effective analysis of accident data based on various

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

factors like location, time, and causes, helping to identify accident-prone areas and peak time periods. The system also supports better data visualization, making it easier to understand patterns and trends. By providing faster and more reliable insights, the proposed system improves decision-making for traffic management authorities and contributes to enhancing road safety measures.

Advantages of proposed system

- Fast and efficient data processing
- Reduces human errors
- Handles large datasets easily
- Automated data analysis
- Provides accurate and reliable results

IV METHODOLOGY

Traffic Accident Data Analysis using Python The methodology of this project explains the step-by-step process used to analyze traffic accident data. It includes data collection, data preprocessing, analysis, visualization, and interpretation of results. The main goal is to extract meaningful insights from accident datasets using Python tools.

1. Data Collection The first step in the methodology is collecting traffic accident data from reliable sources. The dataset may be obtained from platforms such as Kaggle, government transportation websites, or open data portals. The collected data usually contains information such as accident date, location, weather conditions, number of vehicles involved, injuries, and fatalities.

2. Data Preprocessing After collecting the dataset, the next step is data preprocessing. In this stage, the raw data is cleaned and prepared for analysis. This includes removing duplicate records, handling missing values, correcting inconsistent data formats, and converting date and time fields into a suitable format. Data preprocessing improves the quality and accuracy of the dataset.

3. Exploratory Data Analysis (EDA) Exploratory Data Analysis is performed to understand the structure and characteristics of the dataset. In this stage, statistical summaries and basic analysis are conducted to identify patterns and trends in accident data. EDA helps in identifying relationships between different variables such as time, location, and weather conditions.

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

4. Data Analysis In this stage, the cleaned dataset is analyzed using Python libraries. Different factors influencing accidents are studied, such as time of accident, location, road type, weather conditions, and number of vehicles involved. The analysis helps in identifying major causes and patterns of road accidents.

5. Data Visualization After analyzing the data, visualization techniques are used to represent the results in graphical form. Charts and graphs such as bar charts, line graphs, pie charts, and heatmaps are created using visualization libraries. Visualization helps users easily understand accident trends and patterns.

System Architecture

The system architecture of the Traffic Accident Data Analysis project describes the overall structure and workflow of the system. It explains how the traffic accident data is collected, processed, analyzed, and converted into meaningful insights. The system follows a data analysis pipeline that transforms raw accident data into visual reports and useful information. The architecture of this project mainly consists of several layers or modules that work together to perform data analysis.

1. Data Collection Layer This layer is responsible for gathering traffic accident datasets from different sources such as: • Kaggle datasets • Government transportation databases
2. Data Preprocessing Layer In this layer, the raw dataset is cleaned and prepared for analysis. The preprocessing process includes: • Removing duplicate records • Handling missing values • Converting data types
3. Data Analysis Layer This layer performs Exploratory Data Analysis (EDA) on the dataset to identify patterns and relationships between different accident factors. The analysis includes: • Accident frequency analysis • Time-based accident analysis
4. Data Visualization Layer In this stage, the analyzed data is converted into visual representations using graphs and charts. Visualization helps users easily understand accident patterns. Common visualizations include: • Bar charts • Line graph • Pie charts
5. Reporting Layer The final layer of the architecture presents the results and insights obtained from the analysis. The findings are summarized in reports, charts, and conclusions. These results help in: • Identifying accident-prone areas • Understanding major causes of accidents

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

V RESULTS&OUTPUT

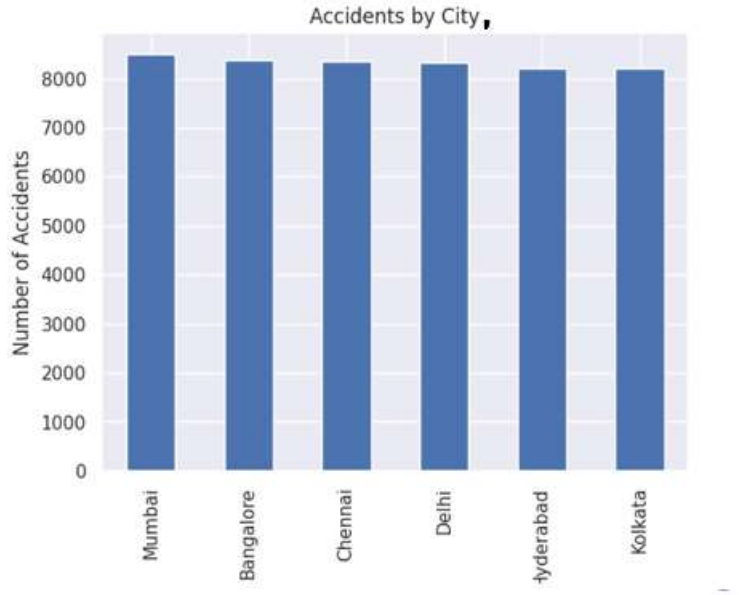
The Traffic Accident Data Analysis project analyzed 50,000 accident records using Python tools like Pandas and Matplotlib. The results show that accidents vary based on city, weather conditions, time of day, and road type. Most accidents occur during evening and night times, and many are of low to medium severity

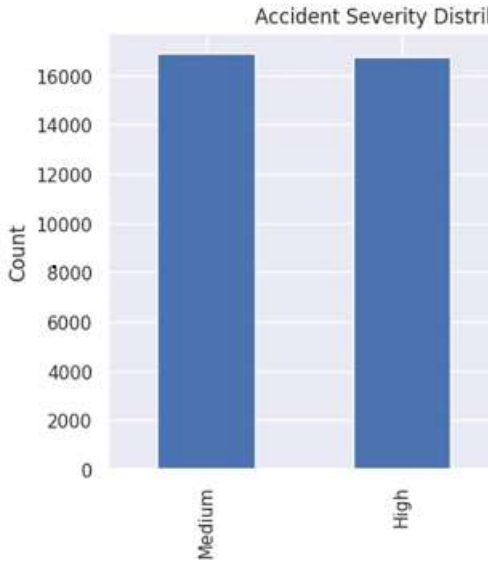
International Journal of

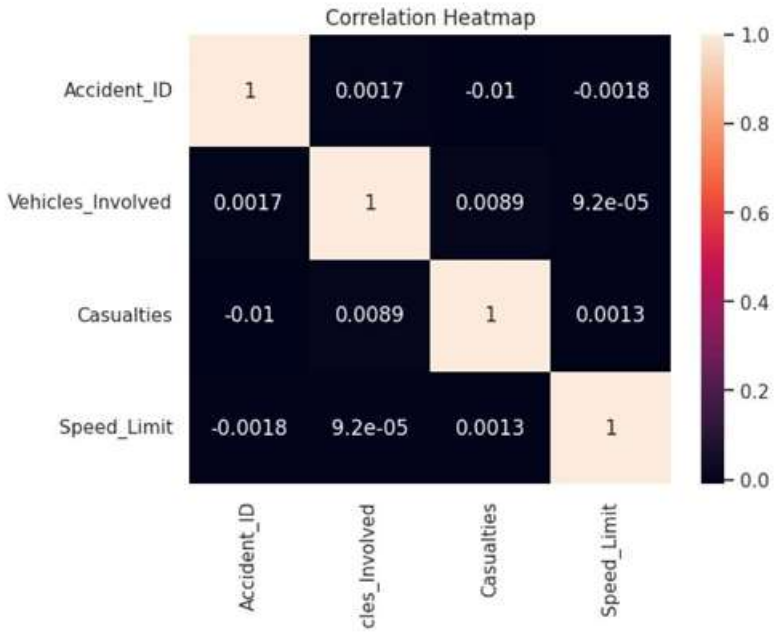
DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper







International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

VI COCNLUSION

The Traffic Accident Data Analysis project was carried out to study and understand the patterns and factors associated with road accidents using data analysis techniques. By using Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn, a dataset containing 50,000 accident records was generated and analyzed to identify meaningful insights.The analysis examined various factors such as city, weather conditions, road type, number of vehicles involved, time of day, and accident severity. Through data cleaning, statistical analysis, and graphical visualizations, the project successfully identified trends and relationships in the accident data. The visualizations helped in clearly understanding how different conditions influence accident occurrence and severity. The results of the analysis indicate that accident frequency varies depending on several factors such as time of day, weather conditions, and traffic density on different road types. The project demonstrates how data analysis techniques can be used to extract useful information from large datasets and support decision-making related to road safety. Overall, this project highlights the importance of data-driven analysis in understanding traffic accident patterns. The insights obtained from this analysis can help traffic authorities, policymakers, and researchers develop strategies to improve road safety and reduce accident risks in the future

REFERENCE

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.