

Improved model for traffic accident management system using KDD and big data: case study Jordan

Faisal Alzyoud

Faisal Alzyoud

Department of Computer Science
Isra Univerity, Jordan
faisal.alzyoud@iu.edu.jo

Abstract

This paper addresses the challenge of the increasing number of traffic accidents resulting from population growth and increased vehicle usage, leading to significant economic and environmental impacts. To address this issue, this paper proposes an improved model for traffic accident management that employs Knowledge Discovery in Databases (KDD) and extensive data analysis techniques to extract the main factors contributing to car accidents. A case study of traffic accidents in Jordan is conducted, utilizing actual data from the Department of Statistics and the Public Security Directorate Records between 2016 to 2020 to identify the primary factors contributing to road accidents and their effects on various types of injuries. The study identifies 11 factors that contribute to car accidents, with driver error being the primary factor for increasing the number of accidents and injuries. Additionally, minor car accidents are found to cause more injuries based on the analysis of the proposed approach. The findings provide a solid foundation for developing an accurate and precise scheme to reduce or eliminate the number of road accidents, which is essential for transitioning to smart cities. Using big data and KDD techniques could significantly improve current traffic accident management practices, informing the development of new policies and regulations to improve road safety and reduce economic and environmental impacts. This paper presents a promising solution for addressing traffic accident management, contributing to safer and sustainable urban environments.

Keywords: artificial intelligence, machine learning, big data, Jordan, KDD, road accidents, traffic management

1 Introduction

The widespread adoption of IoT devices, sensors, smart technologies, and RFID technology in smart cities has led to the generation of large and complex datasets known as big data. These datasets can be leveraged to extract valuable information that can help address many problems, such as traffic congestion [1]. Traffic management is a significant challenge many countries face, as it poses problems related to human factors and economic implications. One of the critical concerns of traffic engineers is identifying the peak hours of congestion on urban and city roads. Numerous policies have been proposed to manage traffic congestion during peak hours [2]. Extensive studies have been

conducted to determine the sources and causes of congestion, such as traffic volume, time of day, weather conditions, and others. Various mitigating measures and solutions have been proposed to reduce traffic congestion during peak hours, including enhancing the roadway capacity, promoting carpooling and public transportation, encouraging flexible working hours, and increasing parking fees in city centres [3].

Road accidents are a significant challenge worldwide, especially in developing countries. Jordan is no exception, with frequent car accidents on rural and suburban roads. As such, previous studies have focused on analyzing accidents, identifying their causes, and analyzing driver faults. Distracted driving is the second-most common driver fault leading to accidents in Jordan, with the number of accidents increasing due to rapid population growth and the number of vehicles, as the number of inhabitants reached 10806000 with a number of recorded vehicles 1502420 [4]. These accidents severely impact society and the economy, with the number of deaths in Jordan reaching 461 in 2020. There is a growing need to develop improved traffic accident management systems in response to this issue.

Numerous endeavours have been undertaken to tackle the problem of road accidents, which are projected to rank as the fifth leading cause of mortality by 2030, as per the World Health Organization's report [5]. The emergence of digital devices and technological advancements have led to the proliferation of the digital economy, which can provide a means of buying and selling goods and services without requiring physical infrastructure. Consequently, it can potentially alleviate traffic congestion [6].

This paper proposes an improved model for traffic accident management using Knowledge Discovery in Databases (KDD) and big data analysis techniques to extract the primary factors contributing to road accidents. A case study is presented using data from the Department of Statistics and the Public Security Directorate Records between 2016 and 2020 to identify the main factors contributing to road accidents and their effects on various types of injuries. The study identifies 11 factors that contribute to car accidents, with driver error being the primary factor for increasing the number of accidents and injuries. The findings provide a solid foundation for developing an accurate and precise scheme to reduce or eliminate the number of road accidents, which is essential for transitioning to smart cities. Using big data and KDD techniques could significantly improve current traffic accident management practices, informing the development of new policies and regulations to improve road safety and reduce economic and environmental impacts.

2 Research Background and Related Work

The process of extracting valuable knowledge from a dataset is known as Knowledge Discovery in Databases (KDD). KDD uses data mining techniques to identify patterns and extract specific information from the data. The importance of KDD has increased since the late 1990s, as companies in the commercial, production, and service sectors have become interested in extracting valuable information from their databases to improve their understanding of consumer preferences and behaviour [7]. KDD is an engineering process that involves several steps, as shown in Figure 1. First, data is entered and aggregated according to association rules. Then, the data discovery process uses statistical or query tools to identify the desired data. The data is reprocessed by developing a model that considers redundant and null attributes to reduce data dimensionality. The data is then analyzed using appropriate methodologies based on the discovery goals. Data science extracts valuable knowledge using reports, actions, and real-time monitoring. A mathematical model can induce reports, actions, and real-time monitoring to extract the desired knowledge. The target and goals of KDD are based on the intended use of the system for data discovery [8]. However, KDD suffers from a high number of redundant records, as it was found that about 78% and 75% of the records are duplicated in KDD train and test sets, respectively [9]. Additionally, KDD has difficulty dealing with unstructured data.

The trend towards developing smart cities that incorporate devices, sensors, and new technology platforms is generating an ever-increasing amount of data. This large and complex data collection, with its diverse types and forms, is commonly called "big data". Big data is characterized by its size and complexity to the extent that it becomes difficult to manage using conventional database management tools or data processing applications [11, 12]. Consequently, big data presents several

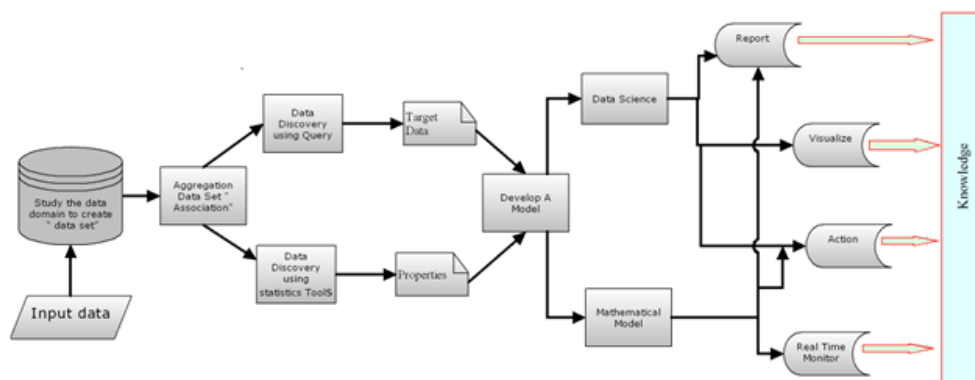


Figure 1: KDD Process Steps

challenges related to data capture, creation, storage, search, participation, transportation, analysis, and visualization [13]. The first level of big data involves the raw data source, which can be captured by collecting data from different devices such as machine sensors or by generating data from trends in additional information derived from the analysis of a single large set of related data, as illustrated in Figure 2.

There are three types of data: structured, semi-structured, and unstructured. Structured data can be easily managed, and any data source problems can be identified during the pre-processing phase, with processing policies varying based on the data category. Semi-structured data can be transformed into a predetermined relational model, such as converting XML documents to free text, although language processing is required to process textual data [14]. Unstructured data, such as images, video, audio, and emails, are more challenging and often processed as text files. These types of data lack correlation and the relation between them. Semi-structured data can be managed using XML files, which are treated as a data set and schema and can be prepared for the next level of information. Information is pooled together and transformed into Meta information level using various techniques, with data cleaning used to eliminate duplication, null fields, compensate empty sets, and boundaries information. Data auditing tools find contradictions by analyzing the data and detecting relationships among data and rules. Ontology matching techniques, such as the automated and robust matching algorithm system 'AgreementMakerLight', can be used for this purpose. It is designed to process huge ontologies while preserving maximum flexibility and expandability of the purebred AgreementMaker structure [15]. The combination of big data and ontology-based services can assist in solving complex problems by extracting useful knowledge, with the Web of Data providing an excellent opportunity for ontology-based services [10].

Big data analysis can be approached in two ways: through data science and mathematical modelling. Data science uses rules to analyze and present knowledge derived from information analysis. On the other hand, mathematical models are used to analyze information mathematically using different mathematical techniques. The ultimate level of big data analysis is knowledge representation, achieved after the presented information has been analyzed and transformed into meaningful information or prediction tools that can be used to predict new knowledge. Another way of using knowledge representation is using measuring tools to check the accuracy of the inferred knowledge. Finally, the transformed data is stored to allow for rapid access to the data. This approach allows for data integration for knowledge sharing and exploration aided by mapping and reduced programming, such as the RDF method. Applying the described technique aims to improve data processing without prior integration [9, 16]. However, Big Data faces several challenges and limitations, such as data acquisition, storage, search algorithms, information sharing, analysis, and visualization. Moreover, computer architecture limitations, such as low-level CPU and I/O, pose a significant challenge. Addressing these challenges can lead to a bright future for Big Data, especially as information continues to exceed our capacity to harness [17, 18]. Table 1 outlines the benefits and challenges of KDD and Big Data. The proposed approach suggests a hybrid approach between KDD and Big Data to enable system discovery from Big Data and overcome the challenges KDD faces.

Big data analysis can be approached in two ways: through data science and mathematical mod-

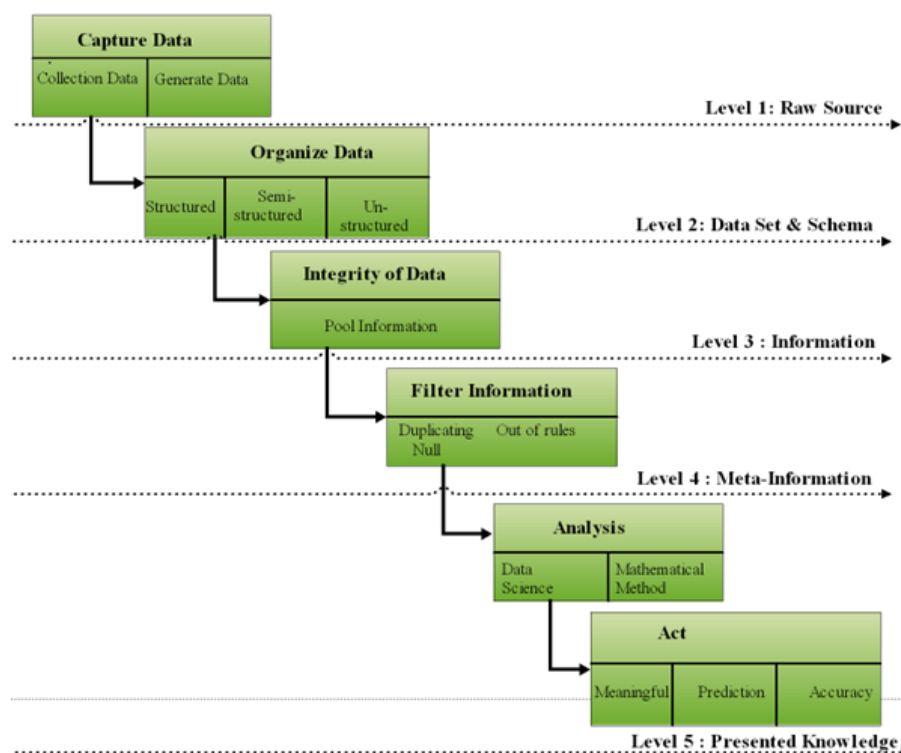


Figure 2: KDD Process Steps

elling. Data science uses rules to analyze and present knowledge derived from information analysis. On the other hand, mathematical models are used to analyze information mathematically using different mathematical techniques. The ultimate level of big data analysis is knowledge representation, achieved after the presented information has been analyzed and transformed into meaningful information or prediction tools that can be used to predict new knowledge. Another way of using knowledge representation is using measuring tools to check the accuracy of the inferred knowledge. Finally, the transformed data is stored to allow for rapid access to the data. This approach allows for the integration of data for knowledge sharing and exploration aided by the use of mapping and reduced programming, such as through the use of the RDF method. Applying the described technique aims to improve data processing without prior integration [9, 16]. However, Big Data faces several challenges and limitations, such as data acquisition, storage, search algorithms, information sharing, analysis, and visualization. Moreover, computer architecture limitations, such as low-level CPU and I/O, pose a significant challenge. Addressing these challenges can lead to a bright future for Big Data, especially as information continues to exceed our capacity to harness [17, 18]. Table 1 outlines the benefits and challenges of KDD and Big Data. The proposed approach suggests a hybrid approach between KDD and Big Data to enable system discovery from Big Data and overcome the challenges KDD faces.

AMOS is a software program used for structural equation modelling (SEM) analysis, which allows users to test and fit various models. With AMOS, users can view the contents of data files and group variables within a database to facilitate testing models involving multiple groups of subjects [27]. AMOS can also handle missing data and empty records, allowing for the estimation of means and intercepts even when the dataset is incomplete. The software is compatible with various data file formats, including Access, Microsoft Excel, FoxPro, Lotus, SPSS, and Comma-delimited text files. SEM is a statistical method used to test theoretical models, which hypothesizes how sets of variables define constructs and how these constructs are related to each other. Researchers typically use sets of indicator variables to define latent variables, independent or dependent variables, and other measurement instruments to obtain observed or latent variables. The Chi-square (χ^2) statistic is an essential indicator used in SEM to evaluate the model's goodness of fit. It indicates the difference between the observed and expected variance-covariance matrices [28].

The Chi-square (χ^2) statistic is an essential tool used in structural equation modelling (SEM) to

	Challenges	Advantages
Big Data	Security and privacy	Usefulness and reliability
	Data heterogeneity [19]	Easy to deal with very large databases [20]
	Ethical issues	Automated update
	Analysis time	Better accuracy
		Support multiple formats
KDD	Not easy to deal with huge databases and unstructured data [21]	Facilitate data mining and implementation of machine intelligence [22]
	Data integrity and missing values	Support decision making [23]
	Data scarcity	Helping the system developer [24]
	Data dimensionality reduction	Provide valuable solutions
	Mathematical model	

Table 1: A comparison between KDD and Big Data

determine the critical N statistic at which the null hypothesis can be rejected. It is used to fit and derive statistics, and the correct Chi-square value represents the model's goodness of fit. Chi-square is a reliable indicator of the robustness of the model-fit measure and is the only statistical test used to test the theoretical model's significance. The Chi-square value ranges from zero for a saturated model with all paths to a maximum value for the independent model with no paths.

In this research, the observed results are almost identical and compatible with the expected results. Therefore, the Chi-square values obtained in this study are almost zero, indicating no significant difference between the observed and expected values.

The application of big data analysis in the study of traffic accidents has facilitated the identification of significant factors that contribute to their occurrence[32]. One research study utilized the UW-DRIVENet transportation big data platform to collect data on over 30,000 traffic crashes in Washington State in 2016. This study employed a joint weighting model based on the interval analytic hierarchy process (IAHP) and the grey relational degree to evaluate the subjective and objective significance of data fields in six dimensions: people, vehicle, road, environment, crash, and time. The model effectively quantified the weight of each dimension, thus providing deeper insights into the factors that contribute to traffic accidents [29].

Additionally, a separate study was conducted to examine the freeway traffic crash data obtained from WDOT (Washington Department of Transportation). To accomplish this, a multi-dimensional multilevel system was developed. To ensure load balancing, the FP-Growth (Frequent Pattern-Growth) algorithm was parallelly optimized on the Hadoop platform, which facilitated efficient and accurate association rule mining calculations on large volumes of traffic crash data [30]. A thorough analysis was conducted using historical crash data to study the impact of environmental and road factors on traffic accidents. The study examined 11 factors that could influence the severity of accidents from both the road and environmental perspectives. These factors were discretized to establish an XGBoost (eXtreme Gradient Boosting) model. To better understand the model's predictions, the SHAP (SHapley Additive exPlanation) value was introduced and used to interpret the XGBoost model. This study aimed to establish a prediction model specifically geared towards determining the severity of traffic crashes on freeways[31].

This study aims to enhance the management of traffic accidents by introducing an improved model that incorporates KDD (Knowledge Discovery in Databases) and big data analysis. To achieve this, a thorough analysis of traffic accidents in Jordan was conducted, which considered several factors contributing to accidents, such as road conditions, weather, and driver behaviour. The proposed model utilizes data mining techniques to scrutinize vast amounts of data and extract patterns that can be utilized to forecast and forestall traffic accidents.

3 Data set

The raw data used in this study was obtained from traffic accident records in Jordan between 2016 and 2020. The data was generated using several methods and includes information on a total of 1,729,343 vehicles, as well as 117,743 transit vehicles and 122,970 accidents by the year 2020. The data was collected from 12 states across Jordan, with three main data centres located in the north, middle, and south regions. The data encompasses a wide range of variables, including vehicle accidents, injuries (age, gender), accident locations, driver information (age, gender), vehicle types, weather conditions, accident images, and road types and conditions [4].

4 Proposed Approach

The rate of information growth in the current century has surpassed Moore's Law due to the diverse aspects of data generation. This study introduces a novel hybrid system integrating knowledge discovery in databases (KDD) and big data analytics to extract valuable insights, as depicted in Figure 3. The proposed approach operates at three distinct levels.

At the initial stage of the proposed hybrid system, the KDD and BD approaches concentrate on the data set, comprising two factors: data capture and data collection. The BD approach is responsible for data capture and collects data in diverse forms, such as structured, semi-structured, and unstructured data, including real-time, streaming, and ACD data. Semi-structured data does not conform to the formal structure of data models used in relational databases and occurs when two data sources use different languages or schema structures, such as XML and OWL, and can include hosted data, mirage data, or data lake. In contrast, structured data is straightforward to handle since it comprises entities and clear relationships between them. The KDD approach is concerned with data collection, and the output is a dataset stored in a structured or semi-structured format. The transformation process identifies the source of data belonging to either the KDD or BD approach and converts unstructured data to structured data, which includes several steps such as acquiring data conditions dependent on the domain of study, determining the path or direction of data, restructuring the data using a tall array or Map-Reduce techniques, and producing a structured dataset. The choice of the technique (tall array or Map-Reduce) depends on factors such as accuracy, performance, and scale. This approach has effectively extracted valuable insights from the data, which is crucial for decision-making in various fields [25].

The second tier is responsible for managing information and making decisions and acts as a mediator between the Pre-Process and Organized levels. To ensure consistency, the transformation process involves comparing the attributes (items) being instantiated with the dataset's content. The tall array confirms the location and content of data, while Map-Reduce confirms the consistency of the item with its content by generating random pairs. Pre-processing uses a set of rules to monitor the entire dataset and clean the data by detecting duplicates, missing values, nulls, empties, and other inconsistencies.

The third layer of the data analysis process pertains to the analysis stage, which involves two main types of analysis: analysis by target and unsupervised analysis. Analysis by target entails generating rules that provide a clear pathway for making accurate decisions based on the domain instructions. To achieve this, a formatting transformation instruction and filter method are required to minimize the number of rules related to the target and/or sub-target and to specify the relationship among decisions. This ensures that the right decision is made for the right target.

On the other hand, unsupervised analysis involves making a cross between properties (items) and target (decision) as presented in the formula below:

$$\alpha = \begin{cases} (\text{sub rules}), & \text{Number of rules} > 1 \\ \text{target A target A}, & \text{Number of rules} = 1 \end{cases} \quad (1)$$

$$\text{Paths} = \begin{cases} (\text{sub rules}), & \text{Number of rules} > 1 \\ \text{target A Rule}, & \text{Number of rules} = 1 \text{ target A} \end{cases} \quad (2)$$

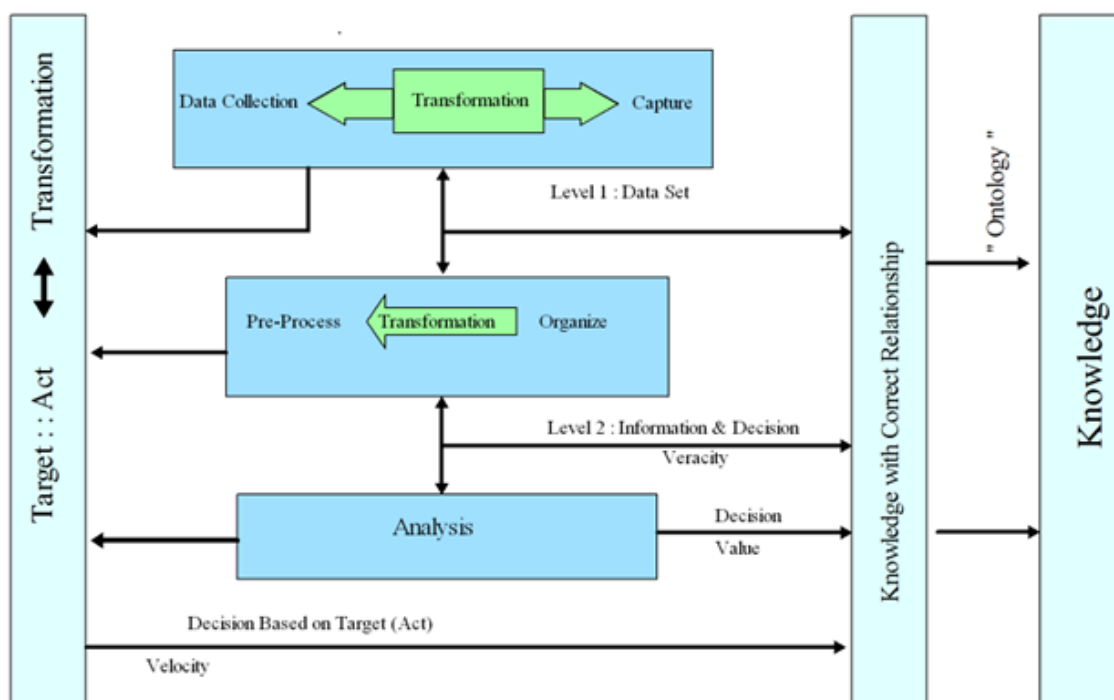


Figure 3: Hybrid approach between KDD and Big Data to extract knowledge

This involves analyzing the relationships among the sub-rules, number of rules, and targets to identify paths to making accurate decisions without needing prior guidance or instruction.

The goal of the analysis stage is to decide based on the target (action) with a short response time and high precision. This requires a trade-off between the number of rules and sub-rules, both of which serve the same purpose of achieving the target. To strike this balance, measures such as time complexity, performance, overhead, capacity, and quality value are adjusted based on the domain studies. This ensures that the analysis process is optimized for efficiency and accuracy while fulfilling the requirements of velocity and veracity.

5 Results and Discussion

The results of this study show that Knowledge Discovery (KDD) offers more effective and valuable detection solutions than traditional Big Data analytics methods. This is because KDD relies heavily on unique and advanced analytical techniques. We combined the two methodologies and utilized tall array techniques to capitalize on the strengths of KDD and Big Data analytics. To test the effectiveness of this approach, we conducted a case study focused on road accidents in Jordan between the years 2016-2020. The results demonstrate that the proposed approach can effectively identify and analyze complex patterns within the data, leading to more accurate and informed decision-making. The data utilized in this study was obtained from traffic accident records in Jordan between 2016-2020. The data includes information on 1,729,343 vehicles, with an additional 117,743 transit vehicles entering Jordan. The analysis reveals that the main factors contributing to accidents in Jordan are driver behaviour, road state, vehicle type and condition, weather condition, and peak hours. Figure 4 illustrates the number of accidents in Jordan between 2016-2020. It is observed that the population and number of vehicles have increased exponentially during the interval, increasing the number of road accidents, injuries, and fatalities. However, in 2020, there was a significant decrease in road accidents due to the country-wide lockdown imposed in response to COVID-19.

The traffic accident data set obtained from the statistical department in Jordan was very large, with different formats and many redundancy and anomalies. To eliminate these defects and extract valuable knowledge, we applied the proposed approach to the dataset using a tall array of Big Data. The result revealed eight main factors responsible for road accidents, as shown in Table 2. These

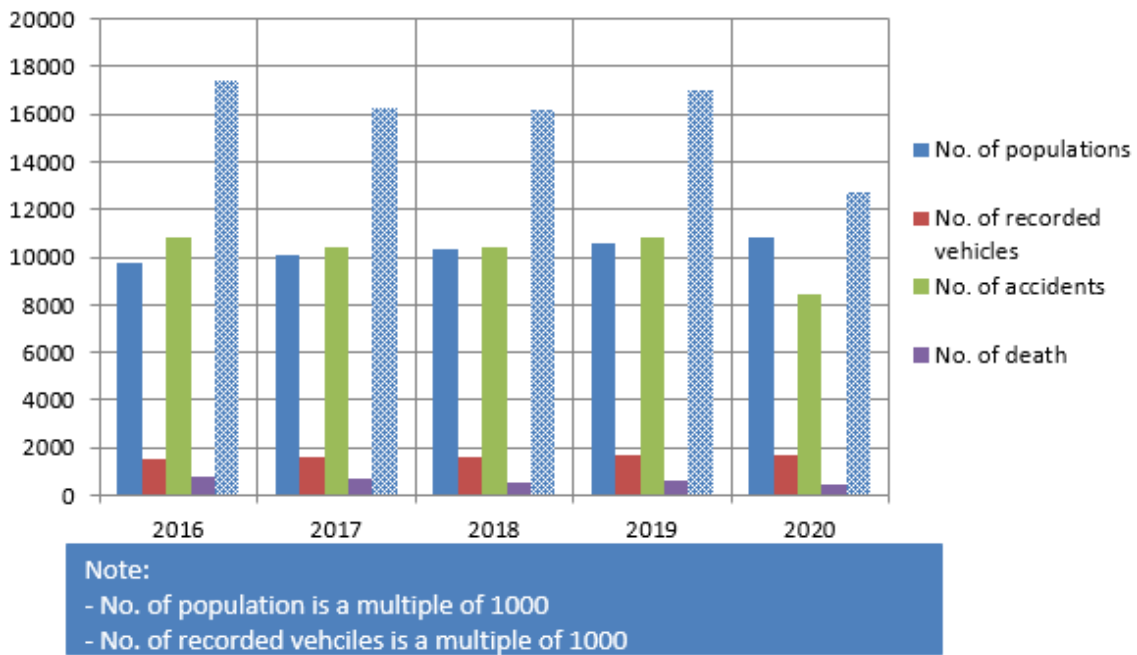


Figure 4: Road accidents information during 2016-2020 in Jordan

factors can be used to build a mathematical and simulation model to allocate the main contributing factors in road accidents and test the optimal parameters with the least property accidents' occurrence using AMOS "SPS", as described above.

The generated data aligned with the expected results generated by the AMOS statistical model. The factors are responsible for road accidents were inferred and presented in Table 2. The analysis revealed that driver mistakes are the main factor responsible for road accidents and cause over 50% of human injuries. One potential solution to this issue is to provide more education and training to drivers using various methods. Additionally, smart cars may have restrictions that reduce driver's mistakes, but some drivers may turn these options off.

Road accidents causes and reasons	Slight injuries	Medium injuries	Severe injuries	Fatal injuries
Traffic light violation	2%	1.30%	0.90%	0%
Road lane changes crashes	23.50%	26.40%	30.50%	34.50%
Speed	1.30%	1.30%	1.60%	3.40%
Drivers' mistakes	43.60%	51.50%	53.70%	50.70%
Roads rules' violations	11.50%	7.80%	4.70%	2.30%
Violation of road traffic signs and regulations	13.50%	8.80%	5.60%	2.80%
Road close follower driving	3.40%	1.20%	1.40%	1.60%
Parking in prohibited areas	0.80%	1.30%	0.70%	3.40%
Others	0.40%	0.40%	0.90%	1.30%

Table 2: Road Accidents Causes and Reasons with Corresponding Injury Levels

We have applied feature selection techniques on the data set to remove non-optimal features and selected only 11 features responsible for causing injuries in traffic accidents. These 11 factors include

the type of vehicles, license driving type, road surface, weather conditions, region, day, time, month, and speed. These factors are summarized in Figure 5. Based on our analysis, we have found that small cars have the highest number of injuries in road accidents. This could be because small car drivers are often unqualified and not properly trained, which increases their likelihood of making mistakes and being unable to avoid injuries during accidents. Other factors, such as weather conditions, daytime running lights, and road surface, significantly reduce the number of injuries in case of road accidents. This suggests that drivers pay more attention to these factors and take appropriate measures to reduce the likelihood of injuries. Based on our analysis, collisions result in more injuries than run-over and deterioration incidents, as shown in Figure 5 -C. We recommend eliminating these incident types in future analyses. We attribute this to the advanced technology used in cars and increased awareness of people in crowded cities. Weekdays and weekends have a similar percentage of injuries, as shown in Figure 5-I, although fewer injuries occur on weekends. However, the difference is not significant. Additionally, July has a higher number of injuries than other months, as shown in Figure 5-H. This may be due to increased traffic from workers returning from neighbouring countries such as Saudi Arabia, Oman, and the UAE. Interestingly, higher speeds over 100 km/h are not a significant factor in increasing the number of injuries. Conversely, accidents at lower speeds result in more injuries, and this is because drivers pay more attention at higher speeds. Most accidents occur between 4 pm and midnight, as shown in Figure 5-G. According to the road department, people get tired after 4 pm and need to rest after a long day, or some start a second job. Identifying 11 factors that contribute to car accidents, with driver error being the most significant factor leading to an increase in the number of accidents and injuries, is a significant contribution of this study. The reliable and valid interpretation of the results through the application of big data and KDD techniques offers the potential for informing the development of new policies and regulations to improve road safety and reduce economic and environmental impacts.

6 Conclusion and Future work

The management of road traffic accidents is a critical aspect of traffic management in most countries, given its significant impact on economies and lives. Despite extensive efforts to study the factors that contribute to road accidents through daily data recording, the increase in data size in volume, velocity, variety, and veracity necessitates the exploration of new trends in Big Data analysis. To extract knowledge that assists decision-makers in minimizing the number of road accidents, this research proposes an approach that merges big data techniques with Knowledge Discovery in Databases (KDD) methods. Real data collected from the Department of Statistics and Public Security Directorate were used to apply feature selection methods, resulting in the selection of 11 factors. The analysis revealed that driver mistakes are the primary factors causing accidents and injuries, with small car accidents causing more injuries. However, some detailed factors, such as road surface and weather conditions, do not affect the increase of injuries and could be eliminated in the future. Meanwhile, other factors have a small percentage and could affect the accuracy of the results. To address these challenges, future work will include generating a simulation model that incorporates these extracted data to determine the best conditions for traffic management through merging Big Data analysis and knowledge data discovery. Additionally, we propose research that utilizes artificial intelligence algorithms such as machine learning and deep neural network to simulate real-world scenarios and explore expected optimal solutions for traffic accident management. Furthermore, new telecommunication technologies and virtual reality approaches will also be incorporated to enhance the accuracy and effectiveness of the simulation model. This research provides valuable insights that will help policy and decision-makers with strategies to reduce the number of road accidents, injuries, and fatalities.

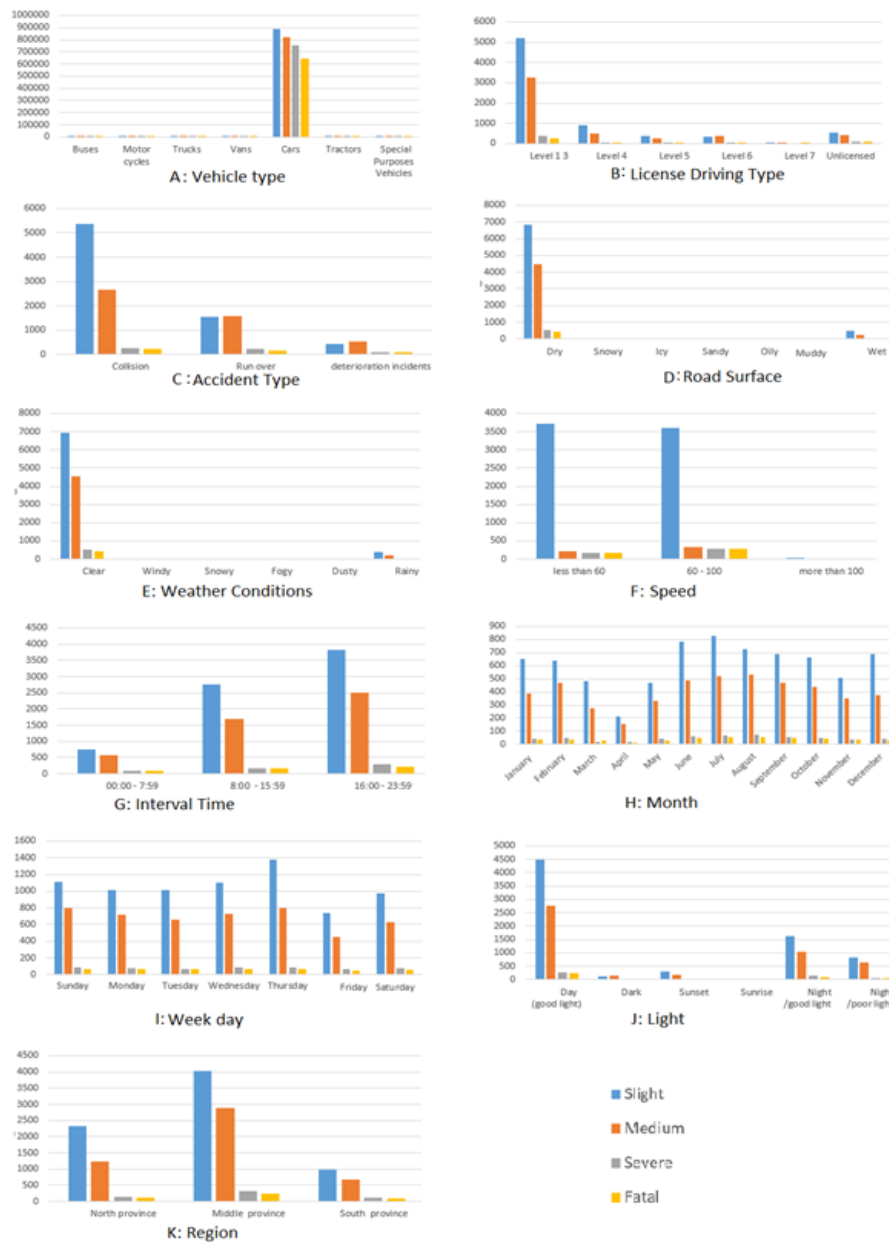


Figure 5: Number of injuries in road accidents based on factor type

References

- [1] Alzyoud, Faisal and Sharman, Nesreen AL and Al-Roosan, Thamer and Alsalah, Yahya (2019). Smart accident management in Jordan using cup carbon simulation. *European Journal of Scientific Research*, 152(2), 128–135, 2019.
- [2] Yan, Ying and Zhang, Shen and Tang, Jinjun and Wang, Xiaofei (2017). Understanding characteristics in multivariate traffic flow time series from complex network structures, *Physica A: Statistical Mechanics and its Applications*, 477, 149–160, 2017.
- [3] Li, Ming and Song, Guohua and Cheng, Ying and Yu, Lei (2015). Identification of prior factors influencing the mode choice of short-distance travel, *Discrete Dynamics in Nature and Society*, 2015, 2015.
- [4] Al-Rousan, Taleb M and Umar, Abdullahi A and Al-Omari, Aslam A (2021). Characteristics of crashes caused by distracted driving on rural and suburban roadways in Jordan, *Infrastructures*, 6(8), 107, 2021.

- [5] Iyanda, Ayodeji E and Osayomi, Tolulope (2021). Is there a relationship between economic indicators and road fatalities in Texas? A multiscale geographically weighted regression analysis, *GeoJournal*, 86(6), 2787–2807, 2021.
- [6] Timotius, Elkana (2021). The implications of digital transformation on developing human resources in business practices in Indonesia: Ana..., *International Journal of Business, Economics, and Management*, 2021.
- [7] Fayyad, Usama M and Pietetsky-Shapiro, Gregory and Smyth, and Padhraic (1996). Knowledge discovery and data mining: Towards a unifying framework, *KDD*, 96, 82–88, 1996.
- [8] Saraswat, Deepak (2017). Knowledge discovery with a hybrid data mining approach, *Agra*, 2017.
- [9] Tavallae, Mahbod and Bagheri, Ebrahim and Lu, Wei and Ghorbani, Ali A (2009). A detailed analysis of the kdd cup 99 data set, *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, 1–6, 2009.
- [10] Guarino, Nicola and Oberle, Daniel and Staab, Steffen (2009). What is an ontology? *Handbook on Ontologies*, 1–17, 2009.
- [11] Abuhammad, Huthaifa and Everson, Richard (2018). Emotional faces in the wild: Feature descriptors for emotion classification, *International Conference Image Analysis and Recognition*, 164–174, 2018.
- [12] Younis, Mohammed Chachan and Abuhammad, Huthaifa (2021). A hybrid fusion framework to multi-modal bio metric identification, *Multimedia Tools and Applications*, 1–24, 2021.
- [13] Singh, Jaskaran and Singla, Varun (2015). Big data: tools and technologies in big data, *International Journal of Computer Applications*, 112(15), 2015.
- [14] Cunningham, Hamish (2002). Gate, a general architecture for text engineering, *Computers and the Humanities*, 36(2), 223–254, 2002.
- [15] Ngo, Duy Hoa and Bellahsene, Zohra (2012). Yam++: A multi-strategy based approach for ontology matching task, *International Conference on Knowledge Engineering and Knowledge Management*, 421–425, 2012.
- [16] Jirkovský, Vaclav and Obitko, Marek and Novák, Petr and Kadera, Petr (2014). Big data analysis for sensor time-series in automation, *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*, 1–8, 2014.
- [17] Ahrens, James and Hendrickson, Bruce and Long, Gabrielle and Miller, Steve and Ross, Rob and Williams, Dean (2011). Data-intensive science in the US DOE: case studies and future challenges, *Computing in Science & Engineering*, 13(6), 14–24, 2011.
- [18] vost and Fawcett, Tom (2013). Data science and its relationship to big data and data driven decision making, *Big Data*, 1(1), 51–59, 2013.
- [19] Bamiah, S. N. and Brohi, Sarfraz N. and Rad, Babak Bashari (2018). Big data technology in education: Advantages, implementations, and challenges, *Journal of Engineering Science and Technology*, 13, 229–241, 2018.
- [20] Zheng, Kangning and Zhang, Zuopeng and Song, Bin (2020). E-commerce logistics distribution mode in big-data context: A case analysis of JD. COM, *Industrial Marketing Management*, 86, 154–162, 2020.
- [21] Zhu, Xingquan and Davidson, Ian, eds. (2007). Knowledge Discovery and Data Mining: Challenges and Realities: Challenges and Realities, *Igi Global*,

- [22] Al-Janabi, Samaher (2021). Overcoming the Main Challenges of Knowledge Discovery through Tendency to the Intelligent Data Analysis, *2021 International Conference on Data Analytics for Business and Industry (ICDABI)*, IEEE, 2021.
- [23] Bhatia, P. (2019). Introduction to Data Mining, *Data Mining and Data Warehousing: Principles and Practical Techniques*, 17–27, Cambridge: Cambridge University Press, doi:10.1017/9781108635592.00, 2019.
- [24] Asrin, Fauzan et al. (2020). Knowledge Data Discovery (Frequent Pattern Growth): The Association Rules for Evergreen Activities on Computer Monitoring, *International Conference on Intelligent and Fuzzy Systems*, Springer, Cham, 2020.
- [25] Al Zyadat, W Jum'ah and Alzyoud, Faisal Y and Alhroob, Aysh M and Samawi, Venus (2018). Securitizing big data characteristics used tall array and mapreduce, *International Journal of Engineering & Technology*, 7(4), 5633–5639, 2018.
- [26] Arbuckle, James and Wothke, Werner (2004). Structural equation modeling using AMOS: An introduction, *EB*, 2004
- [27] Schumacker, Randall E and Lomax, Richard G (2004). A beginner's guide to structural equation modeling, *Psychology Press*,
- [28] Othman, Suad Mohammed et al. (2018). Intrusion detection model using machine learning algorithm on Big Data environment, *Journal of Big Data*, 5(1), 1–12, 2018.
- [29] Yang, Yang and Yuan, Zhenzhou and Meng, Ran (2022). Exploring traffic crash occurrence mechanism toward cross-area freeways via an improved data mining approach, *Journal of Transportation Engineering, Part A: Systems*, 148(9), 04022052, 2022.
- [30] Yang, Yang et al. (2022). A parallel FP-growth mining algorithm with load balancing constraints for traffic crash data, *International Journal of Computers Communications & Control*, 17(4), 2022.
- [31] Yang, Yang et al. (2022). Predicting freeway traffic crash severity using the XGBoost-Bayesian network model with consideration of features interaction, *Journal of Advanced Transportation*, 2022, 2022.
- [32] Tarawneh, Monther and AlZyoud, Fiasal and Sharab, Yousef (2023). Artificial Intelligence Traffic Analysis Framework for Smart Cities, *Computing Conference*, IEEE, 2023.



Copyright ©2023 by the authors. Licensee Agora University, Oradea, Romania.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License.

Journal's webpage: <http://univagora.ro/jour/index.php/ijccc/>



This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE).

<https://publicationethics.org/members/international-journal-computers-communications-and-control>

Cite this paper as:

Faisal.Alzyoud (2023). Improved model for traffic accident management system using KDD and big data: case study Jordan, *International Journal of Computers Communications & Control*, 18(3), 5006, 2023.

<https://doi.org/10.15837/ijccc.2023.3.5006>