(IJRST) 2022, Vol. No. 12, Issue No. II, Apr-Jun **e-ISSN: 2249-0604, p-ISSN: 2454-180X**

Domain-Driven Actionable Knowledge Discovery for Traffic Accidents Using Rules Induction[1]

***Amira Yousif, **Manisha Agarwal, #Vikas Pareek**

**PhD. Scholar, Department of Computer Science, Banasthali Vidyapith*

*** Associate Professor, Department of Computer Science, Banasthali Vidyapith*

 # Professor and Dean, School of CS & ICT, Mahatma Gandhi Central University Bihar

DOI:10.37648/ijrst.v12i02.007

Received: 15 April 2022; Accepted: 01 May 2022; Published: 14 May 2022

ABSTRACT

Due to the limitation of the methodologies of traditional data mining to satisfy business expectations, the shift from mining data-centered hidden patterns to domain-driven actionable knowledge discovery has become a significant direction of KDD research [22]. Traditional data mining algorithms and tools face major obstacles and challenges to solve real-life business problems and issues as they fail to provide actions that can be taken by people in business based on generated rules [22]. A small set of rules are generated by standard classification algorithms to form a classifier, but these classification algorithms use domain independent biases and heuristics [2]. This research aimed to propose a new approach to find actionable rules from sets of discovered rules. It focused on how a combination of traditional classification data mining and domain-driven data mining approach could be applied in solving real-life problems related to the field of traffic accidents in UAE. Real-life data were collected and pre-processed using the user's existing knowledge and needs. Classification using Rules Induction was applied on the domain-driven dataset. The discovered rules from this technique were then summarized, combined, and analyzed. The final set of actionable rules from Classification technique for each class was then generated using a proposed interestingness method. To support such a process, the domain driven in-depth pattern discovery (DDID-PK) framework was followed [9]. Based on experimental results, the extracted domain-driven rules were more interesting and actionable than those produced by the traditional classification technique of data mining. In addition, the integration of data-centered classification technique of data mining to domain-driven approach of data mining and actionable knowledge discovery could help the Dubai police authority to reduce traffic accident severity by formulating new policies and traffic rules based on the domain-driven knowledge extracted from some hidden patterns from real data.

Keywords: *Domain Driven Data Mining, Actionable Knowledge Discovery, Classification, Rules Induction*

INTRODUCTION

Various data mining techniques and algorithms are used to discover knowledge from huge amounts of data and identify understandable patterns from data. Although data mining is considered as one of the most important trends in information technology field in the preceding decade, the recent data mining algorithms, and tools face major challenges to solve real-world business problems [26] as these algorithms fail to provide actions to be taken by businesspeople based on generated rules. Current data mining techniques focus either on data-driven experiments and error detection process [24], or on analyzing business problems on a case-by-case basis. Thus, the patterns discovered from the traditional techniques of data mining do not always satisfy the business requirements and expectations. Due to the limitation of the traditional techniques of data mining to satisfy both business needs and academic attention, shifting from data-centered mining to domain-driven actionable knowledge discovery (AKD) has become a major direction of KDD research [22]. Past research

[1] *How to cite the article:*
Yousif A., Agrawal M., Pareek V., Domain-driven Actionable Knowledge Delivery for Traffic Accidents Using Rules Induction, IJRST, Apr-Jun 2022, Vol 12, Issue 2, 46-71, DOI: http://doi.org/10.37648/ijrst.v12i02.007

has shown this a crucial requirement in real-life applications [22].

Domain-Driven Data Mining (D^3M) can remove the gap between business expectations and academic research by considering real-life aspects such as constraints, human knowledge, and business requirements into account in data mining process [22][24]. D^3M focuses on the development of techniques, methodologies, and tools for Actionable Knowledge Discovery (AKD) [19]. The aim of developing D^3M is to view Knowledge Discovery in Databases (KDD) as Actionable Knowledge Discovery (AKD) problem-solving systems [19]. These systems deliver rules and actions with solid technical significance through developing effective methodologies, methods, and tools helping in decision making [19]. In latest years, a lot of attention has been given to D^3M to make data mining more practical to support decision-making processes in real life business problems in different fields.

Actionable Knowledge Discovery is one of the key challenges for future data mining [22]. According to L. Cao and Y. Zhao, “AKD is a closed optimization problem- solving process from problem definition, framework/model design to actionable pattern discovery, and is designed to deliver operable business rules that can be associated or integrated with business processes and systems” [27]. However, discovering actionable knowledge using traditional KDD methodologies is not a simple task [1] since traditional algorithms of data mining extract/discover patterns through predefined models emphasizing on data-driven experiments and error detection process [24]. The domain driven data mining process should involve key factors such as the business problem domain and KDD interestingness measures [22]. Real-world experience shows the impact of involving domain human experts and constraints factors into the data mining process to develop in-depth patterns [22][26].

Classification rule mining is significant traditional technique of data mining for practical applications. The target for classification rule mining is only one and predetermined known as the class [4]. Classification rule mining algorithms aim to find a small number of rules in the whole database to create an accurate classifier [2]. Classifier performance is usually measured by the overall percent accuracy that is the percentage of the total number of correctly classified observations over the total number of observations [7]. A small set of rules are generated by standard classification algorithms to form a classifier. These algorithms use domain independent biases and heuristics [2].

Interestingness measures of discovered rules are divided into objective measures and subjective measures [5]. Objective measures are used to express rules interestingness by means of mathematical or statistical criteria [5], such as minimum support, minimum confidence, or accuracy. On the other hand, subjective measures of interestingness are used to capture more realistic criteria to be taken into consideration, such as actionability of extracted rules or unexpectedness [31]. In addition to that, interestingness measures of rules, in general, can also be divided into technical interestingness and business ones. Technical interestingness is used to measure how the pattern is interesting from technical perspective [9] whereas business interestingness is determined by some economic and/or domain-oriented social criteria. Actionable pattern can be the pattern that satisfies both technical interestingness and business one [22].

Data mining has widely been implemented in many real-life domains. Many researchers highlighted the significance of applying data mining techniques in the traffic accident field to discover/extract patterns that can reduce the severity of the accidents. Traffic accident/collision is an unexpected or unintended incident that occurs when a road vehicle crashes with pedestrian, animal, geographical or architectural obstacle, or another vehicle [41]. Traffic accidents cause death, injury, or property damage [41]. According to a traffic accidents report generated by the World Health Organization (WHO), millions of road traffic accidents take place worldwide annually, and the fatalities due to these traffic accidents are also in the millions [42]. More than 1.2 million of people are killed yearly due to traffic accidents, and about fifty million get injuries or disabilities [42]. According to the Centers for Disease Control and Prevention (CDC), road traffic accidents cost $100 billion in medical care every year [43]. Almost 40,000 deaths occur due to traffic accidents annually [43] and a car accident victim is treated in an emergency room suffering from accident injuries every 10 seconds [7].

With the increased number of vehicles accompanied by the rapid expanding road construction executed programs, UAE experiences increasing number of road traffic accidents with injuries and deaths causing a major public health problem [6]. In UAE, there are around 600 people killed in traffic accidents each year [11]. Traffic accidents are the second main cause of deaths in the United Arab Emirates [6]. The costs of deaths and injuries due to road traffic accidents have a huge impact on the Emirati society. Previous studies showed that fatality rates in UAE and other Gulf countries are so much higher than those in the developing countries with similar vehicle ownership levels [6].

Dubai, the second largest of the seven emirates in UAE, experienced a loss of Dh4.7 billion due to road traffic accidents in the past 20 years. Traffic accidents are resulting in loss of lives, injuries, and severely affecting the emirate's economy. In 2007, traffic accidents caused an economic loss of around Dh720 million which is about one percent of the gross domestic product (GDP) of the emirate of Dubai [34]. Thus, road traffic accidents problem needs more research to discover associated accidents risk factors and identify new methods to reduce the large number of accidents and fatalities. Applying domain driven data mining into traffic accidents domain still needs further research. Involving the police domain can help to extract interesting actionable rules which are of high significance and need.

This research aims to propose a new approach to remove the big gap between academic research and complicated real-life decision-making problems [26] when conventional data mining applications are deployed. It focuses on how a combination of traditional techniques of data mining and a new approach of domain-driven data mining are applied to solve real-life problems [26] related to road traffic accidents field in UAE. Classification using PART algorithm was applied. This approach can assist in discovering interesting actionable rules from a set of generated rules using classification. It needs to collect real-life data about road traffic accidents from the Dubai police authorities, pre-process the data using the user's existing knowledge and needs, apply Classification data mining techniques based on the selected domain-driven dataset, combine, and summarize the generated classification rules to get the final set of actionable rules for each class. Domain driven data mining approach is used to link recorded accidents factors to accident severity in Dubai using the data mining software, WEKA considering the domain knowledge and needs. The empirical results can assist police decision makers in the formulation of new policies and traffic rules from some hidden in-depth patterns. Conclusions and future work are generated at the end.

RELATED WORK

Based on the literature review, many past and existing studies showed the significance of involving domain experts' knowledge in data mining process [26]. Their inclusion in the stages of traditional techniques of data mining can increase the usefulness of data mining research results when applied to real-life business problems [26] and highlighted the need for more extensive research in Domain Driven Data Mining (D^3M) and Actionable Knowledge Discovery (AKD) areas. Several pieces of research that supports this research approach have been reviewed and covered such as traffic accidents analysis using data mining techniques, domain-driven data mining approach, class association rules, Rule Induction algorithm, objective and subjective interestingness measures of extracted rules, and summary of generated rules.

U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth (1996) presented how knowledge discovery and data mining are related to each other and other domains. They discussed specific data mining techniques, real-life applications, and challenges of knowledge discovery [1].

Bing Liu, Wynne Hsu, and Yiming Ma (1998) proposed to integrate two mining techniques; Classification rule mining and Association rule mining to mine a particular association rules subset, known as Class Association Rules (CARs). [2] Empirical results showed that more accurate classifiers were built than those produced by the classification algorithm C4.5 [2].

Sigal Sahar (1999) proposed a simple method of eliminating many uninteresting association rules outputted by a data mining algorithm. The results of the executions of the algorithm were investigated over three real life databases; web logs, grocery store operations, and adults' survey data [3].

Bing Liu, Wynne Hsu, Shu Chen, and Yiming Ma (2000) presented a new method for helping to identify interesting expected and unexpected association rules. It consisted of an intuitive specification language and an interestingness analysis system. The interestingness analysis system analyzed the generated association rules considering the user's specifications to identify which rules are potentially interesting for the user [4].

T. Brijs, K. Vanhoof, G. Wets (2003) presented an overview of interestingness measures. The researchers focused only on objective measures of interestingness based on mathematical or statistical criteria [5].

Ken Mcgarry (2005) presents a review of literature on the objective and subjective interestingness measures to rank and evaluate the extracted patterns generated by the discovery process. He explored to what extent the two interestingness measures were weak or strong concerning the level of user's incorporation throughout the mining process [8].

Longbing Cao and Chengqi Zhang (2006) proposed a framework called Domain-Driven In-Depth Pattern Discovery (DDID-PD) that presented a domain-driven approach for finding out knowledge that could satisfy business expectations and needs. Its main phases included almost all phases of traditional data mining but were better improved by active collaboration with domain experts. They applied the DDID-PD approach to mine actionable patterns in stock data mining to accurately measure and support trading market [9].

Carlos Ordonez, Norberto Ezquerra, and Cesar A. Santana (2006) investigated how to reduce and summarize association rules. They used greedy algorithm for computing rule covers to summarize rules with the same consequent. The extracted association rules significance was examined using support, confidence, and lift measures. [10].

Roberta Akemi Sinoara and Solange Oliveira Rezende (2006) presented an approach for identifying interesting association rules based on objective and subjective interestingness measures. This approach aimed to take the advantages of both kinds of measures and to involve user's contribution. Objective measures were applied to discover initially interesting rules to be evaluated by the user using the subjective measures [12].

Magaly Lika Fujimoto, Veronica Oliveira de Carvalho, and Solange Oliveira Rezende (2007) proposed an approach that aided the post processing of generalized association rules and facilitated the comprehension and the identification of the interesting ones using objective and subjective measures combined with information visualization techniques, implemented on a system called RulEEGARVis [13].

C. Chien and L. Chen (2008) used data mining techniques to develop specific strategies for human resource and recruitment management in collaboration with domain experts. They successfully applied the results in business [15].

Yanchang Zhao, Huaifeng Zhang, Fernando Figueiredo, Longbing Cao, and Chengqi Zhang (2008) suggested an algorithm to extract novel association rules, combined association rules, which allowed users to perform actions directly. They focused on generating rules using measures of interestingness in combined association rules mining environment. Much interesting actionable knowledge was provided in social security field [16].

Yuming Ou, Longbing Cao, Chao Luo, and Chengqi Zhang (2008) proposed new methods for building microstructure order sequences that involve the business domain to measure the interestingness of patterns based on business perspective. Empirical experiments on an exchange data demonstrated that the generated results using D^3M could satisfy business expectations and helped in taking actions for market surveillance [17].

A. P. Sinha and H. Zhao (2008) presented the effectiveness of involving the domain knowledge when applying the classification data mining technique. Some data mining techniques could perform better than others when involving the domain knowledge [14].

Thomas Piton, Julien Blanchard, Henri Briand and Fabrice Guillet (2009) suggested applying an actionable knowledge discovery approach for marketing by using the right communication channel in contacting the right customer. This approach could satisfy business expectations and improved the efficiency of marketing promotions [21].

S. Sharma and K. Osei-Bryson (2009) investigated the importance of human intelligence in the field of domain driven data mining. The researchers identified twelve processes in data mining that required human intelligence to discover more significant results [20].

H. Zhao, A. P. Sinha and W. Ge Zhao (2009) studied the impact of involving the domain knowledge during feature construction process. Four data mining techniques were applied, and the results showed that involving domain experts could improve classifier performance and the improvement differs across the four techniques [18].

Adeyemi Adejuwon and Amir Mosavi (2010) examined how domain driven data mining approach could be applied to businesses to extract more significant results. They assessed three case studies and showed the usefulness of applying domain knowledge besides the data mining techniques in the business domain [26].

Longbing Cao and Yanchang Zhao (2010) proposed four types of frameworks for Actionable Knowledge Discovery: "Post-analysis-based AKD, Unified-Interestingness-based AKD, Combined-Mining-based AKD, and Multisource Combined-Mining-based AKD (MSCM-AKD)". Experiments showed these frameworks were so flexible and useful to deal with difficult business problems by discovering actionable deliverables for making decisions [27].

Longbing Cao (2010) proposed that ubiquitous intelligence should be incorporated into the mining process including "in-depth data intelligence, human intelligence, domain intelligence, network intelligence, and organizational/social intelligence." He presented theoretical frameworks, architectures, techniques, case studies, and challenges of D^3M [25].

A. Tejaswi, J.N.V.V.S. Prakash, A. Manaswi, G. Sprinivas, and J.N.V.R. Swarup Kumar (2010) introduced the design for intelligent decision support systems to support different e-services business domains. Applying Post Analysis-based AKD (PA-AKD) approach improved the efficiency of the extracted patterns that gave better market value [23].

Ashima Khanna1 and Zoya Siddiqui (2011) aimed to identify several directions, issues for research, and areas of application related to D^3M. Latest methodological, technical, and practical progress in D^3M was reviewed. To prune and filter discovered rules for developing actionable strategies, an interactive approach with domain experts might be required [28].

Mitu Kumari (2011) tried to provide a new approach of data mining known as Domain-Driven Data Mining which was used to promote the need for the shift from data-driven hidden pattern mining to domain-driven actionable data discovery. A methodology for mining actionable knowledge through human collaboration in a repetitive enhancement manner was presented [30].

Ambikavathi.V, Veeraiah.A, and Prabhu.R (2012) viewed actionable knowledge discovery (AKD) and proposed Multisource Combined-Mining-based AKD (MSCM-AKD) framework. A real-life case study of MSCM-based AKD was displayed to discover patterns for debt prevention from social security data [32].

V.Vijay1 and M.Satyanarayana (2012) proposed an Actionable Association Rule (AAR) model called Multisource Combined-Mining-based AAR (MSCM-AAR). A real-life case study of MSCM-based AAR was investigated to discover patterns for debt prevention from social security data. Experiments showed the effectiveness of this model design in finding actionable deliverables for decision making process [33].

P.Sridevi and N.Venkata Subba Reddy (2013) proposed a framework that combined existing methods or algorithms to integrate multiple data sources, multiple features, and multiple mining methods. The patterns combined provided actionable knowledge. A prototype application was developed for testing the proposed framework which combined various mining approaches and merged all the patterns by a component known as pattern merger. The results showed that this approach was efficient and could be applied to real applications [36].

Suvarna R. Bhagwat (2013) reviewed combined mining approach in real life applications. It included designing multi-feature, multi-method, and multi-source approaches. The resultant combined patterns were those evaluated from various sources and included features from sources from where they were discovered. In addition, many methods such as classification, clustering, prediction, association rules mining could be applied to mine the same data [37].

Er. Amarjeet Kaur, Er. Kumar Saurabh, and Er. Gurpreet Singh (2013) proposed an approach using Rule Induction and Association Rule mining in Data Mining to get accurate exact results faster. They suggested using the CN2 as a learning algorithm for rule induction and using the Apriori algorithm for association rules mining. Their research could decrease the number of generated rules and the error rate with better data coverage and fast processing time from a large dataset [40].

Dr.S.S.Dhenakaran and S.Maheswari (2013) proposed a new algorithm for mining actionable patterns from large datasets. Fuzzy based approach was used in the algorithm to get better performance. The researchers compared it with FP-growth, Apriori, and Fuzzy based association rule mining (FARM) and it was performed better than earlier algorithms [35].

K. Priya Karunakaran (2013) reviewed three different Domain Driven Data Mining papers. Domain driven data mining was proposed in the first paper as a methodology and techniques to deliver domain driven actionable knowledge that used to solve business problems. He highlighted while reviewing the second paper how developed domain driven data mining tools and techniques were used for the discovery of actionable knowledge for several data-mining problems. An application to score intelligent credit was reviewed in the third paper using domain driven data mining techniques [38].

Stefan Strohmeier and Franca Piazza (2013) reviewed HR data mining on their research to discuss current research and propose new future research. An initial framework based on domain-driven requirements was discussed and relevant pieces of research were reviewed while discussing this framework. Based on their review, HRM represents an important data mining domain research based on technology and domain requirements [39].

Abdelaziz Araar and Amira A. El Tayeb (2013) analyzed road traffic accidents data in the emirate of Dubai, UAE. A dataset covering accidents between 2008 and 2010 were collected. Four classification methods (Decision trees, Rules induction, BayesNet, and MultilayerPerceptron) were applied and compared. Experimental results showed that accidents could be classified with reasonable accuracy and the neural networks classifier (MutilayerPerceptron algorithm) was the best classifier for all classes [34].

Madeeha Aslam, Ramzan Talib, and Humaira Majeed (2014) reviewed randomly chosen eleven articles in domain driven data mining field. In their review, they emphasized on the need to shift from data centered knowledge discovery to domain-driven knowledge discovery and the role of domain-driven data mining in real world businesses. A case study was reviewed that showed the effectiveness and efficiency of involving domain experts' knowledge in addition to the traditional data mining techniques to improve the results obtained [41].

Amira A. El Tayeb, Vikas Pareek, Abdelaziz Araar (2015) applied Apriori and Predictive Apriori association rules algorithms on a traffic accident dataset collected from the Dubai police authority. They discovered the link between accident severity and accidents' factors by generating two sets of class association rules and summarized them using Rule Covers method to find out the most interesting rules using technical measures. Empirical results showed that the Class Association Rules discovered by Apriori algorithm were more interesting than those generated by Predictive Apriori algorithm since more associations between the severity of accident level and accident factors were discovered [44].

Liling Li, Sharad Shrestha, Gongzhu Hu (2017) analyzed roadway traffic FARS Fatal Accident dataset to discover variables related to fatality. The authors studied the relationship between collision manner, weather, surface condition, light condition, drunk driver, and fatal rate. Association rules, classification, and clustering data mining techniques were used in investigation. Apriori algorithm was used to discover association rules. Naive Bayes classifier was used to build classification model. Simple K-means clustering algorithm was used to form clusters. Based on results obtained from the different techniques, safety driving suggestions were proposed [45].

Vasavi S. (2018) investigated road traffic accident data of major highways of Krishna district for the year 2013. Machine learning techniques were applied into analysis. Environmental conditions such as traffic on the road, type of road, weather, load in the vehicle, driver's health condition, speed, and driver's emotions such as sad, happy, and anger as reasons of accidents were studied. K-medoids algorithm was applied to form clusters and maximization algorithms were analyzed using apriori algorithm to extract hidden patterns. Based on the results, the selected machine learning techniques could discover hidden patterns from the data. Accident data were visualized using density histograms [46].

R. Batra and M. A. Rehman (2019) investigated Actionable Knowledge Discovery approaches to extract and discover the significant business and technical actions/patterns that support decision making to increase enterprise profit. The authors aimed to propose a work that generates efficient actionable patterns. Experiments were carried out on four datasets retrieved from UCI Machine learning repository. Experimental results showed that the proposed work took less time than other two compared methods to extract actions for all datasets. In addition, the number of rules required to generate actions are less than the other two methods [47].

Plotnikova V, Domas M, Milani F (2020) addressed the question of whether data mining methodologies are used 'as-is' or are adapted for specific purposes or problems. The authors covered a literature review of 270 peer reviewed and grey publications. They found that most data mining techniques are applied 'as-is'. They also identified various data mining adaptations. The authors identified technological and organizational adaptations. They suggested that existing data mining methodologies needed refinements to combine data, technological, and organizational aspects to mitigate the gaps [49].

Fakeeha Fatima, Ramzan Talib, M. Hanif, M. Awais (2020) presented their Process-based Domain-Driven Data Mining-Actionable Knowledge Discovery (PD^{3}M-AKD) framework. Their research objective was to improve the actionability of learned rules. This framework included additional factors from five perspectives of the business process in its different phases such as what overall actions were being performed in the business process, the sequence in which tasks were being performed, who performed the tasks, the conditions the tasks were being performed under, and what data was provided. The case study results were evaluated and validated from different real-life domains scenarios such as engineering, education, and business process domains. Results showed that the actionability of learned rules was improved when considering process relevant factors from the above five perspectives of a business process compared to the rules learned from dataset or domain knowledge [48].

Hong Chen, Yang Zhao, and Xiaotong Ma (2020) aimed through their study to minimize the influence of severe traffic accidents in China by analyzing the relations among accident factors that contributed to traffic accidents caused by single-vehicle and multivehicle. The Bayesian network (BN) crash severity model was used. To validate the BN model, severe traffic accident data collected from accident reports published in China were used in the study. The efficiency of model was validated by comparing the conditional probability obtained by the BN model with the actual value. Research results showed that the BN model could reflect the relations among accident factors for the severe traffic accidents in China. In addition, three-factor combination sequences for the number of injuries and five-factor combination sequences for the number of deaths based on BN's junction tree engine were ranked according to the degree of severity to discover the critical reasons and reduce the massive damage of traffic accidents [50].

Antonio Comi, Antonio Polimeni, and Chiara Balsamo (2021) investigated the most effective measures in analyzing accidents to identify and classify the causes that can cause an accident. This study used data mining and clustering approaches for the analysis of accident data for the 15 districts of Rome Municipality. The data collected covered the years 2016 to 2019. The aim of this research was to find out which data mining techniques were more suitable for the analysis of road accidents. Also, a model to predict road accidents was proposed. Results showed that such analyses could be powerful to plan suitable model and measures to predict accidents and reduce them. For data cleaning and preprocessing, spreadsheet software program and geographic information system (GIS) were used. For mining data, R-project software was used [53].

Mohamad Aljaban (2021) aimed to explore the key factors contributing to the increase of the rate of car accidents. In this research, the dataset used was collected from traffic accidents events taken by the department of transportation, law-enforcement agencies, and traffic cameras for years 2016 to 2020 in the United States. To predict the car accidents impact on road traffic, the Naive Bayes and the Random Forest algorithms were implemented focusing on the primary factors contributing to road accidents. It was concluded that Random Forest algorithm outperformed the Naive Bayes algorithm. Research results showed that car accidents rate was affected by population density and work rush hour traffic [57].

Yasin J. Yasin1,2, Michal Grivna1 and Fikri M. Abu-Zidan3 (2021) aimed to review the effects of the COVID-19 pandemic on the incidence, frequency, severity of injury, management, and RTCs outcomes. The authors aimed to give recommendations to improve road safety during the pandemic. A narrative review on the effects of COVID-19 pandemic on RTCs was conducted by the authors and being published in English language using PubMed, Scopus, and Google Scholar without date restriction. They also used Google search engine and websites to retrieve published literature including reports, media news, and discussion papers. Retrieved papers were critically read, and data were summarized and combined. It was noticed that there was a drop in traffic volume during COVID-19 pandemic as well a drop in RTCs globally. In April 2020, there was a decrease in road deaths in 32 out of 36 countries in a comparison with April 2019. There was a reduction of 50% or more in 12 countries, 25 to 49% in 14 countries, and by less than 25% in six countries. In addition, there was a drop in annual road death in 33 out of 42 countries in year 2020 in comparison with 2019, with a decrease of 25% or more in 5 countries, 15 to 24% in 13 countries, and by less than 15% in 15 countries. The number of admitted patients in trauma centers related to RTCs dropped during both periods. Emptier traffic lanes, increase in speeding, reduction in law enforcement, drug abuse, alcohol abuse, and not wearing seat belts occurred [54].

Lei Lin1, Feng Shi2 & Weizi Li3 (2021) investigated the pandemic impact and following mobility changes on road traffic safety. The authors used traffic accident data from Los Angeles and New York Cities. They found that the pandemic impact was not a reduction in traffic and accidents, but rather the ratio of accidents increased unexpectedly for "Hispanic" and "Male" groups, the "hot spots" of accidents shifted in both space and time and were moved from higher-income areas as Lower Manhattan and Hollywood to lower-income areas as southern Brooklyn and southern LA, and the severity level of accidents decreases with the number of accidents regardless modes of transportation. Understanding those variations of traffic accidents not only could shade a light on the diverse impact of COVID-19 across demographic and geographic factors but could also help policymakers and planners design more effective safety policies and interventions during critical conditions such as the pandemic [52].

Nuntaporn Klinjun, Matthew Kelly, Chanita Praditsathaporn, and Rewwadee Petsirasan (2021) presented the result of road traffic investigation reports and determined patterns of risk factor for road traffic injuries in Thailand. This study covered traffic incidents occurring between November 2006 and April 2019. Forensic reports included 25 serious traffic accident events. To analyze risk factors in three phases stratified by four agents, the Haddon matrix was used. The 25 events analyzed involved 47 vehicles and 407 victims. 14.5% of victims died in a total of 65.8% of injured victims. 66.1% of deaths occurred at the scene. Human related factors included drowsiness and speeding. Passenger risks included sitting in cargo area, cab of pickups, and not using the seat belt. Vehicular risks included unsafe cars modifications, overloaded vehicles, having unfixed seats, and no safety equipment. Environmental risks included no traffic lights, no guard rails, fixed objects on the roadside, no traffic signs, and road accident black spots. The outcome of authors' paper identified several risk factors to prevent traffic injury and examined them in conjunction [56].

Abbas Sheykhfard, Farshidreza Haghighi, Eleonora Papadimitriou, and Pieter Van Gelder (2021) investigated pedestrian safety. This study evaluated the methods used by earlier researchers to determine the advantages, characteristics, and limitations of each method. Two analysis approaches (passive and active) were considered to classify 169 previous types of research. The studied methods were based on crash databases, questionnaires, and post-crash field observation data in the passive approach. On the other hand, the studied methods were based on driving simulations in the active approach. The results of the passive approach showed that road users' features, road characteristics, intentional and unintentional violations were among the most significant causes of crashes. The results of the active approach showed that risky behaviors such as unauthorized speeding, illegal crossing, unauthorized overtaking, non-compliance with traffic law are the most significant factor in threatening pedestrian safety. In this study, the findings would lead to better understanding of the road users' behavior for studies on advanced driving assistance systems (ADAS) [55].

Dhanya Viswanath, Preethi K, Nandini R, Bhuvaneshwari R (2021) studied the relationships between road traffic accidents, road condition, and environmental factors in the occurrence of an accident. The authors used data mining techniques to develop an accident prediction model using both Apriori algorithm and Support Vector Machines. Bangalore road accident datasets available on the internet for the years 2014 to 2017 was used for this study. The results obtained from this study can be widely used by the government public work departments, automobile industries, and contractors to better design roads and vehicles [51].

Maowei Chen, Lele Zhou, Sangho Choo, and Hyangsook Lee (2022) analyzed the driver factors, environmental factors, and traffic condition factors contributed to the severity of truck traffic accidents. The authors investigated Truck traffic accidents since the fatality rate was higher than that of other traffic forms. The authors presented solutions for truck safety based on the findings. The study examined the serious injury and fatal traffic accidents of trucks in Incheon which had the highest rate of fatality in Korea's capital area. Based on the regression analysis results, 'Vehicle-single', 'Nighttime', 'Lane violation', 'Vehicle to Pedestrian', and 'Signal violation' contributed to the severity of truck traffic accidents. Moreover, truck traffic accidents which occurred in the logistics influencing area were more serious than those occurred in the non-influencing area. The maximum speed, the number of lanes, and the number of road property-changing nodes were the most significant traffic condition factors. The results provided useful data for setting traffic safety policies in urban areas [58].

SIGNIFICANCE OF THE RESEARCH

The significance of the research is in proposing a new domain-driven actionable knowledge discovery approach that can enhance the traditional data mining techniques results and generate the best actionable patterns those are of great significance to the domain. These patterns can be used in solving real-life problems related to the traffic accident field. Figure 1 depicts the methodology used in the research.

As shown in Figure 1, data related to traffic accidents were collected from the Dubai police authority. Several interviews with expert domain were conducted [9]. Collected data were then cleaned and preprocessed. Data pre-processing is an essential part of data preparation. It involves many tasks such as feature selection and creation and handling of missing values [35]. Most tasks were domain-driven rather than data mining tool-driven. The data mining tool, WEKA, was used to apply classification data mining technique on the dataset. The cleaned dataset to be used contained several accident-related attributes, driver-related attributes, and road-related attributes. Class labels were created based on the real data. The class label ('Accident Severity') has three nominal values: ('Death', 'Severe', and 'Moderate') [34].

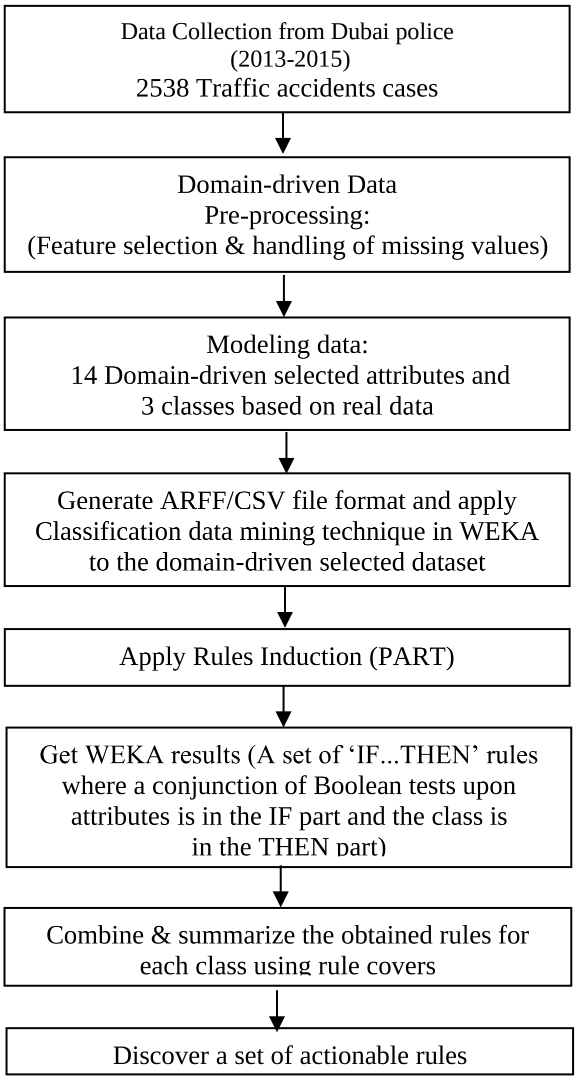

Figure 1. Summary of Research Methodology

DDID-PD FRAMEWORK

Domain-Driven In-Depth Pattern Discovery (DDID-PD) is a framework in domain driven data mining. In this framework, the actionable knowledge discovery is an iterative interactive in-depth pattern discovery process in a specific domain perspective [9][11]. It includes mining constraint-based domain where constraints can be economic, technical, and social aspects involved during the development and deployment of actionable knowledge [9][11]. The involvement of the knowledge of domain experts can help in discovering in-depth patterns which are actionable and more interesting from the domain point of view [9][11]. The domain experts' feedback can iteratively refine the data mining life cycle [9][11]. It includes the following steps as shown in figure 2.1 [9][11]: problem understanding, constraints analysis, data understanding and feature construction, data preprocessing, modeling/in-depth modeling, results evaluation, actionability enhancement, results postprocessing, deployment, and knowledge and report delivery. These steps are iterative and can involve domain experts' knowledge as illustrated in Figure 2 [9].

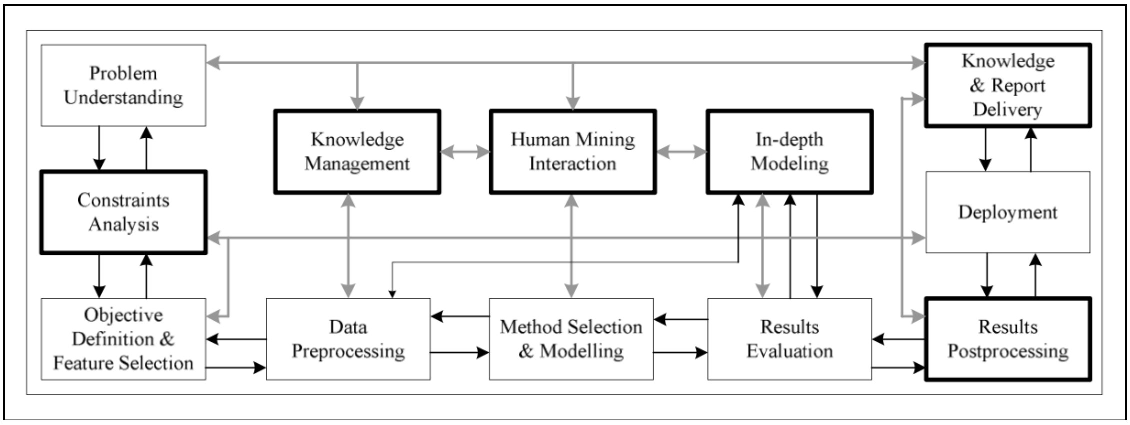

Figure 2. DDID Process Model

RULES INDUCTION CLASSIFIER (PART ALGORITHM)

PART stands for Projective Adaptive Resonance Theory. It is a rule-based algorithm in WEKA that produces a set of if-then rules that can be used to classify data. It is a modification of C4.5 and Repeated Incremental Pruning to Produce Error Reduction (RIPPER) algorithms and draws strategies from both. PART adopts the divide-and-conquer strategy of RIPPER and combines it with the decision tree approach of C4.5.

PART is an indirect method for building classification rules since it extracts rules from other classification models as decision trees not directly from data. To generate a single rule, PART builds a partial decision tree for the current set of instances and chooses the leaf with the largest coverage as the new rule. It is different from C4.5 because the trees built for each rule are partial, based on the remaining set of examples do not complete as in case of C4.5 [29]. In this research, the PART algorithm in WEKA is applied on the accident dataset to represent the knowledge/pattern identified.

EXPERIMENTAL SETUP

Our research is based on real-life traffic accidents collected from Dubai Traffic Department covering the years 2013 to 2015. We collected 2538 records from Dubai police covering all five classes of accidents' severity (death, severe, moderate, minor, and no injury) for the three years 2013 to 2015. However, we only focused in this research on discovering actionable rules related to three types of classes; death, severe, and moderate since these types of traffic accidents led to most death, injuries, and property damage. Best classification rules were generated based on technical interestingness, combined, or summarized to get the final set of actionable rules.

Domain-Driven Data Collection and Preprocessing

Traffic Accidents records were collected for the years 2013, 2014, 2015 from Dubai police authority. Out of 2538 collected records, only 1557 records were selected after ignoring incomplete records during data preprocessing. The collected records covered 98 records for death, 122 records for severe, and 408 records for moderate. All incomplete/ missing data were removed. The total number of accidents cases experimented was 627 records. The records experimented covered 98 records for death, 122 records for severe, and 407 records for moderate cases. Our focus in this research was to cover three classes of severity of accidents; Death, Severe, and Moderate. Table depict description of relevant attributes and class labels respectively. Table 1 depicts number of accidents cases per class experimented for years 2013 To 2015.

Accident Year	Number of Accidents Cases Per Class Experimented			
	Death	**Severe**	**Moderate**	**Total**
2013	23	31	95	**149**
2014	26	27	113	**166**
2015	49	64	199	**312**
Total	**98**	**122**	**407**	**627**

Table 1. Number of Accidents Cases per Class Experimented for Years 2013 To 2015

Selection of Attributes and Classes

All selected attributes and classes are derived from the real data collected from the Dubai police, Traffic Accident Department. Fourteen attributes were selected by the domain to be the focus of our study. These attributes were related to accident, road, and driver status during the accident. Table 2 illustrates all the attributes and the description of each. Table 3 describes the class label (‘Accident Severity’) which has three nominal values: (‘Death’, ‘Severe’, or ‘Moderate’).

Attribute Name	Description
Accid_year	Year of accident
Accid_month	Month of accident
Accid_day	Day of accident
Accid_type	Type of accident
Accid_cause	Cause of accident
Weather	The weather condition (good, rainy, fog, sand, or others)
Road_Surface	The road surface condition (dry, wet, sandy, oily, or others)
Driv_Age	Age of the driver
Driv_Nationality	Nationality of the driver
Driv_Gender	Gender of the driver
Driv_Experience	Driver’s driving experience
Driv_Seatbelt	Driver’s seatbelt status during the accident
Driv_Car_Type	Driver’s vehicle type causing the accident
Car_Prod_Year	Driver’s vehicle production year

Table 2. Description of relevant attributes

Class Label (Accident Severity)	Class Description
Death	One or more persons dies within 30 days of the accident.
Severe	A person is injured and requires intensive care.
Moderate	One or more persons injured and detained in hospital for more than twelve hours.

Table 3. Description of class labels

SUMMARIZING AND COMBINING GENERATED RULES

Best rules were first generated for death, severe, and moderate classes by WEKA. The summarized final sets of rules for each class were then generated using rule covers method and aggregation of rules using OR operation. A rule covers method is applied to summarize the best rules generated by PART algorithm. It summarizes the rules by removing shorter rules covered in bigger rules [10]. Also, rules were aggregated using OR operation when one antecedent in one side of the rules is the same in other rules, other antecedents are the same with different values, and these rules have the same accident class in the other side.

To determine the expectedness and actionability of the rules, the interestingness of the summarized rules from both technical and business perspectives was then determined based on a new proposed interestingness method used in this research to determine which rules were actionable based on number of attributes or accident factors on the antecedents’ side of each rule.

(IJRST) 2022, Vol. No. 12, Issue No. II, Apr-Jun e-ISSN: 2249-0604, p-ISSN: 2454-180X

The rules which had five or more attributes or antecedents were considered as Unexpected Actionable rules. The rules which had four attributes or antecedents were considered as Expected Actionable rules. The rules which had three or less attributes or antecedents were considered as Expected Unactionable rules. This method was based on the concept that the greater number of accidents factors occurred together would be unexpected and actionable from the domain point of view. This method could help Dubai Traffic Accidents department officials set new traffic rule or policies.

EXPERIMENTAL RESULTS

PART is a separate-and-conquer rule learner in WEKA. It goes through much iteration, builds a partial C4.5 decision tree in each one, and makes the “best” leaf into a rule. [59] This algorithm is a combination of C4.5 and RIPPER rule learning. Rules induction (PART) is applied to real life Dubai police dataset covering the years 2013-2015. The accuracy of this algorithm is 63.5196 %. The time taken to test model on training data was 0.04 seconds. PART algorithm was applied to Dubai Accidents dataset covering the years 2013 to 2015. Results obtained from Rules Induction (PART) are described. The set of best rules generated are shown in the following sections for all classes. The numbers in (parentheses) at the end of each leaf or rule tells the number of examples in that leaf. If one or more leaves were not pure (= all of the same class), the number of misclassified examples would be given, after a /slash/. [59]

Analysis of Rules for Death Class

The best rules generated for death class using PART algorithm are shown in table 4. Summarized final rules, number of antecedents, and interestingness of rules are shown in table 5.

Rule#	Best Rules
1	If Driv_Nationality = India AND Driv_Experience = 2_4 years AND Driv_Car_Type = Private ==> Death (4.0/1.0)
2	If Driv_Seatbelt = Fastened AND Driv_Nationality = India ==> Death (5.0/3.0)
3	If Driv_Nationality = Yemen AND Accid_Month = February ==> Death (3.0/1.0)
4	If Driv_Nationality = Iran AND Accid_Type = Vehicle Collision ==> Death (3.0/1.0)
5	If Driv_Nationality = Palestine ==> Death (3.0/2.0)
6	If Driv_Nationality = USA AND Rd_Surface = Asphalt ==> Death (2.0/1.0)
7	If Driv_Nationality = Tunisia AND Driv_Age = 25_29 ==> Death (2.0/1.0)
8	If Driv_Nationality = India AND Driv_Car_Type = Public Transportation ==> Death (3.0/1.0)
9	If Driv_Nationality = Pakistan AND Accid_Cause = Tire Burst ==> Death (4.0/1.0)
10	If Driv_Nationality = Jordan AND Accid_Type = Vehicle Collision ==> Death (4.0/2.0)
11	If Driv_Nationality = Pakistan AND Accid_Cause = Driving in Reverse ==> Death (5.0/3.0)
12	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2013 ==> Death (4.0/2.0)
13	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2003 ==> Death (3.0/1.0)
14	If Driv_Nationality = Pakistan AND Car_Prod_Year = Prod_2000 ==> Death (5.0/2.0)
15	If Driv_Nationality = Syria AND Car_Prod_Year = Prod_2003 ==> Death (3.0/2.0)
16	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2009 ==> Death (9.0/5.0)
17	If Driv_Nationality = Pakistan AND Accid_Cause = Not Keeping Enough Distance AND Car_Prod_Year = Prod_2011 ==> Death (4.0/2.0)
18*	If Driv_Nationality = Pakistan ==> Death (8.0/4.0)
19	If Driv_Nationality = Egypt AND Driv_Age = 40_44 AND Driv_Seatbelt = Unknown ==> Death (3.0)

Table 4. Best Rules for Death Class Using PART Algorithm

Rule#	Summarized Rules	Number of Antecedents	Expectedness & Actionability
1	If Driv_Nationality = India AND Driv_Experience = 2_4 years AND Driv_Car_Type = Private ==> Death (4.0/1.0)	**3**	Expected Unactionable
2	If Driv_Seatbelt = Fastened AND Driv_Nationality = India ==> Death (5.0/3.0)	**2**	Expected Unactionable
3	If Driv_Nationality = Yemen AND Accid_Month = February ==> Death (3.0/1.0)	**2**	Expected Unactionable
4	If Driv_Nationality = Iran AND Accid_Type = Vehicle Collision ==> Death (3.0/1.0)	**2**	Expected Unactionable
5	If Driv_Nationality = Palestine ==> Death (3.0/2.0)	**1**	Expected Unactionable
6	If Driv_Nationality = USA AND Rd_Surface = Asphalt ==> Death (2.0/1.0)	**2**	Expected Unactionable
7	If Driv_Nationality = Tunisia AND Driv_Age = 25_29 ==> Death (2.0/1.0)	**2**	Expected Unactionable
8	If Driv_Nationality = India AND Driv_Car_Type = Public Transportation ==> Death (3.0/1.0)	**2**	Expected Unactionable
9,11	If Driv_Nationality = Pakistan AND Accid_Cause = Tire Burst (4.0/1.0) OR Driv_Nationality = Pakistan AND Accid_Cause = Driving in Reverse (5.0/3.0) ==> Death (9.0/4.0)	**4**	**Expected Actionable**
10	If Driv_Nationality = Jordan AND Accid_Type = Vehicle Collision ==> Death (4.0/2.0)	**2**	Expected Unactionable
12,13, 16	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2013 (4.0/2.0) OR Driv_Nationality = UAE AND Car_Prod_Year = Prod_2003 (3.0/1.0) OR Driv_Nationality = UAE AND Car_Prod_Year = Prod_2009 (9.0/5.0) ==> Death (16.0/8.0)	**6**	**Unexpected Actionable**
14	If Driv_Nationality = Pakistan AND Car_Prod_Year = Prod_2000 ==> Death (5.0/2.0)	**2**	Expected Unactionable
15	If Driv_Nationality = Syria AND Car_Prod_Year = Prod_2003 ==> Death (3.0/2.0)	**2**	Expected Unactionable
17	If Driv_Nationality = Pakistan AND Accid_Cause = Not Keeping Enough Distance AND Car_Prod_Year = Prod_2011==> Death (4.0/2.0)	**3**	Expected Unactionable
19	If Driv_Nationality = Egypt AND Driv_Age = 40_44 AND Driv_Seatbelt = Unknown ==> Death (3.0)	**3**	Expected Unactionable

Table 5. Summarized Rules for Death Class Using PART Algorithm

As stated above, death accidents occurred mostly by drivers from India, Yemen, Iran, Palestine, Syria, Egypt, Pakistan, UAE, Jordan, or Tunisia where seat belt is fastened or unknown, and the accident cause was either Tire Burst, Driving in Reverse, or Not Keeping Enough Distance. The accident type was Vehicle Collision. In addition, death accidents mostly occurred by male Indian drivers driving private vehicles having two to four years of driving experience. The production years for vehicles involved in death accidents were either 2000, 2003, 2009, 2011, or 2013. Drivers' ages were either between 25_29 or between 40_44. When applying rule covers method on the best rules for death class generated using PART algorithm, it was observed that amongst the nineteenth best rules, fifteen summarized rules appeared after eliminating the shorter rules covered in longer ones. One rule was eliminated using rules covers method. Also, five rules were aggregated using OR operation. To determine the number of correctly/incorrectly classified instances in the summarized rule of aggregated rules, this equation was used; Number of correctly/incorrectly classified instances in summarized rule = Sum of correctly classified instances in each aggregated rule/ Sum of incorrectly classified instances in each aggregated rule. Based on proposed interestingness method, there were one Unexpected Actionable rule, one Expected Actionable rule, and thirteen Expected Unactionable rules.

Analysis of Rules for Severe Class

The best rules generated for severe class using PART algorithm are shown in table 6. The set of summarized final rules, number of antecedents, and interestingness of rules are shown in table 7.

Rule#	Best Rules
1	If Accid_Month = October AND Accid_Type = Hitting Road Barrier ➔ Severe (3.0/1.0)
2	If Accid_Month = May AND Accid_Type = Vehicle Collision AND Car_Prod_Year = Prod_2009 ➔ Severe (3.0/1.0)
3	If Driv_Nationality = Russia AND Accid_Day = Saturday ➔ Severe (2.0)
4	If Driv_Nationality = Morocco AND Accid_Year = Year 2015 ➔ Severe (4.0/1.0)
5	If Driv_Nationality = India AND Accid_Month = July ➔ Severe (2.0)
6	If Driv_Nationality = Pakistan AND Accid_Type = Hitting Lamp Post ➔ Severe (5.0/3.0)
7	If Driv_Nationality = Pakistan AND Accid_Cause = Carelessness and Lack of Attention AND Accid_Type = Vehicle Collision ➔ Severe (4.0/1.0)
8	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2014 ➔ Severe (7.0/4.0)
9	If Driv_Nationality = Pakistan AND Accid_Cause = Not Looking Before Entering The Road ➔ Severe (4.0/2.0)
10	If Driv_Nationality = Syria AND Accid_Day = Monday ➔ Severe (2.0)
11	If Driv_Nationality = Pakistan AND Accid_Cause = Sudden Deviation of The Vehicle AND Accid_Month = December ➔ Severe (3.0/1.0)
12	If Driv_Nationality = Pakistan AND Accid_Cause = Not Keeping Enough Distance AND Accid_Month = July ➔ Severe (7.0/3.0)
13	If Driv_Nationality = Egypt AND Driv_Age = 35_39: Severe (3.0)
14	If Driv_Nationality = UAE AND Accid_Cause = Sudden Deviation of The Vehicle AND Accid_Type = Vehicle Collision ➔ Severe (7.0/4.0)

Table 6. Best Rules for Severe Class Using PART Algorithm

Rule#	Summarized Rules	Number of Antecedents	Expectedness & Actionability
1	If Accid_Month = October AND Accid_Type = Hitting Road Barrier ➔ Severe (3.0/1.0)	**2**	Expected Unactionable
2	If Accid_Month = May AND Accid_Type = Vehicle Collision AND Car_Prod_Year = Prod_2009 ➔ Severe (3.0/1.0)	**3**	Expected Unactionable
3	If Driv_Nationality = Russia AND Accid_Day = Saturday ➔ Severe (2.0)	**2**	Expected Unactionable
4	If Driv_Nationality = Morocco AND Accid_Year = Year 2015 ➔ Severe (4.0/1.0)	**2**	Expected Unactionable
5	If Driv_Nationality = India AND Accid_Month = July ➔ Severe (2.0)	**2**	Expected Unactionable
6	If Driv_Nationality = Pakistan AND Accid_Type = Hitting Lamp Post ➔ Severe (5.0/3.0)	**2**	Expected Unactionable
7	If Driv_Nationality = Pakistan AND Accid_Cause = Carelessness and Lack of Attention AND Accid_Type = Vehicle Collision ➔ Severe (4.0/1.0)	**3**	Expected Unactionable
8	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2014 ➔ Severe (7.0/4.0)	**2**	Expected Unactionable
9	If Driv_Nationality = Pakistan AND Accid_Cause = Not Looking Before Entering The Road ➔ Severe (4.0/2.0)	**2**	Expected Unactionable
10	If Driv_Nationality = Syria AND Accid_Day = Monday ➔ Severe (2.0)	**2**	Expected Unactionable

11	If Driv_Nationality = Pakistan AND Accid_Cause = Sudden Deviation of The Vehicle AND Accid_Month = December (3.0/1.0) ➔ Severe (3.0/1.0)	**3**	Expected Unactionable
12	If Driv_Nationality = Pakistan AND Accid_Cause = Not Keeping Enough Distance AND Accid_Month = July (7.0/3.0) ➔ Severe (7.0/3.0)	**3**	Expected Unactionable
13	If Driv_Nationality = Egypt AND Driv_Age = 35_39 ➔ Severe (3.0)	**2**	Expected Unactionable
14	If Driv_Nationality = UAE AND Accid_Cause = Sudden Deviation of The Vehicle AND Accid_Type = Vehicle Collision ➔ Severe (7.0/4.0)	**3**	Expected Unactionable

Table 7. Summarized Rules for Severe Class Using PART Algorithm

As stated above, severe accidents occurred mostly by drivers from Russia, Morocco, India, Pakistan, UAE, Syria, Egypt where the accident cause was either Carelessness and Lack of Attention, Not Looking Before Entering The Road, Sudden Deviation of The Vehicle, or Not Keeping Enough Distance. The accident type was either Vehicle Collision or Hitting Lamp Post. In addition, severe accidents mostly occurred during the months of October, May, or July and Accident Day was either Saturday or Monday. The ages of Egyptian Drivers were between 35_39. When applying rule covers method on the best rules for severe class generated using PART algorithm, it was observed that amongst the fourteen best rules, all rules appeared. No shorter rules were eliminated by being covered in longer rules. Also, no rules were aggregated using OR operation. Based on proposed interestingness method, there were no Unexpected Actionable rules, no Expected Actionable rules, and fourteen Expected Unactionable rules.

Analysis of Rules for Moderate Class

The best rules generated for moderate class using PART algorithm are shown in table 8. The set of summarized final rules, number of antecedents, and interestingness of rules are shown in table 9.

Rule#	**Best Rules**
1	If Accid_Month = October AND Accid_Type = Hitting Lamp Post ➔ Moderate (5.0/1.0)
2	If Accid_Month = May AND Driv_Car_Type = Motorcycle ➔ Moderate (10.0/2.0)
3	If Accid_Month = May AND Car_Prod_Year = Prod_1999 ➔ Moderate (3.0/1.0)
4	If Driv_Nationality = Sri Lanka AND Driv_Experience = 2_4 years ➔ Moderate (4.0/1.0)
5*	If Accid_Month = October ➔ Moderate (15.0/9.0)
6	If Driv_Nationality = Palestine AND Accid_Day = Monday ➔ Moderate (4.0/1.0)
7=	If Driv_Nationality = India AND Car_Prod_Year = Prod_2011 ➔ Moderate (7.0/3.0)
8-	If Driv_Nationality = Oman ➔ Moderate (17.0/9.0)
9	If Driv_Nationality = India AND Accid_Cause = Sudden Deviation of The Vehicle AND Driv_Experience = 2_4 years ➔ Moderate (13.0/2.0)
10-	If Driv_Nationality = Philippines ➔ Moderate (6.0/2.0)
11=	If Driv_Nationality = India AND Car_Prod_Year = Prod_2012 ➔ Moderate (5.0/2.0)
12*	If Accid_Month = May ➔ Moderate (27.0/16.0)
13-	If Driv_Seatbelt = Fastened AND Driv_Nationality = Lebanon ➔ Moderate (7.0/3.0)
14-	If Driv_Seatbelt = Fastened AND Driv_Nationality = Iraq ➔ Moderate (4.0/2.0)
15-	If Driv_Nationality = Afghanistan ➔ Moderate (4.0/2.0)
16	If Driv_Nationality = Yemen AND Accid_Year = Year 2015 ➔ Moderate (3.0)
17	If Driv_Nationality = Nepal AND Car_Prod_Year = Prod_2014 ➔ Moderate (4.0/1.0)
18-	If Driv_Nationality = KSA ➔ Moderate (3.0/1.0)
19-	If Driv_Nationality = USA ➔ Moderate (2.0)
20	If Driv_Nationality = Bangladesh AND Accid_Cause = Sudden Deviation of The Vehicle ➔ Moderate (14.0/7.0)
21	If Driv_Nationality = Bangladesh AND Car_Prod_Year = Prod_2013 ➔Moderate (5.0/2.0)
22	If Driv_Nationality = India AND Accid_Day = Tuesday ➔ Moderate (7.0/2.0)
23-	If Driv_Nationality = Tunisia ➔ Moderate (3.0/1.0)
24	If Driv_Nationality = India AND Accid_Year = Year 2015 ➔ Moderate (69.0/41.0)

25	If Driv_Nationality = Egypt AND Driv_Seatbelt = Not Fastened AND Driv_Experience = <=1 year ➔Moderate (6.0)
26	If Driv_Nationality = Egypt AND Accid_Type = Vehicle Overturn AND Accid_Year = Year 2015 ➔ Moderate (6.0/2.0)
27	If Driv_Nationality = UAE AND Accid_Type = Hitting a Wall ➔ Moderate (6.0/3.0)
28	If Driv_Nationality = Syria AND Accid_Type = Hitting Road Barrier ➔ Moderate (4.0/1.0)
29==	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2006 ➔ Moderate (11.0/4.0)
30	If Driv_Nationality = UAE AND Accid_Month = June AND Car_Prod_Year = Prod_2009 ➔ Moderate (3.0/1.0)
31	If Driv_Nationality = Pakistan AND Accid_Type = Hitting Cement Barrier ➔ Moderate (9.0/4.0)
32==	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2007 ➔ Moderate (9.0/4.0)
33	If Driv_Nationality = Pakistan AND Accid_Type = Iron Barrier Hit AND Driv_Seatbelt = Unknown ➔ Moderate (4.0/1.0)
34=	If Driv_Nationality = Pakistan AND Driv_Seatbelt = Not Fastened AND Accid_Month = April ➔ Moderate (5.0/1.0)
35=	If Driv_Nationality = Pakistan AND Driv_Seatbelt = Not Fastened AND Accid_Month = January ➔ Moderate (3.0)
36	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2008 AND Accid_Cause = Not Keeping Enough Distance AND Driv_Experience = 5_7 years ➔Moderate (4.0/1.0)
37	If Driv_Nationality = Pakistan AND Accid_Cause = Not Keeping in Lane AND Accid_Month = February ➔ Moderate (3.0)
38=	If Driv_Nationality = Syria AND Car_Prod_Year = Prod_2006 ➔ Moderate (4.0/2.0)
39=	If Driv_Nationality = Syria AND Car_Prod_Year = Prod_2005 ➔ Moderate (2.0/1.0)
40	If Driv_Nationality = Pakistan AND Accid_Cause = Overspeeding AND Accid_Month = August ➔ Moderate (3.0/1.0)
41---	If Driv_Nationality = Pakistan AND Accid_Cause = Sudden Deviation of The Vehicle AND Car_Prod_Year = Prod_2000 ➔ Moderate (3.0/1.0)
42---	If Driv_Nationality = Pakistan AND Accid_Cause = Sudden Deviation of The Vehicle AND Car_Prod_Year = Prod_2008 ➔ Moderate (4.0/2.0)
43---	If Driv_Nationality = Pakistan AND Accid_Cause = Sudden Deviation of The Vehicle AND Car_Prod_Year = Prod_2013 ➔ Moderate (3.0/1.0)
44**	If Driv_Nationality = Pakistan AND Car_Prod_Year = Prod_2003 ➔ Moderate (7.0)
45==	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2015 ➔ Moderate (5.0/2.0)
46==	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2011 ➔ Moderate (8.0/4.0)
47	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2008 AND Accid_Cause = Not Keeping Enough Distance AND Driv_Experience = 8_10 years AND Driv_Gender = Male ➔ Moderate (3.0/1.0)
48	If Driv_Nationality = Pakistan AND Driv_Experience = 8_10 years ➔ Moderate (10.0/4.0)
49**	If Driv_Nationality = Pakistan AND Car_Prod_Year = Prod_2009 ➔ Moderate (6.0/3.0)
50	If Driv_Nationality = Pakistan AND Accid_Cause = Sudden Deviation of The Vehicle AND Accid_Month = September ➔ Moderate (3.0)
51==	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2008 ➔ Moderate (7.0/4.0)
52	If Driv_Nationality = Syria AND Accid_Day = Tuesday ➔ Moderate (2.0/1.0)
53*	If Driv_Nationality = Syria ➔ Moderate (4.0/1.0)
54*	If Driv_Nationality = Pakistan AND Accid_Cause = Sudden Deviation of The Vehicle ➔ Moderate (9.0/5.0)
55	If Driv_Nationality = Pakistan AND Accid_Cause = Not Keeping Enough Distance AND Driv_Age = 25_29 ➔ Moderate (7.0/3.0)
56*	If Driv_Nationality = Pakistan AND Accid_Cause = Not Keeping Enough Distance ➔ Moderate (19.0/11.0)
57*	If Driv_Nationality = Egypt ➔ Moderate (4.0/1.0)

Table 8. Best Rules for Moderate Class Using PART Algorithm

Rule#	Summarized Rules	Number of Antecedents	Expectedness & Actionability
1	If Accid_Month = October AND Accid_Type = Hitting Lamp Post ➔ Moderate (5.0/1.0)	**2**	Expected Unactionable
2	If Accid_Month = May AND Driv_Car_Type = Motorcycle ➔ Moderate (10.0/2.0)	**2**	Expected Unactionable
3	If Accid_Month = May AND Car_Prod_Year = Prod_1999 ➔ Moderate (3.0/1.0)	**2**	Expected Unactionable
4	If Driv_Nationality = Sri Lanka AND Driv_Experience = 2_4 years ➔ Moderate (4.0/1.0)	**2**	Expected Unactionable
6	If Driv_Nationality = Palestine AND Accid_Day = Monday ➔ Moderate (4.0/1.0)	**2**	Expected Unactionable
7,11	If Driv_Nationality = India AND Car_Prod_Year = Prod_2011 (7.0/3.0) OR Driv_Nationality = India AND Car_Prod_Year = Prod_2012 (5.0/2.0) ➔ Moderate (12.0/5.0)	**4**	**Expected Actionable**
8,10,15, 18,19,23	If Driv_Nationality = Oman (17.0/9.0) OR Driv_Nationality = Philippines (6.0/2.0) OR Driv_Nationality = Afghanistan (4.0/2.0) OR Driv_Nationality = KSA (3.0/1.0) OR Driv_Nationality = USA (2.0) OR Driv_Nationality = Tunisia (3.0/1.0) ➔ Moderate (35.0/15.0)	**6**	**Unexpected Actionable**
9	If Driv_Nationality = India AND Accid_Cause = Sudden Deviation of The Vehicle AND Driv_Experience = 2_4 years ➔ Moderate (13.0/2.0)	**3**	Expected Unactionable
13,14	If Driv_Seatbelt = Fastened AND Driv_Nationality = Lebanon (7.0/3.0) OR Driv_Seatbelt = Fastened AND Driv_Nationality = Iraq (4.0/2.0) ➔ Moderate (11.0/5.0)	**4**	**Expected Actionable**
16	If Driv_Nationality = Yemen AND Accid_Year = Year 2015 ➔ Moderate (3.0)	**2**	Expected Unactionable
17	If Driv_Nationality = Nepal AND Car_Prod_Year = Prod_2014 ➔ Moderate (4.0/1.0)	**2**	Expected Unactionable
20	If Driv_Nationality = Bangladesh AND Accid_Cause = Sudden Deviation of The Vehicle ➔ Moderate (14.0/7.0)	**2**	Expected Unactionable
21	If Driv_Nationality = Bangladesh AND Car_Prod_Year = Prod_2013 ➔ Moderate (5.0/2.0)	**2**	Expected Unactionable
22	If Driv_Nationality = India AND Accid_Day = Tuesday ➔Moderate (7.0/2.0)	**2**	Expected Unactionable
24	If Driv_Nationality = India AND Accid_Year = Year 2015 ➔ Moderate (69.0/41.0)	**2**	Expected Unactionable
25	If Driv_Nationality = Egypt AND Driv_Seatbelt = Not Fastened AND Driv_Experience = <=1 year ➔ Moderate (6.0)	**3**	Expected Unactionable
26	If Driv_Nationality = Egypt AND Accid_Type = Vehicle Overturn AND Accid_Year = Year 2015 ➔ Moderate (6.0/2.0)	**3**	Expected Unactionable
27	If Driv_Nationality = UAE AND Accid_Type = Hitting a Wall ➔ Moderate (6.0/3.0)	**2**	Expected Unactionable
28	If Driv_Nationality = Syria AND Accid_Type = Hitting Road Barrier ➔ Moderate (4.0/1.0)	**2**	Expected Unactionable
29,32,45, 46,51	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2006 (11.0/4.0) OR If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2007 (9.0/4.0) OR If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2015 (5.0/2.0) OR If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2011 (8.0/4.0) OR	**10**	**Unexpected Actionable**

	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2008 (7.0/4.0) ➔ Moderate (40.0/18.0)		
30	If Driv_Nationality = UAE AND Accid_Month = June AND Car_Prod_Year = Prod_2009 ➔ Moderate (3.0/1.0)	**3**	Expected Unactionable
31	If Driv_Nationality = Pakistan AND Accid_Type = Hitting Cement Barrier ➔ Moderate (9.0/4.0)	**2**	Expected Unactionable
33	If Driv_Nationality = Pakistan AND Accid_Type = Iron Barrier Hit AND Driv_Seatbelt = Unknown ➔ Moderate (4.0/1.0)	**3**	Expected Unactionable
34,35	If Driv_Nationality = Pakistan AND Driv_Seatbelt = Not Fastened AND Accid_Month = April (5.0/1.0) OR Driv_Nationality = Pakistan AND Driv_Seatbelt = Not Fastened AND Accid_Month = January (3.0) ➔ Moderate (8.0/1.0)	**6**	**Unexpected Actionable**
36	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2008 AND Accid_Cause = Not Keeping Enough Distance AND Driv_Experience = 5_7 years ➔ Moderate (4.0/1.0)	**4**	**Expected Actionable**
37	If Driv_Nationality = Pakistan AND Accid_Cause = Not Keeping in Lane AND Accid_Month = February ➔ Moderate (3.0)	**3**	Expected Unactionable
38,39	If Driv_Nationality = Syria AND Car_Prod_Year = Prod_2006 (4.0/2.0) OR Driv_Nationality = Syria AND Car_Prod_Year = Prod_2005 (2.0/1.0) ➔ Moderate (6.0/3.0)	**4**	**Expected Actionable**
26	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2008 AND Accid_Cause = Not Keeping Enough Distance AND Driv_Experience = 8_10 years AND Driv_Gender = Male ➔ Moderate (3.0/1.0)	**5**	**Unexpected Actionable**
40	If Driv_Nationality = Pakistan AND Accid_Cause = Overspeeding AND Accid_Month = August ➔ Moderate (3.0/1.0)	**3**	Expected Unactionable
41,42,43	If Driv_Nationality = Pakistan AND Accid_Cause = Sudden Deviation of The Vehicle AND Car_Prod_Year = Prod_2000 (3.0/1.0) OR Driv_Nationality = Pakistan AND Accid_Cause = Sudden Deviation of The Vehicle AND Car_Prod_Year = Prod_2008 (4.0/2.0) OR Driv_Nationality = Pakistan AND Accid_Cause = Sudden Deviation of The Vehicle AND Car_prod_Year = Prod_2013 (3.0/1.0) ==> Moderate (10.0/4.0)	**9**	**Unexpected Actionable**
44,49	If Driv_Nationality = Pakistan AND Car_Prod_Year = Prod_2003 (7.0) OR Driv_Nationality = Pakistan AND Car_Prod_Year = Prod_2009 (6.0/3.0) ➔ Moderate (13.0/3.0)	**4**	**Expected Actionable**
47	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2008 AND Accid_Cause = Not Keeping Enough Distance AND Driv_Experience = 8_10 years AND Driv_Gender = Male ➔ Moderate (3.0/1.0)	**5**	**Unexpected Actionable**
48	If Driv_Nationality = Pakistan AND Driv_Experience = 8_10 years ➔ Moderate (10.0/4.0)	**2**	Expected Unactionable
50	If Driv_Nationality = Pakistan AND Accid_Cause = Sudden Deviation of The Vehicle AND Accid_Month = September ➔ Moderate (3.0)	**3**	Expected Unactionable
52	If Driv_Nationality = Syria AND Accid_Day = Tuesday ➔ Moderate (2.0/1.0)	**2**	Expected Unactionable

55	If Driv_Nationality = Pakistan AND Accid_Cause = Not Keeping Enough Distance AND Driv_Age = 25_29 ➔ Moderate (7.0/3.0)	**3**	Expected Unactionable

Table 9. Summarized Rules for Moderate Class Using PART Algorithm

As stated above, moderate accidents occurred mostly by male drivers from Sri Lanka, Palestine, India, Oman, Philippines, Afghanistan, KSA, USA, Tunisia, Yemen, Nepal, Bangladesh, Egypt, UAE, Syria, or Pakistan where the accident cause was Sudden Deviation of The Vehicle, Not Keeping Enough Distance, Not Keeping in Lane, or Overspeeding. The accident type was Hitting Lamp Post, Vehicle Overturn, Hitting a Wall, Hitting Road Barrier, Hitting Cement Barrier, or Iron Barrier Hit. In addition, moderate accidents mostly occurred during the months of May, September, or October. The Accident Day was either Monday or Tuesday. The ages of Pakistani Drivers were between 25 and 29. Driver's car type appeared in one rule and was Motorcycle during the month of May. Cars' production years were 1999, 2003, 2005, 2006, 2007, 2008, 2009, 2011, 2012, 2013, 2014, or 2015. The drivers' experience was between 2 to 10 years. When applying rule covers on the best rules generated using PART algorithm for moderate class, it was observed that amongst the best fifty-seven rules, only thirty-six summarized rules appeared. Six shorter rules were eliminated by being covered in longer rules using rules covers methods. Also, twenty-four rules were aggregated using OR operation. To determine the number of correctly/incorrectly classified instances in the summarized rule of aggregated rules, this equation was used; Number of correctly/incorrectly classified instances in summarized rule = Sum of correctly classified instances in each aggregated rule/ Sum of incorrectly classified instances in each aggregated rule. Based on proposed interestingness method, there were six Unexpected Actionable rules, five Expected Actionable rules, and twenty-five Expected Unactionable rules.

COMPARISON OF GENERATED RULES FOR EACH CLASS

In this section, number of rules generated for PART algorithms is illustrated. Table 10 illustrates the number of best rules generated by PART algorithm and number of summarized rules after applying Rules Covers method and/or Aggregation of Rules using OR operation. As stated above, best, and summarized rules were generated for death, severe, and moderate classes when applying rules Induction technique using PART algorithm. Table 10 and Figure 3 depict best and summarized rules for classes using PART algorithm.

Accident Class	Number of Best Rules	Number of Summarized Rules	Number of Rules Eliminated	Number of Rules Aggregated
Death	19	15	1	5
Severe	14	14	0	0
Moderate	57	36	6	24

Table 10. Best and Summarized Rules for Classes Using PART Algorithm

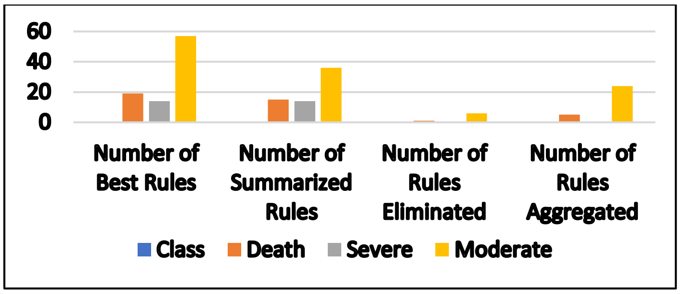

Figure 3. Best and Summarized Rules for Classes Using PART Algorithm

CATEGORIES OF INTERESTINGNESS OF RULES

When presenting the generated rules to the domain experts, it was hard for them to finalize which rules were interesting in terms of expectedness and actionability of rules. The interestingness of the general rules generated by the two data mining techniques was then analyzed and evaluated based on the dataset given from domain experts and classified based on a new proposed method to determine interestingness and actionability of rules into three categories: (1) Unexpected Actionable rules, (2) Expected Actionable rules, and (3) Expected Unactionable rules. These categories lead to the set of actionable and unactionable rules for each class which could be used as indication of how Rules Induction technique could lead to better results based on technical and business interestingness measures. The set of actionable rules from PART algorithm for each class was then summarized to derive the final set of interesting actionable rules/patterns.

The new proposed interestingness method used in this research was used to determine which rules were actionable based on a greater number of attributes or antecedents. The rules which had five or more attributes or antecedents were considered as Unexpected Actionable rules. The rules which had four attributes or antecedents were considered as Expected Actionable rules. The rules which had three or less attributes or antecedents were considered as Expected Unactionable rules.

SN	Category of Interestingness	Number of Antecedents
1	Unexpected Actionable	5+ Antecedents
2	Expected Actionable	4 Antecedents
3	Expected Unactionable	3 or less Antecedents

Table 11. Number of Antecedents for the Five Categories of Interestingness

Table 11 illustrates the number of antecedents suggested in the proposed method for the five categories of interestingness of rules based on the summarized rules for the different classes. This proposed method could help the domain experts to finalize which rules could be put into action afterwards.

To determine interestingness and actionability of rules, table 12 and figure 4 illustrate the number of interesting rules generated by PART algorithm for the five categories.

Summarized Rules	Unexpected Actionable	Expected Actionable	Expected Unactionable
Death	1	1	13
Severe	0	0	14
Moderate	6	5	25

Table 12. Number of Summarized Rules in all Categories Per Class Using PART

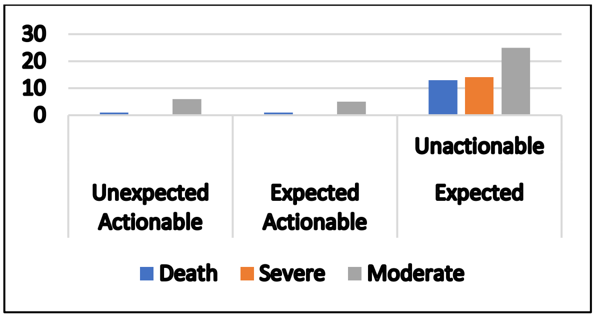

Figure 4. Number of Summarized Rules in all Categories Per Class Using PART

ACTIONABLE RULES USING PART ALGORITHM

To get the final actionable rules generated by PART algorithm, Unexpected Actionable and Expected Actionable Rules for each class were considered. Following sections illustrate all actionable rules for each class using PART algorithm.

Actionable Rules for Death Class

After applying proposed interestingness method on the summarized rules generated for death class using PART algorithm, two rules were actionable. One rule was Unexpected Actionable, and one rule was Expected Actionable. Table 4.25 illustrates actionable rules for death class using PART algorithm.

Rule#	Summarized Rules	Number of Antecedents	Expectedness & Actionability
9,11	If Driv_Nationality = Pakistan AND Accid_Cause = Tire Burst (4.0/1.0) OR Driv_Nationality = Pakistan AND Accid_Cause = Driving in Reverse (5.0/3.0) ➔ Death (9.0/4.0)	**4**	**Expected Actionable**
12,13,16	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2013 (4.0/2.0) OR Driv_Nationality = UAE AND Car_Prod_Year = Prod_2003 (3.0/1.0) OR Driv_Nationality = UAE AND Car_Prod_Year = Prod_2009 (9.0/5.0) ➔ Death (16.0/8.0)	**6**	**Unexpected Actionable**

Table 13. Actionable Rules for Death Class Using PART Algorithm

Actionable Rules for Severe Class

After applying proposed interestingness method on the summarized rules generated for severe class using PART algorithm, no rules were actionable since number of attributes or antecedents was less than 4 for all generated summarized rules. So, no rules were Expected Actionable, and no rules were Unexpected Actionable.

Actionable Rules for Moderate Class

After applying proposed interestingness method on the summarized rules generated for moderate class using PART algorithm, eleven rules were actionable. Six rules were Unexpected Actionable, and five rules were Expected Actionable. Table 4.27 illustrates actionable rules for moderate class using PART algorithm.

Rule#	Summarized Rules	Number of Antecedents	Expectedness & Actionability
7,11	If Driv_Nationality = India AND Car_Prod_Year = Prod_2011 (7.0/3.0) OR Driv_Nationality = India AND Car_Prod_Year = Prod_2012 (5.0/2.0) ➔ Moderate (12.0/5.0)	**4**	**Expected Actionable**
8,10,15, 18,19,23	If Driv_Nationality = Oman (17.0/9.0) OR Driv_Nationality = Philippines (6.0/2.0) OR Driv_Nationality = Afghanistan (4.0/2.0) OR Driv_Nationality = KSA (3.0/1.0) OR Driv_Nationality = USA (2.0) OR Driv_Nationality = Tunisia (3.0/1.0) ➔ Moderate (35.0/15.0)	**6**	**Unexpected Actionable**
13,14	If Driv_Seatbelt = Fastened AND Driv_Nationality = Lebanon (7.0/3.0) OR Driv_Seatbelt = Fastened AND Driv_Nationality = Iraq (4.0/2.0) ➔ Moderate (11.0/5.0)	**4**	**Expected Actionable**
29,32,45, 46,51	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2006 (11.0/4.0) OR If Driv_Nationality = UAE	**10**	**Unexpected Actionable**

	AND Car_Prod_Year = Prod_2007 (9.0/4.0) OR If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2015 (5.0/2.0) OR If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2011 (8.0/4.0) OR If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2008 (7.0/4.0) ➔ Moderate (40.0/18.0)		
34,35	If Driv_Nationality = Pakistan AND Driv_Seatbelt = Not Fastened AND Accid_Month = April (5.0/1.0) OR Driv_Nationality = Pakistan AND Driv_Seatbelt = Not Fastened AND Accid_Month = January (3.0) ➔ Moderate (8.0/1.0)	**6**	**Unexpected Actionable**
36	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2008 AND Accid_Cause = Not Keeping Enough Distance AND Driv_Experience = 5_7 years ➔ Moderate (4.0/1.0)	**4**	**Expected Actionable**
38,39	If Driv_Nationality = Syria AND Car_Prod_Year = Prod_2006 (4.0/2.0) OR Driv_Nationality = Syria AND Car_Prod_Year = Prod_2005 (2.0/1.0) ➔ Moderate (6.0/3.0)	**4**	**Expected Actionable**
26	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2008 AND Accid_Cause = Not Keeping Enough Distance AND Driv_Experience = 8_10 years AND Driv_Gender = Male ➔ Moderate (3.0/1.0)	**5**	**Unexpected Actionable**
41,42,43	If Driv_Nationality = Pakistan AND Accid_Cause = Sudden Deviation of The Vehicle AND Car_Prod_Year = Prod_2000 (3.0/1.0) OR Driv_Nationality = Pakistan AND Accid_Cause = Sudden Deviation of The Vehicle AND Car_Prod_Year = Prod_2008 (4.0/2.0) OR Driv_Nationality = Pakistan AND Accid_Cause = Sudden Deviation of The Vehicle AND Car_prod_Year = Prod_2013 (3.0/1.0) ==> Moderate (10.0/4.0)	**9**	**Unexpected Actionable**
44,49	If Driv_Nationality = Pakistan AND Car_Prod_Year = Prod_2003 (7.0) OR Driv_Nationality = Pakistan AND Car_Prod_Year = Prod_2009 (6.0/3.0) ➔ Moderate (13.0/3.0)	**4**	**Expected Actionable**
47	If Driv_Nationality = UAE AND Car_Prod_Year = Prod_2008 AND Accid_Cause = Not Keeping Enough Distance AND Driv_Experience = 8_10 years AND Driv_Gender = Male ➔ Moderate (3.0/1.0)	**5**	**Unexpected Actionable**

Table 14. Actionable Rules for Moderate Class Using PART Algorithm

COMPARISON OF ACTIONABLE RULES

In this section, number of actionable rules generated for each class using PART algorithms is illustrated. Table 15 illustrates the number of Unexpected Actionable and Expected Actionable rules generated by PART algorithm.

Class	Number of Unexpected Actionable Rules	Number of Expected Actionable Rules
Death	1	1
Severe	0	0
Moderate	6	5

Table 15. Number of Expected and Unexpected Actionable Rules Using PART

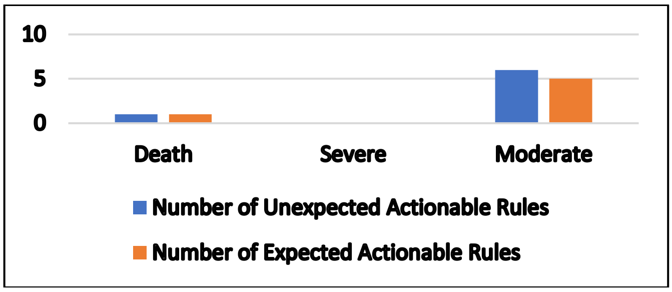

Figure 5. Number of Expected and Unexpected Actionable Rules Using PART

Based on number of Unexpected Actionable and Expected Actionable rules, it was observed that when applying PART algorithm, actionable rules were generated for death and moderate classes. No actionable rules were generated for Severe class.

CONCLUSIONS AND FUTURE WORK

In our research, the focus was on the analysis of road traffic accidents during the years 2013-2015 in Dubai and the investigation of the performance of Rules Induction (PART) algorithm in domain-driven data mining approach. This research presented an approach of using domain-driven data mining techniques to study road traffic accidents. Empirical results showed that PART algorithm could offer sufficient insight and more interesting actionable rules into the problem being studied. Rules generated by the proposed approach were more interesting and actionable than those produced by the traditional classification data mining techniques based on objective interestingness measures. These actionable patterns could assist police decision makers in the formulation of new policies and traffic rules from some hidden in-depth patterns. As future work, this work could be extended to other emirates to assist the UAE police authority in reducing the general traffic accidents impact. The proposed approach and proposed interestingness method could also be applied to other practical domains to generate actionable rules that allow the domain to make decisions and take actions based on the discovered patterns or knowledge. This work can also be extended to study the suitability of domain-driven data mining approach and proposed interestingness method in different domains, investigate the capability of domain experts in analyzing summarized rules to discover interesting actionable rules in other domains, study the traffic accidents in other Emirates and assist the police authorities in reducing the traffic accidents impact in UAE in general, use other domain-driven data mining techniques and algorithms for mining road traffic accidents data in Dubai, other Emirates, and other domains.

ACKNOWLEDGMENT

We would like to thank Dubai Traffic Accident Department staff for their cooperation and providing us with traffic accidents data covering the years 2013 to 2015.

REFERENCES

1) U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, “From Data Mining to Knowledge Discovery in Databases”, AI Magazine Vol. 17 No. 3, 1996, pp. 37–54.
2) Bing Liu, Wynne Hsu, Yiming Ma, “Integrating Classification and Association Rule Mining”, KDD-98, New York, Aug 27-31, 1998.
3) Sigal Sahar, “Interestingness via what is not interesting”, KDD '99 Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 1999, pp. 332-336.
4) Bing Liu, Wynne Hsu, Shu Chen and Yiming Ma, “Analyzing the Subjective Interestingness of Association Rules”, IEEE Intelligent Systems, volume 15 Issue 5, September 2000.
5) T. Brijs, K. Vanhoof, G. Wets, “Defining Interestingness for Association Rules”, International Journal "Information Theories & Applications", Vol.10, 2003.
6) Bener, D. Crundall, “Road traffic accidents in the United Arab Emirates compared to Western countries”, Advances in Transportation Studies an international Journal Section A6, 2005.

7) David A Landgrebe, “Signal Theory Methods in Multispectral Remote Sensing”, 2005.
8) Ken Mcgarry, “A Survey of Interestingness Measures for Knowledge Discovery”, the Knowledge Engineering Review Volume 20 Issue 1, March 2005, Pages 39 – 61.
9) Cao, L., Zhang, C., “Domain-Driven Actionable Knowledge Discovery in the Real World”, PAKDD2006, LNAI 3918, 821-830, Springer , March 2006.
10) Carlos Ordonez, Norberto Ezquerra, Cesar A. Santana, “Constraining and summarizing association rules in medical data”, Knowledge and Information Systems, Volume 9 Issue 3, March 2006, Pages 259 – 283.
11) Longbing Cao and Chengqi Zhang, “Domain-Driven Data Mining: A Practical Methodology”, International Journal of Data Warehousing & Mining, Volume 2, Issue 4, pp.49-65, October-December 2006.
12) Roberta Akemi Sinoara, Solange Oliveira Rezende, “A methodology for identifying interesting association rules by combining objective and subjective measures”, Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial. No. 32, December 2006, pp. 19-27.
13) Magaly Lika Fujimoto, Veronica Oliveira de Carvalho, Solange Oliveira Rezende, “Evaluating Generalized Association Rules Combining Objective and Subjective Measures and Visualization”, In proceeding of: IASTED International Conference on Artificial Intelligence and Applications, part of the 25th Multi-Conference on Applied Informatics, Innsbruck, Austria, February 12-14, 2007.
14) P. Sinha and H. Zhao, “Incorporating Domain Knowledge into Data Mining Classifiers: An Application in Indirect Lending”, Decision Support Systems, Vol.46, 2008, pp.287–299.
15) Chien and L. Chen, “Data Mining to Improve Personnel Selection and Enhance Human Capital: A Case Study in High Technology Industry, Expert Systems with Applications, Vol. 34, 2008, pp. 280–290.
16) Yanchang Zhao, Huaifeng Zhang, Fernando Figueiredo, Longbing Cao, Chengqi Zhang, “Combined Association Rule Mining”, PAKDD'08 Proceedings of the 12th Pacific-Asia conference on Advances in knowledge discovery and data mining, 2008.
17) Yuming Ou, Longbing Cao, Chao Luo, and Chengqi Zhang, “Domain-Driven Local Exceptional Pattern Mining for Detecting Stock Price Manipulation”, in proceeding of: PRICAI 2008: Trends in Artificial Intelligence, 10th Pacific Rim International Conference on Artificial Intelligence, Hanoi, Vietnam, December 15-19, 2008.
18) H. Zhao, A. P. Sinha and W. Ge, “Effects of Feature Construction on Classification Performance: An Empirical Study in Bank Failure Prediction”, Expert Systems with Applications, Vol. 36, 2009, pp. 2633–2644.
19) Longbing Cao, Philip S. Yu, Chengqi Zhang, Huaifeng Zhang, “Data Mining for Business Applications”, Springer Science+Business Media, LLC, 2009.
20) S. Sharma and K. Osei-Bryson, “Role of Human Intelligence in Domain Driven Data Mining”, Data Mining for Business Applications, New York, Springer Science+Business Media, 2009, pp 53-61.
21) Thomas Piton, Julien Blanchard, Henri Briand, Fabrice Guillet, “Domain Driven Data Mining to Improve Promotional Campaign ROI and Select Marketing Channels”, The 18th ACM Conference on Information and Knowledge Management, Hong Kong, 2009.
22) Longbing Cao, Philip S. Yu, Chengqi Zhang, Yanchang Zhao, “Challenges and Trends”, Domain Driven Data Mining”, Springer Science+Business Media, LLC, January 2010, p.1-25.
23) Tejaswi, J.N.V.V.S. Prakash, A. Manaswi, G. Sprinivas, J.N.V.R. Swarup Kumar,” Intelligent Decision System Based on PAAKD Approach of D3M”, International Journal of Engineering Science and Technology, Vol.2 (3), March 2010.
24) Jiying Li., “A Survey on Actionable Knowledge Discovery Applications”, 2nd International Workshop on Intelligent Systems and Applications, 05/2010.
25) L. Cao,” Domain-Driven Data Mining: Challenges and Prospects”, IEEE transactions on knowledge and data engineering, Vol.22, No. 6, June 2010.
26) Adeyemi Adejuwon, Amir Mosavi, “Domain Driven Data Mining- Application to Business”, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No.2, July 2010.
27) L. Cao, Y. Zhao, “Flexible Frameworks for Actionable Knowledge Discovery”, IEEE transactions on knowledge and data engineering, Vol. 22, No.9, September 2010.
28) Ashima Khanna, Zoya Siddiqui, “Domain Driven Data Mining (D3M)”, Proceedings of the 5th National Conference; INDIACom-2011 Computing for Nation Development, March 10-11, 2011.
29) S.Krishnaveni, Dr.M.Hemalatha, “A Perspective Analysis of Traffic Accident using Data Mining Techniques”, International Journal of Computer Applications, June 2011.

30) Mitu Kumari, "Data Driven Data Mining to Domain Driven Data Mining", Global Journal of Computer Science and Technology, Volume 11 Issue 23 Version 1.0, December 2011.
31) C K Bhensdadia, Y P Kosta, "An Efficient Algorithm for Mining Frequent Sequential Patterns and Emerging Patterns with Various Constraints", International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Volume-1, Issue-6, January 2012.
32) Ambikavathi.V, Veeraiah.A, Prabhu.R, "Actionable Knowledge Discovery", International Journal of Computational Engineering Research, Vol. 2 Issue No.1, Jan-Feb 2012.
33) V.Vijay1,M.Satyanarayana, "Actionable Knowledge discovery using MSCAM", International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 6, August 2012.
34) Abdelaziz Araar, Amira A. El Tayeb, "Mining Road Traffic Accident Data to Improve Safety in Dubai", Journal of Theoretical and Applied Information Technology, Vol. 47 No.3, 31st January 2013.
35) Dr. S.S. Dhenakaran and S.Maheswari, "Fuzzy based combined pattern mining algorithm for Domain Driven Data Mining (DDDM)", International Journal of Knowledge Engineering and Research, Vol 2 Issue 1 January 2013.
36) P. Sridevi, N. Venkata Subba Reddy, "Informative Knowledge Discovery using Multiple Data Sources, Multiple Features and Multiple Data Mining Techniques", IOSR Journal of Engineering (IOSRJEN), Vol. 3, Issue 1, January 2013.
37) Suvarna R. Bhagwat, "Combined Mining and Actionable Pattern Discovery Using DDID-PD Framework: A Review", International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue 2, February 2013.
38) K. Priya Karunakaran, "Review of Domain Driven Data Mining", International Journal of Innovations in Engineering and Technology (IJIET), Vol. 2 Issue 3, June 2013.
39) Stefan Strohmeier and Franca Piazza, "Domain driven data mining in human resource management: A review of current research", Expert Systems with applications: An International Journal, Volume 40, Issue 7, June 2013.
40) Er. Amarjeet Kaur, Er. Kumar Saurabh, Er. Gurpreet Singh, "A Combined Approach of Data Mining Algorithms Based on Association Rule Mining and Rule Induction, International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Volume-3, Issue-5, November 2013.
41) Madeeha Aslam, Ramzan Talib, Humaira Majeed, "A Review on the Role of Domain Driven Data Mining", International Journal of Computer Science and Mobile Computing, Vol.3 Issue.5, May 2014, pg. 708-712.
42) http://www.articlesbase.com/cars-articles/the-worlds-worst-drivers-car-accident-statistics-from-around-the-world-609862.html. [Last accessed on 24th August 2014].
43) http://www.marylandinjurylawyerblog.com/2010/09/car_accident_statistics_from_t.html. [Last accessed on 24th August 2014].
44) Amira A. El Tayeb, Vikas Pareek, Abdelaziz Araar, "Applying Association Rules Mining Algorithms for Traffic Accidents in Dubai", International Journal of Soft Computing and Engineering (IJSCE), Volume-5 Issue-4, September 2015.
45) Liling Li, Sharad Shrestha, Gongzhu Hu, "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques", IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), 7-9 June 2017.
46) Vasavi S., "Extracting Hidden Patterns Within Road Accident Data Using Machine Learning Techniques". In: Mishra D., Azar A., Joshi A. (eds) Information and Communication Technology. Advances in Intelligent Systems and Computing, vol 625, Springer, Singapore, 2018.
47) R. Batra and M. A. Rehman, "Actionable Knowledge Discovery for Increasing Enterprise Profit, Using Domain Driven-Data Mining", in IEEE Access, vol. 7, 2019, pp. 182924-182936.
48) Fakeeha Fatima, Ramzan Talib, M. Hanif, M. Awais, "A Paradigm-Shifting from Domain-Driven Data Mining Frameworks to Process-Based Domain-Driven Data Mining-Actionable Knowledge Discovery Framework", 2020.
49) Plotnikova V, Dumas M, Milani F., "Adaptations of Data Mining Methodologies: A systematic Literature Review", PeerJ Comput Sci. 2020; 6:e267, 2020 May 25, doi:10.7717/peerj-cs.267.
50) Hong Chen, Yang Zhao, and Xiaotong Ma, "Critical Factors Analysis of Severe Traffic Accidents Based on Bayesian Network in China", 16 November 2020.
51) D. Viswanath, P. K, N. R and B. R, "A Road Accident Prediction Model Using Data Mining Techniques," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), April 2021, pp. 1618-1623, doi: 10.1109/ICCMC51019, 2021, 9418336.
52) Lei Lin1, Feng Shi2 & Weizi Li3, "Assessing inequality, irregularity, and severity regarding

road traffic safety during COVID-19", Sci Rep 11, 13147, June 23, 2021. https://doi.org/10.1038/s41598-021-91392-z

53) Antonio Comi, Antonio Polimeni, and Chiara Balsamo, "Road Accident Analysis with Data Mining Approach: evidence from Rome", 24th EURO Working Group on Transportation Meeting, EWGT 2021, 8-10 September 2021, Aveiro, Portugal.
54) Yasin J. Yasin1,2, Michal Grivna1 and Fikri M. Abu-Zidan3, "Global Impact of Covid-19 Pandemic on Road traffic collisions", World Journal of Emergency Surgery, 28 September 2021.
55) Abbas Sheykhfard, Farshidreza Haghighi, Eleonora Papadimitriou, Pieter Van Gelder, "Review and assessment of different perspectives of vehicle-pedestrian conflicts and crashes: Passive and active analysis approaches", Journal of Traffic and Transportation Engineering (English Edition), 8(5), 681-702. 21Oct 2021. https://doi.org/10.1016/j.jtte., 2021, 08.001
56) Nuntaporn Klinjun; Matthew Kelly; Chanita Praditsathaporn; Rewwadi Petsirasan," Identification of Factors Affecting Road Traffic Injuries Incidence and Seve9rity in Southern Thailand Based on Accident Investigation Reports. Sustainability 11, November 2021, 13, 12467. https://doi.org/10.3390/ su132212467.
57) Mohamad Aljaban, "Analysis of Car Accidents Causes in the USA", Thesis, Rochester Institute of Technology, 19 December 2021.
58) Maowei Chen, Lele Zhou, Sangho Choo, and Hyangsook Lee, "Analysis of Risk Factors Affecting Urban Truck Traffic Accident Severity in Korea.", Sustainability 2022, 14, 2901. https:// doi.org/10.3390/su14052901.
59) http://mydatamining.wordpress.com