Predicting Road Traffic Accident Severity using Decision Trees and Time-Series Calendar Heatmaps

Charith Silva
School of Computing, Science and Engineering
University of Salford
Greater Manchester, M5 4WT, UK
h.c.e.silva@edu.salford.ac.uk

Mo Saraee
School of Computing, Science and Engineering
University of Salford
Greater Manchester, M5 4WT, UK
m.saraee@salford.ac.uk

Abstract— **The European Commission estimates that around 135,000 people are seriously injured on Europe's roads each year. The road traffic injuries are a significant but neglected global general public health problem, needing rigorous attempts for effective and workable prevention. One of the ways to decrease the amount of traffic accidents is to conduct an in-depth assessment on the historically documented road traffic incident data and understand the cause of the accidents and factors associated with incident severity. It may provide crucial information for emergency services to evaluate the severity level of accidents, estimate the potential impacts of the casualties, and ultimately it might help to improve the road safety. In this study author is trying to identify the factors that correlate with the slight and serious (including fatal) Road Traffic Accident using Decision Tree classification algorithms using UK STATS19 dataset. Also, author is exploring the possibility of enhancing the knowledge gain from Decision Tree classification algorithms using Time-Series Calendar Heatmap in order to identify hidden temporal patterns. The methodology described in this study offers significant advantages over understanding correlation between hour and month of the accident and the severity of the accident. Although this study is based on a region in North of England, the approach can be applicable to other areas in UK and globally with similar kind of road side accident data. This study found out that combining classification methods like decision tree and time-series calendar heatmaps cam be a useful tool for accurately classifying roadside traffic accidents according to their injury severity.**

Keywords— Road Traffic Accidents, Road Safety, Data Mining, Decision Trees, Time-Series Calendar Heatmaps

I. INTRODUCTION

Road traffic accident is one of the undesirable events that is uncertain. Road safety is a major societal issue in worldwide and road accident is one of the main causes of unnatural fatalities, disability, and property harm [1]. Apart from fatalities, the road traffic crashes also cause thousands of slight and serious injuries annually [2]. European roads remain the safest in the world. In 2017, the EU counted 49 road fatalities per one million inhabitants, against 174 deaths per million globally [4]. Sweden, the Netherlands, the UK, and Malta have the lowest reported road fatality rate, below 30 dead per million Inhabitants [2]. According to data from the World Health Organisation, about 1.3 million people die each year on the world's roads, of which 25,300 lost their lives in the EU last year [4]. For every person killed in traffic crashes, about five more suffer serious injuries with life-changing consequences. Serious injuries often costly to the society because of long-time rehabilitation and healthcare needs.

Every day, millions of citizens use road infastructre for walking, riding or driving. Any kind of serious accident on our roads is a disaster. As a responsible member of the community, it is our duty and responsibility to take the road safety measures more seriously to protect the community. In some cases, victims of a road side accident can fully recover from their injuries within a short period of time, while other victims become permanently disable[6]. Road traffic injuries are majorly ignored public health problem, requiring rigorous efforts to provide effective and sustainable protection[7]. One way to reduce the number of traffic accidents is to analyse traffic accidents recorded in depth and to understand the causes of the accidents. Understanding where traffic accidents most happened and identify any peak time widows are crucial for improving road safety and road traffic enforcement officers. One of the most common methods of analysing road traffic accidents is descriptive statistics and time series analysis. But this study author is trying to use Decision Tree algorithm and Time-Series Calendar Heatmaps technique to understand severity of the accident. In addition, this study seeks to evaluate existing data and interpret road side accidents from multiple angles, at various levels and more broadly to discover potentially useful rules to improve road safety. Currently, improving road safety is the main goal of smart transport system and it leads to the importance of safety of the road infrastructure [20]. Accident severity prediction information can provide rescuers with valuable information in response to the severity of accidents and other associated risks [8]. The author believes that closely studying the road traffic accident data with the accident severity and time components of the accident (Month, Weekday, and Hour) can help to identify any hidden temporal patterns.

II. OBJECTIVE OF THE STUDY

1. To identify the factors that correlate with the slight and serious road traffic accident using Decision Tree classification.

2. Explore the possibility of enhancing the knowledge gain from Decision Tree Classification algorithms using Time-Series Calendar Heatmaps.

III. RELATED WORK

In many countries, road accident data is being investigated to examine risk factors that leading to road accidents. Academic and industrial researchers and road safety authorities seek to identify patterns and relationships between risk factors and severity. Among the important attributes that are usually studied in road accident data are: driver characteristics, vehicle characteristics, road condition variables, and weather and environmental factors [16]. Lee et al. [18] investigated the relation between fatality rate and road accident factors with Weather conditions. This study applied Structural Equation Modeling (SEM) to analyse the relationship between the accident severity and rain-related factors. In particular, rain-related factors are an important cause of traffic accidents due to the poor visibility and reduced friction resulting from slippery road conditions. The

study found out that traffic, environmental, and human factors; rain and water depth factors; and road factors are mutually correlated with the level of accident severity. But any of date or time factors like month or time of the accident has not been considered in this study. Isra et al. [16] used Classification to study and understand the factors leading to car accidents severity in Riyadh. Three classification techniques were used: CHAID, J48, and Naive Bayes. The performance of all the obtained models are evaluated and compared to identify the best model. The resulting models highlight the danger of distraction while driving. Age of the car is also found to be an important factor. The study discovered, accidents with older cars are more likely to result in injuries or deaths. But data or time factors of the accident have not been considered in this study. Kashyap et al. [17] also conducted a study on traffic accident data in Uttar Pradesh, India. They have applied data mining technologies to link recorded road characteristics to accident severity and developed a set of rules that could be used by the Indian Traffic Agency to improve safety and could help to save precious life. Naïve Bayes classifier was employed and its accuracy is variable based on the number of attributes that feeds to the data mining tool. Even they consider various road-related factors like environment, weather conditions, condition of vehicle, engine problems that have impact on the accident severity. But important factors like month of the incident happened or time has not been considered in this study.

IV. METHODOLOGY

In this study, author is attempting understand most significant factors that contribute to roadside accident severity, predict road traffic accident severity with high accuracy using Decision Tree classification algorithm and enhance the knowledge gain from the visualisation techniques. Classification and visual data exploration techniques were performed in order to achieve the objectives. In order to achieve the outcomes, we have followed the following steps.

A. Classification:

Classification is a form of data analysis where a model or classifier is constructed to predict categorical labels. The objective of the predicting using classification technology is to accurately forecast the target class variable for each new data point, e.g., predict roadside accident severity for new accident which is missing in the data.

Decision Tree Classifiers: Decision trees are an unsupervised machine learning method for data classification, they separate data into predefined groups or classes to create a model which can be used to predict future classification cases. Decision Trees follow a tree-like structure, containing root nodes, internal nodes, and leaf nodes. The root node is the very first node at the top of the tree, this is the first variable that the data will split on, attributes used as nodes for splitting are determined via which will present the highest information gain. Decision tree can be illustrated as a form a rule such as "IF predictive information THEN target," as in "IF Road_Type=Motorway and Vehicle_Type=Van THEN the probability of Accident Severity=Serious is 0.01." Decision tree rules provide transparency that ensure the end user, such as a business user, marketing analyst or business analyst, can easily understands the model's predictions, and therefore freely applies and explains them to other people [9].

B. Visual data exploration:

Data exploration is a recommended first step in any analysis, but analysts often look at the figures, summary statistics such as average, median, and distribution. They don't always engage in visual data exploration. The study of visual data exploration is certainly a necessary first step if further formal analysis is conducted. When combined with descriptive numbers, visualization provides a powerful means of recognizing the abstract, structure, relationships, differences, and anomalies of the data [11]. In this study calendar heatmap based visualisation will be used to support the knowledge discovery and pattern recognition combine with Decision tree calcification. The key difference is that visual data exploration is a purely human process, while data mining algorithms can automatically analyze data and may find any hidden useful information [13]. The degree of automation of data mining algorithms varies considerably as different levels of human guidance and interaction are usually required, but still the algorithm, not the user, is the one that is to look for patterns [12]. Below shown (Figure 1) Visual Analytics process was introduced by Kohlhammer and his team [13] to describe the process of visual analysis. The process begins with data acquisition and transformation, and then an automated approach to data analysis, such as statistical modeling, data mining and machine learning. The next step is data visualization, it implies that users collaborate with the visual interface to extract knowledge.
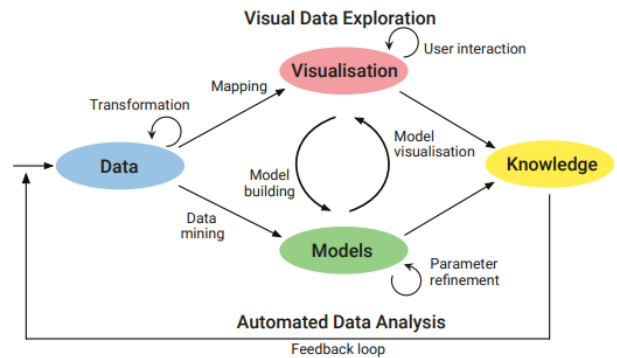

FIGURE 1: VISUAL ANALYTICS PROCESS FOR KNOWLEDGE DISCOVERY [13]

The process of visual analysis combines the methods of mechanical and optical analysis with closely link through human interaction to obtain knowledge from the raw data. The knowledge obtained from the proposed combined models and its findings can be analysed and used to increase the road safety. Also, the knowledge gain from both serious accidents and slight accidents can then be correlated and identify any factors that most significant for serious accidents. Thus, this knowledge can be merged through its interlinkages and relationships, creating comprehensive knowledge for roadside accident management in the road networks.

V. CASE STUDY

A. Case study overview

This case study provides the details about road traffic accidents in Great Britain that were reported by the police during 2007 and 2016 using the STATS19 reporting system. The data used in this investigation was collected by police forces, either through officials who take part in the scene or from members of the public reporting the incident in police stations after the incident. Accidents that happened on private land or car parks are not included in the study. Damage only accidents that do not result in personal injury

are also excluded from this study. Fatal accident defined as a road side accident that at least one person is killed. Serious accident defined as that at least one person is seriously injured but no one involved in the accident is killed. Slight accident defined as that at least one person is injured but not requiring medical treatment.

B. Focus study area

In this case study, the area known as the North West of England is used as the focus study area. The study area covers an area of 14,000 square kilometers. It is the fourth smallest area in England. The population of the study area is approximately 9 million.

VI. EXPERIMENTAL EVALUATION

A. Data acquisition

The dataset for this research has been obtained from UK Department of Transport's STATS19 Road Safety Data source [16]. For this study, only North-West England data has been extracted from whole UK dataset. Extracted dataset contains 323,058 road side accidents records from Jan 2007 – Dec 2016. Several data transformation has been applied to some attributes such as Fatal and Serious accidents have been combined to simplify the study. Sample dataset shows in table 1.

TABLE 1: Sample dataset

Driver	Day of Week	Vehicle	Area	Road	Road Surface Conditions	Light Condition	Accident Severity
Male	Weekend	Car	Rural	A	Dry	Daylight	Slight
Male	Weekend	Car	Rural	A	Dry	Daylight	Slight
Male	Weekend	Car	Rural	A	Dry	Daylight	Slight
Male	Weekday	Car	Rural	A	Wet or damp	Daylight	Serious
Male	Weekend	Car	Rural	Motorway	Dry	Daylight	Serious
Male	Weekend	Car	Rural	Motorway	Dry	Daylight	Serious
Male	Weekday	Car	Rural	A	Wet or damp	Daylight	Slight
Male	Weekend	Car	Rural	A	Dry	Daylight	Slight
Female	Weekday	Bus	Rural	A	Dry	Daylight	Slight
Male	Weekend	Car	Rural	C	Dry	Daylight	Slight

B. Data Filtering and pre-processing

Prior to the creation of a data mining models, it is vital to inspect the data and identify any areas which may require modification. Typical methods used to clean data include dealing with null values, outlier detection and treatment. The datasets used in this study, each require different levels of cleaning.

This phase is one of other most important phrase of data mining projects. This includes the selection of relevant data attribute to performing the required data mining tasks. The dealing with missing and unknown data is required step which might affect the efficiency of prediction models. The deriving new data or variable by merging datasets also part of this phase. Data filtering and pre-processing phase can take considerable amount of time of any data analytic project.

C. Predictive modelling using Decision Tree

A Decision Tree is a structure that can be used to address classification problems. One of the most widely adopted algorithms in Decision Trees is ID3 which adopts Entropy and Information Gain to calculate node impurity. The study dataset was split into two datasets. The first subset, which contain 80% of the records, was used to train the decision tree model. The remaining 20% was used to evaluate the performance of the decision tree model.

D. Decision Tree without Month and Hour

Below Decision Tree model shows how accidents are classified by each status of the class variable (slight accidents vs. serious accidents) with 6 independent variables. Month and Hour variables have been eliminated from this model. But Month and Hour will be added to the next decision tree model (shows in figure 2) and later results will be compared to understand the impact of the date and time factors like Month and Hour to the accident severity (Shows in Table 8).
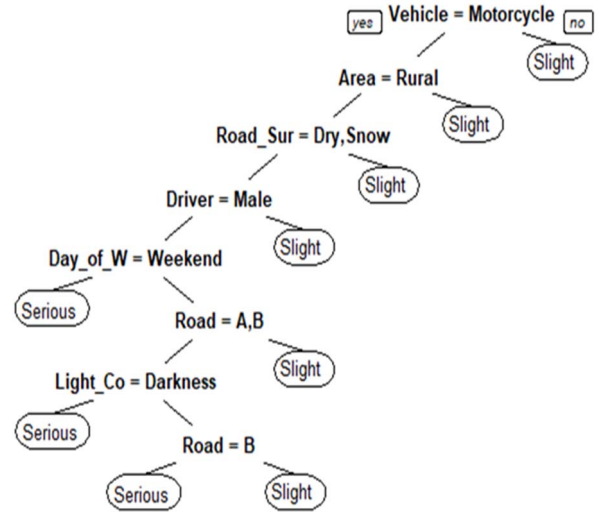

FIGURE 2: DECISION TREE WITHOUT MONTH AND HOUR

When rules are obtained from a Decision Tree like above (Figure 2), they are determined by the variable that is used as a root node. See below four sample rules out of nine rules obtained from above Decision Tree.

Rule number: 1 [Accident_Severity=Slight]
WHEN Vehicle= Van,Bus,Car,Other vehicle,Pedal cycle

Rule number: 2 [Accident_Severity=Slight]
WHEN Vehicle=Motorcycle
AND Area=Urban

Rule number: 7 [Accident_Severity=Serious]
WHEN Vehicle=Motorcycle
AND Area=Rural
AND Road_Surface_Conditions=Dry,Snow
AND Driver=Male
AND Day_of_Week=Weekday
AND Road=A,B
AND Light_Condition=Daylight
AND Road=B

Rule number: 8 [Accident_Severity=Serious]
WHEN Vehicle=Motorcycle
AND Area=Rural
AND Road_Surface_Conditions=Dry,Snow
AND Driver=Male
AND Day_of_Week=Weekday
AND Road=A,B
AND Light_Condition=Darkness

E. Performance Evaluation

In order to evaluate the performance of the developed model, using a confusion matrix the results of correctly and incorrectly predicted records for each class are illustrated in Table 2:

	Predict: Serious	**Predict: Slight**
Actual: Serious	256	9906
Actual: Slight	200	54247

TABLE 2: CONFUSION MATRIX OF THE VALIDATION DATASET

Based on the confusion matrix, prediction has been evaluated using four measures: Accuracy, Precision, Recall and F-measure.

Accuracy is the percentage of correctly classified instances in the validation dataset. It can be calculated as the number of all correct predictions divided by the total number of the data points. The best accuracy is 1, whereas the worst is 0.

$$Accuracy = \frac{True\ Positive + True\ Negative}{Positive + Negative}$$

The precision is defined as the proportion of positive examples that are truly positive.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Recall is also known as the sensitivity rate.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

The f-measure is a measure that combines both precision and recall. The F-score acts as a compromise even though it ignores that one might be more important than the other. It is a good measure because it is the harmonic mean of the two competing measures, that is combining recall and precision.

$$F - measure = \frac{2 * recall * precision}{recall + precision}$$

The results that have been obtained in each performance evaluation method with the decision tree algorithm discussed above are shown in Table 3 below:

TABLE 3: PERFORMANCE EVALUATION RESULTS

Model	Accuracy	Class	Precision	Recall	F-measure
Without Hour or Month	84.36%	Slight	99.63%	84.56%	91.48%
		Serious	2.52%	56.14%	4.82%

F. Decision Tree with Month and Hour

Figure 3 shows the decision tree generated with Month and Hour variables. As shown in Figure 3, due to large number of leaf nodes and rules, decision tree is very convoluted and not very clear.
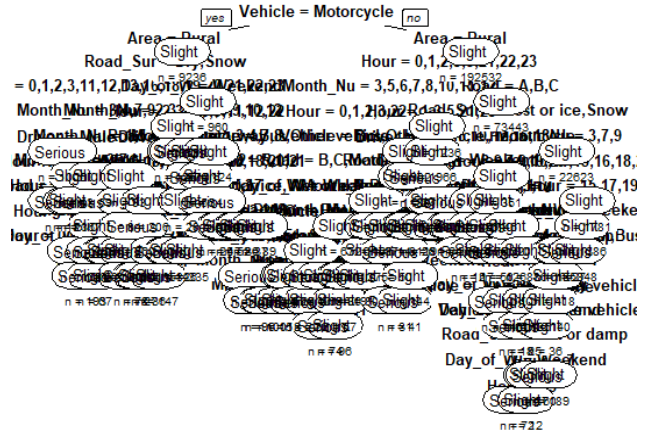

FIGURE 3: DECISION TREE WITH MONTH AND HOUR

There were 116 rules were generated with above decision tree (Figure 3). See below one of a sample rule generated,

Eg: Rule number: 102 [Accident_Severity=Serious]
WHEN Vehicle= Van,Bus,Car,Other vehicle,Pedal cycle
AND Area=Rural
AND Hour=4,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20
AND Road=Motorway
AND Road_Surface_Conditions=Dry,Flood over 3cm. deep,Wet or damp
AND Month_Number=3,7,9
AND Hour=7,9,10,11,13,16,18,20
AND Hour=7,10,11,13,16,18,20
AND Month_Number=3
AND Road_Surface_Conditions=Wet or damp
AND Hour=16
AND Day_of_Week=Weekday

Table 4 shows the confusion matrix for validation dataset which was generated from decision tree with Month and Hour variables.

TABLE 4: CONFUSION MATRIX OF THE VALIDATION DATASET

	Pred: Serious	**Pred: Slight**
Actual: Serious	556	9606
Actual: Slight	252	54196

Below table 5 summarises the evaluation measures from the confusion matrix shows in table 6.

TABLE 5: PERFORMANCE EVALUATION RESULTS

Model	Accuracy	Class	Precision	Recall	F-measure
With Hour & Month	84.74%	Slight	99.54%	84.94%	91.66%
		Serious	5.47%	68.81%	10.14%

G. Results comparison of both the performance evaluation measures

Table 8 shows the difference between performance evaluation measures of the two decision tree models. Based on the values in the table, impact of the date and time factors like Month and Hour are not much contribute to accuracy of accident severity prediction.
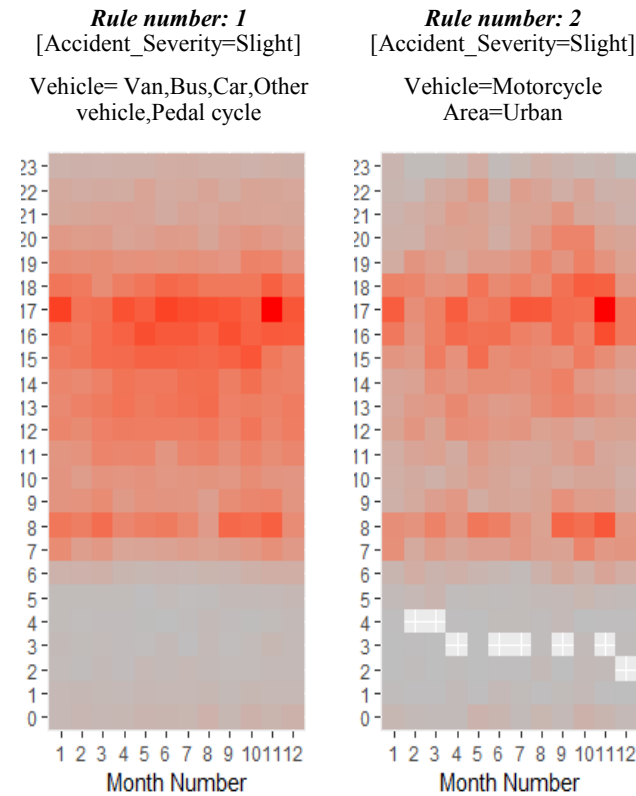
TABLE 8: COMPARISON OF PERFORMANCE EVALUATION RESULTS

Accuracy	Class	Precision	Recall	F-measure
0.38%	Slight	-0.09%	0.38%	0.18%
	Serious	2.95%	12.67%	5.32%

H. Enhance the results using Time-Series Calendar Heatmap

The temporal aspects of data presentation are of paramount importance in road traffic accidents, especially when the accident data is more complex than normal time series data due to the factors associated with the dataset. Time series data is series of data that indexed in chronological order. The order of time can be given in days, weeks, months, quarters or years. The most common way to display time series data is to use a simple line graph, where the horizontal axis represents the time and the vertical axis represents the measured variable. However, in this study, a heatmap form of the time series calendar was used to identify temporal patterns. The calendar heatmap is a variation of the traditional heatmap, in which the data is organized in a calendar format. This is an alternative representation for the analysis of time series data. Typically calendar heatmap use to represent activity over a long period of time, such as months or years.

Calendar heat maps are simple, beautiful, effective and simple to understand even though they represent a lot of data in a compact visual. Aggregating and visualising hourly accident data on calendar heatmap gives much more granular overview of the road traffic accident patterns across the 24 hour windows. The colour scale represents the number of roadside accidents. In this study, grey represent low numbers and red represents high number of accidents. Still need to evaluate the distribution and magnitude of colours.

Rule number: 1
[Accident_Severity=Slight]

Vehicle= Van,Bus,Car,Other vehicle,Pedal cycle

Rule number: 2
[Accident_Severity=Slight]

Vehicle=Motorcycle
Area=Urban

Rule number: 7
[Accident_Severity=Serious]
Vehicle=Motorcycle
Area=Rural
Road_Surface_Conditions=Dry,Snow
Driver=Male
Day_of_Week=Weekday
Road=A,B
Light_Condition=Daylight
Road=B

Rule number: 8
[Accident_Severity=Serious]Vehicle=Motorcycle
Area=Rural
Road_Surface_Conditions=Dry,Snow
Driver=Male
Day_of_Week=Weekday
Road=A,B
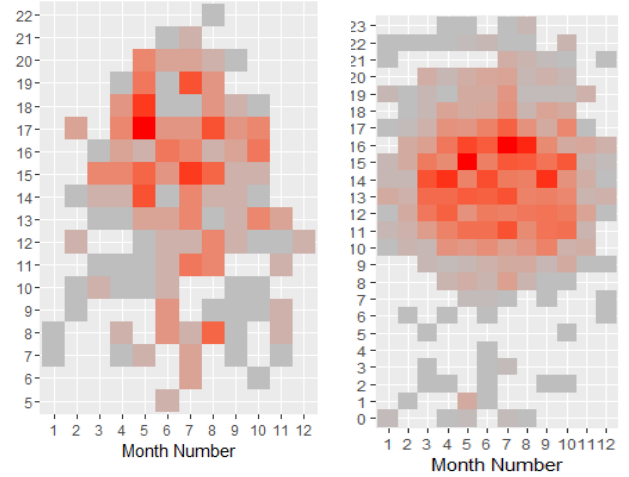Light_Condition=Darkness

FIGURE 4: TIME-SERIES CALENDAR HEATMAPS

As shown in Figure 4, Nine different Time-Series Calendar Heatmaps have been built for nine rules which derived from the decision tree shown in Figure 2. Aggregating and visualising hourly accident data on calendar heatmap gives much more granular overview of the road traffic accident patterns across the 24 hour windows over 10 years study window. The colour scale represents the number of roadside accidents. Few temporal patterns can be clearly identified by looking at the above calendar heatmaps. Based on the values in the table 10, impact of the date and time factors like Month and Hour not much contributing to accuracy of the accident severity prediction. But calendar heatmaps shows in Figure 4 clearly shows that few obvious and important rules exists in the data.
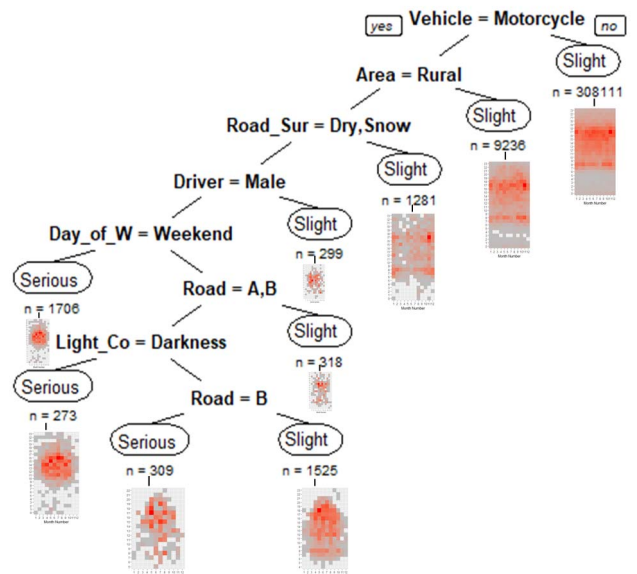

FIGURE 5: ENHANCED DECISION TREE WITH TIME-SERIES CALENDAR HEATMAP

In this study, author proposed to build decision tree model without time and date factors first, then as a second stage, add Time-Series Calendar Heatmaps in to the leaf nodes. Final enhanced decision tree is look like as shown in Figure 5. Proposed enhanced decision tree can be used to predict the accident severity with temporal patterns. But, temporal patterns only can be added to the generated decision tree rules by visual data exploration.

I. Visual data exploration

The visual analytics process combines automatic and visual analysis methods with a tight coupling through human interaction in order to gain knowledge from data [13]. In this study, visual analysis process has been used to enhance the decision tree rules with temporal pattern through human interaction. Let's take rule no. 8 as an example, according to the rule 8, serious accident happened in below conditions.

Rule number: 8 [Accident_Severity=Serious]
WHEN Vehicle=Motorcycle
AND Area=Rural
AND Road_Surface_Conditions=Dry,Snow
AND Driver=Male
AND Day_of_Week=Weekday
AND Road=A,B
AND Light_Condition=Darkness

But important date and time variables like month and the hour are missing from the rules. So, using knowledge gain from the Time-Series Calendar Heatmaps shown in Figure 5, rule can be rewritten as below.

Rule number: 8 [Accident_Severity=Serious]
WHEN Vehicle=Motorcycle
AND Area=Rural
AND Road_Surface_Conditions=Dry,Snow
AND Driver=Male
AND Day_of_Week=Weekday
AND Road=A,B
AND Light_Condition=Darkness
AND Month BETWEEN 3 AND 10
AND Hour BETWEEN 10 AND 17

VII. Conclusion

Identifying the temporal, environment and behavioral factors leading to determine the severity of accident is an important for authorities who work in road safety. In this study, the decision tree and calendar heatmap was used to examine the influence on several factors with the injury severity of the road traffic accidents in UK. Several indicators such as Accuracy, Precision, Recall and F-measure have been used in order to evaluate the performance of the built decision tree models. Nine significant validated rules were obtained from the model that built without time factors like month and hour on traffic accident data from North West Region of England. But 116 rules were obtained from the model built on same dataset with time factors like month and hour. So, more than 10 times of rules were generated with time components and distorted the usability of the rules and knowledge gained. This study proposed novel approach of combining decision tree and calendar heat map to extract the knowledge from the accident datasets with time factors. Based on the decision tree models and evaluation measures, highlights that the Month and Hour of the accident is not much contribute to the model accuracy, because both factors didn't show substantial impact on crash severity. But, by looking at the calendar heatmaps, some specific time and month related patterns can be clearly identified. The approach described in this paper offers significant advantages over understanding correlation between hour and month of the accident and the severity of the accident. Although this study is based on a region in North of England, the approach can be applicable to other areas in UK and globally with comparable data. So, it could be concluded that combining method of decision tree and time-series calendar heatmaps might be a useful tool for classifying traffic accidents according to their injury severity.

References

[1] Panda, B., Sharma, S., & Roy, N. R. (2018). Data science and analytics: 4th International Conference on Recent Developments in Science, Engineering and Technology, REDSET 2017, Gurgaon, India, October 13-14, 2017, revised selected papers. Singapore: Springer.

[2] Road safety in the European Union trends, statistics and main challenges, March 2015. (2015). Brussels: European Commission.

[3] Road Traffic Collision. (n.d.). Retrieved February 21, 2019, from https://www.north-wales.police.uk/contact/minor-incident-reporting/road-traffic-collision

[4] 2017 road safety statistics: What is behind the figures? (Rep.). (2018, April 10). Retrieved February 21, 2019, from European Commission website: europa.eu/rapid/press-release_MEMO-18-2762_en.pdfs

[6] WEIJERMARS,W. et al.,2017.Physical and psychological consequences of serious road traffic injuries,deliverable 7.2 of the H2020 project Safety Cube.Loughborough:Loughborough University

[7] Peden, M. (2004). World report on road traffic injury prevention: Information kit. Geneva, Switzerland: World Health Organization.

[8] T. K. Bahiru, D. Kumar Singh and E. A. Tessfaw, "Comparative Study on Data Mining Classification Algorithms for Predicting Road Traffic Accident Severity," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 1655-1660.

[9] Data Mining Concepts. (2005, June 16). Retrieved February 22, 2019,

[10] I. Al-Turaiki, M. Aloumi, N. Aloumi and K. Alghamdi, "Modeling traffic accidents in Saudi Arabia using classification techniques," 2016 4th Saudi International Conference on Information Technology (Big Data Analysis) (KACSTIT), Riyadh, 2016, pp. 1-5.

[11] Visual Data Exploration. (n.d.). Retrieved February 22, 2019, from https://uc-r.github.io/2018/01/22/visual-data-exploration/

[12] M. C. Ferreira de Oliveira and H. Levkowitz, "From visual data exploration to visual data mining: a survey," in IEEE Transactions on Visualization and Computer Graphics, vol. 9, no. 3, pp. 378-394, Sept. 2003.

[13] D.A. Keim, J. Kohlhammer, G. Ellis, F. Mannsmann (Eds.), Mastering the Information Age. Solving Problems with Visual Analytics, Eurographics Association, Goslar, 2010.

[14] North West England. (2019, February 20). Retrieved February 22, 2019, from https://en.wikipedia.org/wiki/North_West_England#cite_note-Manchester_population-1

[15] Young, R., & Sly, F. (2010). Portrait of the North West (Rep.). Retrieved February 22, 2019, from Office for National Statistics.

[16] Al-Turaiki, Isra, et al. "Modeling Traffic Accidents in Saudi Arabia Using Classification Techniques." 2016 4th Saudi International Conference on Information Technology (Big Data Analysis) (KACSTIT), 2016, doi:10.1109/kacstit.2016.7756072.

[17] J. Kashyap and Asst. Prof. Chandra Prakash Singh, "Mining road traffic accident data to improve safety on road-related factors for classification and prediction of accident severity", International Research Journal of Engineering and Technology, Volume: 03, Issue: 10, 2016.

[18] Abellán, Joaquín, et al. "Analysis of Traffic Accident Severity Using Decision Rules via Decision Trees." Expert Systems with Applications, vol. 40, no. 15, 2013, pp. 6047–6054., doi:10.1016/j.eswa.2013.05.027.

[19] Cuenca, Laura Garcia, et al. "Traffic Accidents Classification and Injury Severity Prediction." 2018 3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE), 2018, doi:10.1109/icite.2018.8492545.

[20] "An Angle Displacement Sensor Using a Simple Gear." NeuroImage, Academic Press, 30 Dec. 2017, www.sciencedirect.com/science/article/pii/S0924424717312797.

[21] Saraee, M., & Silva, C. (2018). A new data science framework for analysing and mining geospatial big data. Proceedings of the International Conference on Geoinformatics and Data Analysis - ICGDA 18. doi:10.1145/3220228.3220236