

ROAD SAFETY ANALYSIS BY USING K-MEANS ALGORITHM

¹G. PRIYANKA and ²Dr. R.JAYAKARTHIK

¹Research Scholar, Dept. of Computer Science, VELs Institute of Science, Technology and Advanced Studies (VISTAS), Chennai.priyankaharitsa@gmail.com

²Asst. Professor, Dept. of Computer Science, VELs Institute of Science, Technology and Advanced Studies (VISTAS), Chennai.jayapranav15@gmail.com

ABSTRACT

This paper predicts the road safety analysis using K-means algorithm. Road safety is one of the major issue to be taken care in our daily life. The aim of the paper is to analyse the data of the road condition and reasons behind the accidents and to cluster it by using k-means clustering algorithm by using WEKA tool. The proposed method will show the instance of risk for each and every attributes in the data set.

KEYWORDS: Road safety, k means clustering and WEKA tool.

INTRODUCTION

This paper predicts the road safety analysis using with one of the latest data mining algorithm. Road safety is very dangerous in now days. So am interested in this area and we want give analysis relates with road safety precautions. we chose one of the latest data mining algorithms for the prediction and we select WEKA tool for the implementation. Hence we can able to get the results and discussion easily.

II. LITERATURE SURVEY

Present study shows that [1] road safety improvement using data mining techniques by using 3 database like Elements of Improvement (ESM), road, hazard sections.[2] A CASE STUDY OF LIBYA'S ROAD TRAFFIC used data mining techniques for analyzing road traffic incidents. They proposed theory to control the traffic management with the application of Data Mining (DM) functionalities by combining intelligent agent based system. [3] An overview of data mining in road traffic and accident analysis

by using some of the data mining techniques, tools, applications and search engines for accident investigation and traffic analysis.[4]characterizing road accident locations using data mining approach to find that road accident occurrences are more frequent at certain specific locations. The analysis of these locations can help in identifying certain road accident features that make a road accident to occur frequently in these locations. [5] Analysis of road traffic fatal accidents using data mining techniques apply statistics analysis and data mining algorithms on the FARS Fatal accident dataset as an attempt to address traffic safety problem. [6] A data mining framework to analyse road accident data by proposing a framework that used K-modes clustering technique as a preliminary task for segmentation of

11,574 road accidents on road network of Dehradun

(India) between 2009 and 2014 (both included). Next, association rule mining are used to identify the various circumstances that are associated with the occurrence of an accident for both the entire data set (EDS) and the clusters identified by K-modes clustering algorithm. [7]

Role of Road- elated Factors on Accident Severity in Ethiopia for road safety improvement using data mining to link recorded road characteristics to accident severity in Ethiopia, and developed a set of rules that could be used by the Ethiopian Traffic Agency to improve safety.

III. DATA AND METHODOLOGY

A. Data

	A	R	C	H	F	F	F	H
1	Road condition	curvy roads	laneways in sign	low-light	burners in area	curvy	no	maximal
2	day	no	yes	no	no	yes	no	yes
3	good	yes	yes	no	no	no	no	no
4	bad	no	no	yes	yes	no	no	yes
5	bad	no	yes	yes	yes	no	no	yes
6	bad	no	no	yes	yes	no	yes	no
7	bad	no	no	no	no	yes	yes	yes

	A	B	C	D	E	F	G	H
1	light	speed	damage	sign board	water	drink	ant	right
2	Dark	normal	yes	normal	no	yes	good	no
3	Day light	high	no	leg	no	no	bad	no
4	Dark	low	no	normal	yes	yes	good	no
5	Dark	high	yes	normal	yes	yes	bad	no
6	Dark	high	yes	leg	no	yes	good	no
7	Dark	normal	no	normal	no	yes	good	no
8	Day light	high	no	leg	yes	no	bad	yes

In this paper we are going to have two dataset which is made by the help of survey papers.

1. Road status: where we can get the status of the road. For example road condition, damage in sign board, narrow road etc.

We are having eight numbers of attributes in both the data set, which will help in merging the data set in WEKA.

B. Methodology

K-means Clustering

The main goal of cluster analysis is to group objects that are similar in one cluster and dissimilar to other. K-means clustering algorithm is one of the most popular clustering methods. The classification of the object is based on predefined number of clusters which is given by user. Random cluster centers are chosen for each cluster. These centers are preferred to be as far as possible from each other.

In this algorithm mostly Euclidean distance is used to find distance between data points and centroids. The

Euclidean distance between two multi-dimensional data points $X = (x_1, x_2, x_3... x_m)$ and Y



= $(y_1, y_2, y_3, ..., y_m)$ is described as follows

$$D(X, Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_m - y_m)^2}$$

Algorithm:

K-Means clustering

1. The algorithm selects k points arbitrarily as the initial cluster centers.
2. Each point is assigned to the closed cluster in the data-set, based upon the Euclidean distance between each cluster center and each point.
3. Each cluster center is recalculated as the average of the points in that cluster.
4. Steps 2 and 3 repeat until the clusters coincide.

IV. IMPLEMENTATION OF K-MEANS CLUSTER

In this section we are going to explain how the dataset has been pre-process and the algorithm used to get the result by working on WEKA.

In this paper we are going to have our own sample data set which has been created by the reference of survey papers [1]. We created two data sets, one contains the road conditions and another contains





3) K-means clustering:

We need to do cluster in cluster panel. We mine the current dataset by k means using WEKA. When we click on k-means cluster we will get a dialog box has shown in the fig (10) there we need to specify how many cluster we require. Usually I will depend upon the size of the dataset. We sixty data in our data set so I will give cluster number have five.

V. EXPERIMENTAL RESULT

Here the experimental result if k-means algorithm is given. This proposed methodology is experimented with two different data sets to predict the road safety by k-means clustering algorithm using WEKA. We merge two data sets in simple CLI to get the result of both road status and reasons for accident data set.

The following fig (12) graphically represents the result after applying k-means algorithm. In this we can select different x axis and y axis to visualize cluster representation.

The conclusion of this paper is to experiment the datasets using WEKA by using clustering to find the cluster of interest. We analyze the road safety prediction using with different attributes related with the data sets and clustered the dataset automatically by using WEKA. In this paper we studied K – means clustering algorithm to casualties' future hazard.

```
> java weka.core.Instances merge F:/road.arff F:/reason.arff > F:/merge.arff
```

Finished redirecting output to 'F:/merge.arff'.

REFERENCES

- [1] Luis Martín, Leticia Baena, Laura Garach, Griselda López and Juan de Oña, "Using data mining techniques to road safety improvement in Spanish roads," Peer-review under responsibility of CIT 2014.doi: 10.1016/j.sbspro.2014.12.174
- [2] Sayed Mujahed Hussain , Faraj A. El-Mouadib, "APPLICATION OF DATA MINING TECHNIQUES FOR ANALYZING ROAD TRAFFIC INCIDENTS A CASE STUDY OF LIBYA'S ROAD TRAFFIC"
- [3] K. Jayasudha Dr. C. Chandrasekar, "AN OVERVIEW OF DATA MINING IN ROAD TRAFFIC AND ACCIDENT ANALYSIS ", Journal of Computer Applications, Vol – II, No.4, Oct – Dec 2009
- [4] Sachin Kumar, Durga Toshniwal , "A data mining approach to characterize road accident locations". Received: 20 August 2015 / Revised: 5 January 2016 / Accepted: 6 January 2016 / Published online: 11 February 2016
- [5] Liling Li, Sharad Shrestha, Gongzhu Hu, "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques", 978-1-5090-5756-6/17/\$31.00 ©2017 IEEE SERA 2017, June 7-9, 2017, London, UK
- [6] Sachin Kumar^{1*} and Durga Toshniwal. Kumar and Toshniwal, "A data mining framework to analyse road accident data"Journal of Big Data (2015) 2:26 DOI 10.1186/s40537-015-0035
- [7] Tibebe Beshah, Shawndra Hill, "Mining Road Traffic Accident Data to Improve Safety: Role of Road- related Factors on accident Severity in Ethiopia"
- [8] Data mining with WEKA, Part 1 Introduction and regression Michael Abernethy Published on April 27, 2010. <https://www.ibm.com/developerworks/library/os-weka1/>

