International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT228637

A Novel Framework for Analysis of Road Traffic Information for Decision Support by Using Data Mining Techniques

Dr. Mahesh Kotha, Sathini Santhosh Kumar, A Deepika, M Swathi Reddy

[1]Associate Professor, Department of CSE (AI&ML), CMR Technical Campus, Hyderabad, Telangana, India

[2]Assistant Professor, Department of CSE, Vignan's Institute of Management and Technology for Women, Hyderabad, Telangana, India

[3]Assistant Professor, Department of CSE (AI&ML), CMR College of Engineering and Technology, Hyderabad, Telangana, India

[4]Assistant professor, CSE(AIML) Department, Sri Indu College of Engineering and Technology, Hyderabad, Telangana, India

Article Info

Publication Issue :
Volume 8, Issue 6
November-December-2022

Page Number : 268-276

Article History
Accepted: 10 Nov 2022
Published: 23 Nov 2022

ABSTRACT

As we know we want traffic data analysis, data mining measures are recommended to help manage and traffic information decision making. Such as this limit, the road type is added intelligence analysis and decision support are another gave. This model has a data collection system and the data warehouse at its heart, and works well mathematical intelligence techniques and the invention of decision support system. As the main part of this report, the main parts of the model, the main process and an implementation plan is developed. An intelligent urban traffic decision analysis system is a system designed to solve open road problems, improve road safety and quality, and improve road quality work. The application of big data mining technology can improve various capabilities types of analysis and processing of traffic information, spread rapidly and provide a decision reference for the application of intelligent transport systems, which will be carried out to ensure smooth and safe roads. This topic will introduce the concept of large data mining technology and intelligent transport systems, and review the architecture of smart city road systems based on big data technology and design implementation of the mining process.

Keywords : Data mining, Association Rules, Intelligent Traffic, Traffic Information, Integration.

I. INTRODUCTION

The development process of urban road intelligent decision-making is the concept of intelligence, scheduling, communication and social services. Traffic control methods are also is increasing. The company that these services provide for public transport is increasing. In the development of electronic information technology, especially after the rapid development of Cloud computing and the

Internet of things, big data mining began to be applied in the field of smart open travel. In the same way, it can use the feature of real time, large data concurrency, scalability and security to increase traffic collection and analysis; On the other hand, big data and traffic data management can be applied to access urban road project objectives. Brings the benefits of urban transportation to the full game process, it will help improve the smoothness of the work of the media system, improve efficiency of road services, reduction of safety accidents and pollution issues, traffic optimization type of organization and improvement of travel service level. Transportation is the main industry of this national economy. In previous years, the transport system is already very large amount of data.

Now, however, these data are only used in to research and statistical work to get information up. And information who can explain the full characteristics of this data and can predict the behavior and evolution of this data do not exploit, because this message is deep hidden in this big data. How to use this data to determine the understanding and improve the traffic Effective and scientific management is now problem immediately solved. Data mining, one of the most popular topics in the area such as data protection, artificial information, statistics, etc., can identify useful information data association status and development through extraction many data warehouses. From the data Mining process, traffic data can be become useful information for decision making, and decision makers can be approved it with other approvals to make it valid resolution [1].

II. RELATED WORK

Roadway traffic safety is a major concern for transportation governing agencies as well as ordinary citizens. Data Mining is taking out of hidden patterns from huge database. It is commonly used in a marketing, surveillance, fraud detection and scientific discovery. In data mining, machine learning is mainly focused as research which is automatically learnt to recognize complex patterns and make intelligent decisions based on data. Globalization has affected many countries. There has been a drastic increase in the economic activities and consumption level, leading to expansion of travel and transportation [3]. The increase in the vehicles, traffic lead to road accidents. Considering the importance of the road safety, government is trying to identify the causes of road accidents to reduce the accidents level. The exponential increase in the accidents data is making it difficult to analyse the constraints causing the road accidents. The paper describes how to mine frequent patterns causing road accidents from collected data set. We find associations among road accidents and predict the type of accidents for existing as well as for new roads. We make use of association and classification rules to discover the patterns between road accidents and as well as predict road accidents for new roads.

There are a lot of vehicles driving on the roadway every day, and traffic accidents could happen at any time anywhere. Some accident involves fatality, means people die in that accident. As human being, we all want to avoid accident and stay safe. To find out how to drive safer, data mining technique could be applied on the traffic accident dataset to find out some valuable information, thus give driving suggestion. Accidents happened due to the negligence of driving vehicle on the roads. There are various reasons responsible for the accident like abandon of traffic rules but road conditions and the traffic are considered the one of prime cause of fatality and causality across the globe[5].

These accidents occur due to dynamic design and development of automobile industries. A traffic crash happens due certain reasons like smashes of two vehicles on road, walking person, animal, or any other natural obstacles. It could result in injury,

property damage, and death. Traffic accident analysis required study of the various factor affecting behind them. In survey it's seen that approximate 1.2 million death and 50 million injuries estimated worldwide every year. The
approximate estimation of causality and injuries due to poor road infrastructure is a big challenge before the living beings. The order to deal with the problem, in computational science, we can adopt data mining model for different scenario. In any vehicle accident, it studies about the driver's behaviour, road infrastructure and possibilities of weather forecast that could be somewhere connected with different accident incidents. The main problem in the study and analysis of accident data is its mix heterogeneous environment and data segmentation which is used widely to overcome accident problem. [2,5,7]

Data Mining is a computational technique to deal with large and complex data set and these data sets can be of normal, nominal and mixed. It is quite easy to use in variety of domain belong to science and management; also, it could be used in fraud identification and many more scientific cases as well as in accident severity problem. Partition of objects in a group of clusters or in a homogeneous set is a fundamental operation of data mining.

Clustering is a method to partition objects in a similar group. The k-means algorithm having a good efficiency for clustering large data sets but restricted in forming clusters for real word data while working only on numerical data because it helps in reducing the cost function by altering the meaning of the clusters [1,3].

Data mining technique is recognized as reliable technique for analysis of traffic accident severity problem and finding factors behind them. Damage like property, people due to road accident is undesirable. Happened that road accident incidents are more common at certain places that can help in identifying factors behind them. power based multi mobile charger system is implemented in the proposed system which can be used in public places like railway stations, bus stands, hospitals and parks etc[11].

III. PROPOSED APPROACH

Roadway traffic safety is a major concern for transportation governing agencies as well as ordinary citizens so, for that purpose we are introducing an analytic tool in which Genetic algorithm will be used for classification. We are taking dataset of a country and analyzing that database month wise for one year, To find out which states are similar to each other considering fatal rate, and which states are safer or more risky to drive, clustering algorithm was performed on the fatal accidents dataset. Before applying the algorithms, the tuples with missing value in chosen attributes were removed. The proposed work is planned to be carried out in the following manner.

Data preparation was performed before each model construction. All records with missing value (usually represented by 99 in the dataset) in the chosen attributes were removed. All numerical values were converted to nominal value according to the data dictionary in attached user guide.

The primary objectives of this study can be summarized as follows:

1. To process the dataset.
2. To cluster the dataset based on Parameters and apply k-means for clustering and Bayesian algorithm for classification.
3. To analyzed the graph and Predict the Road Accidents based on various parameters.
4. To develop automatic analytical tool which mines road wise accident patterns
5. To develop a Decision support system for general public and for government which will predict the

possibilities of accidents and spread awareness about the roads and take decision about redevelopment of roads

There are a lot of vehicles driving on the roadway every day, and traffic accidents could happen at any time anywhere. Some accident involves fatality, means people die in that accident. As human being, we all want to avoid accident and stay safe. To find out how to drive safer, data mining technique could be applied on the traffic accident dataset to find out some valuable information, thus give driving suggestion. Accidents happened due to the negligence of driving vehicle on the roads. There are various reasons responsible for the accident like abandon of traffic rules but road conditions and the traffic are considered the one of prime cause of fatality and causality across the globe. These accidents occur due to dynamic design and development of automobile industries.

Correlation coefficient	-0.1539	
Mean absolute error	94965961788.9088	
Root mean squared error	109620175687.726	
Relative absolute error	100	%
Root relative squared error	100	%
Total Number of Instances	144	

Fig 1. Cross Validation values of traffic dataset

A traffic crash happens due certain reasons like smashes of two vehicles on road, walking person, animal, or any other natural obstacles. It could result in injury, property damage, and death. Traffic accident analysis required study of the various factor affecting behind them. In survey it's seen that approximate 1.2 million death and 50 million injuries estimated worldwide every year.

	Pedestrians & Cyclists	Two-wheeled vehicles	Four-wheeled vehicles	Public services Vehicles	Heavy Vehicles	Total
Single vehicle accidents	91	2,918	11,703	319	561	15,592
Pedestrians & Cyclists	23	1790	5825	163	394	8195
Two-wheeled vehicles		895	17568	258	1182	19903
Four-wheeled vehicles			41745	1904	7339	50988
Public vehicles				23	142	165
Heavy vehicles					347	347
Total	114	5603	76841	2667	9965	95190

Table 1. Summary of various accidents

Type of accident	2018 [%]	2019 [%]	2020 [%]	2022(Apr)[%]
Collision with obstacle	26.19	23.06	27.42	25.93
Rear-end collision	15.37	15.94	14.99	15.33
Side collision	22.54	23.65	22.64	23.62
Head on collision	25.72	26.86	24.85	24.99
Rollover	1.02	0.97	0.89	0.92
Pedestrian hit	6.38	6.66	6.46	6.65
Sudden braking	1.27	1.18	1.3	1.07
Vehicle fall	0.67	0.76	0.64	0.58
Run-off roadway	0.84	0.92	0.81	0.91

Table 2. Types of road accidents.

The approximate estimation of causality and injuries due to poor road infrastructure is a big challenge before the living beings. The order to deal with the problem, in computational science, we can adopt data mining model for different scenario. In any vehicle accident, it studies about the driver's behaviour, road infrastructure and possibilities of weather forecast that could be somewhere connected with different accident incidents. The main problem in the study and analysis of accident data is its mix heterogeneous environment and data segmentation which is used widely to overcome accident problem. [5] Data Mining is a computational technique to deal with large and complex data set and these data sets can be of normal, nominal and mixed. It is quite easy to use in variety of domain belong to science and management; also, it could be used in fraud identification and many more scientific cases as well as in accident severity problem. Partition of objects in a group of clusters or in a homogeneous set is a fundamental operation of data mining.

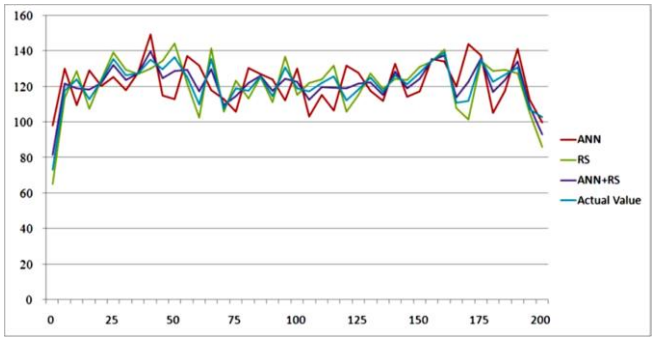

Fig 2. The estimation results and the actual values of the traffic flow

Clustering is a method to partition objects in a similar group. The k-means algorithm having a good efficiency for clustering large data sets but restricted in forming clusters for real word data while working only on numerical data because it helps in reducing the cost function by altering the meaning of the clusters [1,3].

Data mining technique is recognized as reliable technique for analysis of traffic accident severity problem and finding factors behind them. Damage like property, people due to road accident is undesirable. Happened that road accident incidents are more common at certain places that can help in identifying factors behind them. power based multi mobile charger system is implemented in the proposed system which can be used in public places like railway stations, bus stands, hospitals and parks etc. provide scientific and accurate support for dynamic traffic management.

Implementation of Intelligent Traffic Data Mining System The intelligent traffic data mining system has the characteristics of high reliability, high extensibility, and high efficiency. The computation mode is mainly batch processing and stream processing. The structure can be divided into three layers : data source , big data mining platform , and user layer . During the construction of the system, traditional databases and processing tools, graph parallel computing, and in-memory computing are also integrated into the platform. The functions are rich and varied. Relying on supporting technologies,

complex data can be processed in real time and quickly. In the following, the internal working flow will be introduced to the hardware and software environment of the system, system data preprocessing , system data storage , system data calculation and analysis , and system data display to analyze the data processing method and features of intelligent traffic data mining system based on big data mining .

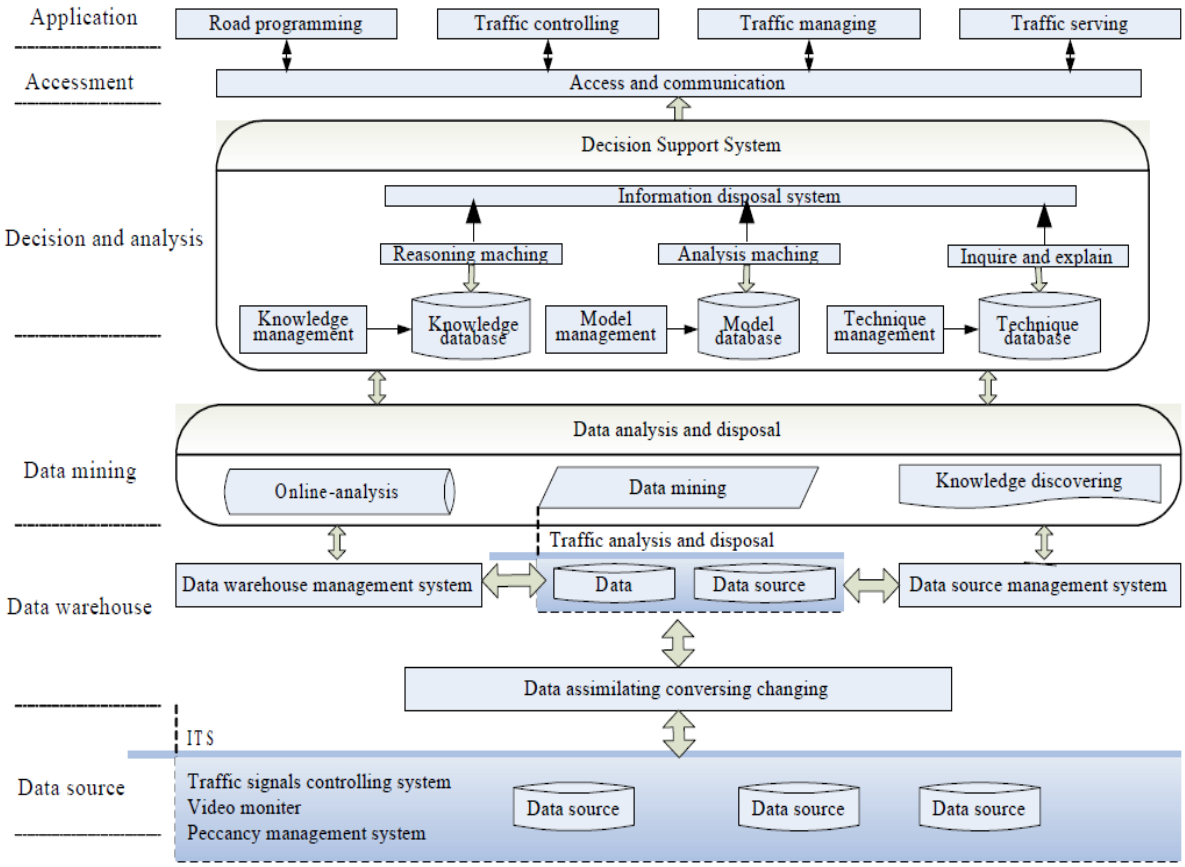

Figure 3. Model of analysis and decision of traffic information

The hardware and software environment of the intelligent traffic data mining system includes several subsystems such as data area, traffic application area, communication area, GIS area and terminal area. A database server is arranged in the data area, a traffic application server is arranged in the traffic application area, a GIS server is arranged in the GIS area, a communication server is arranged in the communication area, and a plurality of operation terminals are arranged in the terminal area. The big data mining theory is used to build the intelligent traffic big data mining system. The system has comprehensive features.

The system design includes a cube customization module, an OLAP analysis mode selection module, a graph display module, and a data mining model selection module. The cube customization module can set the dimensions of the cube and then import the required analysis dimensions or filter out unnecessary dimensions; the OLAP analysis selection module is actually a tool that enables users to use multidimensional views for analysis; The function of the display module is that the user can set the type, vertical and horizontal coordinates of the chart, set the font, name, size, and legend of the chart, and

display the chart results; the function of the data mining model selection module is based on the required analysis of data-reactor characteristics and analysis goals selection and determination of mining models [3].

IV. SYSTEM DESIGN AND IMPLEMENTATION

When designing an intelligent transportation system, it needs to meet the development needs of the transportation business, and also has certain degree of expansion and adaptability. The main application should be to achieve good results in the following aspects: First, GPS data, bayonet data, and aircraft can be used. The data of vehicle traffic flow detection and other data are combined with the road network model to calculate and analyze the traffic operation index through the spatial correlation analysis algorithm, so as to reflect the traffic conditions of the entire city traffic, the degree of congestion, etc., which can provide decisions for traffic planning and congestion management. Reference; Secondly, the use of video surveillance equipment to collect vehicle traffic information, through the establishment of prevention and control database for the dynamic tracking of the target vehicle, can achieve automatic alarm, identify abnormal driving behavior, traffic safety risk assessment and suspects early warning and other functions.

Road users	No of persons killed
Pedestrians	15746
Bicycles	2586
Two wheelers	52500
Auto rickshaws	7150
Cars, Taxis,Vans,Bus	26923

Table 4. Accident on based on type of vehicle

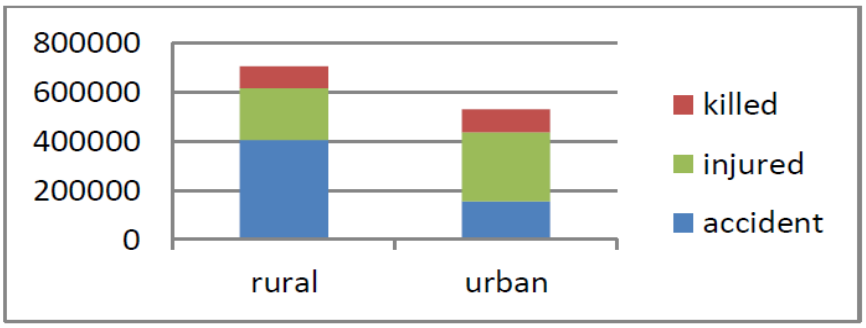

Figure 4. x- axis: accidents,y -axis:person killed on road type

Thirdly, statistics on the traffic flow of road sections can be calculated and forecasted through technologies such as intelligent video processing, neural networks, and association rules. Then, the dangers of each road section can be predicted based on the weather and the number of lanes, and the road sections where traffic accidents may occur are predicted. And alarms to improve the level of traffic management and reduce the chance of accidents. In addition, according to the data of traffic management business, the comprehensive processing of traffic flow, traffic accidents and other information is implemented, and the travel laws[5] are analyzed through the establishment of OD matrix model, thus providing supportive data for decision-making of traffic management organizations.

Road environment information includes mainly about the road network topology, road surface information such as lane number, lane width, intersection drainage, traffic capacity, transport facilities such as signal, lights, speed limit signs, abnormal events information such as construction information, temporary closures and control measures, weather conditions, public transportation information. Some road environment information could not be obtained from the existing system. It will be collected by hand or from other systems. Road traffic data is chronologically sampled numerical data sequence. It is the main management and control object of ITS [1-3]. The vehicles on the road become the traffic flow. Traffic continuous flow is formed by the motor vehicles travelling on the road within a certain time without influence or no horizontal cross sections. Typically traffic continuous flow includes the off-ramp of the supreme viaduct segments, cross-river tunnels and highways. Motor vehicles at a road intersection are controlled by traffic lights, so the traffic flow is showing non-contiguous states. Existing traffic models have inconsistency issues between microscopic models and macroscopic models, which hinder the development and application of traffic theories and models.

Macroscopic traffic model is still lack of theoretical foundation. It has only the empirical regression model. Many transportation phenomena cannot yet be explained. Because of the complexity of traffic system, complete and accurate physical model for the road traffic system cannot be established based on the traditional physics, mathematics, statistical methods [6-10]. Traffic characteristics is determined by the irrational behaviors of human, so there is a big gap between the simplified and assumed traffic model and the transportation system in real world. Real-world traffic behavior does not fully in accordance with the assumed model; therefore, it is unable to effectively analyze traffic conditions and is of significant limitations.

Data mining technology provides a new analytical tool for the processing of traffic data, but common data mining algorithms cannot meet the specific application requirements of ITS for traffic data mining analysis. Direct application of the existing data mining analysis methods to the analysis of traffic data cannot get a good effect. Traffic data is of time characteristics and strong spatial correlation. Various traffic information collection equipment have accumulated massive complex temporal data sets [4]. These data present new data mining needs for ITS, such as complex temporal and spatial queries of road network traffic flow, association between road traffic conditions, and analysis of traffic flow congestion model.

V. CONCLUSION

For the sake of validly analysis of great deals of isomerous data which are accumulated continuously in the application of traffic information system and on order to deal with the challenge from simple static management to intelligent dynamic management, this paper puts forward a model of traffic information intelligence analysis and decision support and further provides a objective and scientific basis for dynamic traffic management. From the data mining perspective, further research should investigate whether limiting the analysis to fatal accidents would simplify the task of data mining techniques in recognizing accident patterns without the “noise” probably created by considering also severe and light injury accidents. A limitation of the approach (as emerged in the predictive analysis) is that data mining fails in capturing patterns not observed previously or the impacts due to the implementation or relaxation of enforcement.

VI. REFERENCES

[1]. WANG Ya-jun, A Data-mining-based Study on Road Traffic Information Analysis and Decision Support.

[2]. Analysis of Road Accidents Using Data Mining Techniques, International Journal of Engineering & Technology, 7 (3.10) (2018) 40-44.

[3]. Ravindra Changala, "Retrieval of Valid Information from Clustered and Distributed Databases" in Journal of innovations in computer science and engineering (JICSE), Volume 6, Issue 1,Pages 21-25, September 2016.ISSN: 2455-3506.

[4]. Ravindra Changala, "Data Mining Challenges With Big Data" International Journal for Research in Applied Science and Engineering Technology (IJRASET)" with Impact Factor 1.241, ISSN: 2321-9653,Volume 3 Issue VI, June 2015.

[5]. S. Shanthi, R. Geetha Ramani, "Classification of Vehicle Collision Patterns in Road Accidents using Data Mining Algorithms", International Journal of Computer Applications Vol-35, December 2011.

[6]. Dheeraj Khera, Williamjeet Singh, "Prediction and Analysis of Injury Severity in Traffic System using Data Mining Techniques", International Journal of Computer Applications, 2015.

[7]. S. Shanthi, R. Geetha Ramani, "Feature Relevance Analysis and Classification of Road Traffic Accident Data through Data Mining Techniques", Proceedings of the World Congress on Engineering and Computer Science 2012.

[8]. Ravindra Changala, "Statistical Models in Data Mining: A Bayesian Classification" in International Journal of Recent Trends in Engineering & Research (IJRTER), volume 3, issue 1, pp.290-293. in 2017.

[9]. Ravindra Changala, "Secured Activity Based Authentication System" in " in Journal of innovations in computer science and engineering (JICSE), Volume 6, Issue 1,Pages 1-4, September 2016.ISSN: 2455-3506.

[10]. Lu Biao, Li Yue, Zhang Wanli. Research and design of intelligent traffic data analysis platform system based on big data technology[J]. Journal of Hubei Institute of Science and Technology, 2016(05):6-9.

[11]. Wu Weiqiang. Intelligent Traffic Decision Analysis System Based on Big Data Mining[J]. Mechatronic Engineering Technology, 2017(s2).

[12]. Dikaiakos M. D.. Minersoft: ACM Transactions on Internet Technology, v.12, n.1, June 2012.

[13]. Seng Dewen, Shu Yueqing. Advances in Intelligent Systems and Computing, p. 393-400, 2013.

[14]. Du P. J, Liu S.C., Xia J. S. Information Fusion, v.14, n.1, pp.19-27, January 2013.

Cite this article as :

Dr. Mahesh Kotha, Sathini Santhosh Kumar, A Deepika, M Swathi Reddy "A Novel Framework for Analysis of Road Traffic Information for Decision Support by Using Data Mining Techniques ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 6, pp. 268-276, November-December 2022. Available at doi : https://doi.org/10.32628/CSEIT228637

Journal URL : https://ijsrcseit.com/CSEIT228637