

Driving Safety Risk Analysis using Naturalistic Driving Data

F. Mary Harin Fernandez¹, S.Chithra², G.Mounisha³

¹Assistant Professor, Dept. of Computer Science and Engineering, Jeppiaar SRR Engineering College, Chennai.

^{2,3}Final Year Student, Dept. of Computer Science and Engineering, Jeppiaar SRR Engineering College, Chennai.

Abstract :- Driving risk varies well among drivers. The elaborated driving knowledge on a few number of areas which might be collected by victimization big data Technology. With these dataset's a potential-crash info is made that contains vehicle standing, driving setting, road type, climatic conditions and driver details. By victimization these details, differing kinds of risk levels is analysed. About 6 June 1944 of drivers were known as high-hazard and eighteen of driver as high/moderate risk drivers. To recognizing and predicting high-hazard drivers can tremendously profit the event of proactive driver teaching programs and safety countermeasures. The results indicate that speed once braking, Age, temperament characteristics and Environmental conditions have robust relationship to the high-hazard driving or aggressive driving.

Key Words: Potential crash info, Driving risk, Aggressive driving, Road type.

1. INTRODUCTION

It is broadly speaking acknowledged that accidents square measure principally wherever the deaths square measure surprising inside life. Accident is Associate in Nursing unforeseen or surprising event that generally cause convenient or nasty consequences, different times be vain. The phrase implies that such a happening may not be preventable since its antecedent circumstances go unrecognized and not addressed . Most of the scientists who study unintentional wounds avoid exploitation the term, "accident" and focus on factors that increase the chance of severe wound that reduce wound incidence and severity. Driving accidents square measure getting to be the foremost wherever the injuries square measure severe or even cause death. There square measure numerous characteristics of driving accidents. Rash or aggressive driving conjointly enclosed. Age, temperament characteristics and Environmental conditions have a powerful relationship with the high-hazard driving or aggressive driving. By using these information safety route is going to be analysed.

2. RELATED WORKS

L.Qi [1] "Research on intelligent transportation system technologies and applications" :World population increasing at a bigger pace alit crossed the digit of 7billion; simultaneously the planet economy is additionally growing. individuals are wont to the bigger mobility and

thus once it involves quality Transportation particularly road transportation is that the one that is definitely accessible to everybody. there's little question in higher the individuals victimization the facility a lot of are the transportation conflicts (accidents), and thus there comes the demand of correct systematic demand for facility that is capable of handling giant mass of individuals on wheels safely and it's created positive that it's surroundings friendly yet. Vehicle to vehicle communication, vehicle to infrastructure communication electronic fees assortment are a number of the highly regarded comes undergoing worldwide. once it involves the developing countries like Bharat, Intelligent facility is in primary stage of development. every nation whether developed or developing, once implement the intelligent technologies the surface facility are safest, economical and last however not the smallest amount Environment friendly.

S.-H. An, B.-H. Lee, and D.-R. Shin [2] "A survey of intelligent transportation systems" :Transportation or transport sector may be a legal supply to require or carry things from one place to a different. With the passage of your time, transportation faces several problems like high accidents rate, holdup, traffic amp; carbon emissions pollution, etc. In some cases, transportation sector long-faced assuaging the brutality of crash connected injuries in accident. because of such complexness, researchers integrate virtual technologies with transportation that called Intelligent Transport System. The concept of virtual technologies integration may be a novel in transportation field and it plays an important half to beat the problems in international world. This paper tackles the great kind of Intelligent Transport System applications, technologies and its completely different areas. the target of this literature review is to integrate and synthesize some areas and applications, technologies talk over with all prospects. what is more, this analysis focuses on a good field named Intelligent Transport Systems, discussed its wide applications, used technologies and its usage in several areas severally.

N. Mohamed and J. Al-Jarood [3] "Real-time big data analytics: Applications and challenges":The big information application refers to the distributed applications that are typically massive in scale and typically works with massive volume of data sets. however it's tough for the standard processing applications to handle such an outsized and sophisticated information sets, that triggers the event of massive information applications . however if the info analytics may be tired

time period, a big quantity of benefits can be achieved. That's why, in recent time, a time period massive information application have gained a heavy attention for generating a timely response. A time period massive information associate degree analytic application is an programme that method among a timeframe and generate a quick response (real-time or nearly time period response). Example of massive data analytics application may be within the space of transportation, financial service like exchange, military intelligence, resource management natural disaster, numerous events/festivals, etc. The latency of this kind of application typically measured in milliseconds or seconds however really for many application it may be measured in minutes.

Y. Liu, X. Weng, J. Wan, X. Yue, and H. Song [4] "Exploring data validity in transportation systems for smart cities": A new framework for emulating the practicality of a device by victimisation multiple on the market soft sensors and machine intelligence algorithms. As a case study, the localization of town buses during a sensible town setting is investigated by victimisation the measuring system and microphones of the passengers and a Support Vector Machine (SVM) running within the cloud; in this application, the GPS practicality is emulated by victimisation these two soft sensors. What makes such Associate in Nursing emulation possible is that the statistical dependence of the placement knowledge (which would usually be obtained from a GPS) on the measuring system and mike data whereas accelerometers capture knowledge that relate to the everyday stop start patterns of the buses, mike capture enter/exit patterns of the passengers through the sound levels within the bus we tend to assess our planned theme through simulations and show that the planned framework will operate with over 0% accuracy in estimating the placement of public buses whereas preserving the particular location privacy of the smartphone users. This approach leads to smartphone battery energy savings of 8-46% (as compared to GPS-based approaches) because of the elimination of the power-hungry GPS devices.

Chang YU, Zhao-Cheng HE [5] "Spatial-temporal daily frequent trip pattern of public transport passengers using smart card data": As the basic travel service for urban transit, bus services carry the bulk of urban passengers. A better understanding of transit riders' travel characteristics will give a first-hand reference for the analysis, management and coming up with of urban conveyance system. Over the past twenty years, knowledge from good cards have become a replacement supply of travel survey knowledge, providing a lot of comprehensive spatial-temporal data about urban conveyance visits. during this paper, a strategy for mining positive identification knowledge is developed to recognize the travel patterns of transit riders. a wise card dataset is 1st processed to get the trip information when reconstructing the transit trip chains from the trip data,

this paper adopts the density based abstraction bunch of application with noise (DBSCAN) rule to mine the historical travel patterns of each transit riders. additionally, a sensitivity analysis is conducted to judge the optimum parameters. In case study the analysis of travel pattern characteristics is conducted specializing in the transit riders of port City, China.

Marco Di Felice, Rahman Doost-Mohammady [6] "Smart Radios for Smart Vehicles": The recent strides created in vehicular networks have enabled a brand new category of in car entertainment systems and increased the flexibility of emergency responders mistreatment opportunist spectrum usage enabled by psychological feature radio (CR) technology. These CR-enabled vehicles (CRVs) have the flexibility to use additional spectrum opportunities outside the IEEE 802.11p specified standard 5.9-GHz band. The aim of this text is to produce taxonomy of the present literature on this fast-emerging application space of CRV networks, highlight the key analysis that has already been undertaken, and point toward the open issues. We explore completely different architectures [i.e., fully suburbanized as well as base station (BS) supported], the sensing schemes suited for extremely mobile eventualities with stress on cooperation, and spectrum access strategies that assure the provision of the required quality of service (QoS). Moreover, we tend to describe the planning of a brand new machine tool that's ready to merge data from real world street maps with authorized user activity patterns, there by resulting in a strong platform for testing and analysis of protocols for CRVs.

Teresa Orlowska-Kowalska, Senior Member, IEEE, Mateusz Korzonek, Student Member, IEEE, Grzegorz Tarchała [7] "Stabilization Methods of Adaptive Full-Order Observer for Sensor less Induction Motor Drive - Comparative Study": The Adaptive Full-order Observer (AFO) is that the hottest speed estimator for the induction motor (IM) drives. Therefore, due to its instability in the regenerating mode a few stabilization methods has been proposed in the literature. However, there is a lack of fair comparison of all advantages and drawbacks of these proposals. Thus, in this paper the classical AFO and its all known stabilization methods are compared with a recent new proposal, which is based on the introduction of an auxiliary adaptive variable in the observer state matrix. The stability analysis of this new proposal is presented in details in this paper.

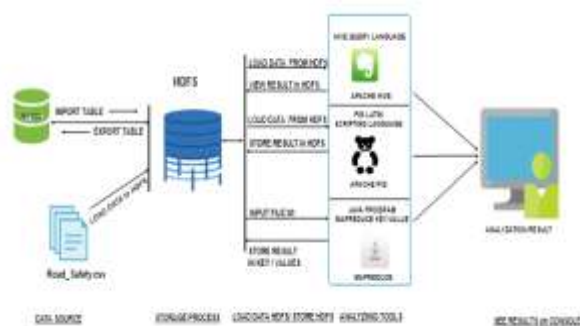
4. PROPOSED SYSTEM

In Road safety what precautions got to be compelled to soak up order to decrease the accidents occurring in cities. By observing the road traffic routine wise what's happening in the universe, meanwhile crossing each stage of traffic issues which could be analyzed into a few years and implemented in Traffic analysis wing (TRW). Planned thought deals with providing data by using the Hadoop

tool. The Hadoop tool contains two things one is MapReduce and another is HDFS. In Hadoop, their square measures techniques like joins, partitions and bucketing. This planned system would help to analyze the protection route by collected data. It's a question primarily based analysis in huge information technology. Through the planned system the protection routes are progressing to be analyzed. It helps to the strangers to know the route that is safe from driving risks. No data loss problem. Efficient data processing. Getting results with less time, high throughput and maintenance cost is very less. There is no limitation of data and simple add number of machines to the cluster.

4. PROPOSED ARCHITECTURE

System architecture can comprise system components, this internally works in the form of java code, that will work together to implement the overall system for large data handling processing and cost is less.



5. METHODOLOGY

5.1. MAPREDUCE

Map stage – The map or mapper’s job is to method the computer file. Typically the computer file is within the kind of file or directory and is hold on within the Hadoop classification system. The computer file is passed to the plotter operate line by line. The plotter processes the information and creates many little chunks of information.

Reduce stage – This stage is that the combination of the Shuffle stage and also the reduce stage. The Reducer’s job is to method the information that comes from the plotter. When process, it produces a replacement set of output, which can be hold on within the HDFS.

5.2. HDFS

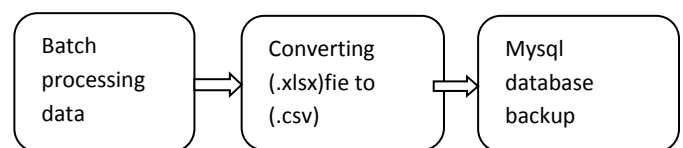
Hadoop classification system was developed victimisation distributed classification system style. it’s run on trade goods hardware. In contrast to different distributed systems, HDFS is very faulttolerant and designed victimisation inexpensive hardware. HDFS

holds terribly great amount of information and provides easier access. To store such immense information, the files are hold on across multiple machines. These files are hold on in redundant fashion to rescue the system from attainable information losses just in case of failure. It is appropriate for the distributed storage and process. Hadoop provides a command interface to act with HDFS. The intrinsic servers of namenode and datanode facilitate users to simply check the standing of cluster. Streaming access to classification system information. HDFS provides file permissions and authentication.

6. MODULES AND DESCRIPTION

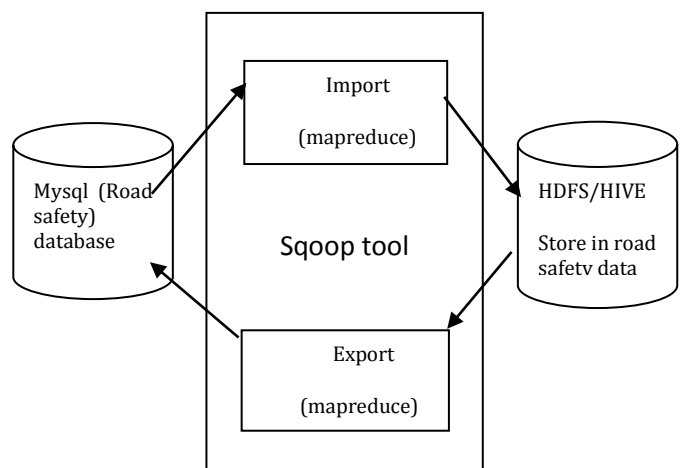
6.1. PREPROCESSING ROAD ACCIDENT DATABASE

In this work, analyzing the data with different kinds of fields in Microsoft Excel then it converted into comma delimited format which is said to be csv(comma separator value) file and moved to MySQL backup through Database.



6.2. DATA STORAGE IN HDFS

In this work, getting all those backup data which we have stored in MYSQL and importing all those data by use of sqoop commands to HDFS (Hadoop Distributed File System). Now all the data are stored in HDFS were it is ready to get processed by use of hive.



6.3. ANALYSE HIVE QUERY:

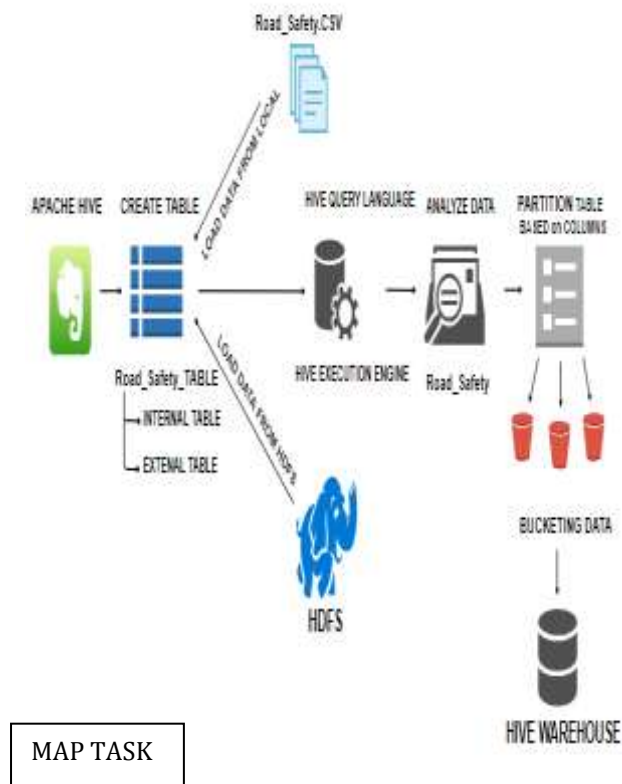
In this work, getting all those data from HDFS to HIVE by use of sqoop import command .where hive is ready to analyze. Here in HIVE user can process only structured data to analyze. By extracting only the meaningful data

and neglecting un-cleansed data user can analyze the data in more effective manner by use of hive.

As soon as enter a Load statement in the Grunt shell, its semantic checking will be carried out. To see the contents of the schema, Dump operator is used. The MapReduce job for loading the data into the file system will be carried out only after performing the dump operation. Pig provides many built-in operators to support data operations like grouping, filters, ordering, etc.

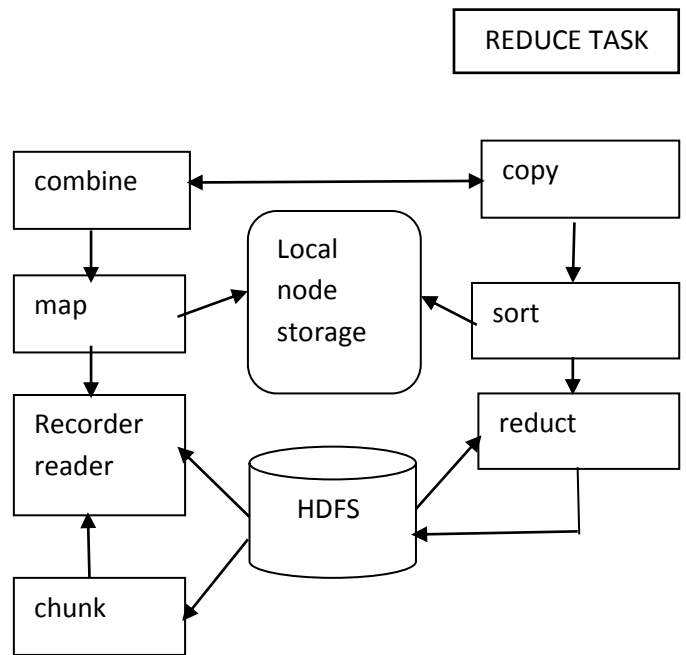
6.5. PARALLEL PROCESS:

Map reduce may be a framework victimization that users are able to write applications to method vast amounts of Road traffic knowledge, in parallel, on massive clusters of artefact hardware in a very reliable manner. Map reduce may be a process technique and a program model for distributed computing supported java. The Map reduce rule contains 2 necessary tasks, particularly Map and cut back. Map reduce program executes in 3 stages, particularly map stage, shuffle stage, and reduce stage.



6.4. SCRIPTING PROCESS(PIG):

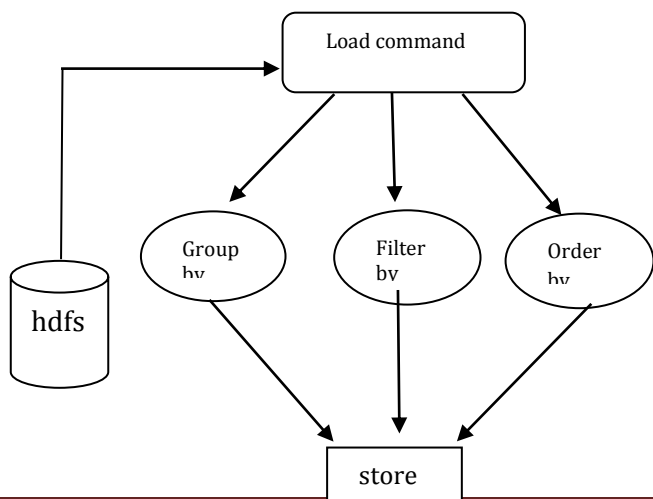
To analyze Road safety using Pig, programmers need to write scripts using Pig Latin language and execute them in interactive mode using the Grunt shell. All these scripts are internally converted to Map and Reduce jobs. After invoking the Grunt shell, it runs the Pig scripts in the shell. Except the commands LOAD and STORE, while performing all other operations, Pig Latin statements take a relation as input and produce another relation as output.



The map or mapper's job is to method the input file. Typically the input file is within the style of file or directory and is hold on within the Hadoop classification system (HDFS). The input data is passed to the plotter perform line by line. The plotter processes the info and creates many tiny chunks of information. This stage is that the combination of the Shuffle stage and also the reduce stage. The Reducer's job is to method the info that comes from the plotter. Once process, it produces a brand new set of output, which can be hold on within the HDFS.

7. RESULT

The query analysisation produce the result which contains the route details as shown in the diagram.



```
File Edit View History Bookmarks Tools Help
http://localhost:50075/browseBack.jsp?blackid=87435602839055105836&backSize
HDFS:usehivewarehouse@hive...
File: /user/hive/warehouse/drive.db/saf_partroad_class=4/060000_0
Get: /user/hive/warehouse/drive go
Go back to driving
Accessed via terminal session
View Next chunk
121009,430708,430730,1,8:40 AM,Darkness: street lights present and lit.off.Fine without high winds,1,Driver,Slight,Male,
1211459,428367,434449,2,5:40 PM,Darkness: street lights present and lit.off.Fine without high winds,1,Driver,Slight,Male,
1211579,434145,430427,2,6:40 PM,Darkness: street lights present and lit.off.Fine without high winds,1,Driver,Slight,Male,
Traffic_rama],Egnera
1230039,428677,431760,2,8:40 PM,Darkness: street lights present and lit.off.Fine without high winds,1,Driver,Slight,Male,
1230433,429070,429670,2,8:40 PM,Daylight: street lights present,on.Fine without high winds,1,Driver,Slight,Male,36,Motorc
Traffic_Kattuparam
1230257,422143,426046,2,5:40 AM,Darkness: street lights present and lit.off.Fine without high winds,2,Passenger,Slight,M
1230444,421474,465809,2,8:40 AM,Daylight: street lights present,on.Fine without high winds,1,Driver,Slight,Male,26,Car,56
1230781,438776,435205,2,8:40 AM,Daylight: street lights present,on.Fine without high winds,1,Driver,Slight,Male,36,Taxi,PP
1230446,434833,434423,2,8:40 PM,Daylight: street lights present,on.Fine without high winds,2,Passenger,Slight,Male,21,Tax
1230446,434833,434423,2,8:40 PM,Darkness: street lights present and lit.off.Fine without high winds,1,Driver,Slight,Male,
1230446,434833,434423,2,5:40 PM,Darkness: street lights present and lit.off.Fine without high winds,1,Passenger,Slight,M
1230239,439845,435322,2,8:40 PM,Daylight: street lights present,on.Fine without high winds,1,Driver,Serious,Male,45,Motor
traffic_warehouse
```

8. CONCLUSION

In this paper, The Road accident report is generated by using real-time data backup hardly taken two years of data is presented. The Report contains the details about the road condition, weather condition, driving details, safe path etc., Due to advancement of big data analytics its processing speed is fast and useful for prediction purposes which indicate to perform a proper maintenance.

9. FUTURE ENHANCEMENT

The transport research wing able to find issues behind the road traffic and also it will be helpful to find the reasons beforehand by predict with the help of big data analytic report. The spark implementation is used for 100 times faster further. Apache Spark is an open source processing engine built around speed, simple use, and analytics. If the large amounts of data that requires low latency processing that a typical Map Reduce program cannot provide, Spark is the alternative. Spark provides in-memory cluster computing for fast speed.

10. REFERENCES

[1] J.R. Treat, N. S. Tumbas, S. T. McDonald, D. Shinar, R. D. Hume, and R. E. Mayerm, "Tri-level study of the causes of traffic accidents: Interim report I, volume I research findings," NHTSA, Washington, DC, USA, Tech. Rep. DOT HS-805 085, 1979.

[2] J. Buckley and I. James, "Linear regression with censored data," *Biometrika*, vol. 66, no. 3, pp. 429-436, 1979.

[3] D. L. Hendricks, J. C. Fell, and M. Freedman, "The relative frequency of unsafe driving acts in serious traffic crashes," NHTSA, Washington, DC, USA, Tech. Rep. DTNH22-94-C-05020, 1999.

[4] M. Smith and H. Zhang. (Task 9): A Literature Review of Safety Warning Countermeasures. Safety Vehicles Using Adaptive Interface Technology.2004

[5] G. Rodriguez, "Parametric survival models," Lectures Notes, Princeton University, 2005

[6] Institute for Road Safety Research. SWOW Fact Sheet. Naturalistic Driving: Observing Everyday Driving Behavior.2012.

[7] V. Beanland, M. Fitzharris, K.L. Young, M.G. Lenne Driver inattention and driver distraction in serious casualty crashes: data from the Australian National Crash in-depth study *Accid. Anal. Prev.*, 54 (2013)

[8] R. Yua and M. Abdel-Atya, "Multi-level Bayesian analyses for single-and multi-vehicle freeway crashes," *Accident Anal. Prevention*, vol. 58, pp. 97-105, Sep. 2013.

[10] Yu WD, Pratiksha C, Swati S, et al. "A Modeling Approach to large data based Recommendation Engine in fashionable Health Care Environment", portable computer package and Applications Conference, IEEE portable computer Society, pp.75-86, 2015.

[11] Vancampfort D, Mugisha J, Hallgren M, et al. "The prevalence of DM kind a combine of in of us with alcohol use disorders: a scientific review and large scale meta-analysis", *psychopathology analysis*, vol. 246, pp. 394-400, 2016.

[12] port A-C, Ziefle M, Verbert K, et al. "Recommender Systems for Health Informatics: progressive and Future Perspectives", *Machine Learning for Health science*, Springer International industrial enterprise, 2016.

[13] M. Alaa and M. van der Schaar, "Deep multi-task mathematician processes for survival analysis with competitive risks," in *Proceedings of the thirtieth Conference on Neural science Systems*, 2017.

[14] N. Arbabzadeh and M. Jafari, "An information-Driven Approach for Driving Safety Risk Prediction victimization Driver Behavior and route information information," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 2, pp. 446-460, Feb. 2018.