

Driving Safety Risk Analysis Using Naturalistic Driving Data

*P. Sathya Siva Reddy, ²N. Deepa

**UG Scholar, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India*

²Assistant professor, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India

**satya90009@gmail.com, ²ndeepa.sse@saveetha.com*

Article Info

Volume 82

Page Number: 10441 - 10447

Publication Issue:

January-February 2020

Abstract

Driving danger differs generously among drivers. Notwithstanding the genuine car crash information, naturalistic driving information can permit specialists gain bits of knowledge into the elements that reason chance/danger circumstances. Utilizing Big Data Technology, we are gathering nitty gritty driving information on a couple of regions. With these dataset's we are building a potential-crash database which contains Vehicle status, driving condition, street type, climate conditions and driver subtleties in which we investigate to get various kinds of hazard levels. Surmised 6% of drivers were recognized as high-hazard and 18% of driver as high/moderate hazard drivers. Recognizing and foreseeing high-chance drivers will significantly profit the improvement of proactive driver training projects and security countermeasures. The outcomes show that Velocity while braking, Age, Personality qualities and Environmental conditions have solid relationship to the high-chance driving or forceful driving.

Article History

Article Received: 18 May 2019

Revised: 14 July 2019

Accepted: 22 December 2019

Publication: 19 February 2020

Keywords: Driving, safety, naturalistic driving data, hazards.

1. Introduction

Progressive thickness based grouping is a useful asset for exploratory information investigation. Notwithstanding, its pertinence to enormous datasets is restricted on the grounds that the computational intricacy. Information bunching is a significant information mining innovation that assumes an essential job in various logical applications. In any case, it is trying because of the size of datasets has been developing quickly to extra-huge scale in reality. Guide Reduce is a well-known programming model to accelerate information mining and AI calculations working on enormous, perhaps appropriated datasets. In the interim, Map Reduce is an alluring parallel programming stage that is broadly applied in sorts

of information process fields. At last, we investigate our work of the enormous scale datasets utilizing Hadoop instrument. The outcomes uncover that the work can be accelerated and scaled up utilizing Big Data Technology. It will be a lot quicker subsequent to executing the future proposed innovation Spark.

2. Literature Review

Total populace expanding at a greater pace alit crossed the digit of 7billion [1]; at the same time the planet economy is also developing. people are wont to the greater versatility and consequently once it includes quality Transportation especially street transportation is that the one that is certainly open to everyone. There's little question in higher

the people exploitation the office a great deal of are the transportation clashes (mishaps), and accordingly there comes the interest of right deliberate interest for office that is equipped for taking care of mammoth mass of people on wheels securely and it's made constructive that its surroundings neighbourly yet. Worldwide various social orders and affiliations are arrangement for the occasion of shrewd transportation framework, beginning was arrangement in 1991 by North American nation Department of Transportation: alongside this numerous models are arranged in setting for the indistinguishable, just barely any authorized. Vehicle to vehicle correspondence, vehicle to foundation correspondence electronic charges collection are some of the exceptionally respected comes experiencing around the world. When it includes the creating nations like Bharat, Intelligent office is in essential phase of improvement. Each country whether created or growing, when execute the canny advancements the surface office are most secure, efficient and last anyway not the littlest sum Environment benevolent.

Transportation or transport area might be a lawful inventory to require or convey things from one spot to an alternate [2]. With the entry of your time, transportation faces a few issues like high mishaps rate, burglary, traffic amp; carbon discharges contamination, and so forth. Sometimes, transportation area since a long time ago confronted mitigating the mercilessness of crash associated wounds in mishap. As a result of such multifaceted nature, specialists incorporate virtual innovations with transportation that called Intelligent Transport System. The idea of virtual innovations joining might be a novel in transportation field and it plays a significant half to beat the issues in universal world. This paper handles the incredible sort of Intelligent Transport System applications, innovations and its totally various regions. The objective of this writing survey is to coordinate and combine a few regions and applications, innovations talk over with all

possibilities. In addition, this investigation centers around a decent field named Intelligent Transport Systems, examined its wide applications, utilized innovations and its utilization in a few territories severally.

The large data application alludes to the circulated applications that are regularly huge in scale and ordinarily works with huge volume of informational indexes [3]. Anyway it's intense for the standard preparing applications to deal with such an outsized and advanced data sets that triggers the occasion of monstrous data applications. Be that as it may if the data examination might be worn out timeframe, a major amount of befits can be accomplished. That is the reason, in ongoing time, a timespan gigantic data application have increased an overwhelming consideration for producing a convenient reaction. A timeframe monstrous data partner degree application is a program that strategy among a time span and create a speedy reaction (constant or about timeframe reaction). Case of huge information investigation application might be inside the space of transportation, money related assistance like trade, military insight, asset the board cataclysmic event, various occasions/celebrations, and so on. The inertness of this sort of use commonly estimated in milliseconds or seconds anyway truly for some application it might be estimated in minutes.

Another structure for copying the common sense of a gadget by exploitation various available delicate sensors and machine insight calculations. As a contextual analysis, the limitation of town transports during a reasonable town setting is explored by exploitation the estimating framework and mouthpieces of the traveler [4] and a Support Vector Machine (SVM) running inside the cloud; in this application, the GPS reasonableness is copied by exploitation these two delicate sensors. What makes such Associate in Nursing imitating conceivable is that the factual reliance of the arrangement information (which would for the most part be gotten from a GPS) on

the estimating framework and mine information while accelerometers catch information that identify with the regular stop start examples of the transports, mine catch enter/leave examples of the travellers through the sound levels inside the transport we will in general survey our arranged topic through re-enactments and show that the arranged system will work with over 90% precision in evaluating the position of open transports though saving the specific area security of the cell phone clients. This methodology prompts cell phone battery vitality reserve funds of 8–46% (when contrasted with GPS-based methodologies) on account of the end of the eager for power GPS gadgets.

As the fundamental travel administration for urban travel [5], transport administrations convey the heft of urban travellers. A superior comprehension of travel riders' movement attributes will give a direct reference for the examination, the board and thinking of urban

transport framework. In the course of recent years, information from great cards have become a substitution supply of movement study information, giving a ton of far reaching spatial-transient information about urban transport visits. During this paper, a methodology for mining positive distinguishing proof information is created to perceive the movement examples of travel riders. a shrewd card dataset is first prepared to get the outing data while reproducing the travel trip chains from the outing information, this paper receives the thickness based deliberation pack of use with commotion (DBSCAN) rule to mine the recorded travel examples of each travel riders. Furthermore, an affectability examination is directed to pass judgment on the ideal parameters. On the off chance that review the examination of movement design attributes is led gaining practical experience in the travel riders of port City, China.

3. Proposed Approach

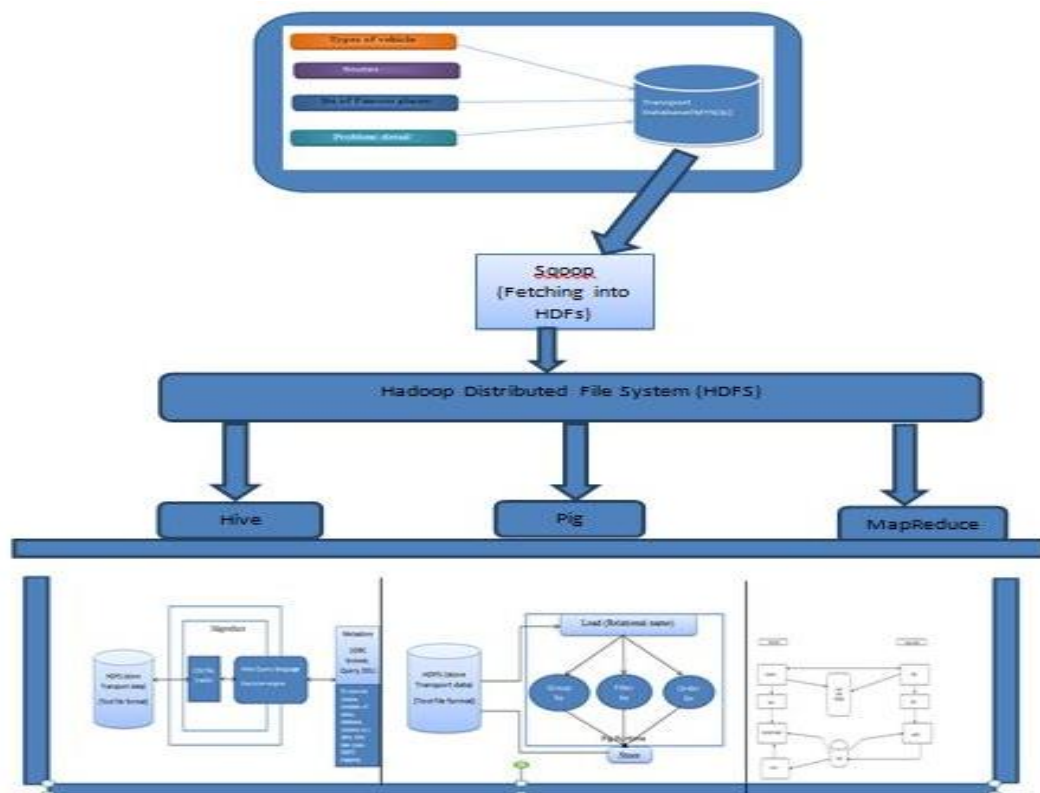


Figure 1: Proposed Approach

Pre-processing Transport System Database

In this module, examining the information with various types of fields in Microsoft Excel then it changed over into comma delimited arrangement which is said to be csv (comma separator esteem)

record and moved to mysql reinforcement through Database.

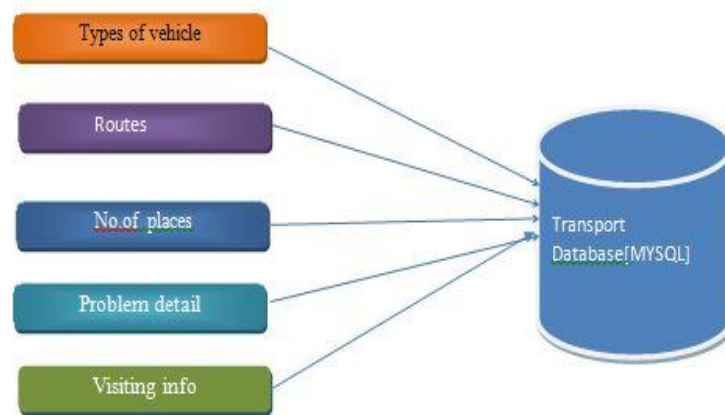


Figure 2: Database

Here by getting the historical data we have to convert those historical batch processing data from (.xlsc) format to (.csv) format and by taking backup of all those data in MYSQL Database to avoid loss of data.

In this module we are 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 where it is ready to get processed by use of hive.

Storage

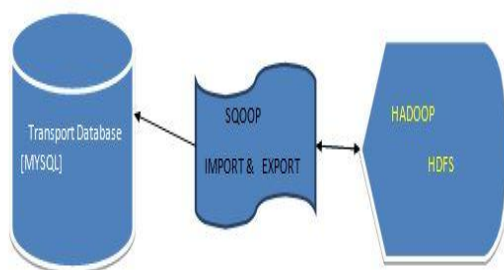


Figure 3: Storage

Analyze Query

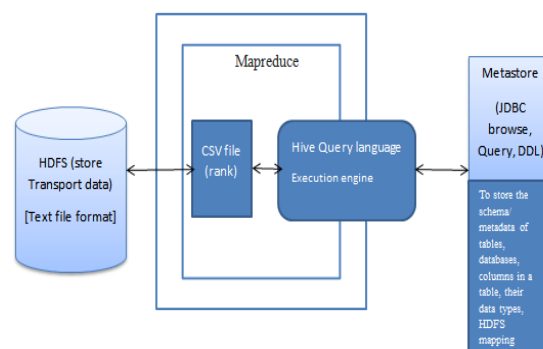


Figure 4: Query Analysis

In this module we are getting all those data from HDFS to HIVE by use of sqoop import command. Where hive is ready to analyze. Here in HIVE we

can process only structured data to analyze. By extracting only the meaningful data and neglecting unclenched data we can analyze the data in more effective manner by use of hive.

Analysis Latin Script (Pig): To analyse User Transport System utilizing Pig, software engineers need to compose contents utilizing Pig Latin language and execute them in intuitive mode utilizing the Grunt shell. Every one of these contents are inside changed over to Map and Reduce tasks. After summoning the Grunt shell, you can run your Pig contents in the shell. Except LOAD and STORE, while playing out every single other activity, Pig Latin articulations accept a connection as information and produce another connection as yield. When you enter a Load articulation in the Grunt shell, its semantic looking freely is conveyed. To see the substance of the pattern, you have to utilize the Dump administrator. Simply in the wake of playing out the landfill activity, the MapReduce work for stacking the information into the record framework will be done. Pig gives many inherent administrators to help information tasks like gathering, channels, requesting, and so on.

Processing (Mapreduce): MapReduce is a structure utilizing which we can compose applications to process gigantic measures of Transport System, in parallel, on huge groups of product equipment in a dependable way. MapReduce is a preparing strategy and a program model for appropriated registering dependent on java. The MapReduce calculation contains two significant errands, specifically Map and Reduce. MapReduce program executes in three phases, in particular guide organize, mix arrange, and decrease stage. The guide or mapper's main responsibility is to process the information. By and large the info information is as document or index and is put away in the Hadoop record framework (HDFS). The information record is passed to the mapper work line by line. The mapper forms the information and makes a few little pieces of data. This arranges is the mix of the Shuffle organize and the Reduce arrange. The Reducer's main responsibility is to process the information that originates from the mapper. Subsequent to handling, it creates another arrangement of yield, which will be put away in the HDFS.

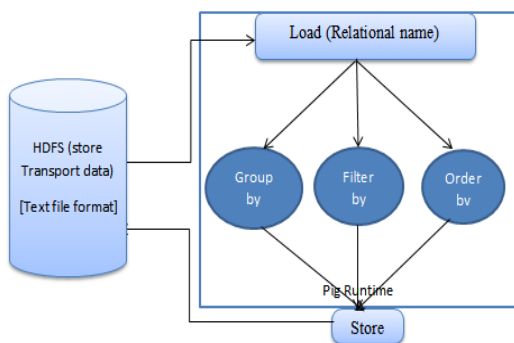


Figure 5: HDFS

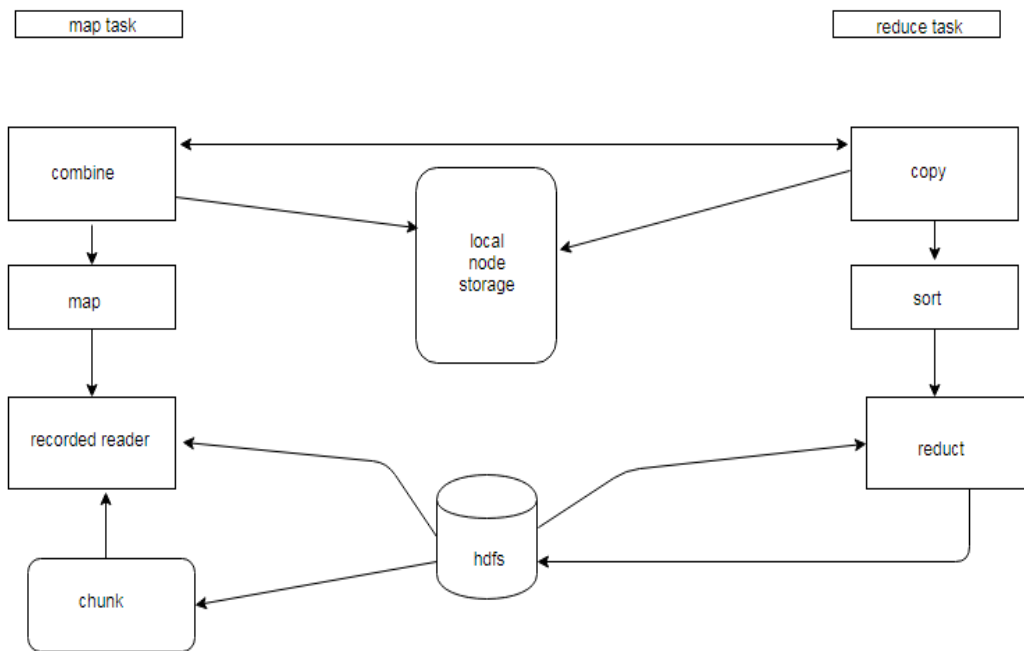


Figure 6: Node Storage

4. Result

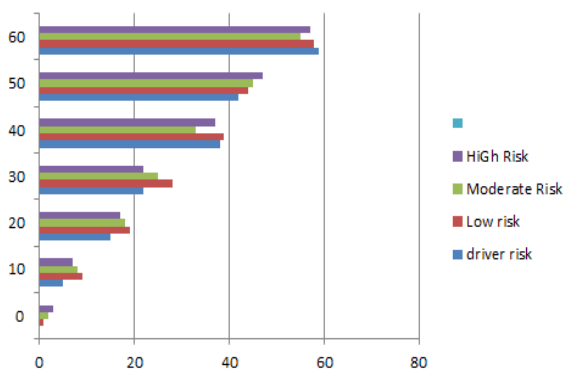


Figure 7: Comparison Bar chart for risk Analysis data.

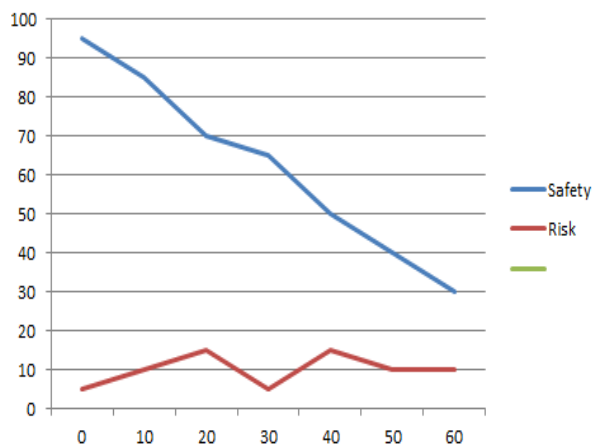


Figure 8: Measuring safety and risk tasks

5. Conclusion

Apache Spark is an open source preparing motor worked around speed, instance of utilization, and examination. On the off chance that you have a lot of information that requires low idleness preparing that an ordinary Map Reduce program can't give, Spark is the other option. Sparkle gives in-memory group figuring to extremely quick speed and supports Java, Scala, and Python APIs for simplicity of improvement. In this paper, we exhibited an investigation on Transport System is help to offer attention to choose best course among choices what we have in datasets. To examine the Transport System information in Hadoop environment. Hadoop environment is hive, pig, mapreduce. In future the flash multiple times quicker than hadoop, it is effectively investigation quicker.

References

- [1] G. Bello-Orgaz, J. J. Jung, and D. Camacho, "Social big data: Recent achievements and new challenges," *Inf. Fusion*, vol. 28, pp. 45–59, Mar. 2016.

- [2] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Netw. Appl.*, vol. 19, no. 2, pp. 171–209, Apr. 2014.
- [3] H. Chen, R. H. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact," *MIS Quart.*, vol. 36, no. 4, pp. 1165–1188, 2012.
- [4] T. B. Murdoch and A. S. Detsky, "The inevitable application of big data to health care," *JAMA*, vol. 309, no. 13, pp. 1351–1352, 2013.
- [5] M. Mayilvaganan and M. Sabitha, "A cloud-based architecture for bigdata analytics in smart grid: A proposal," in *Proc. IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC)*, Dec. 2013, pp. 1–4.
- [6] L. Qi, "Research on intelligent transportation system technologies and applications," in *Proc. Workshop Power Electron. Intell. Transp. Syst.*, 2008, pp. 529–531.
- [7] S.-H. An, B.-H. Lee, and D.-R. Shin, "A survey of intelligent transportation systems," in *Proc. Int. Conf. Comput. Intell.*, Jul. 2011, pp. 332–337.
- [8] N.-E. El Faouzi, H. Leung, and A. Kurian, "Data fusion in intelligent transportation systems: Progress and challenges—A survey," *Inf. Fusion*, vol. 12, no. 1, pp. 4–10, 2011.
- [9] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data driven intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, Dec. 2011.
- [10] Q. Shi and M. Abdel-Aty, "Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 380–394, Sep. 2015.