A survey paper on Harnessing Big data for Estimation of Driving Events through Sensor data using Hadoop Ecosystem

[1]Pooja M. Gupta and [2] Shrikant.D.Zade

[1]M.Tech. Scholar, [2]Associate Prof. (CSE)
PIET, Nagpur, India

Abstract: **In this paper we are using some machine learning approaches for the estimation of driving patterns and driving behavior of respective drivers. We collect all the data of a respective vehicle and make computations using hadoop tools. The first step is to built database to estimate the electrical vehicles energy consumption from driving logs for examine a data analysis framework using key points curb weight , motors capacity, speed acceleration , cargo weight , battery or fuel consumption ,transit time . We propose the business objective, to better understand the risk is under from fatigue of drivers, over-used trucks, and the impact of various trucking events on risk such as risk factor, driver mileage, truck mileage, average mileage. This use case is to analyze various parameter of a truck fleet. Each truck has been equipped to log location and event data. We propose the use of sensor big data to detect these events. We focus on two main infrastructure systems, transportation and energy. The goal of a project is to get familiar with end to end implantation of a big data project using various tools in Hadoop ecosystem. Collected geo-location and truck data has been provided for the estimation of proper solutions for improving the driving range anxiety.**

IndexTerms **- Electric vehicle, Drivers behavior , machine learning , single node cluster, Energy consumption rate , range anxiety , driving environment , GPS**

I. INTRODUCTION

A big data information is the collection of data for extracting the different operations from vehicle such as state of battery, driving condition, speed, time which is used to analyze the driving patterns of Electric vehicle owners.

We used big data to understand the relationship between the driving patterns, energy consumption of electric vehicle applications such as mileage, risk factor, range anxiety etc. The large amount of data is generated by Electric vehicle that is not only useful useful for the drivers but also able to benefit for the automakers. The detailed data records are to be maintain by recording state of health and operating log. By using GPS tracker we can extract the real time data of a driver's behavior, energy consumption and real time information about distance, time and location.

The paper represents the three essential major factors of electric vehicle as shown in fig 1[1]- 1. Machine learning approach 2. Pattern recognisation 3. Sensor information

The growth of electric vehicles (EV) depends on the way people transit. The major issue of this EV usage is still range anxiety. To overcome this problem the sensor can be used to sense the real time operations on EV. Most of the production EV models take time to enhance the driver's confidence. For instance, drivers to be concern about distance travelled and about fuel consumption and battery. The data can be extract online inspection with the internet of vehicle (IOV) . Each EV contains three domain knowledge powertrain design, battery management and information technology which are required to consider every respect of the system. We attempted to transform the driving log into recognized driving patterns. Range anxiety is the biggest problem issue for drivers that fear of losing power and seeing EV shut down in the middle of a long-distance drive.

A quick way to overcome from this problem increasing the EV battery capacity , use of battery data on long trip optimal energy planning , power grid information on development of battery charging stations and GPS(Global Positioning System)data for vehicle mass estimation to optimize the development of charging station.[1] Big data collected EV's from creates an unprecedented opportunity for developing ways for transportation and information exchange with more vehicle connected to the EV data cloud . The goal of this work is to develop an approach offering a sensible energy management scheme for automaker to estimate the range anxiety for the drivers.

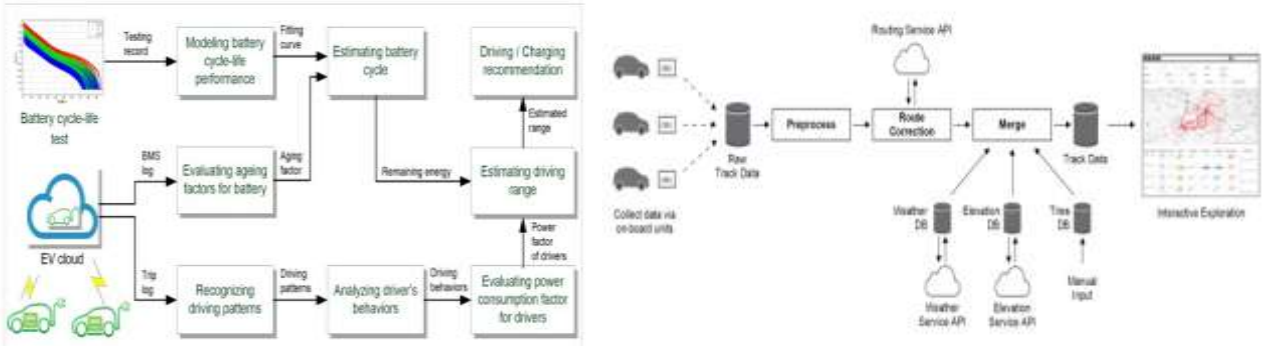

Figure 1: Research framework

In Transport Network Company (TNC) driers third issue of a EV comes at the transportation time for long distance. Where drivers and automakers don't know the present state of health of driver as well as vehicle. To overcome from this problem they need to use sensor and GPS. The present state of health and location, speed and time to reach from source to destination are the key factors should be known.

II. LITERATURE REVIEW

In the existing research we have to focus on the three issues Battery State of health (SOH) estimation, driving patterns, fuel consumption. Unlike traditional ICE vehicles, the big data associated with EV operations to analyze the driving condition, condition of battery, fuel consumption etc. can be used to analyze the patterns of EV owners. The huge amount of information is generated by EV which is not only useful for the drivers but also able to benefit for the engineers and vendors. The detailed data of an EV such as State of health (SOH) and error log like accidents, unpredictable conditions able to recorded and transmitted over a massive scales. In some cases, to adapt and react to changes in the environment automatically as the cloud computing model for managing the real-time streams of smart grid data and big data application for business trend [1].

We performed the five types of big data were considered for estimating the driving range in including route information, weather information, electrical vehicle modeling data ,driving behavior data. The proposed idea is used the sensor to overcome from the problems of SOH condition and range anxiety. Some researchers have proposed the different methods but we can use genetic algorithm (GA), neural network and single node cluster. However most of work only used single cell as testing. Driving habits and fuel consumption affects the driving range of the EV's but improved the driving skills with constant velocity. We applied EV big data to analyze the driving patterns, energy efficiency and SOH by using machine learning approaches. In order to understand the driving patterns of each truck. In this work, an unsupervised clustering approach is applied to formulate its driving behavior. We performed an analysis on the effect of range anxiety quantifying the state of charge (SOC) where drivers appear reluctant to start an electric journey and the effect of SOC has efficiency. It helps to reduce the modify their driving style to conserve the energy when SOC reduces.

III. EXISTING SYSTEM

The major issue for automakers is of range anxiety which can be overcome by combining EV battery modeling method and driving pattern analysis to improve the accuracy. It is necessary to collect the related systematic information which affects the performance of an EV. For better understanding of this problem we are using truck as a electric vehicle and performing the hadoop tools for estimation of different operations of truck fleet and drivers behavior. The main focus on the proposed work to estimate the SOH, SOC and mileage of respective drivers and vehicle. Each truck will equipped with log location and event data. The data can be stored into RDBMS and by using log and current scenario of a respective truck will appear in the record which will helpful for the transport company for the estimation of performance of each truck and for the enhancement in the drivers driving skills. It also beneficial for the business trend.

Disadvantages of Existing system

1. Driver doesn't know about the upcoming weather information, SOH and error logs.
2. The problem of range anxiety occurs.
3. Does not have any idea about current location and time to reach during transportation.

4. Automakers don't have any idea about the current situation of the drivers speed, fuel consumption, risk factors and mileage of vehicle.
5. Most of the company faces the problem to pay the maximum level of premium for long-distance transportation.

IV. PROPOSED SYSTEM

The system consists of the single node clustering technique in hadoop. It consist of single node as master as well as slave which can extract the data from different vehicle and which can be estimate and record can be stored into a single machine.

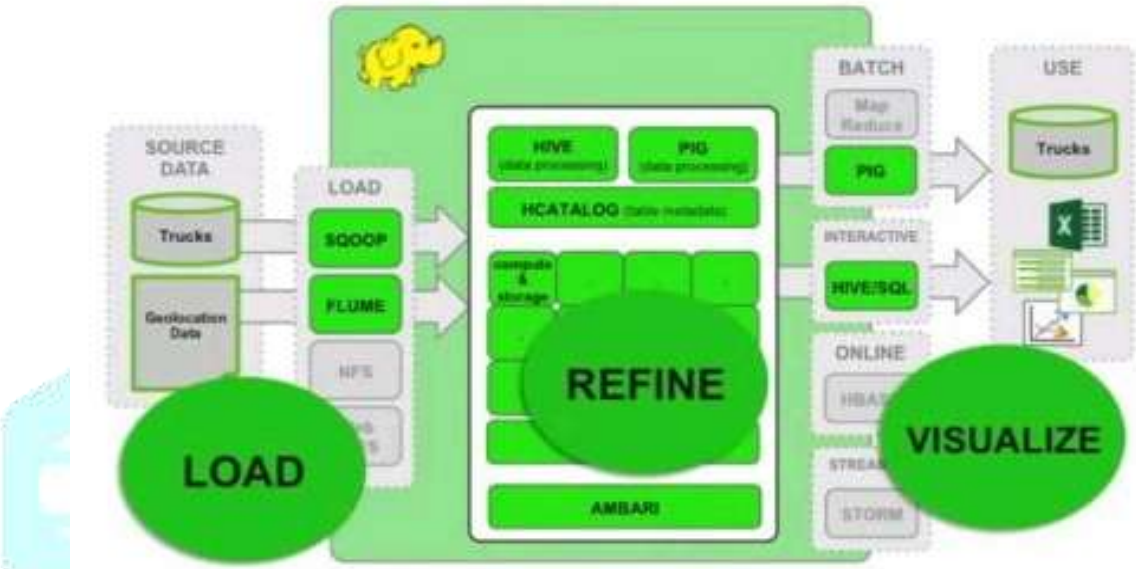

Figure 3: System Architecture

4.1 Advantages of Proposed System

1. It provides the faster results for the estimation of accuracy of a vehicle and fuel consumption.
2. It reduces the risk factor for drivers.
3. It helps to automakers to judge each drivers driving skill.
4. For long distance transportation upcoming weather information, fuel consumption and the issue of range anxiety can be remove.
5. It is the best approach for the transportation network company (TNC) for maintenance of a premium.
6. Improve the accuracy and efficiency of a vehicle.
7. The real time scenario, time, speed and location can be able to track.

4.2 Flowchart of the system

In this project first we have to collect the data through the sensor which is first loaded into the RDBMS. After this it uses the truck data from RDBMS to HDFS or hive. The data which is collected from the each sensor helps to estimate the operations performed in truck. The data which is collected from the RDMS, data analysis continue in hive by writing pig script it will estimate the drivers mileage and truck mileage and drivers Risk Factor.

The number of estimations gives the final refine sensor data which will be stored in excel sheet .It provides the information about each truck with log location and drivers behavior. The overall scenario helps to find the real time estimations as per as the

GPS tracking device gives the current location and time to reach the destination. It helps to reduce the driver's risk factors and accuracy of a vehicle.

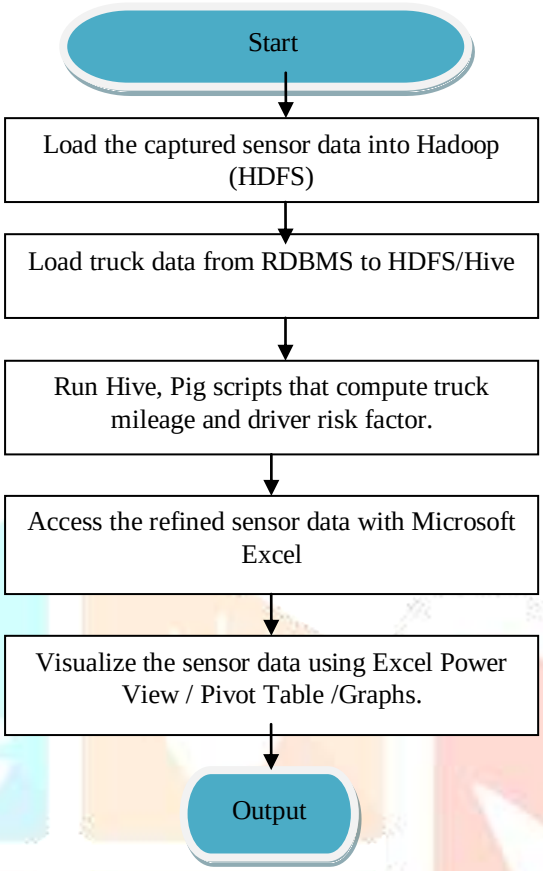

Fig 2.flow of system

The data visualization gives the overall tape record detail about the driver and vehicle (truck) in a pivot table or excel power view or graph representation.

V. METHODOLOGY

In this project we are using machine learning approaches for estimation of driving patterns and drivers behavior. For the analysis of the sensor data we are using single node clustering technique.

5.1 A. Single Node Cluster

In this project several operations of a vehicle can be evaluated by using the single node hadoop cluster.

In this technique Hadoop can run on one physical machine using the local file system, not the Hadoop file system (HDFS). In this we are using single machine as a master and slave. The use of this cluster can be provide the refine data from the sensor. This is also known as pseudo-distributed mode, because all the services are not distributed across multiple machines (nodes).

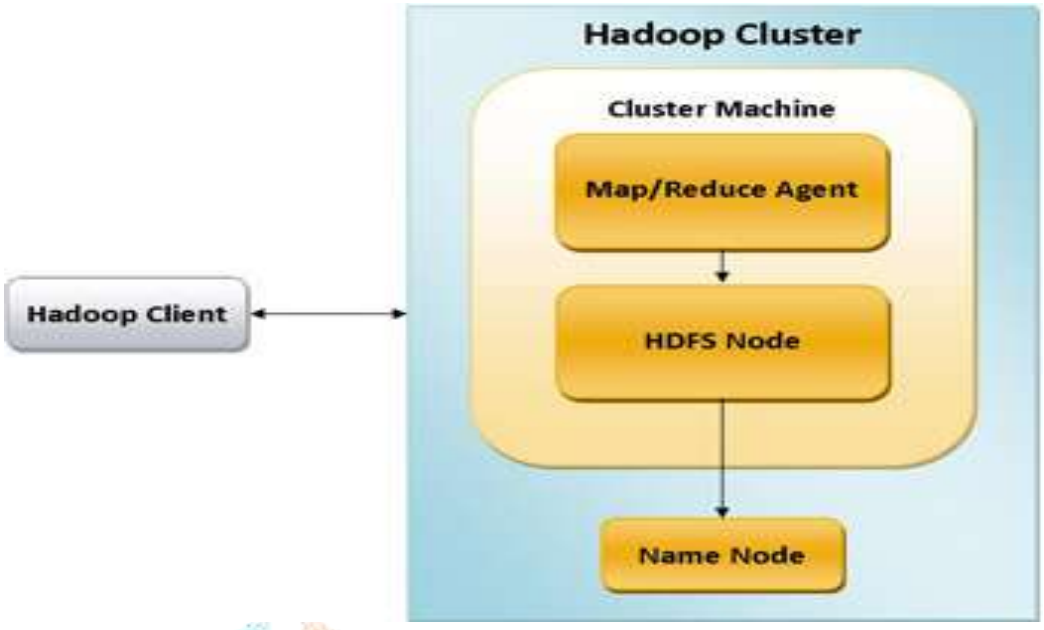

Fig 4 : Single node Cluster

Advantages:

1. Lower cost
2. Reduces complexity
3. Creates a base for growth

We can also use multi-node cluster technique if we need to extract abundant amount of data. The project consist of three sections which includes data collection, data analysis and data visualization.

1. Data Collection

Four data sources were used for this paper :

a. Driving patterns collected from GPS data loggers installed.
b. Drivers Characteristics
c. Road characteristics collected from the map matching of the GPS data with the road network.
d. Weather information, location and time to reach.

2. Data Analysis

First we have to calculate the ECR of vehicle which represents the performance of BEV's. ECR is calculated as the ratio between the power consumed and distance traveled for different models in different environment. The lower is the ECR; the better is the energy efficiency. And also used to compute the driving range of a vehicle. The overall performance is not only depend upon the driving range also depends on the performance. We also estimate the driver mileage, truck mileage and risk factor according to the behavior and performance of the truck.

3. Data Visualization

The representation of chart, graph and pivot table can be form as a record to store all the finest detail in Excel sheet. GPS maps the current location and distance covered by vehicle. The overall performance of truck can be log into the record with driver's performance.

VI. RESULTS AND DISCUSSION

In this paper we are performing the operations on a truck data which can be extract by using sensor data. It provides run time information and operations performed in truck. It is helpful for enhancing the performance of a vehicle and driver skills as well as it is beneficial for business point of view. It reduces the cost, provides the faster result at the run time. In this project we used the single node cluster in future if we need faster execution at a time able to use muti-node clustering techniques.

VII. ACKNOWLEDGMENT

I would like to thank my author Dr. Shrikant D. Zhade of RTMN University for the support and encouragement for this research and providing me his valuable guidance and assistance.

REFERENCES

[1] Chung-Hong Lee, Chih-Hung Wu," An incremental learning technique for detecting driving behaviors using collected EV big data ",ASE BD & SI'15 Proceedings of the ASE Big Data & Social Informatics, 2015, Article No. 10,pp. October 07 - 09, 2015

[2] Chung-Hong Lee, Chih-Hung Wu," A Novel Big Data Modeling Method for Improving Driving Range Estimation of EVs", IEEE acess, VOl.3. pp. October 2015

[3] Chi-Kin Chau, Khaled Elbassioni, Chien-Ming Tseng,"Fuel Minimization of Plug-in Hybrid Electric Vehicles by Optimizing Drive Mode Selection", ACM, Article No. 13, ISBN ,pp.21-24 ,june2016

[4] Gebeyehu M. Fetenea, Carlo G. Pratob, Sigal Kaplanc , Stefan L.Mabitd, Anders F. Jensen," Harnessing Big-Data for Estimating the Energy Consumption and Driving 2 Range of Electric Vehicles", pp10-14,2016

[5] Benjamin Pichler, Andreas Riener," Evaluation of Historical Electric Vehicle (EV) Driving Data to Suggest Improvements in Driving Efficiency", pp.130-135, 2015

[6] Naehyuck Chang, Donkyu Baek and Jeongmin Hong," Power Consumption Characterization, Modeling and Estimation of Electric Vehicles", IEEE/ACM publication, INSPEC No**. :** 14841450, January 2015

[7] Daiki Kawanuma, Yuta Kashiwabara, Toshiaki Uemura," Data Analysis Framework for Visualizing Correlation of Energy Consumption and Transit Time in Road Sections using the ECOLOG database",ACM,No.1,pp. 207-213, 2016

[8] Toshiaki Uemura, Yuta Kashiwabara Daiki Kawanuma, Takashi Tomii,"Accuracy Evaluation by GPS Data Correction for the EV Energy Consumption Database",pp.213-218 , Dec 2016

[9] Jie Liu , Bodhi Priyantha, Ted Hart, Heitor S. Ramos,Ted Hart, Qiang Wang, Antonio A.F.Loureiro," Energy efficiency GPS Sensing with cloud offloading", 2012

[10] Vladimir Vukadinovic, , Stefan Mangold,"Performance of Collaborative GPS localization in Pedstrain ad hoc network ", pp. 45-52,2012