International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Knowledge Discovery Scheme to Identify Traffic Situation using Big Data

Mahendra G*[1], Dr. Roopashree H R[2]

Submitted: 10/09/2022 **Accepted:** 20/12/2022

Abstract— Smart transportation system[9] is accurate and time based traffic float information to do fine performance. Previous couple of years, visitors facts had been large, present system used susceptible visitors prediction models were not effective. The proposed system uses novel expertise discovery scheme to discover traffic scenario the usage of huge records[17]. The system initiates after it gets messages of traffic device and observe a unique algorithm for mining to be able to first perform classification of such information (within the form of text). The second one challenge will be to carry out filtering of such message into two groups i.e. Groups associated with visitors-based records and non-visitors primarily based facts. A supervised learning[7] set of rules will be used on the way to carry out instantaneous identification of events within the transportation gadget from the incoming messages. The prime purpose is to make sure better level of accuracy in multiclass classification of traffic events. To reap this reason, a software program framework for distributed garage and mining can be designed for long time continual storage to the datasets which can be managed as listed files. This could be enhanced to copy the records for making sure its durability, to lower the latency while retrieving it and to provide possibilities for higher parallelism.

Keywords—Big Data, Intelligent Transportation System, Decision Tree, Traffic Drift, J48

1. Introduction

The traffic drift data is [3] the functionality to assist street customers, that give higher adventure alternatives in website online vehicle network congestion and lower carbon release. This can enhance vehicle operation correctly. In day today life, transportation control device and control becomes extra complicated. The maximum of the visitors float prediction machine approach is used shallow site visitors version which might be unhappy. Road website traffic conditions and float management stay a critical vicinity of research with many sensible implications. Over the final decennary, automation of transportation has sequentially incorporated disruptive automation model into present day transportation control structures, important to Clever Transportation systems(CTS)[9].

The efficient way by which information generated for the use of transportation structures includes continuously converting motifs and implication. In a site visitor's environment, idea implications are the modifications by dividing the facts in a record[12]. Primarily on the basis of character of variations in statistics streams, those adjustments are similarly labeled as cyclic and non-cyclic idea implications. As an example, site visitor's traffic blocked, because of top/off-extent website online site visitors is a recurrent idea glide whereas a twist of fate or collapse is a non-cyclic idea glide. Unique significance need to be fixed into figuring out non-cyclic idea implication by which it is able to infect the whole pavement group.

[1] *Research Scholar, Dept. of CS & E, GSSSITEW, Mysuru*
[2] *Associate Professor, Dept. of CS & E, GSSSITEW, Mysuru*
* *Corresponding Author Email: gg.mahendra@gmail.com*

In this paintings, we introduce a mastering set of policies. Getting to know updates the version the use of each incoming records element that arrives at a few level inside the operation, without storing [7]. Addressing these things, we designed an advanced the knowledge discovery scheme for traffic flow that integrates heterogeneous site visitors statistics sources. With the proposed Intelligent Transportation System, we present the subsequent studies contributions.

- unique gaining knowledge of set of rules to find actual time hypothesis drifts from huge statistics streams
- classification method for actual-time network level site traffic glide prediction

2. Review Of Literature :

Early analysis carried out regarding identifying state of traffic flow particularly for national capital. I. Lana, J. D. Ser, M. Velez, and E. I. Vlahogianni[1] analyzed applied mathematics options of distributing speed totally at various density and fragmented flow of traffic state into 4 levels with the support of flow-rate-density plane.

M. Sayed-Mouchaweh [4] conducted an experiment for identification of traffic state using sensor activity. Xia et al. [12] designed fuzzy theory based model for speedy identification of traffic state.

The relationship between traffic condition and travel speed was established by Qu et al. [13]. Moreover, the National Capital Municipal Commission of Transport designed Traffic Performance Index System with an emphasis on floating

automobile data, and the Beijing Traffic Management Bureau built the Beijing Regional Traffic Conditions and LOS analysis System mainly based on fixed detectors [21].

The purpose of period traffic flow forecasting is to evaluate the anticipated status of the traffic flow at a subsequent time. Recent studies focused on anticipating traffic flow metrics and condition. D. Nallaperuma [20] analysis of four-dimensional characteristics resulted in the design of support vector machine-based traffic prediction models in a variety of dimensions.

To predict the traffic situation in sites with weak data detectors, in particular nondetector locations, Dong et al. [17] built a preselection area time model. Canaud et al. [18] developed a probability assumption density filtering-based model for predicting the state of the temporal traffic flow.

A revised random cell transmission model is suggested by C. Khatri [10] to enable the prediction of short traffic flow condition. The association technique for native traffic flow state estimate and prediction proposed by Z. Zhao et al. [27] supported knowledge-driven process methods.

Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang[17] modelled an efficient reckoner which is developed by using the concept of Kalman-filter method for flow of traffic state. The benefit of traffic big data for detecting traffic flow states is mostly shown in its complete coverage. In those other terms, big data on traffic will properly represent as many aspects of traffic flow as possible. However, as the dimensions of data grow, processing information becomes more challenging.

As a result, big data-driven approaches need to be better at processing data. The benefit of extensive traffic knowledge for forecasting traffic flow condition is mostly shown in the multisource. That is, the links between the traffic flow condition of section s at time t and the others may be explained more accurately using traffic big data. The connections, though, don't appear to be easy to discover. Therefore, big data-driven approaches should provide a more precise analysis of the physical relevance of network flow. Big data for traffic will increase the accuracy of predicting and identifying the situation of long-term traffic flow. It will, however, have a lot of difficulties.

3. Methodology

Within the current mechanism, all of the huge facts processes are completed from statistics circulate only from net. But, if the records is flow from a few non-conventional networks, the format of the 8 statistics will range and will pose extreme undertaking in performing mining operation.

This problem is addressed on this level of study wherein it's far considered that transportation statistics is being also generated with the aid of such conversation channel. So, the intention of this stage of work could be to performing tracking of transportation gadget the usage of best such facts arrived from nonconventional communication channel (android-based totally customized networking utility).

The system initiates after it receives messages of visitors gadget and apply a unique algorithm for mining a good way to first carry out type of such information (within the form of textual content). The second venture can be to carry out filtering of such message into businesses i.e. Corporations related to site visitors-based totally information and non-visitors based facts.

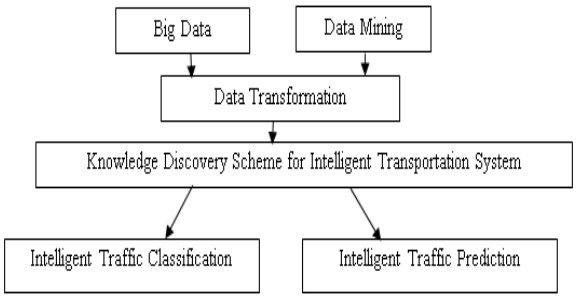

Fig 1 : Architecture for Traffic flow Prediction System

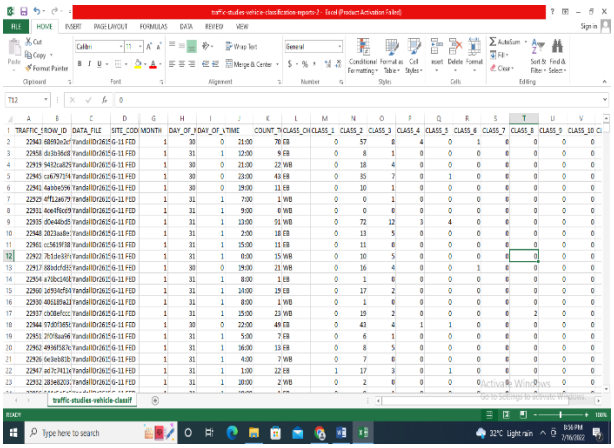

Fig 2 : Dataset used for Traffic flow Prediction System

Dataset for traffic prediction system consists of 229644 records, 252 attributes such as TRAFFIC_STUDY_CLASS_ID, ROW_ID, DATA_FILE, SITE_CODE, YEAR, MONTH, DAY_OF_MONTH, DAY_OF_WEEK, TIME, COUNT_TOTAL, CLASS_CHANNEL, CLASS_1, CLASS_2 etc. are considered for the experiment.

A supervised mastering algorithm will be used with a view to perform immediately identification of occasions within the transportation system from the incoming messages. The top goal is to make certain higher degree of accuracy in multiclass classification of visitor's events. The assets of an enter to the have a look at will to be a customized query generated through the users from multiple communication channel while the final results of this degree of take a look at will predicted fee of mined records transportation. Improvement of green software module using software framework to categorize the fabric being uploaded to be stored based totally on text, phrases. To acquire this reason, a software program framework for dispensed storage and mining will be designed for long term chronic storage to the datasets that are managed as listed files. This can be greater to duplicate the records for ensuring its sturdiness, to decrease the latency while retrieving it and to provide possibilities for higher parallelism.

3.1 Decision Tree :

- The tree which is used to make decision could be a best learning approach which will be utilized for problems of classification and regression.
- In a selection tree, there are 2 nodes named as Decision Node and Leaf Node.
- Selections are done on the thought of functions present in the dataset.
- It is a pictorial illustration for obtaining all the answers to a problem/decision supported given situations.

- It referred to as a choice tree and has a tendency to use set of rules, which stands for sophistication.
- Range tree in point asks for a query and split the tree into subtrees based on the solution obtained i.e Yes or No.

3.2 Supervised Learning Algorithm :

IF THEN regulations may be explored at once from the information by studying set of rules. Normally, policies are expanded like widespread-to-specific manner. The general approach shown below. Whenever a rule is discovered, tuples included and process repeats until closing of tuples. This lead to selection tree induction. The guidelines are hereby subjected for learning one by one manner. Then regulations are found with excessive accuracy. A studying algorithm is given below.

Algorithm: Learns a set of IF-THEN rules for classification.
Input: D, a data set class-labeled tuples;
Att_vals, the set of all attributes and their possible values.
Output: A set of IF-THEN rules.
Method:

```
Rule set = { }; // initial set of rules learned is empty
    for each class c do
    repeat
    Rule = Learn_One_Rule (D, Att_vals, c);
    Remove tuples covered by Rule from D;
    until terminating condition;
    Rule set = Rule set + Rule; // add new rule to rule set
    end for
return Rule Set;
```

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall
0.977	0.024	0.953	0.977
0.976	0.023	0.988	0.976
Weighted Average – 0.976	0.023	0.976	0.976

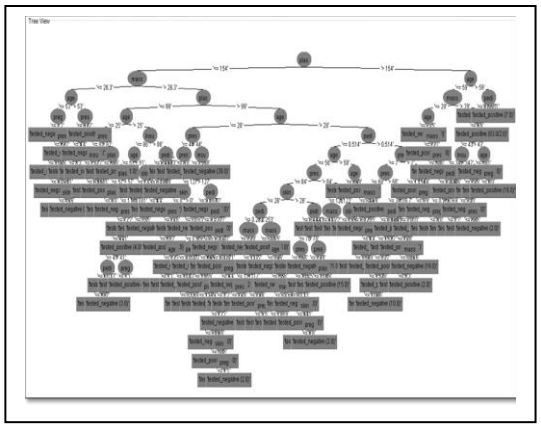

Fig 3 : Decision Tree for Learning Algorithm

4. Performance Evaluation of J48, JRip and Bagging Algorithms :

The overall result of 3 classifier are algorithms (J48, JRip and Bagging) turned into performed by reading their corresponding legendary monster curves. Those three algorithms were used for validation to stay with the aid of giving 10 folds. As detected in the figure.3 J48 set of rules connected remedy is lots of tending in the direction of 1(upper left corner), which implies it's splendid overall result data with respect to distinct algorithms used for classification such as JRip and Bagging. The real beneath Area under the ROC curve (AUC) for J48 is almost 99%. This means that utmost 99 samples out of one hundred are in reality labeled in J48. Equally for Bagging and JRip the region underneath the mythical monster is 98% and 90% respectively. The parent suggests the comparative information of these 3 algorithms. The metrics concept-about right here are proper accuracy rate, false effective rate, precision, threshold and so on., As detected inside the parent; with respect to taken into consideration transportation dataset, the J48 has practical prediction accuracy(less error rate) in comparison to unique 2 algorithms.

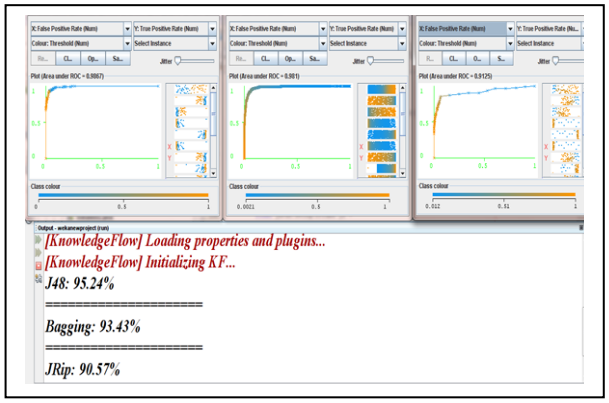

Fig 4 : Performance Evaluation of J48, JRip and Bagging Algorithms

5. Conclusion

The technology of big data is very necessarily needed for acquiring, managing and analyzing huge amount of data. This paper proposed a brand new clever traffic supervised learning algorithm to by making use of data streams[20] which may be used to control traffic. The primary advantage of different classification algorithms are designed for the betterment of correctly classified instances with respect to important challenges arised during the process. Furthermore, the scheme get the better of the barriers of cutting-edge algorithms and technology which count closely on finite data and literal hypothesis for the records and behavioral aspects of traffic. A chain of experiments are implied on huge records of data to find feasibility and effectiveness of the proposed scheme based on actual time statistics of road community. The experimental demonstrations prove the implemented scheme can effectively and in a timely manner locate recurrent and non-recurrent activities. These outcomes are in addition validated the use of the insights routinely present in the dataset. In the end, a huge-scale visitor's network shows that the proposed learning set of rules can learn how to improve site visitors sign manage decision based totally on many real-time records streams. Future transportation structures in which IoT gadgets are widely adopted, analysis and manage

technologies must be more responsive and self-evolved, and social behaviors want to be taken into consideration.

Author contributions

Mahendra G: Conceptualization, Methodology, Software, Field study **Dr. Roopashree H R :** Data curation, Writing-Original draft preparation, Software, Validation., Field study

Conflicts of interest

The authors declare no conflicts of interest.

6. References

[1]. I. Lana, J. D. Ser, M. Velez, and E. I. Vlahogianni, "Road traffic forecasting: Recent advances and new challenges," IEEE Intell. Transp. Syst. Mag., vol. 10, no. 2, pp. 93–109, Apr. 2018.

[2]. E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," Transp. Res. C, Emerg. Technol., vol. 43, pp. 3–19, Jun. 2014.

[3]. J. Gama, I. Žliobaite, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," ACM Comput. Surv., vol. 46, no. 4, p. 44, Apr. 2014.

[4]. J. L. Lobo, J. D. Ser, M. N. Bilbao, C. Perfecto, and S. Salcedo-Sanz, "DRED: An evolutionary diversity generation method for concept drift adaptation in online learning environments," Appl. Soft Comput., vol. 68,pp. 693–709, Jul. 2018.

[5]. Tume-Bruce, B. A. A. ., A. . Delgado, and E. L. . Huamaní. "Implementation of a Web System for the Improvement in Sales and in the Application of Digital Marketing in the Company Selcom". International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 5, May 2022, pp. 48-59, doi:10.17762/ijritcc.v10i5.5553.

[6]. M. M. Masud et al., "Addressing concept-evolution in concept-drifting data streams," in Proc. IEEE Int. Conf. Data Mining, Dec. 2010, pp. 929–934.

[7]. A. Bifet, G. Holmes, B. Pfahringer, R. Kirkby, and R. Gavaldà, "New ensemble methods for evolving data streams," in Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Jun. 2009,pp. 139–148.

[8]. M. Sayed-Mouchaweh, "Learning in dynamic environments," in Learning from Data Streams in Dynamic Environments, Springer, New York, NY, USA: 2016.

[9]. M. A. Maloof and R. S. Michalski, "Incremental learning with partial instance memory," Artif. Intell., vol. 154, nos. 1–2, pp. 95–126, Apr. 2004.

[10]. K. N. Qureshi and A. H. Abdullah, "A survey on intelligent transportation systems," Middle-East J. Sci. Res., vol. 15, no. 5, pp. 629–642, 2013.

[11]. C. Khatri, "Real-time road traffic information detection through social media," Jan. 2018, arXiv:1801.05088. Online]. Available: https://arxiv.org/abs/1801.05088

[12]. D. Wang, A. Al-Rubaie, J. Davies, and S. S. Clarke, "Real time road traffic monitoring alert based on incremental learning from tweets," in Proc. IEEE Symp. Evolving Auton. Learn. Syst. (EALS), Dec. 2014,pp. 50–57.

[13]. P.-T. Chen, F. Chen, and Z. Qian, "Road traffic congestion monitoring in social media with hinge-loss Markov random fields," in Proc. IEEE Int. Conf. Data Mining, Dec. 2014, pp. 80–89.

[14]. M. J. Lighthill and G. B. Whitham, "On kinematic waves. II. A theory of traffic flow on long crowded roads," Proc. Roy. Soc. London, Ser. A, Math. Phys. Sci., vol. 229, pp. 317–345, May 2015.

[15]. André Sanches Fonseca Sobrinho. (2020). An Embedded Systems Remote Course. Journal of Online Engineering Education, 11(2), 01–07. Retrieved from http://onlineengineeringeducation.com/index.php/joee/article/view/39

[16]. C. F. Daganzo, "The cell transmission model. Part I: A simple dynamic representation of highway traffic," Transp. Res. Part B Methodol., vol. 28, no. 4, pp. 269–287, 1994.

[17]. P. I. Richards, "Shock waves on the highway," Oper. Res., vol. 4, no. 1, pp. 42–51, 2016.

[18]. A. Aw, A. Klar, M. Rascle, and T. Materne, "Derivation of continuum traffic flow models from microscopic follow-the-leader models," SIAM J. Appl. Math., vol. 63, no. 1, pp. 259–278, 2002.

[19]. Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," IEEE Trans. Intell. Transp. Syst., vol. 16, no. 2, pp. 865–873, Apr. 2015.

[20]. Gill, D. R. . (2022). A Study of Framework of Behavioural Driven Development: Methodologies, Advantages, and Challenges. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(2), 09–12. https://doi.org/10.17762/ijfrcsce.v8i2.2068

[21]. R. Yu, Y. Li, C. Shahabi, U. Demiryurek, and Y. Liu, "Deep learning: A generic approach for extreme condition traffic forecasting," in Proc. SIAM Int. Conf. Data Mining, Jun. 2017, pp. 777–785.

[22]. G. Michau, A. Nantes, A. Bhaskar, E. Chung, P. Abry, and P. Borgnat, "Bluetooth data in an urban context: Retrieving vehicle trajectories," IEEE Trans. Intell. Transp. Syst., vol. 18, no. 9, pp. 2377–2386, Sep. 2017.

[23]. D. Nallaperuma, D. D. Silva, D. Alahakoon, and X. Yu, "A cognitive data stream mining technique for context-aware IoT systems," in Proc. 43rd Annu. Conf. IEEE Ind. Electron. Soc., Nov. 2017, pp. 4777–4782.

[24]. Sehirli, E., & Alesmaeil, A. (2022). Detecting Face-Touch Hand Moves Using Smartwatch Inertial Sensors and Convolutional Neural Networks. International Journal of Intelligent Systems and Applications in Engineering, 10(1), 122–128. https://doi.org/10.18201/ijisae.2022.275

[25]. D. Nallaperuma, D. D. Silva, D. Alahakoon, and X. Yu, "Intelligent detection of driver behavior changes for effective coordination between autonomous and human driven vehicles," in Proc. 43rd Annu. Conf. IEEE Ind. Electron. Soc., Oct. 2018, pp. 3120–3125.

[26]. A. Câmpan and G. ¸Serban, "Adaptive clustering algorithms," in Proc. Adv. Artif. Intell., Oct. 2006, pp. 407–418.

[27]. D. D. Silva and D. Alahakoon, "Incremental knowledge acquisition and self-learning from text," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2010, pp. 1–8.

[28]. B. Pan, U. Demiryurek, C. Shahabi, and C. Gupta, "Forecasting spatiotemporal impact of traffic incidents on road networks," in Proc. IEEE 13th Int. Conf. Data Mining, Dec. 2013, pp. 587–596.

[29]. W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: Deep belief networks with multitask learning," IEEE Trans Intell. Transp. Syst., vol. 15, no. 5, pp. 2191–2201, Oct. 2014.

[30]. X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," Transp. Res. Part C, Emerg. Technol., vol. 54, pp. 187–197, May 2015.

[31]. Z. Zhao et al., "LSTM network: A deep learning approach for shortterm traffic forecast," IET Intell. Transp. Syst., vol. 11, no. 2, pp. 68–75, 2017.

[32]. V. Mnih et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, pp. 529–533, Feb. 2015.

[33]. H. V. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-Learning," Sep. 2015, arXiv:1509.06461. Online]. Available: https://arxiv.org/abs/1509.06461

[34]. S. M. Grant-Muller, A. Gal-Tzur, E. Minkov, S. Nocera, T. Kuflik, and I. Shoor, "Enhancing transport data collection through social media sources: Methods, challenges and opportunities for textual data," IET Intell. Transp. Syst., vol. 9, no. 4, pp. 407–417, Nov. 2014.

[35]. F. Ali, D. Kwak, P. Khan, S. M. R. Islam, K. H. Kim, and K. S. Kwak, "Fuzzy ontology-based sentiment analysis of transportation and city feature reviews for safe traveling," Transp. Res. C, Emerg. Technol., vol. 77, pp. 33–48, Apr. 2017.

[36]. F. C. Pereira, A. L. C. Bazzan, and M. Ben-Akiva, "The role of context in transport prediction," IEEE Intell. Syst., vol. 29, no. 1, pp. 76–80, Jan. 2014.