IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Data Analytics Approach for Train Time Table Performance Measure Using Automatic Train Supervision Data

K.RAJALAKSHMI [1] and MRS.R.SUBBULAKSHMI [2]

[1]Final Year PG Student, [2]Assistant Professor.

Department Of Computer Science,

G.Venkataswamy Naidu College,

Kovilpatti,

Tamil Nadu, India.

ABSTRACT:

Passenger train delay significantly influences riders to choose rail transport as their mode choice. This paper proposes real-time passenger train delay prediction (PTDP) models using machine learning techniques. In this article, the impact on PTPD models using real-time with real-time-based data-frame structure (RT-DFS) and history-based data-frame structure (RWH-DFS) is investigated. The results show that PTDP models using MLP with RWH-DFS outperformed all other models. The influence of external variables such as historical delay profiles (HDPD), ridership and population, day of the week, geography and weather information on real-time PTPD models is further analyzed and discussed. This system is very important for improving airport traffic efficiency to improve accuracy in predicting train arrival delay time. In our process, we need to take the input as a time series dataset. After that, machine learning algorithms like logistic regression and random forest should be implemented. Experimental results show that each algorithm has accuracy and error values. The model has good predictive accuracy and can track the trends of many delay indicators well.

KEYWORDS : RANDOM FOREST, LOGISTICS REGRESSION.

INTRODUCTION:

Transport systems are critical pieces of infrastructure and they have substantially increased in size in many countries worldwide. This includes rail transport systems that have evolved significantly, including to provide long-distance travel services. In Sweden, the total distance travelled by trains increased by 8% between 2013 and 2016. In the United States (U.S.), ridership on state supported routes increased by more than 10%, making it the fastest growing segment of Amtrak's services. On long-distance routes, both ridership and revenue increased in fiscal year 2018 by 6.2% and 7.3%, respectively. To sustain its competitiveness and attract more riders, ensuring a high on-time performance is critical. Poor on-time performance can impact passenger trust and their satisfaction, and it may result in a shift to other modes of transport, especially private vehicles and air transport.

Service disruption is a root cause of lower rail punctuality and customer satisfaction. Major Service disruptions result from various conditions or factors such as accidents, problems in train operation, malfunctioning or damaged equipment, routine maintenance, construction, passenger boarding or alighting, and even extreme weather conditions.

Rail service disruptions directly affect scheduled timetable and inevitably cause train delay. Significant train delay can eventually lead to service loss or even cancellation.

In addition, train delay can also negatively affect connecting trains and passengers' journeys or activities. Thus, delay estimations or predictions can help train operators develop better plans to manage, reschedule, or adjust the timetable of the current and consecutive trains more effectively, as well as to inform passengers in advance so they themselves can adjust their travel plans in time. Using or referring to historical average delay is insufficient to estimate future train delay as passenger train can potentially be affected by different factors such as ridership, accumulated delay from prior trains, or weather conditions.

Passenger train delay prediction (PTDP) has been made and modeled in several ways using a variety of approaches and techniques. A fuzzy Petri net (FPN) model to estimate train delay of the Belgrade rail service (the train primary delays were simulated by a fuzzy Petri net module. It used dispatch simulation software to simulate traffic volume and estimate train delays on single and double track rail lines. Wang and Work indicated that although simulation methods can be used to estimate complex train operations, they require tremendous effort to configure parameters such as dispatching rules as well as calibrating the models for complex train systems.

SYSTEM PROPOSAL

EXISTING SYSTEM:

Currently, the suspension of passenger trains will a ffect passengers' choice of rail as a mode of travel. This article presents a real-
time train delay model (PTDP) using the following machine learning techniques: Random Forest (RF) , Gradient Boosting Machine (GBM), and Multilay er Perceptron (MLP). In this article, the effect of us ing time-based data frame structure (RT-
DFS) and historical data frame duration (RWH-

DFS) of the PTPD model is evaluated. The results show that the PTDP model using MLP and RWH-DFS outperforms the other models. The effect of e xternal variables such as site historical latency (HD PD), cycling, population, day of the week, geograp hy and weather on the PTPD time model.

DISADVANTAGES:

- It performed low accuracy.
- It does not efficient for large number of data's.
- Theoretical limits.

PROPOSED SYSTEM:

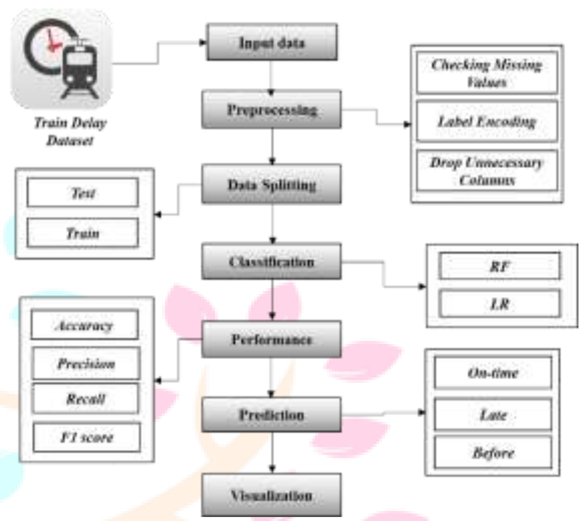
Train delay is a major problem in the aviation sector. In proposed system, we have to use the train delay dataset. After that, we have to implement the pre-processing step. In this step, we have to implement the handling missing values for avoid wrong prediction and label encoder for machine readable. After that, we have to implement the different classification algorithms such as Random forest and logistic Regression for analysing or forecasting the train delay. Finally, the experimental results shows that the accuracy, precision, recall and f1 score. Then, we can predict the train is arrived (on-time or before or late) effectively.

ADVANTAGES:

- High accuracy is performed for both supervised.
- It also display the visual graphs. (i.e) comparison graph.
- It is efficient for large number of dataset.
- The process is implemented with removing unwanted data.

SYSTEM DIAGRAMS

SYSTEM ARCHITECTURE:



3SEQUENCE DIAGRAM:

Sequence diagrams document the interactions between classes to achieve a result, such as a use case. Because UML is designed for object-oriented programming, these communications between classes are known as messages. The Sequence diagram lists objects horizontally, and time vertically, and models these messages over time.

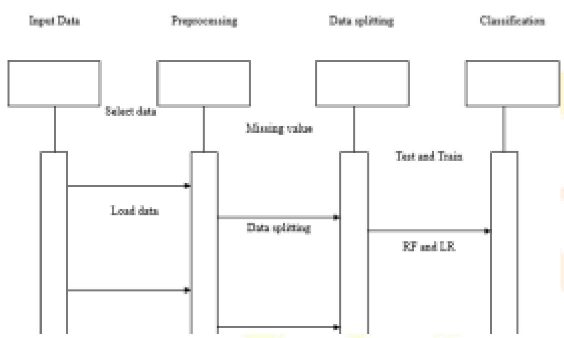
Graphical Notation: In a Sequence diagram, classes and actors are listed as columns, with vertical lifelines indicating the lifetime of the object over time.

Object: Objects are instances of classes, and are arranged horizontally. The pictorial representation for an Object is a class (a rectangle) with the name prefixed by the object.

Lifeline: The Lifeline identifies the existence of the object over time. The notation 2for a Lifeline is a vertical dotted line extending from an object.

Activation: Activations, modelled as rectangular boxes on the lifeline, indicate when the object is performing an action.

Message: Messages, modelled as horizontal arrows between Activations, indicate the c.



MODULES:

- Input Data
- Preprocessing
- Data splitting
- Classification
- Performance metrics
- Prediction

USER ACCEPTANCE TESTING:

User acceptance of the system is the key factor for the success of the system. The system under consideration is tested for user acceptance by constantly keeping in touch with prospective system at the time of developing changes whenever required.

OUTPUT TESTING:

After performing the validation testing, the next step is output asking the user about the format required testing of the proposed system, since no system could be useful if it does not produce the required output in the specific format. The output displayed or generated by the system under consideration. Here the output format is considered in two ways. One is screen and the other is printed format. The output format on the screen is found to be correct as the format was designed in the system phase according to the user needs. For the hard copy also output comes out as the specified requirements by the user. Hence the output testing does not result in any connection in the system.

CONCLUSION

We conclude that, the input dataset was taken from dataset repository. We are developed the different classification algorithms such as logistic regression and random forest.Finally, the result shows that some performance metrics such as Accuracy, precision, recall and f1 score. Then, we are forecast or analyzed the train delayand visualization.

FUTURE ENHANCEMENT

In the future, we should like to hybrid the two different machine learning. In future, it is possible to provide extensions or modifications to the proposed clustering and classification algorithms to achieve further increased performance. Apart from the experimented combination of data mining techniques, further combinations and other clustering algorithms can be used to improve the detection accuracy.

REFERENCES

[1] S. Derrible, Urban Engineering for Sustainability. Cambridge, MA, USA: MIT Press, 2019.

[2] R. Nilsson and K. Henning. Predictions of Train Delays Using Machine Learning. 2018. [Online]. Available: http://urn. kb.se/resolve?urn=urn:nbn:se:kth:diva-230224 (Accessed: Jul. 27, 2019).

[3] Amtrak Five Year Service Line Plans FY20-24, Amtrak, Washington, DC, USA, 2019.

[4] N. O. E. Olsson and H. Haugland, "Influencing factors on train punctuality—Results from some Norwegian studies," Transp. Policy, vol. 11, no. 4, pp. 387–397, Oct. 2004, doi: 10.1016/j.tranpol.2004.07.001.

[5] W. Peetawan and K. Suthiwartnarueput, "Identifying factors affecting the success of rail infrastructure development projects contributing to a logistics platform: A Thailand case study," Kasetsart J. Social Sci., vol. 39, no. 2, pp. 320–327, 2018, doi: 10.1016/j.kjss.2018.05.002.

[6] P. Wang and Q. Zhang, "Train delay analysis and prediction based on big data fusion," Transp. Safety Environ., vol. 1, no. 1, pp. 79–88, Jul. 2019, doi: 10.1093/tse/tdy001.

[7] L. Oneto et al., "Train delay prediction systems: A big data analytics perspective," Big Data Res., vol. 11, pp. 54–64, Mar. 2018, doi: 10.1016/j.bdr.2017.05.002.

[8] Z. Alwadood, A. Shuib, and N. A. Hamid, "Rail passenger service delays: An overview," in Proc. IEEE Bus. Eng. Ind. Appl. Colloquium (BEIAC), Apr. 2012, pp. 449–454, doi: 10.1109/BEIAC.2012.6226102.

[9] S. Milinkovic, M. Markovi ´ c, S. Veskovi ´ c, M. Ivi ´ c, and ´ N. Pavlovic, "A fuzzy Petri net model to estimate train ´ delays," Simulat. Model. Pract. Theory, vol. 33, pp. 144–157, Apr. 2013.

[10] B. W. Schlake, C. P. L. Barkan, and J. R. Edwards, "Train delay and economic impact of in-service failures of railroad rolling stock," Transport. Res. Rec., vol. 2261, no. 1, pp. 124–133, Jan. 2011, doi: 10.3141/2261-14.

[11] R. Wang and D. B. Work, "Data driven approaches for passenger train delay estimation," in Proc. IEEE 18th Int. Conf. Intell. Transport. Syst., Sep. 2015, pp. 535–540, doi: 10.1109/ITSC.2015.94.

[12] L. Oneto et al., "Advanced analytics for train delay prediction systems by including exogenous weather data," in Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA), Oct. 2016, pp. 458–467, doi: 10.1109/DSAA.2016.57.

[13] A. Gal, A. Mandelbaum, F. Schnitzler, A. Senderovich, and M. Weidlich, “Traveling time prediction in scheduled transportation with journey segments,” Inf. Syst., vol. 64, pp. 266–280, Mar. 2017, doi: 10.1016/j.is.2015.12.001.

[14] A. Estes, M. O. Ball, and D. Lovell, “Predicting performance of ground delay programs,” presented at the 12th USA/Europe air traffic management R&D seminar, Seattle, WA, USA, 2017.

[15] R. Gaurav and B. Srivastava, “Estimating train delays in a large rail network using a zero shot Markov model,” Jun. 2018, arXiv:1806.02825.

[16] R. Nair et al., “An ensemble prediction model for train delays,” Transport. Res. C, Emerg. Technol., vol. 104, pp. 196–209, Jul. 2019, doi: 10.1016/j.trc.2019.04.026.

[17] P. Taleongpong, S. Hu, Z. Jiang, C. Wu, S. Popo-Ola, and K. Han, “Machine learning techniques to predict reactionary delays and other associated key performance indicators on British railway network,” J. Intell. Transport. Syst., vol. 26, no. 3, pp. 311–329, Dec. 2020, doi: 10.1080/15472450.2020.1858822.

[18] C. M. Zappi, Amtrak Host Railroad Report Card 2019, Amtrak, Washington, DC, USA, Jan. 2020.

[19] “Amtrak Host Railroad Report Card 2020.” Amtrak. Apr. 2021. [Online]. Available: https://www.amtrak.com/content/dam/projects/ dotcom/english/public/documents/corporate/HostRailroadReports/ Amtrak-2020-Host-Railroad-Report-Card-FAQs.pdf (Accessed: Aug. 17, 2021).

[20] P. Lapamonpinyo, S. Derrible, and F. Corman. “A Python Tool and Database of Amtrak Departure and Arrival Times with Weather Information.” Oct. 19, 2021. [Online]. Available: https://engrxiv.org/index.php/engrxiv/preprint/view/1960