ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

Train Delay Prediction Systems Using Big Data Analytics

Jasti Satyakrishna [1], Raj Kumar Sagar [2]

M.Tech (CS&E), Department of Computer Science and Engineering, AMITY University, Noida, India[1]

Assistant Professor, Department of Computer Science and Engineering, AMITY University, Noida, India[2]

ABSTRACT: A noteworthy part of the present day economy, Indian Railways represents 7% of the GDP in may create nations. In India, most of the people are travelling by train in one city to another city. Indian railway is one of the world's largest railway networks. Revenue of Indian railway is 165,068 crores. Train Delay Prediction System has used to find the delay of the train time. Train Delay is depending on the train movements and weather conditions like Temperature, Humidity, Wind, Rain and Solar Radiation. The Train follows the fixed schedule called Nominable Timetable. Nominable Timetable and an actual timetable has used to find the delay of the train duration. [1] In this paper using Big Data Analytics and Techniques are applying in the Train Delay Prediction System. The methods Shallow and Deep Extreme Learning Machines have used in this paper.

KEYWORDS: Big Data Analytics, Railway Networks, Train Delay Prediction System, Extreme Learning Machines, Shallow Architecture, Deep Architecture.

I. INTRODUCTION

Data, which is available in abundance, can be streamlined for growth and expansion in technology as well as business. When data is analysed successfully, it can become the answer to one of the most important questions: how can business acquire more customers and gain business insight? The key to this problem lies in being able to source, link, understand, and analyse data. Analytical approaches are predictable analysis, Behavioural analysis and data interpretation. [2] Predictive Analysis is how a business can use the available data for predictable and real-time analysis across its difference domains. How can a business leverage complex data in order to create new models for driving business outcomes, decreasing business costs, driving innovation in business strategy, improving overall customer satisfaction, converting an audience to a customer. According to Atual Butte, Standford, "Hiding within those mounds of data is knowledge that would change the life of a patient or change the world. So the real power of Big Data lies in its analysis. [3] Processing studying, and implementing the conclusions derived from the analysis of big Data help you to collect accurate data, take timely and more informed strategic decisions, target the might right set of audience and customers, increase benefits, and reduce wastage and costs. The right analysis of the available data can improve major business in various ways.

II. LITERATURE SURVEY

In 2018, Indian population is 132 crore people. Comparative to the world, second in most of the trains operated by India. Every day at least 70 million people travel by train. British Government has started the train services in major parts of India. In British time, train runs depend on the coal. Trains useful to the people for travelling one city to another city. It fair is less. Total number of trains in India has 20 thousand trains and operated by Indian Railways. One of the major problems in Indian Railway Department is delay of train. The train delay prediction system has used in recent years ago [4]. Advantage of this prediction system has used to find the status of train. The train delay prediction system is useful to the Indian railways. Indian Railways covers many states overall India. This is the big achievement. Train Delay Prediction has used to evaluate the delay of train in hours or minutes [5]. Train Delay systems are handle

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

by the railway staff. Many rail information systems didn't collect the old history from different information sources. The dependent of on the present running train information is vulnerable to incorrect prediction. Most of the people

travelling by train. Many Scientists and Research Companies are working in railway and they are designing new technology are based on the line characteristics, train characteristics and overall statistics data. It is targeting at work out the amount of time required to complete a particular part of travel and run on it for predictions. Railway Traffic Management is developing in furtherance administration of the natural unpredictability of railway services. It is proving an incorporated and all-encompassing way of operational execution. Train Delay has depended on the train movements and weather conditions. Weather conditions have Temperature, Humidity, Wind, Rain and Solar Radiation. In India, most of the trains can't run the current time [6]. It is one of the world's largest networks. The purpose of this paper is finding the delay of train. Every year central government increases the budget for railways. The railways income plays a major role in India Economy.

III. METHODS

A. *Extreme Learning Machines*
B. *Shallow Extreme Learning Machines*
C. *Deep Extreme Learning Machines*

A. Extreme Learning Machines

Extreme learning machines are feed forward neural systems for characterization, relapse, grouping, scanty estimate, pressure and highlight learning with a solitary layer or various layers of shrouded hubs, where the parameters of concealed hubs (not only the weights interfacing contributions to shrouded hubs) require not be tuned. These concealed hubs can be arbitrarily allotted and never refreshed (i.e. they are arbitrary projection however with nonlinear changes), or can be acquired from their progenitors without being changed [7]. As a rule, the yield weights of shrouded hubs are generally learned in a solitary advance, which basically sums to taking in a direct model. The name "Extreme learning machine" (ELM) was given to such models by its primary innovator Guang-Bin Huang. As indicated by their makers, these models can create great speculation execution and learn a great many circumstances quicker than systems prepared utilizing back propagation. In writing, it likewise demonstrates that these models can beat super vector machines (SVM) and SVM gives problematic arrangements in both order and relapse applications.

B. *Shallow Extreme Learning Machines*

SELM were originally developed for the single hidden layer feed-forward neural networks.

$f(x) = \sum_{i=1}^{h} wig_i(x)$ (1)

Where $g_i: R^d \to R$, $i \in (1, \ldots \ldots, h)$ is the hidden layer is output corresponding to the input sample $x \in R^d$, and $W \in R^h$ is the output weight vector between the hidden layer and the output layer [8]. In this case, the input layer has d neurons and connects to the hidden layer through a set of weights $W \in R^{h(0,\ldots d)}$ and a nonlinear activation function, $\phi: R \to R$.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

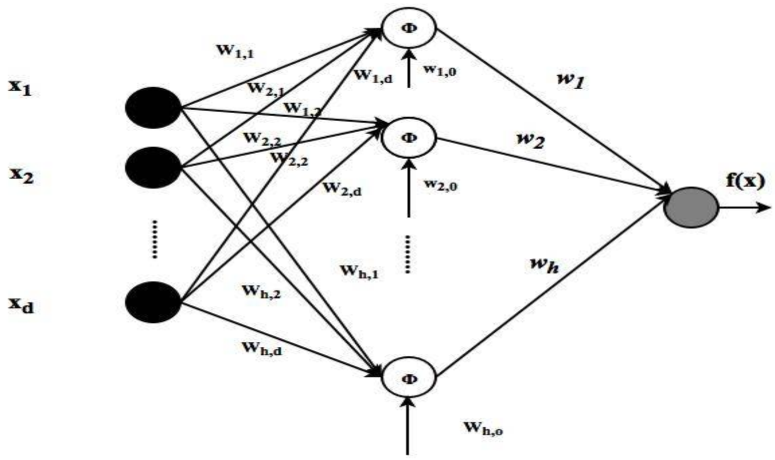

Fig. Shallow Extreme Learning Machines Structure

D. *Deep Extreme Learning Machines*

Due to its shallow architecture, feature learning using SELM may not be effective even with a larger number of hidden nodes. Since feature learning is often useful to improve the accuracy of the final model, multilayer solutions are usually needed [9]. In a multilayer learning architecture is developed using ELM based autoencoder as its building block, which results in a sort of “Deep”. The original inputs are decomposed into multiple hidden layers l, each one composed of $h_{i\in\{1,..,l\}}$ hidden neurons, and the outputs of the previous layer are used as the inputs of the current one.

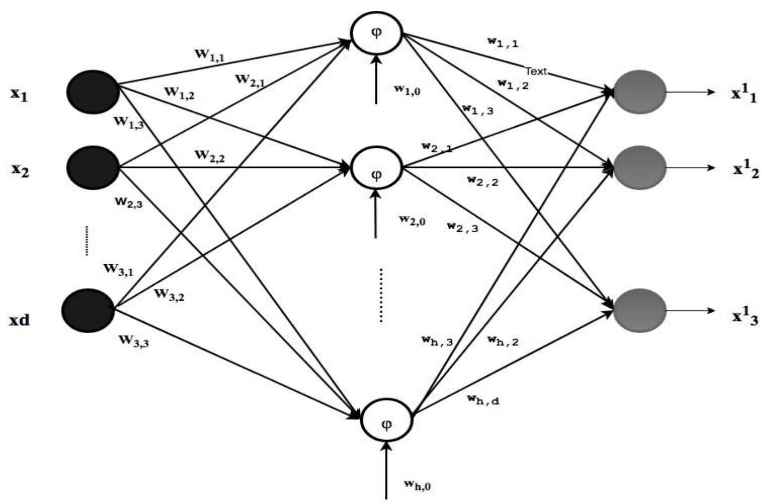

Fig. Deep Extreme learning Machines Structure

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

IV. ALGORITHMS

ALGORITHM 1 SGD FOR SELM

INPUT: D_n, Λ, T , n_{iter}

OUTPUT: W

READ D_n:

COMPUTE A;

W=0;

FOR T $\longleftarrow$ 1 TO n_{iter} DO

$W = W - \frac{r}{\sqrt{r}} \frac{a}{aw} [||AW - y||^2 + \Lambda||W||^2]$

RETURN (W,B);

ALGORITHM 2: SGD FOR SELM ON SPARK ($d \geq h$)

INPUT: D_n, Λ, T , n_{iter}

OUTPUT: W

COMPUTE A/ * COMPUTE THE PROJECTION

W=0;

FOR T $\longleftarrow$ 1 TO n_{iter} DO

$g = (A, y).map(Gradient())$

/* COMPUTE THE GRADIENT FOR EACH SAMPLE

.REDUCE(SUM())

/* SUM ALL THE GRADIENTS OF EACH SAMPLE

$W = w - \frac{r}{\sqrt{r}} g;$

RETURN W;

ALGORITHM 3: SGD FOR SELM ON SPARK ($d \leq h$)

INPUT: D_n, Λ, T , n_{iter}

OUTPUT: W

READ D_n;

W=0;

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

FOR T $\leftarrow$ 1TO n_{iter} DO

$$g = D_n.map(\phi\&Gradient())$$

/ * COMPUTE BOTH THE PROJECTION ϕ AND THE GRADIENT

FOR EACH SAMPLE */

.REDUCE(SUM())

/* SUM ALL THE GRADIENTS OF EACH SAMPLE */

$$W = W - \frac{r}{\sqrt{r}}g;$$

RETURN W;

V. RESULTS

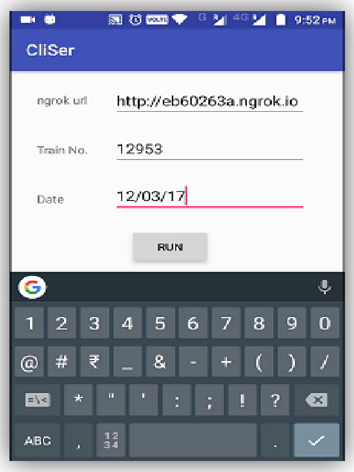

Fig. Entering the Train number

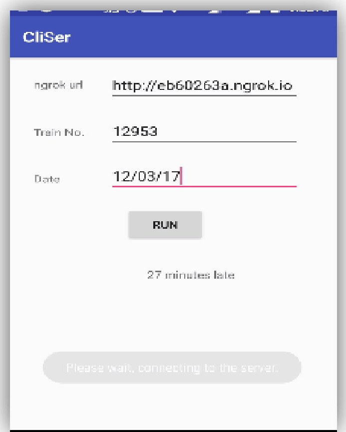

Fig. Delay of Train Time

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

VI. CONCLUSION AND FUTURE WORK

In this paper is the Train Delay Prediction System Using Big Data Analytics. In India train delay is the major problem and Government of India can't solve this problem. Train delay is depending on the train movements and weather conditions. Weather conditions like Temperature, Humidity, wind, Rain and Solar Radiation. The main purpose of this paper is finding delay of train time. We are using the Shallow Extreme Learning Machines and Deep Extreme Learning Machines. More often than not we hear about communal riots, strikes, railway vandalism and disasters that really affect the running schedule of the trains passing by concerned regions [10]. We have pointed it as X-factor because if this is included in the model it would make the application more useful and convincing. To add such a feature we would need to keep a parameter X which will initially be just null. We will do data crawling on some known news websites of the country and look for any news that relates with factors that might cause trains to get delayed. This would not be trivial but some existing procedures can be used [11]. If there are such news available related to a particular region we could set the parameter X for the region and the model will this knowledge to predict delay.

- **Widening the scope of Region/ Station: -** Right now are having only considered trains which commences from Delhi or its vicinity. We would to increase the number of regions and concerned cities from which the trains depart. It is an essential extension to the model and it comes with inclusion of a lot more trains [12]. To add the functionality we would add a parameter for the train departing region. We might be able get some interesting correlations between the departing and ending region.

- **Using Railway Infrastructure: -** It is very well known that number of track lanes differ a lot in different regions. Their number increases on busy routes and decreases on usually less-busy routes. We would like to

  Incorporate this information in our current model. We have made a valid assumption that the routes which has less track lanes, encounters a train delay then the other trains which are going through the route might get some drift [13]. In case of the routes which have s good number of track lanes available this assumption might not help but it will impact to a certain extent.

- **Let's look at it from other side: -** We can use the predicted/ real delays to tell the problems with the railways infrastructure which could be come out to be a good tool for the infrastructure engineers [14]. We can certainly find correlations between parameters and suggest them where do they need to take care or add something to the infrastructure.

REFERENCES

1. H. Li, D. Parikh, Q. He, B. Qian, Z. Li, D. Fang, and A. Hampapur, "Improving rail network velocity: A machine learning approach to predictive maintenance," *Transportation Research Part C: Emerging Technologies*, vol. 45, pp. 17–26, 2014.
2.Y. Bai, T. K. Ho, B. Mao, Y. Ding, and S. Chen, "Energy-efficient locomotive operation for chinese mainline railways by fuzzy predictive control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 3, pp. 938–948, 2014.
3. J. F. Cordeau, P. Toth, and D. Vigo, "A survey of optimization models for train routing and scheduling," *Transportation science*, vol. 32, no. 4, pp. 380–404, 1998.
4.T. Dollevoet, F. Corman, A. D'Ariano, and D. Huisman, "An iterative optimization framework for delay management and train scheduling," *Flexible Services and Manufacturing Journal*, vol. 26, no. 4, pp. 490–515, 2014.
5. S. Milinkovi´c, M. Markovi´c, S. Veskovi´c, M. Ivi´c, and N. Pavlovi´c, "A fuzzy petri net model to estimate train delays," *Simulation Modelling Practice and Theory*, vol. 33, pp. 144–157, 2013.
6. R. M. P. Goverde, "A delay propagation algorithm for large-scale railway traffic networks," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 3, pp. 269–287, 2010.
7. Xavier Delorme, Xavier Gandibleux, and Joaqu´ın Rodriguez. Stability evaluation of a railway timetable at station level. *European Journal of Operational Research*, 195(3): 780–790, 2009.
8. C. Aytekin, Y. Rezaeitabar, S. Dogru, and I. Ulusoy, "Railway fastener inspection by real-time machine vision," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 7, pp. 1101–1107, Jul. 2015.
9. H. Feng *et al.*, "Automatic fastener classification and defect detection in vision-based railway inspection systems," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 4, pp. 877–888, Apr. 2014.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

10. E. Fumeo, L. Oneto, and D. Anguita, “Condition based maintenance in railway transportation systems based on big data streaming analysis,” in *Proc. INNS Big Data Conf.*, San Francisco, CA, USA, 2015, pp. 437–446.
11. H. Li *et al.*, “Improving rail network velocity: A machine learning approach to predictive maintenance,” *Transp. Res. C Emerg. Technol.*, vol. 45, pp. 17–26, Aug. 2014.
12.Y. Bai, T. K. Ho, B. Mao, Y. Ding, and S. Chen, “Energy-efficient locomotive operation for Chinese mainline railways by fuzzy predictive control,” *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 3, pp. 938–948, Jun. 2014.
13. M. Ma, P. Wang, C.-H. Chu, and L. Liu, “Efficient multipattern event processing over high-speed train data streams,” *IEEE Internet Things J.*, vol. 2, no. 4, pp. 295–309, Aug. 2015.
14. K. Noori and K. Jenab, “Fuzzy reliability-based traction control model for intelligent transportation systems,” *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 43, no. 1, pp. 229–234, Jan. 2013.