Building Energy Model for Citypulse Smart City Traffic using Trees.M5P

Renuka sagar, U Eranna

Abstract: The main objective of WSN is to increase network lifetime. Energy models are required to achieve this objective[1]. These models reduces no of sensors in any physical condition and achieve higher accuracy. In this paper, we present regression model to identify a relationship between a target variable and attributes in the dataset. We also present statistical[2] relations between target variable and observed variable. Simulation results show that trees M5P builds energy model and classifies the dataset with the Average measured time, Average speed and vehicle count to extend the network lifetime .

Keywords: Linear Regression, Energy model, average speed, vehicle count.

I. INTRODUCTION

ML belongs to a category of algorithms that allows for more accurate prediction without being explicitly programmed. It is similar to data mining and predictive modeling[3], especially where all the techniques require searching for patterns and adjusting program actions. Machine learning algorithms are classified into supervised and unsupervised learning algorithms. In the case of supervised learning algorithm[4], data scientists have a set of training data through which model can be built. The model is tested with testing data. In unsupervised algorithms, the model is built by not using the training dataset but by discovering new patterns.

Supervised learning is the learning in which the supervisor teaches or trains a machine using labeled data with correct answers[5]. After that, the machine is provided with a new set of data so that the supervised learning algorithm can produce the correct outcome from the labeled data with the help of training examples. Regression analysis falls under supervised machine learning. Here, the system tries to predict a value for an input based on previous information. The characteristics of regression are:

1. The responses obtained from the model are always quantitative in nature.
2. The model can be constructed only if past data is available.

Our major contributions in this work are as follows. First, we have provided detailed state-of-the-art related to vehicle datatraffic dataset. second, the performance of the intelligent classifiers is examined using correlation coefficient.

Third, we have proposed block diagram for our proposed energy model for our proposed system. Finally, we have proposed[6] an intelligent model for efficient energy management in WSNs. This model has been addressed not only at the classification level, but also at the processing level, where we employed the statistical ranking and selection methods.

These two methods, in addition to using Linear-regression as an intelligent classifier, are core elements in building the model.

II. DATASET

We have chosen a vehicle traffic dataset provided by city of Aarhus in Denmark. A collection of datasets of vehicle traffic, observed between two points[7] for a set duration of time over a period of 6 months (80 observation points in total). The data is available in raw (CSV) and semantically annotated format using the citypulse information model. There are 9 attributes and 80 instances.

I. Block Diagram of the proposed system
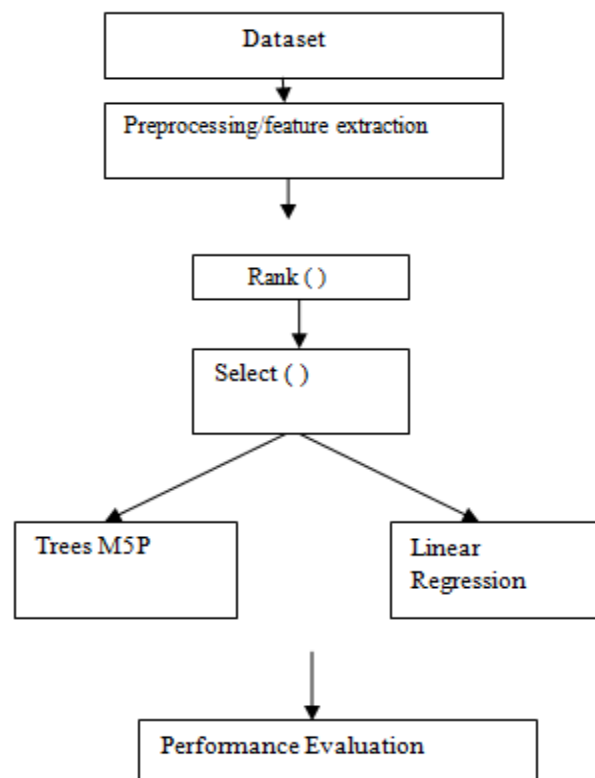
III. EXPERIMENTAL WORK

Experiment was conducted on the vehicle traffic dataset. In this experiment, we have adopted trees.M5P algorithm

Table 1: trees M5P classification algorithm attributes

Average measured time	Average speed	Number of instances(AMT)	Number of instances(AS)
<=49.5	<=53	18	7
>49.5 <=56.5	>53	23	9
>=56.5 <=61.5	<=64	22	22
>61.5 <=69.5	>64	10	14
>69.5	>58	7	11

to build an intelligent classification algorithm. In addition, for all experiments, we have used cross validation set with 10 folds. Also, WEKA[8] simulator is used in the classification process. For the algorithms, Table 1 lists the values of the important parameters such as min value, max value, mean & , std deviation of avg measured time, avg speed, vehicle count, id. All other parameters were set to their default settings in WEKA. These are the inferences observed after feature extraction[9] and data preprocessing. The M5P model tree algorithm is implemented having a batch size 100, min number of instances 4.0 with pruned tree.

Table2: summary of Data preprocessing Average measured time,average speed and vehicle count

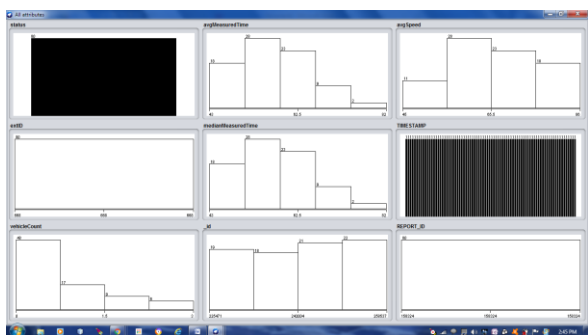| S.No	Min	Max	Mean	stdDev
Avg measured time	43	82	56.338	9.198
Avg speed	45	86	67.063	11.38
Vehicle count	0	3	0.663	0.954
Id	225471	258537	242673.56	9818.969

Fig1 is the visual representation of all the attributes for the vehicle traffic dataset.

The dataset is passed through trees M5P and Linear regression with 10folds crossvalidation[10]. Based on the avg measured time[11] and the average speed. The no of instances classified are as follows. The followingconclusions are made during the experimental study[12].
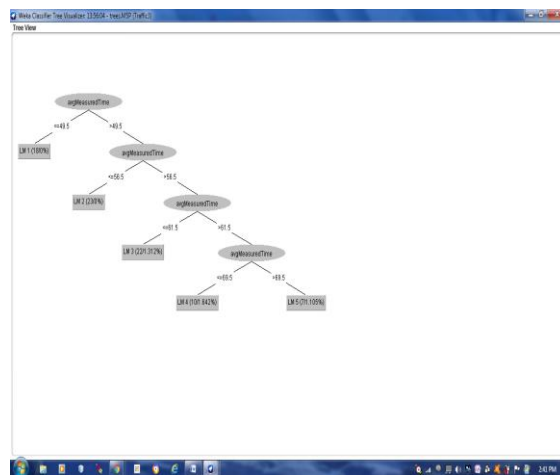

Fig2: visualize tree M5P classifier

IV. RESULT ANALYSIS

Table3: Result Analysis of vehicle traffic dataset

Average speed(km/hr)	Actual vehicle count	Predicted vehicle count	Remarks
45-55	48	11	No severity in Accident
55-65	17	28	Severity in Accident
65-75	9	23	Major Accident
75 & above	6	18	Major Accident

V. CONCLUSION OF THE RESEARCH WORK:

This paper aims to build the most intelligent classification model for efficient energy management using ML. In order to evaluate the performance of the classifier, different experiments were conducted on the classification algorithms , namely, vehicle traffic datasets. These datasets correspond to different WSNs that are used in various types of applications. From the Simulation results on vehicle traffic datasets, we can obtain the simple linear regression which is a statistical method that helps to summarize and study relationships between two continuous variables. Statistical relations can be highly useful even though they do not have the exactitude of a functional relation. This regression model identifies a functional relationship between a target variable and a subset of remaining attributes contained in the dataset. We (a)carried out the experiment of gathering of a sample of observed values of avg speed and corresponding vehicle count,(b) creating a relationship model,(c) finding the coefficients from the model created and establishing the mathematical equation using these,(d) getting the summay of the relationship model to know the avg error in prediction(e) predicting the value of an unknown sample.Time taken to build the model 0.33 sec.The karl pearson's coefficient of correlation is a helpful statistical formula that quantifies the strength of two variables.

This coefficient value helps in determining how strong that relationship is between the two variables.

Table4: Cross validation summary of trees M5P classification algorithm

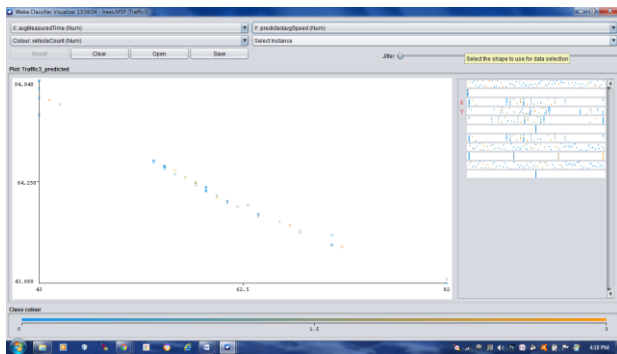|	
Correlation coefficient	0.9811
MAE	1.5086
RMSE	2.8221
RAE	17.31%
Total no of instances	80

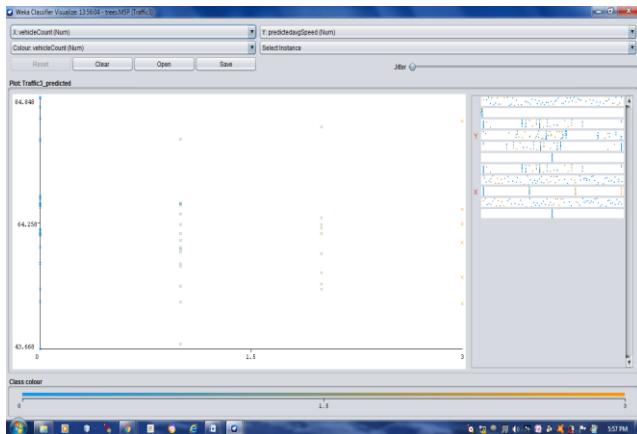**Fig3:Avg measured time vs predicted avg speed**

Fig4: vehicle count vs predicted avg speed

REFERENCES

1. Han,J. and kamber, M. (2015) Data Mining: concepts and Techniques, McGraw-Hill Inc., New York.
2. Harrington, P. (2015) Machine learning in Action, Maning publications, Newyork.
3. Hartman, G.(2016) Fundamentals of Matrix Algebra. Create Space independent Publishing Platform.
4. Marsland, S. (2016) Machine learning: An algorithmic perspective,CRC Press, Florida.
5. Mitchell,T.(2017) Machine learning, Mc Graw-Hill Inc., NewYork.
6. https://machinelearningmastery.com/practical-machine-learning-problems/
7. http://www.analyzemath.com/statistics/linear_regression.html
8. http://www.stst.wmich.edu/wang/160/egs/slr.pdf
9. https://www.google.com/search?q=citypulse+vehicle+traffic+data+in+weka&source=lmns&bih=508&biw=1093&hl=en&ved=2ahUKEwiT9fTblKPmAhVexzgGHV_ODboQ_AUoAHoECAEQAA
10. https://www.researchgate.net/publication/330865819_Data_mining_approach_for_predicting_the_daily_Internet_data_traffic_of_a_smart_university
11. http://epubs.surrey.ac.uk/813389/1/Manuscript.pdf
12. https://tandfonline.com/doi/abs/10.1080/15472450.2015.1004063?src=recsys&journalCode=gits20

AUTHORS PROFILE

Renuka sagar, research scholar, completed MTech(CNE) under VTU. Working as AP under VTU affiliated college (BITM).With teaching experience of 10 years. My areas of interest in research field are Machine learning, Artificial intelligence, Internet of things, Wireless sensor network, develops projects under IOT, AI and ML . works with advanced level machine learning projects with simpliLearn team. Mail id: renukasagar83@gmail.com

Dr U Eranna, prof & HOD, Dept of ECE ,BITM , VTU. Completed B.Tech ECE,M.E Automatic control systems,Ph.D. Has Teaching experience of 23 years .His research interest is in the fields of communication, IOT, Image processing . also industry consultant member. He ha s published 8 conference papers and 20 journal papers. Delivered invited lectures on antennas & wavepropagation,signals & systems,industrial automation in various colleges Mail Id:jayaveer_88@yahoo.com