ISSN 2278-3091
Volume 9, No.1.4, 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse2491.42020.pdf
https://doi.org/10.30534/ijatcse/2020/2491.42020

Machine Learning Decision Tree Classifier and Logistics Regression Model

Amit Sagu[1], Nasib Singh Gill[2]
[1]Department of Computer Science & Applications, MaharshiDayanand University, Rohtak, Haryana, India, saguamit98@gmail.com,
[2]Department of Computer Science & Applications, MaharshiDayanand University, Rohtak, Haryana, India, nasibsgill@gmail.com

ABSTRACT
Machine learning is being used in several aspects and revolutionized the method of believing. It is branch of Artificial Intelligence that takes input data for training the models. In present paper we focus on widely used model of machine learning i.e. Decision tree classifier and Logistics Regression, additionally tracking their performance along with their accuracy. Machine learning is emerging notion in IoT environment as security of IoT devices are crucial due to their resource constrained properties.

Key words: Machine Learning, Decision Tree classifier, Logistics Regression, IoT

1.OVERVIEW

Machine learning mainly holds three kinds specifically supervised, unsupervised and reinforcement. All the varieties come up with their distinct learning ways. This paper spot light on supervised learning, and in what manner it can be valuable in IoT environment. Supervised learning has two categories, leading is Classification and another is regression. In classification, outputs are generally two classes, described as binary classification, if it possesses more than two class, it well-known as multiple class classification. On other hand we come up with regression, which delivers the output in continuous values as prices in rupees, temperature in Celsius, possibility of any event to occur etc. Ahead of time we will get to understand that regression can be applied as classification too.

We employ Python 3.7 and PyCharm editor for illustrating the implementation of use case. Python language remarkably skilled and user salutary while working in machine learning, data mining and in data science. It encompasses a sizable community of developer along with a gigantic database of libraries. Scikit Learn or sklearn is extremely worth while library for machine learning task which offers the developer various form of classifier.

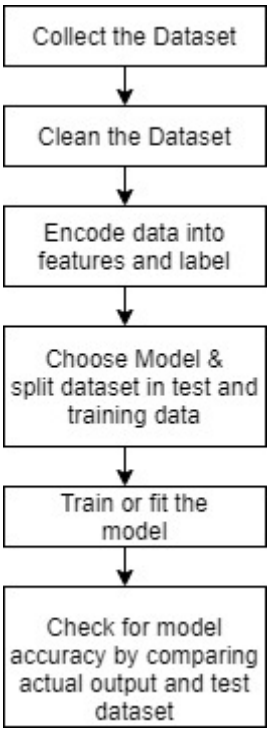

Figure 1: Machine Learning Use Case

(Figure 1) While working with machine learning models, there are certain stages we must go across by. The early phase is collecting the dataset. Dataset can be in continues values, discrete values or even in string or characters. Several machine learning models needed discrete values, and a few expected continues values. In adjoining step, we clean the data by eliminating the null value rows, duplicate row and encode the dataset into features and label. Features are those column values which we are providing the model as input and label are the actual output.

```
0       play
1    no play
2    no play
3       play
4       play
5    no play
Name: match, dtype: object
```

Figure 4: Dataset Labels

2. MACHINE LEARNING CLASSIFIER

One of the types of supervised learning. The classification model has classes in their output labels and having discrete values. At the stop of processing it come to select one of them. Decision tree and random forest are the largely applied classifier. We will be using decision tree classifier for implement the use case however it can be applied in regression tasks too.

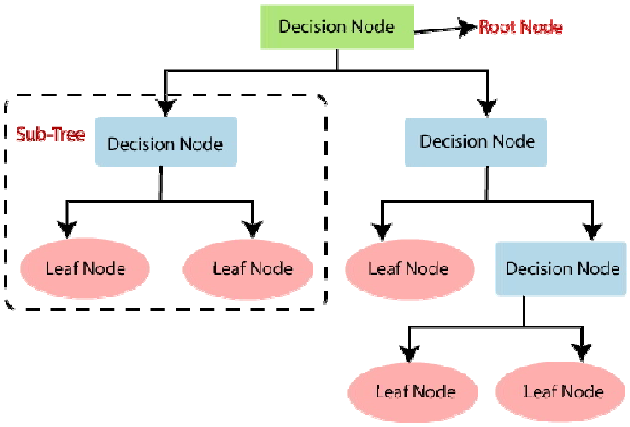

Figure 2: Machine Learning Classifier (Source: Https://Static.Javatpoint.Com/Tutorial/Machine-Learning)

Implementing the Use Case

We are conducting cricket match prediction as our use case scenario. We get features or circumstances that can influence the cricket match, these are weather conditions, 'Sunny', 'Rainy', and 'Windy' along with label column i.e. if the match will be played or not in the .csv (Coma Separated Values) file. Since we are working with binary classifier, all the feature must be encoded in binary format. As can be seen in fig 2, we have encoded our features in 0 or 1. We interpret 0 as false/absence/not-happening and 1 vice versa.

	A	B	C	D	E
1	match	sunny	rainy	wind	
2					
3	play	0	0	1	
4	no play	0	1	0	
5	no play	0	1	1	
6	play	1	0	0	
7	play	1	0	1	
8	no play	1	1	0	
9	no play	1	1	1	
10	play	0	0	0	
11					
12					

Figure 3: Use Case Dataset

Implementation Approach

In the original dataset we have 8 rows and 4 columns. We set aside certain portion of data for testing model. We are holding 5 rows out of 8 and split dataset in respect of input and output. We are benefiting Pandas library for manipulate the dataset. As can be seen in fig 3 and fig 4 we had declined the match column which was output.

```
   sunny  rainy  wind
0      0      0     1
1      0      1     0
2      0      1     1
3      1      0     0
4      1      0     1
5      1      1     0
```

Figure 4: Features

Since we have both input and output, now we call the decision tree classifier. We import the sklearn library for use of classifier. For testing and training model we use 80:20 rule, it means 80 percent of the dataset will be used for training and rest 20 percent will be used for testing model. Accuracy is evaluated by comparing the predicted output by the model with testing data's output.

Testing Accuracy Score

For testing the accuracy score we use up python library function.

```
fromsklearn.metricsimport accuracy_score
```

By running testing, we got 1.0 score, since dataset was small, it will give more accurate result if we gradually increase the dataset then accuracy score may have change. after testing accuracy let's check prediction test. By giving the model input [1,1,1] and [0,0,0] model give out result as ['no play' 'play'] respectively.

By [1,1,1] mean here ['Sunny' 'Rainy' 'Wind'] and [0,0,0] mean ['Not Sunny' Not Rainy' 'Not Wind']

Machine Learning Regression Model

Regression is kind of predictive model to investigate the relationship between dependent (Y) and independent variable (X). Although there are large number of regression models, but we dedicated on broadly applied i.e.Logistics and Linear regression. This technique is finding casual effect relationship between the variables, for instance relation between the weight and height of any human.

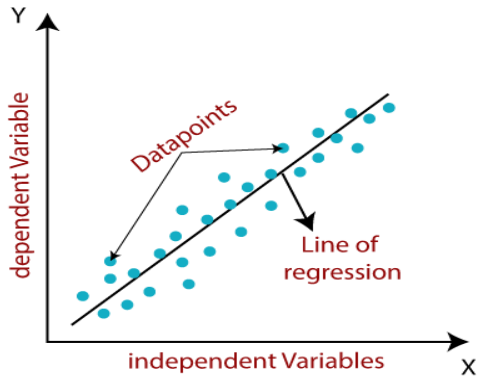

Figure 5: Regression Model (Source: Https://Www.Javatpoint.Com/Linear-Regression-In-Machine-Learning)

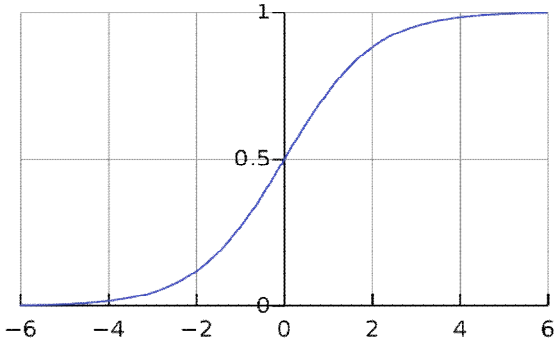

Figure 7: Sigmoid Function Curve

In Linear regression dependent variable is always continuous values like price, age etc. and independent variable can be continuous or discrete. The general approach of this model to find the best fit line which shows the relationship between X and Y.

In (1) B0 is intercept and B1 is slope just like straight line equation

$$Y = B0 + B1 * X \quad (1)$$

Logistic regression is same as linear regression but bound within the certain range of values i.e. 0 to 1. Whereas linear regression gives value in continuous fashion, logistics regression gives values within 0 and 1. Linear regression can be changed into logistic regression by logistic function which is sigmoid function.

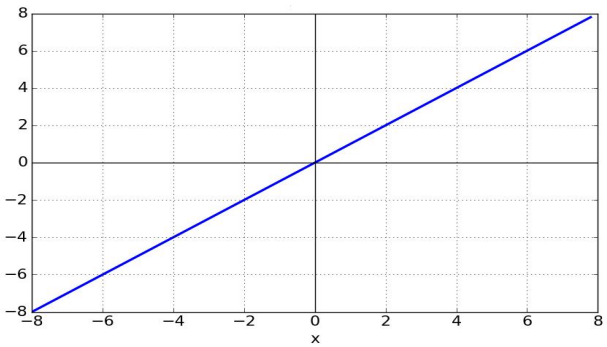

Figure 6: Linear Function Curve

Sigmoid function used in logistics regression. The main characteristic of sigmoid function (2) is that it always gives values between 0 and 1.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

With the help of sigmoid function model would give output in two classes i.e. binary classification.

3.RESULTS

For evaluating the result, we took related dataset for both model and run with identical data sample.

Following the accuracy score of decision tree classifier and logistic regression, we inferred that decision tree is more accurate as it gave 1.0 accuracy, on the other hand logistic regression was not steady in their accuracy score and oscillate between 0 to 1.Sometime accuracy may depend upon the volume of dataset.

If we have independent variable in discrete values like 0 or 1, TURE or FALSE then we should use Decision tree classifier, when we have independent variable in continuous form that is price, age or and statistics then we should go for logistics regression.

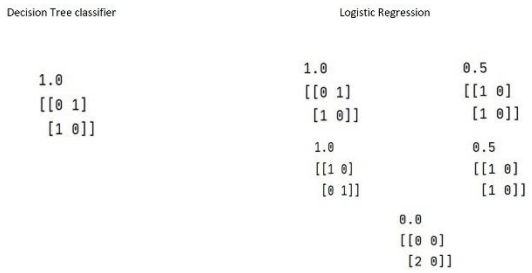

Figure 8: Accuracy and Confusion Matrices

Using Machine Learning Model in Iot

Predictive ability always helpful in all the way. Machine learning can be applied where desired outcome is known, or data is unknown. if we are talking about IoT, machine learning can be applied in security context as devices are very resource constrained and more vulnerable to attack like DDOS or Botnet. With assist of machine learning, malicious traffic can be recognized and stop before entering in IoT environment. By extracting the attack characteristics, we can easily feed the machine learning model. DDOS attack are mostly known attack for IoT environment, in which attacker continuously send the request to IoT server resulting overburden the server. In botnet attack, hacker machine (bot) try to exploit IoT device by entering the common credential details.

Machine learning can be very significant for securing the IoT environment.

4.CONCLUSION

In present paper we centered on machine learning classifier and regression model. We spotlight decision tree and logistics regression expressly by taking exemplary dataset, test their accuracy and performance. Following performance testing, we inferred that for binary class result decision tree was more accurate than logistics regression as accuracy score of logistics regression was fluctuating. It likewise depends upon the volume and complexity of dataset, which would be more fair comparison. In our future work we would try to study their performance on more complex and capacious data.

REFERENCES

1. Hyo-Sik Ham & et al. “Linear SVM-Based Android Malware Detection for Reliable IoT Services” Journal of Applied Mathematics Volume 2014, Article ID 594501, 2014.
2. Bhunia& et al. “Dynamic Attack Detection and Mitigation in IoT using SDN” 2017, 27th International Telecommunication Networks and Applications Conference, IEEE.
3. Deepa, S., and R. Umarani. "Steganalysis on images based on the classification of image feature sets using SVM classifier." International Journal of Computer Science and Engineering (IJCSE) 5.5 (2016): 15-24.
4. Mehdi Nobakht et al. “A Host-based Intrusion Detection and Mitigation Framework for Smart Home IoT using OpenFlow” 2016 11th International Conference on Availability, Reliability and Security IEEE.
5. Miettinen& at el. “IOT SENTINEL Demo: Automated Device-Type Identification for Security Enforcement in IoT” 2017, IEEE 37th International Conference on Distributed Computing Systems.
6. Available online at https://www.arduino.cc/en/Guide/ArduinoUno
7. MehmentcanGule& at el. “Android Based WI-FI Controlled Robot using Raspberry Pi” 2017 IEEE.
8. Janice Canedo& et al. “Using Machine Learning to Secure IoT Systems” 2016 14th Annual Conference on Privacy, Security and Trust, PST, 2017, IEEE.
9. PATEL, AJAY M., A. PATEL, and HIRAL R. PATEL. "COMPARATIVE ANALYSIS FOR MACHINE LEARNING TECHNIQUES APPLIANCE ON ANOMALY BASED INTRUSION DETECTION SYSTEM FOR WLAN." (2013).International Journal of Computer Networking, Wireless and Mobile Communications (IJCNWMC) 3.4, Oct 2013, 77-86
10. Cote & at el. “Using Machine Learning in Communication Networks” J. OPT. COMMUN. NETW. /VOL. 10, NO. 10/OCTOBER 2018, IEEE
11. Fanzeng Xia & et al. “Securing the wireless environment of IoT”, 2018 IEEE International Conference of Safety Produce Informatization (IICSPI).
12. TagyAldeen& at el. “Towards Machine Learning Based IoT Intrusion Detection Service” 2018, Springer.
13. Doshi& at el. “Machine Learning DDoS Detection for Consumer Internet of Things Devices” 2018, IEEE Symposium on Security and Privacy Workshops.
14. Chien& at el. “Machine Learning Techniques for Recognized IoT Devices” 2019, Springer.
15. Danthala, S. W. E. T. H. A., et al. "Robotic Manipulator Control by using Machine Learning Algorithms: A Review." International Journal of Mechanical and Production Engineering Research and Development 8.5 (2018): 305-310.
16. Ring & at el. “A Survey of Network-based Intrusion Detection Data Sets” computer & security, vol. 77, 2019
17. Roopak& at el. “Deep Learning Models for Cyber Security in IoT Networks” 2019, IEEE.
18. Durgabai, R. P. L., and P. Bhargavi. "Pest Management using Machine Learning Algorithms: A Review." International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR) 8.1 (2018): 13-22. https://doi.org/10.24247/ijcseitrfeb20182
19. T. Lu el al, "Future internet: The Internet of Things," in Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on, 2010, pp. V5-376-V5-380.