Landslide Prediction using Machine Learning on the Edge Node in Pervasive Internet of Things

Saniya Zahoor[1], Roohie Naaz Mir[1]

[1]Department of Computer Science and Engineering, NIT Srinagar, India

{saniyazahoor,naaz310}@nitsri.net

Abstract. In today's world, everything is connected via internet, but Internet of Things will change our life in the future. For, this to happen, large amount of data has to be generated, processed and captured by IoT and are considered to be highly useful and contain valuable information. The critical role in making things smart is through Machine Learning techniques. Machine Learning will play important role in extracting data and knowledge from the connected things after that constructing smart system that provides convenient services. The data aggregated from the IoT nodes in an IoT environment has to be analyzed in a time bound manner by using different mining algorithms such as KNN, Naïve Bayes, LDA, SVM, C4.5. So, in this paper we focus on comparative analysis of data mining algorithm on internet of things data using MATLAB. We are using various data mining software because there are resource constraints in an IoT environment, and due to resource constraints we cannot use them at device level. In order to use them at edge level, we will analyze the data using MATLAB software tools as if they are to be used at the edge level. At last our preliminary results on real IoT datasets shows that C4.5 has better accuracy and have relatively higher processing speeds in all tools we used.

Keywords: Pervasive, Internet of Things, Machine Learning.

1 Introduction

The main goal of IoT is to form a network of day-to-day life objects and make them programmable using wireless and sensor technologies, and pervasive connectivity. The pervasiveness of IoT eases the everyday activities such as data exchange by objects while sensing, reacting to events and interacting with the environments. The main aim of IoT is that in the future, the everyday objects would be accessible and interconnected inside the global structure of Internet [1]. Figure 1 gives the system model of IoT network. The IoT system reduces the complexity and simplifies the development of applications of various domains. The life cycle of IoT consists of capturing, processing and presenting the data to the outside world either directly or indirectly depending on the connecting capabilities of IoT devices. The cloud offers storage of data, seamless connectivity and provides services for data analysis, thus acts as a backbone of the IoT system. At the edge or cloud level, the data can be analyzed with the help of data mining techniques.
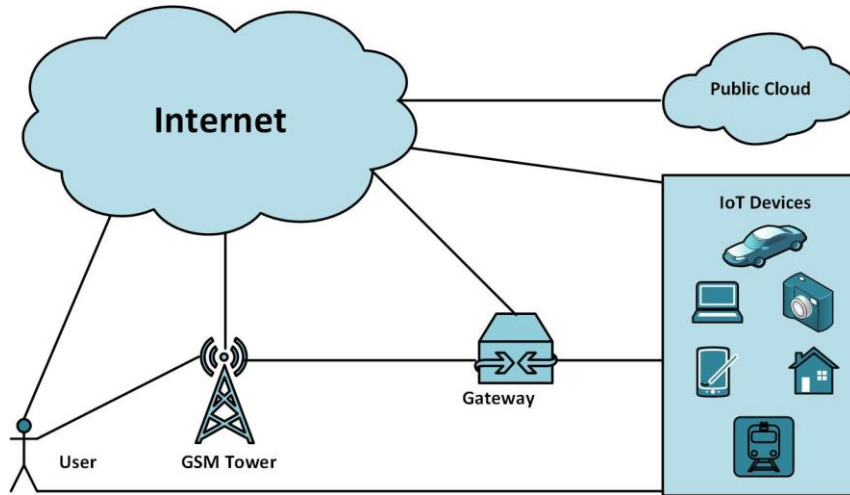
Fig.1. Model of IoT System

In Edge-IoT applications, various data mining techniques such as Support Vector Machine (SVM), K-nearest neighbor (KNN), Nave Bayes (NB), C4.5 and Linear discriminant analysis (LDA) are employed to reduce the complexity of data sensed by IoT devices [2]. This paper presents prediction of landslides using Machine Learning techniques on the sensed data collected on edge node.

The paper is organized as follows: Section 2 discusses literature survey, Section 3 presents case study of landslide prediction, Section 4 presents results and Section 5 gives conclusion.

2   Literature Review

In Internet of Things, IoT devices sense, gather and generate large amount of data about the environment. The important role in making IoT smart is through Machine learning. Machine learning extracts the data and knowledge from these devices by providing convenient services in IoT environments.

2.1   Internet of Things

Internet of Things encompasses large number of IoT devices, densely deployed in a geographical area, and an edge device (e.g., gateway). The basic components of an IoT system includes an assembly of IoT nodes (localized or distributed nodes), sink nodes (cluster heads, aggregator nodes, edge nodes) that help in gathering information from the neighboring nodes, an interconnecting network and computing resources (node or network resources) that help in local processing (e.g., data correlations, fusion, redundancy elimination, etc). One of the interesting characteristics of IoT is that it often allows smart in-network processing wherein intermediate nodes do not only forward packets but also processes the packet content going through them to optimize the utilization of resources.

IoT applications are heavily affecting our daily lives in various domains that range from smaller-sized wearable IoT devices to large industrial IoT systems. The complexity of the pervasive IoT applications can be understood by considering the

nature of resources at the node and network levels. In such IoT networks, the nodes are resource-constrained micro-systems that employ sensors to monitor their environments; as such the main trade-off in such IoT applications is to have automated environments against constrained resources.

The wireless and embedded technologies aim at creating tiny, low-cost and efficient devices. But the tiny size impacts the hardware and software designs at various levels. Such hardware constraints also impact the design of various algorithms and protocols e.g., data aggregation algorithms executed in resource-constrained IoT networks. Various architectural and node level solutions need to be designed to operate efficiently on the small-sized IoT devices. Edge devices can also improve the resource utilization to some extent in pervasive IoT environments by local data aggregation that minimizes the latency and optimizes the bandwidth utilization. Further, various data mining and machine learning techniques can be employed to reduce the complexity of data in such pervasive environments.

2.2 Machine Learning

Machine learning deals with the construction of systems that can learn from data. Benefits include efficient decision making, higher efficiencies, automation productivity, accuracy, etc [2]. Various data mining algorithms include K-NN, Naïve Bayes, LDA, SVM and C4.5.

2.3.1 K-nearest neighbor (KNN)

K-NN is a non-parametric method used for classification and regression of data [3]. The output for classification is a class membership and for regression, the output is the property value for the entity.

2.3.2 Naïve Bayes

Naive Bayes require a number of parameters in a learning process. Maximum-likelihood training can be done by evaluating a closed-form expression. Benefit of this technique is that a less number of training data is required for estimation of the necessary classification parameters [4].

2.3.3 Linear Discriminant Analysis

LDA works when the measurements made on independent variables for each observation are continuous quantities [5].

2.3.4 Support Vector Machine (SVM)

Given a training set, SVM builds a model that assigns new examples to one category or the other. It also works efficiently in a non-linear classification [6].

2.3.5 C4.5

From the training data set, a decision tree is built by C4.5 using information entropy concept [7]. At each node of the tree, C4.5 uses the concept of splitting criterion based on the difference in entropy.

3 Case Study: Landslide Prediction

This paper mainly focuses on the comparative analysis of data mining algorithm on internet of thing data using MATLAB. The rainfall data gathered using IoT nodes on the edge is used to predict landslides using machine learning approaches. We have data of 30-day dataset based on the rainfall and land susceptibility. We are using MATLAB software for analyzing various mining algorithms like KNN, Naïve bayes, Linear Discriminant Analysis (LDA), Support vector Machine (SMO), C4.5.

4 Results

The dataset used for landslide prediction is the sensed data in National highway-21(NH-21) extending from Mandi to Manali. It is a collection of 5 years dataset from year 2011 to 2015 and contains more than 10 thousands of rows and 32 columns. From the datasets, landslide prediction is done on the basis of 30-day rainfall and land susceptibility.
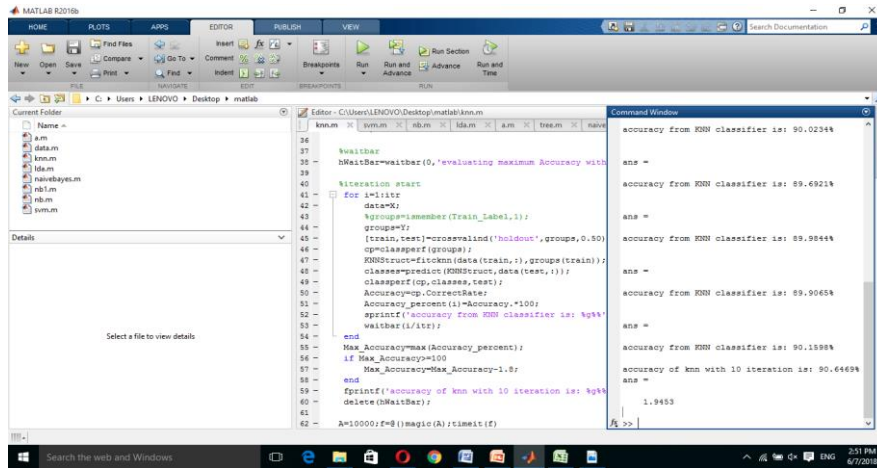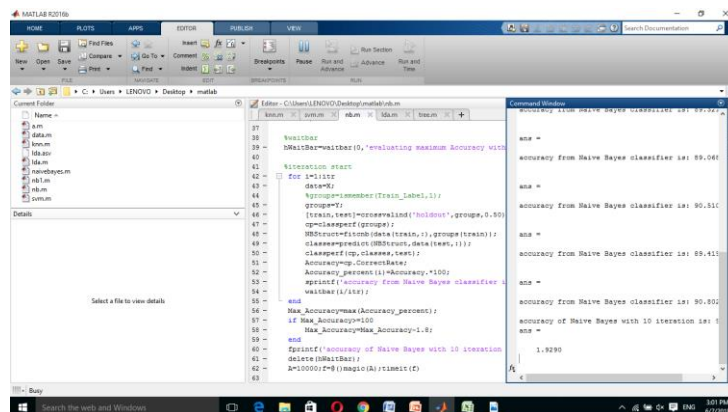

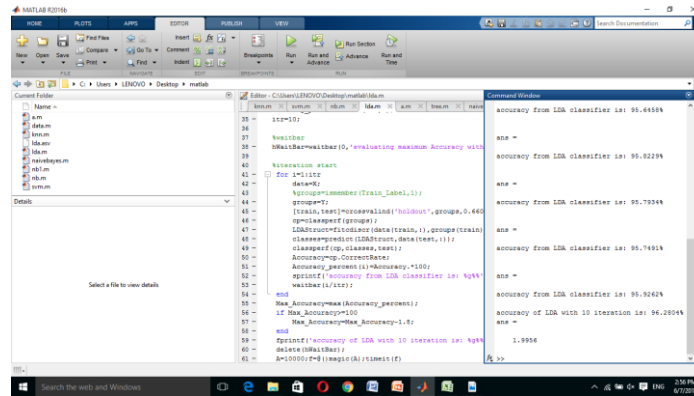**Fig.2.** KNN in MATLAB

Fig.3. Naïve Bayes in MATLAB
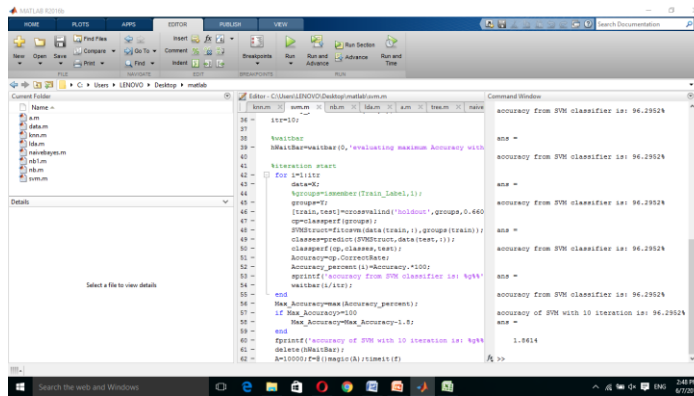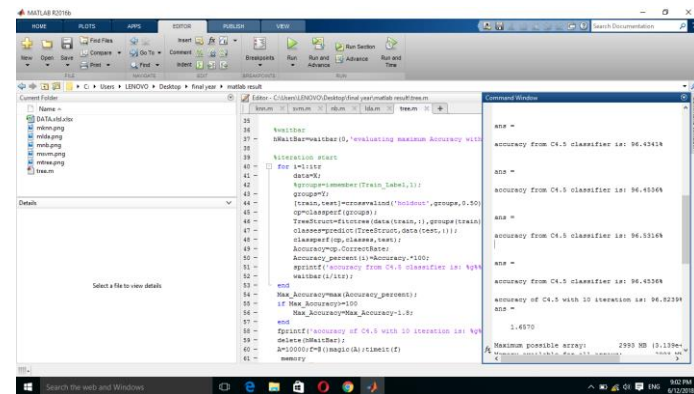

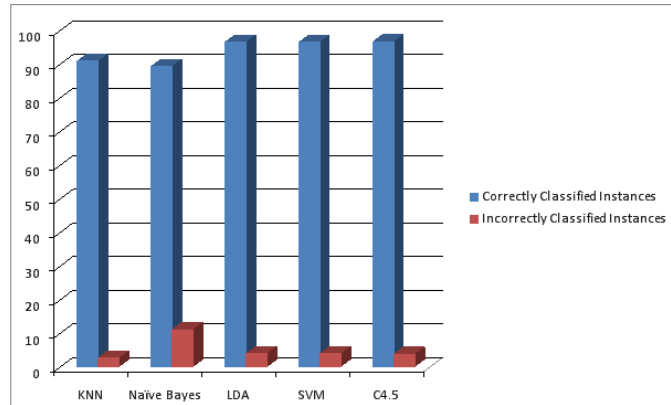**Fig.4.** LDA in MATLAB

Fig.5. SVM in MATLAB

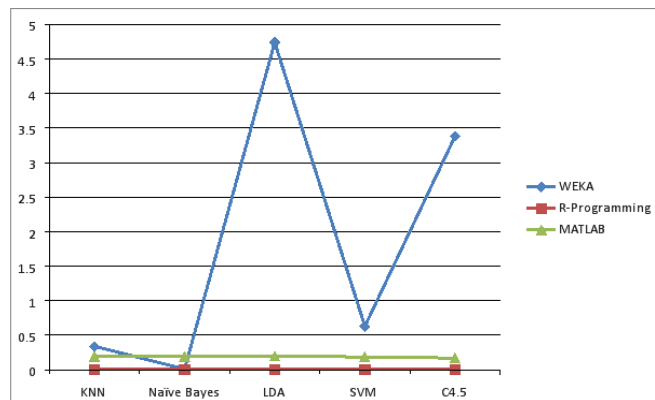**Fig.7.** Comparison of Algorithms in MATLAB

Fig.8. Comparison of Algorithms in WEKA, R-programming and MATLAB

We have analyzed the data on Weka and R-Programming as well. Figure 8 shows the time taken by different algorithms in these softwares.

5   Conclusions

From the above results, following points can be concluded:

- C4.5 gives best results of accuracy among all the implemented algorithms across all platforms
- Naïve gives less accuracy among all platforms across all   platforms.
- If we take processing time into account, R gives better results and thus can be used in time-sensitive IoT applications.
- Based on the above analysis, we infer that R is recommended for IoT environments with limited resources such as processing.

References

1. Ansari, S., Aslam, T., Poncela, J., Otero, P. and Ansari, A., 2020. Internet of Things-Based Healthcare Applications. In IoT Architectures, Models, and Platforms for Smart City Applications (pp. 1-28). IGI Global.
2. Sen, P.C., Hajra, M. and Ghosh, M., 2020. Supervised Classification Algorithms in Machine Learning: A Survey and Review. In Emerging Technology in Modelling and Graphics (pp. 99-111). Springer, Singapore.
3. Shahabi, H., Shirzadi, A., Ghaderi, K., Omidvar, E., Al-Ansari, N., Clague, J.J., Geertsema, M., Khosravi, K., Amini, A., Bahrami, S. and Rahmati, O., 2020. Flood Detection and Susceptibility Mapping Using Sentinel-1 Remote Sensing Data and a Machine Learning Approach: Hybrid Intelligence of Bagging Ensemble Based on K-Nearest Neighbor Classifier. Remote Sensing, 12(2), p.266.
4. Soni, J., Mathur, K. and Patsariya, Y.S., 2020. Performance Improvement of Naïve Bayes Classifier for Sentiment Estimation in Ambiguous Tweets of US Airlines. In Data Engineering and Communication Technology (pp. 195-204). Springer, Singapore.
5. Nie, F., Wang, Z., Wang, R., Wang, Z. and Li, X., 2020. Adaptive Local Linear Discriminant Analysis. ACM Transactions on Knowledge Discovery from Data (TKDD), 14(1), pp.1-19.
6. Pradhan, B. and Sameen, M.I., 2020. Modeling traffic accident severity using neural networks and support vector machines. In Laser Scanning Systems in Highway and Safety Assessment (pp. 111-117). Springer, Cham.
7. Cuan, Y., Wang, Z. and Han, J., 2020, January. Research on TV imaging casing damage detection and classification method based on C4. 5 decision tree. In Journal of Physics: Conference Series (Vol. 1437, No. 1, p. 012132). IOP Publishing.