

# **A COMPARATIVE STUDY OF KNN AND SVM ALGORITHMS IN IOT**

Mr.Rakesh Kumar B<sup>1</sup>, Maria Telma Dsouza<sup>2</sup> & Praveen U.C<sup>3</sup>

**Abstract-** Internet of Things (IoT) is going to make major changes in all aspects of our lives. IOT devices generates enormous data having useful, valuable and highly accurate data. The huge amount of data obtained by the Internet of Things (IoT) are considered to be useful Information for the business use as well as technological developments. This large amount of data is difficult to process and extract knowledge hence Data Mining algorithms should be used. When a large number of devices connected to IoT, enormous data is generated that should be analyzed, also the existing algorithms should be modified to apply for such big data analytics . To make IoT efficient, smatter lot of technologies are introduced into it; one of the most important technology is Data Mining. Data Mining will play important role in constructing the smart and efficient system that provides convenient services. It is required to extract data and knowledge from the connected things. For this purpose, various data mining techniques are used. We are mainly Concentrating on two algorithms namely Support Vectors Machine and K-nearest Neighbor. We compare these two algorithms and focus on how these algorithms help to achieve the purpose and also identifv which one these is most accurate and efficient.

## **1. INTRODUCTION**

The Internet of Things (IoT) is "a worldwide infrastructure for the data society, empowering propelled benefits by interconnecting (physical and virtual) things in light of existing and advancing interoperable data and communication technologies"<sup>1</sup>. IoT<sup>2,3</sup>, is one of the major technological developments of our times given its potential is fully realized. It is set to make major changes in all aspects of our lives, be it work, entertainment and social interaction., A large amount of data are emerging day by day to be part of the IoT infrastructure and the number of information connected to IoT is expected to reach 50 billion by 2020<sup>4</sup>. This would produce huge amounts of critical information. A significant goal of IoT is to make nature around us more smarter and simpler, by giving the condition the data it needs, through ongoing and memorable information sustains, and apply computational insight on the data to settle on better and keen choices consequently. This would upgrade our capacity to deal with our nation, cities, streets, health, homes, forests and much more. We will have better crowd and traffic management, emergency predictions, better prediction of accidents and crimes, etc.

The data's from the IoT gadgets will be gathered and used to control complex conditions around us, and empower to comprehend settle on better basic leadership, profitability, more noteworthy robotization, exactness, higher efficiencies and riches age. The most troublesome test in these procedures is investigations of tremendous measures of information (i.e. huge information) to deliver exceptionally precise and solid bits of knowledge and choices so IoT could satisfy its guarantee. Machine learning is among the best techniques to increase concealed experiences from IoT information. The point of this examination is to check whether the conventional data mining algorithms would moreover work for the IoT datasets, or new gatherings of data mining algorithms are required. To this end, this paper gives a preparatory investigation on inspecting the appropriateness of two surely understood data mining algorithms to genuine IoT datasets. We have utilized two data mining calculations that are Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). The fundamental commitment of this work is the examination of the viability and efficiency of these two calculations as far as their precision. Section 2 provides a review of the relevant literature. Our experimental methodology is explained in Section 3. Simulation results and analysis are reported in Section 4. Conclusions are drawn in Section 5.

## **2. BACKGROUND MATERIAL AND LITERATURE REVIEW**

Now a days data mining tasks are much more difficult because of an anomalous increment in the amount and multifaceted design of data<sup>8,9</sup>. As the IoT is improving worldview, a totally new level of difficulties have been added to the data mining domain<sup>5,6,9</sup>. Support Vector Machine (SVM), K-nearest neighbor (KNN), Linear discriminant analysis (LDA), C4.5, Nave Bayes (NB), C50 and ANNs are widely used in the field of data mining. In the beginning SVM was designed to address bias variance tradeoff, over-fitting and capacity control<sup>10</sup>. Burges<sup>10</sup> also stated that accuracy of the SVM mainly depends on the quality of machine capacity and training data. The use of SVM in the further stages extended from classification to regression

<sup>1</sup> Department of Computer Applications, Asst. Professor, St Aloysius College, AIMIT, Mangaluru, Karnataka, India

<sup>2</sup> Department of Computer Applications, MCA V Semester, St Aloysius College, AIMIT, Mangaluru, Karnataka, India

<sup>3</sup> Department of Computer Applications, MCA V Semester, St Aloysius College, AIMIT, Mangaluru, Karnataka, India

and element ranking. SVM is an exceptionally proficient device to work with in complex and noisy domains. The Computational wastefulness is one of the real downside of SVM, however a few advancements have been done to decrease its computational cost and to build its capacity of a system<sup>9,11</sup>. A lazy learner algorithm known as KNN is one of the most straightforward available classifier which is easy to understand and implement. As a result of the simplicity of KNN, a few issues arise that limit its performance for example the selection of right distance measure, number of neighbors and larger part vote to combing class labels is not always effective<sup>5,9</sup>. KNN is likewise utilized for different tasks in wireless sensor networks (WSN) and IoT domain for intrusion detection<sup>12</sup>, indoor positioning systems<sup>13</sup> and activity recognition<sup>14</sup>. For binary classification problems, SVM is a one of the best options.

However, versatility is dependably an issue<sup>9,10,11</sup>. The approach taken by the KNN and SVM algorithms to understand the data mining task is totally unique to the ones we have examined in earlier section. The SVM algorithm is discover hyper-plane isolating the distinctive classes of the training occurrences with the most outrageous mistake margin and KNN determine the closest k training cases to your target case. KNN is a robust and simple classifier like Nave Bayes. SVM have extraordinary learning, and deliver enormous amount of data, and produce highly accurate results which are impractical with other conventional machine learning and data mining algorithms.

### 3. EXPERIMENTAL METHODOLOGY

In this paper We have considered two surely understood data mining algorithms. Every calculation are performed utilizing the R platform. For the trails, we used three real sensor datasets from the UCI data repository<sup>15,16,17</sup>. Datasets are gathered by utilizing sensors and accelerometers and are utilized to characterize human exercises, robot route, body stances and developments. Before re-enacting the figuring's, we pre-processed the datasets to make them reasonable for the classifiers. This is a preliminary analysis and hence, we have only used partial datasets. Our experimental methodology is depicted in Fig. 1.

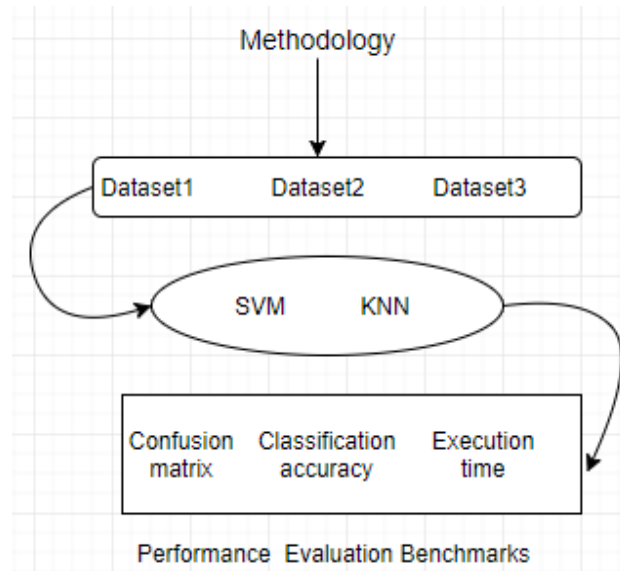


Fig 1

### 4. RESULTS AND ANALYSIS

We now show a preparatory investigation of the two settled data mining algorithms as specified in Fig. 1. The experiments have been done on the Aziz supercomputer. The Aziz supercomputer is Fujitsu made and can convey top execution of 230 teraflops. It has an aggregate of 11,904 centers in 496 hubs. Aziz was positioned number 360 in the June 2015 Top500 rivalry (<http://www.top500.org/>), at present, it is at number 491 (November 2015). For execution assessment of the algorithms, we have summarized the outcomes in form of confusion matrix (CM). With the help of CM, we are able to know the total number of instances rightly and wrongly classified. The class to which the wrongly classified instances belong to can also be identified with CM. The CMs of these two algorithms simulated on three different datasets<sup>15,16,17</sup> are given in Fig. 2 to Fig. 4. Moreover, Table 1, classification accuracy (CA) percentage and elapsed time is mentioned. Concerning Fig. 2 to Fig. 4 and Table 1, we take note of that SVM algorithms performed far superior than KNN. When considering the classification accuracy table 92.25% of average accuracy (AAC) obtained by SVM, it performs slightly better than 88.16% AAC of the KNN algorithm. SVM algorithms has higher classification accuracy than KNN. All the datasets<sup>15,16,17</sup> are multi-label as said in Fig.2 to Fig.4. Thus, SVM demonstrates its shortcomings towards multi-label data classification as compared to binary classification problem where its execution is extraordinary compared to other. The average elapsed time for SVM is higher compare to KNN. SVM performed superior to KNN with 4.09% higher AAC. The decision of k-neighbors and distance measure influences the CA of KNN.

#### 4.1 Execution Time–

KNN has slightly less processing times than SVM. KNN won't have higher CA if prejudicial data is in the variance. SVM has slow processing speed and uses a lot of system resources. KNN is lighter and has low execution times as specified in Table.1. For IoT, there can be problems where high CA does not make a difference much but rather processing time matters;

Actual/Predicted	Sitting	Sitting Down	Standing	Standing Up	Walking	Actual/Predicted	Sitting	Sitting Down	Standing	Standing Up	Walking
Sitting	50594	3	1	33	1	Sitting	46023	457	3885	258	8
Sitting Down	12	11523	139	103	50	Sitting Down	1078	6838	3084	174	653
Standing	2	16	47127	82	143	Standing	306	614	43852	146	2452
Standing Up	48	260	267	11806	34	Standing Up	1099	2733	5117	1658	1808
Walking		106	979	85	42220	Walking	588	2127	8820	2623	29232

a b  
Fig. 2. Confusion matrix of (a) SVM; (b) KNN of Dataset1;

Actual/Predicted	Brushing Teeth	Climbing stairs	Combing	Eating Meat	Pouring Water	Actual/Predicted	Brushing Teeth	Climbing stairs	Combing	Eating Meat	Pouring Water
Brushing Teeth	1991	15	73	43	105	Brushing Teeth	1762	121	151	114	79
Climbing Stairs	48	705	161	85	2	Climbing Stairs	17	836	39	59	50
Combing	162	92	1234	100	61	Combing	280	477	718	105	69
Eating Meat	5		6	4210	29	Eating Meat	3		224	3902	121
Pouring Water	26	1	18	327	518	Pouring Water	76	1	20	389	404

a b  
Fig. 3. Confusion matrix of (a) SVM; (b) KNN of Dataset2;

Actual/Predicted	Brushing Teeth	Climbing stairs	Combing	Eating Meat	Pouring Water	Actual/Predicted	Brushing Teeth	Climbing stairs	Combing	Eating Meat	Pouring Water
Brushing Teeth	1991	15	73	43	105	Brushing Teeth	1762	121	151	114	79
Climbing Stairs	48	705	161	85	2	Climbing Stairs	17	836	39	59	50
Combing	162	92	1234	100	61	Combing	280	477	718	105	69
Eating Meat	5		6	4210	29	Eating Meat	3		224	3902	121
Pouring Water	26	1	18	327	518	Pouring Water	76	1	20	389	404

a b  
Fig. 4. Confusion matrix of (a) SVM; (b) KNN of Dataset3;

Dataset <sup>15</sup>			Dataset <sup>16</sup>		Dataset <sup>17</sup>	
Algorithms	Accuracy	Elapsed Time	Accuracy	Elapsed Time	Accuracy	Elapsed Time
SVM	98.57%	2350.1	86.43%	5.2	91.75%	3.12
KNN	98.94%	450.6	86.88%	0.88	78.67%	1.32

Table 1. Classification accuracy in % and elapsed time in seconds for all the algorithms

## 5. CONCLUSION

The IoT pattern brings new arrangements of information those are primarily gathered from sensors and accelerometers gadgets. To recognize this concealed knowledge pattern from IoT data is a challenging and difficult task in data mining. Some scientists argue that a new set of data mining algorithms are expected to deal IoT data. In our work, we analyzed the appropriateness of a portion of the entrenched data mining algorithms like SVM and KNN. With our preliminary examination ,SVM can give moderately higher precision outcomes. SVM demonstrates its shortcomings towards multi-label data classification as compared to binary classification problem where its execution is extraordinary compared to other .SVM performed superior to KNN with 4.09% higher AAC. The choice of k-neighbors and distance measure impacts the CA of KNN. We intend to direct a definite report on bigger and diverse IoT datasets later on with some data mining algorithms like Linear discriminant analysis (LDA), Nave Bayes (NB) ,C50 C4.5 and ANNs.

## 6. REFERENCES

- [1] ITU. Recommendation ITU-T Y.2060 (06/2012), <http://handle.itu.int/11.1002/1000/11559>, 2012.
- [2] Atzori L, Iera A, Morabito G. The Internet of Things: A survey. *Computer Networks*. 2010;54(15):2787-2805.
- [3] Ma H. Internet of Things: Objectives and Scientific Challenges. *J Comput Sci Technol*. 2011;26(6):919-924.
- [4] Evans D, The Internet of Things: How the Next Evolution of the Internet Is Changing Everything, Cisco. April 2011. Retrieved 11 June 2016.
- [5] Chen F, Deng P, Wan J, Zhang D, Vasilakos A, Rong X. Data Mining for the Internet of Things: Literature Review and Challenges. *International Journal of Distributed Sensor Networks*. 2015;2015:1-14.
- [6] Ngai E, Xiu L, Chau D. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*. 2009;36(2):2592-2602.
- [7] S. Cuomo, P. D. Michele, A. Galletti and F. Piccialli, "A Cultural Heritage Case Study of Visitor Experiences Shared on a Social Network," 2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), Krakow, 2015, pp. 539-544.
- [8] Angelo Chianese, Fiammetta Marulli, Francesco Piccialli, Paolo Benedusi, Jai E. Jung, An associative engines based approach supporting collaborative analytics in the Internet of cultural things, *Future Generation Computer Systems*, 2016.
- [9] Wu X, Kumar V, Ross Quinlan J, Ghosh J, Yang Q, Motoda H et al. *Top*
- [10] algorithms in data mining. *Knowledge and Information Systems*. 2007;14(1):1-37. 10. BURGESS C. A Tutorial on Support Vector Machines for Pattern Recognition. Bell Laboratories and Lucent Technologies. 1998.
- [11] Burbidge RBuxton B. An Introduction to Support Vector Machines for Data Mining. UCL: Computer Science Dept. 2001.
- [12] Li W, Yi P, Wu Y, Pan L, Li J. A New Intrusion Detection System Based on KNN Classification Algorithm in Wireless Sensor Network. *Journal of Electrical and Computer Engineering*. 2014;2014:1-8.
- [13] Li D, Zhang B, Li C. A Feature Scaling based k-Nearest Neighbor Algorithm for Indoor Positioning Systems. *IEEE Internet of Things Journal*. 2015;;1-1.
- [14] Bourobou SYoo Y. User Activity Recognition in Smart Homes Using Pattern Clustering Applied to Temporal ANN Algorithm. *Sensors*. 2015;15(5):11953-11971.
- [15] Ugulino W, Cardador D, Vega K, Velloso E, Milidi? R, Fuks H. Wearable Computing: Accelerometers Data Classification of Body Postures and Movements. *Advances in Artificial Intelligence - SBIA 2012*. 2012;;52-61.
- [16] Lichman M. UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science. 2013.
- [17] Freire A, Barreto G, Veloso M, Varela A. Short-term memory mechanisms in neural network learning of robot navigation tasks: A case study. 2009 6th Latin American Robotics Symposium (LARS 2009). 2009