Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 5, No. 4, May : 2023

A COMPARATIVE STUDY OF VARIOUS CLASSIFICATION MACHINE LEARNING ALGORITHMS IN FOG COMPUTING: TASK SCHEDULING

Suraj Rajaram Nalawade, Research Scholar, Pacific Academy of Higher Education and Research University, Udaipur, (Rajasthan)
Prof. Ashok Kumar Jetawat, Professor, Pacific Academy of Higher Education and Research University, Udaipur (Rajasthan)

Abstract
The study focuses on addressing the issue of task scheduling and resource allocation in Fog computing environment. In this research work various machine learning techniques such as classification algorithms are being used for effective task scheduling and resource allocation. The two main objectives of the study are firstly the comparative analysis of machine learning classification algorithms based on various performance measures and secondly to Identify the most appropriate classification algorithm for task scheduling and resource allocation based on performance measure accuracy. Based on accuracy performance measure it can be concluded that at 25-fold cross validation the K-star and Logistic Regression are the most appropriate algorithms for task scheduling with accuracy level of 90% and 89%. Similarly, overall the most appropriate algorithm being identified is K-Star with overall accuracy level of 90.74%.

1. Introduction

Fog computing systems generate large volumes of data, therefore more apps and services are being developed. Machine learning (ML), a vital field, has made significant advances in robotics, neuromorphic computing, computer graphics, natural language processing (NLP), decision-making, and speech recognition. ML-based fog computing problem-solving research has been proposed. ML has been used more and more to improve fog computing applications and deliver fog services including resource management, security, latency reduction, energy savings, and traffic modelling. No work has examined the role of ML in fog computing, to our knowledge. Hence, our study illuminated fog computing ML functions. ML fog computing applications provide deep insights and smarter task replies. The study examines the newest ML strategies for task allocation and resource management, accuracy, and security in fog computing.

2. Background

The most important part of machine learning is the classification ML algorithm. As a supervised learning algorithm, it uses training data sets that already exist to build a model that can predict the categories of new data sets. By looking at the training data set, it can find rules for classifying data and guess what new data types will be. There are two parts to a classification algorithm: building the model and using the model. In the first step, it looks at the existing training data set, builds a model that fits it, and then comes up with some rules for how to classify things. In the second step, it sorts new data sets into groups based on the classification model it built in the first step. [1]

Major classification algorithms include random decision forests, the decision tree algorithm, the Bayes algorithm, the genetic algorithm, the artificial neural network algorithm, and the classification algorithm based on association rules. Classification algorithms are used a lot in wireless sensor networks, detecting network intrusions, looking at call logs, and figuring out how safe a bank is. In this paper, we introduce a classification algorithm based on association rules and improve and evaluate the Apriori algorithm. [2]

Apriori is a well-known algorithm for classifying things based on rules of association (CBA). It uses an iterative process to make many sets of items. There are two steps to the Apriori algorithm. First, it finds frequent item sets from a known transaction where the frequency is greater than or equal to the minimum support threshold by pruning and connecting frequent item sets. Then, it makes association rules based on the most common sets of items and the lowest level of confidence. [3,4,5]

3. Methodology
Figure below shows that the Fog computing system has three levels in a hierarchy network. IoT devices, which act as user interfaces and send requests from users through WiFi access points or the Internet, make up the front-end tier. IoT devices always have to work with limited resources, like CPU, memory, and a very complex application when they are running. The fog tier, which is made up of a group of near-end fog nodes, gets some of the requests from users and works on them. Most of the time, the fog tier is set up near IoT terminals, giving users limited access to computing resources. Users can directly use the computing resources in the fog tier, so there are no extra delays in communication. The cloud tier is made up of several servers, which are called cloud nodes. The remote cloud can offer a lot of computing resources, but it is far away from the users and takes a long time to send information. [6]

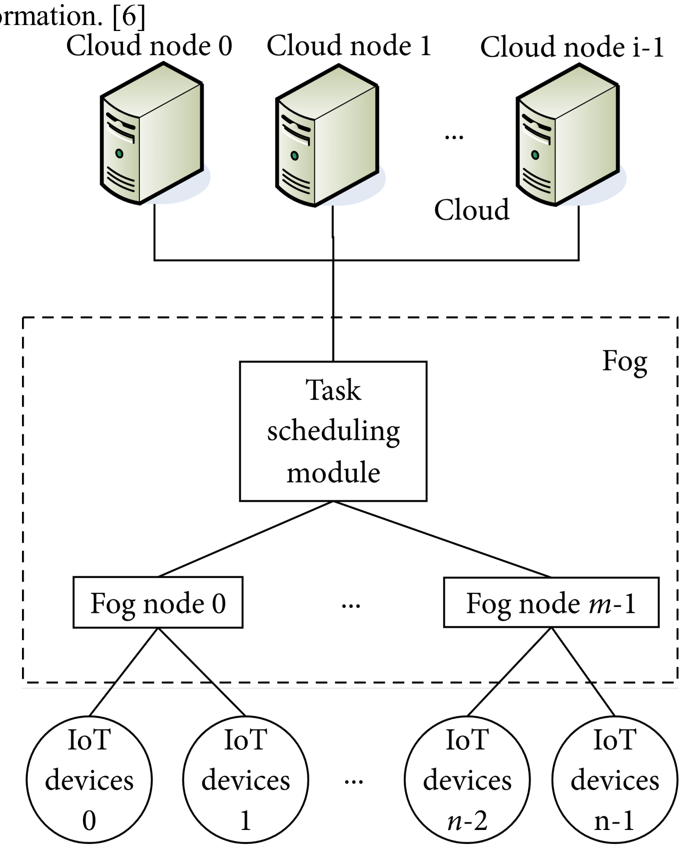

Figure 3.1: Fog-Cloud System
(Source: Lindong Liu et al., 2018).

Objectives:

1. Comparative analysis of machine learning classification algorithms based on various performance measures.
2. Identify the most appropriate classification algorithm for task scheduling and resource allocation.

Hypotheses:

H_0 1: There is no significant difference between various classification algorithms based on performance measure accuracy.

H_a 1: There is significant difference between various classification algorithms based on performance measure accuracy.
The proposed system is put into place with the help of IfogSim and CloudSim version 3.0. The WEKA tool to choose a machine learning classifier. The simulation is done on a computer with Windows 7 as its operating system.
Dataset Description:
This dataset is generated by taking 13 nodes out of which 10 are fog nodes and the rest 3 are cloud nodes. It contains 2 files based on the number of tasks starting from 40 which increases and goes up to 160.

4. Result and Discussion
Classification-based techniques were mostly employed for task scheduling in supervised learning. Logistic Regression, IBK, K-Star, and AdaBoostM1 were the task scheduling methods under consideration.

Comparative Analysis of Classification Algorithms:

Due to algorithm complexity and performance disparities, choosing the optimum classification method is difficult. This study quantified assessing and selecting algorithms by comparing four supervised machine learning classification algorithms on 2 datasets related to Fog Computing. The study indicated that no classifier performed better than others in terms of measure accuracy without advanced approaches when applied to distinct datasets. Second, as the data set grew, most algorithms fared better. Lastly, the research revealed that a classifier's success depends on the dataset, specifically the number of characteristics.

Experiment 1: Number of Task: 40 and Nodes: 4

Cross-validation – 10 folds:

Table 4.1: Classification Algorithm Performance Analysis
(Configuration Settings: 10 folds, Number of Task: 40 and Nodes: 4)

Performance Measure	Logistic Regression	K-Star	IBK	AdaBoostM1
Accuracy	0.88	0.91	0.58	0.60
Precision	0.887	0.927	0.658	-
Recall	0.880	0.910	0.585	0.600
F-Measure	0.871	0.903	0.517	-
ROC Area	0.988	1.000	0.602	1.000
Mean absolute error	0.0599	0.0488	0.2114	0.3188
Execution Time Model Building	15ms	10ms	10ms	30ms

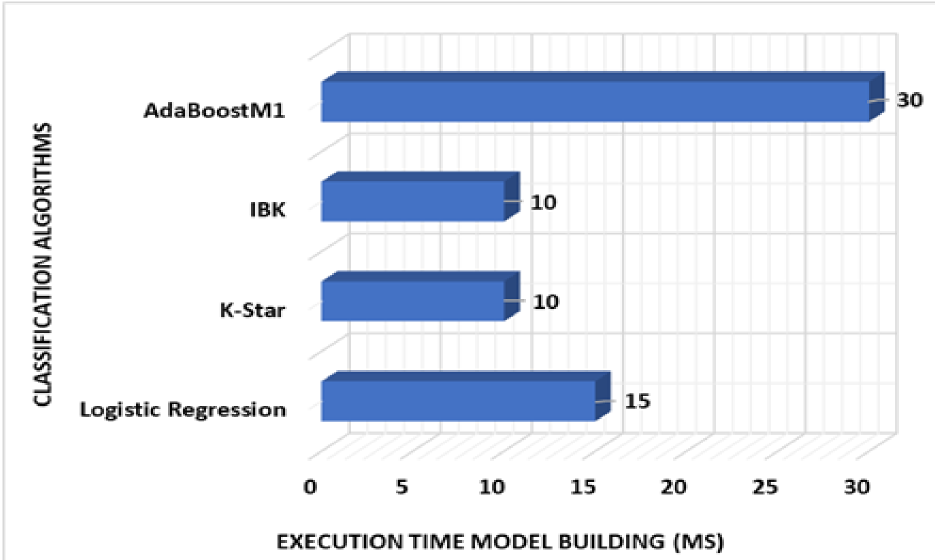

Figure 4.1: Average Execution Time (ms): 10 folds

The figure above confirms that based on average execution time the most appropriate algorithms for task scheduling are IBK, and K-Star with average execution time of 10 millisecond each which is quite less as compared with other algorithms. The next most appropriate algorithm is Logistic Regression with 15 millisecond execution time. The AdaBoostM1 classification algorithm has shown poor performance based on average execution time.

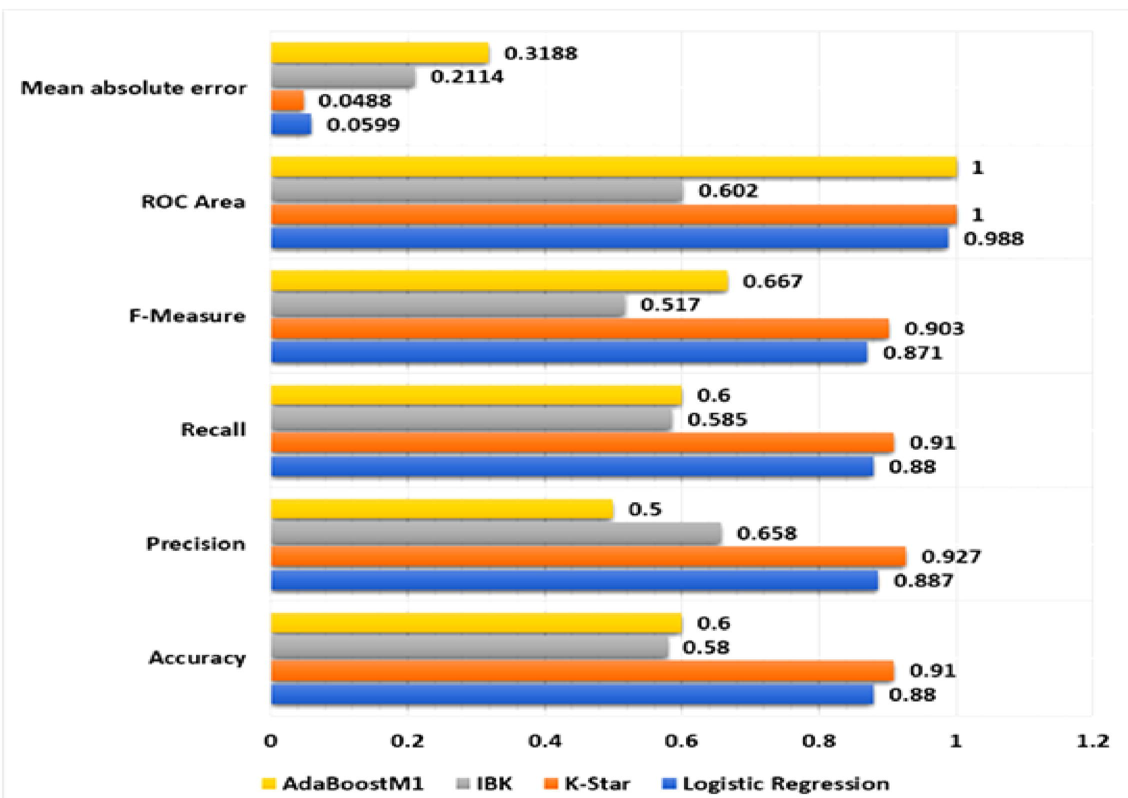

Figure 4.2: Performance measure & classification algorithms at 10-fold cross validation

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 5, No. 4, May : 2023

From the figure above it is clear that K-star and Logistic Regression are the most appropriate algorithms for task scheduling while considering the configuration setting; cross validation 10 folds.

Cross-validation – 25 folds:

Table 4.2: Classification Algorithm Performance Analysis
(25 folds, Number of Task: 40 and Nodes: 4)

Performance Measure	Logistic Regression	K-Star	IBK	AdaBoostM1
Accuracy	0.89	0.92	0.64	0.52
Precision	0.911	0.937	0.680	0.459
Recall	0.895	0.925	0.640	0.525
F-Measure	0.888	0.920	0.560	0.460
ROC Area	0.990	1.000	0.488	0.955
Mean absolute error	0.0526	0.044	0.1844	0.3186
Execution Time Model Building	15ms	10ms	10ms	35ms

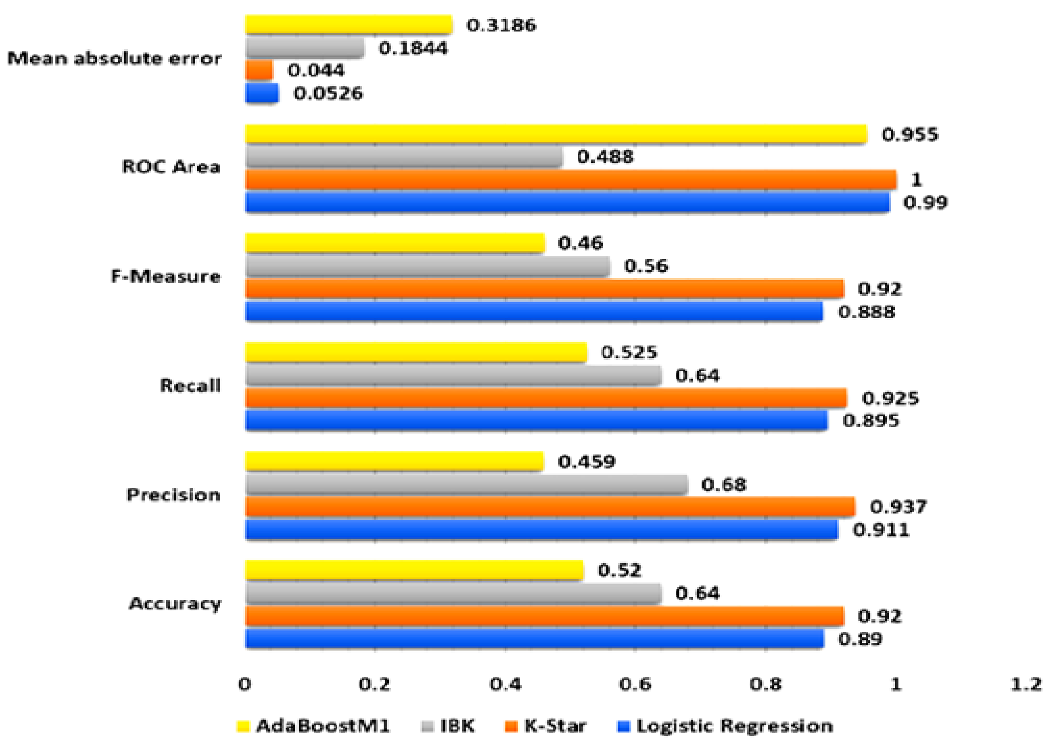

Figure 4.3: Performance measure & classification algorithms at 25-fold cross validation

From the figure above it is clear that Logistic Regression and K-star are the most appropriate algorithms for task scheduling while considering the configuration setting; cross validation 25 folds.

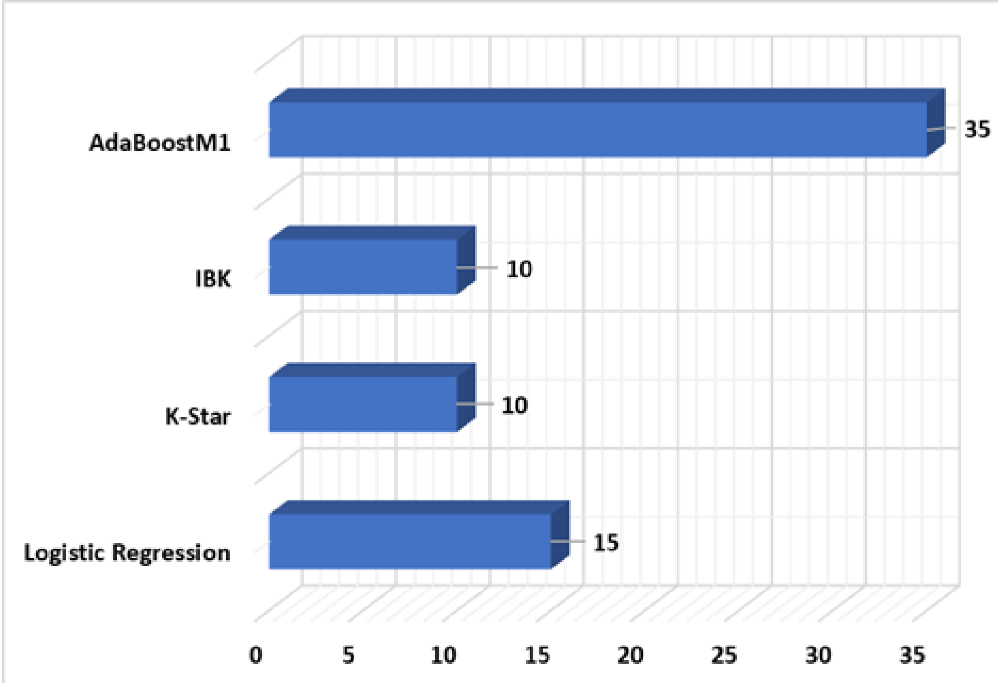

Figure 4.4: Average Execution Time (ms): 25 folds

IBK, K-Star, and Logistic Regression were determined to be the best suited algorithms when considering 40 tasks, 4 nodes, and cross validation 25 folds. This conclusion was reached based on the average execution time of each method as shown above in the figure.

Experiment 2: Number of Task: 160 and Nodes: 4

Cross-validation – 10 folds:

Table 4.3: Classification Algorithm Performance Analysis
(10 folds, 160 number of tasks and Nodes: 4)

Performance Measure	Logistic Regression	K-Star	IBK	AdaBoostM1
Accuracy	0.81	0.90	0.25	0.50
Precision	0.833	0.906	0.257	-
Recall	0.814	0.903	0.255	0.500
F-Measure	0.816	0.904	0.255	-
ROC Area	0.953	0.963	0.503	0.833
Mean absolute error	0.0931	0.0691	0.3727	0.25
Execution Time Model Building	1660ms	20ms	20ms	25ms

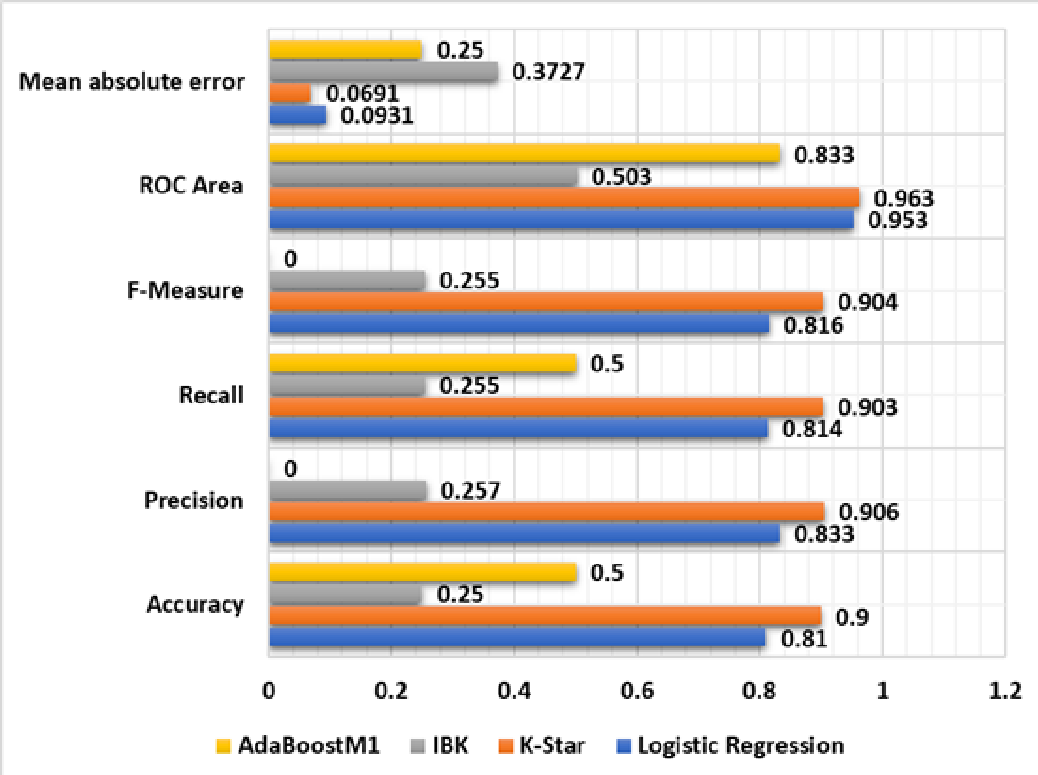

Figure 4.5: Performance measure & classification algorithms at 10-fold cross validation (Number of Task: 160 and Nodes: 4)

From the figure above it is clear that Logistic Regression and K-star are the most appropriate algorithms for task scheduling while considering the configuration setting; cross validation 10 folds and 160 number of tasks and 4 nodes.

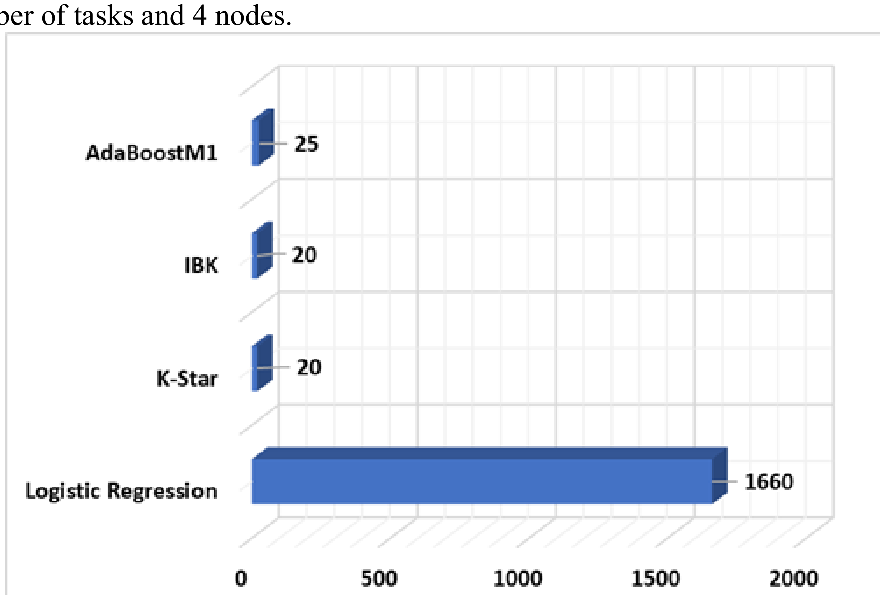

Figure 4.6: Average Execution Time (ms): 10 folds

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 5, No. 4, May : 2023
(Number of Task: 160 and Nodes: 4)

Based on average execution time the most appropriate algorithms are found to be IBK and K-Star while considering 160 number of tasks and 4 nodes and cross validation 10 folds.

Cross-validation – 25 folds:

Table 4.4: Classification Algorithm Performance Analysis
(25 folds, 160 number of tasks and Nodes: 4)

Performance Measure	Logistic Regression	K-Star	IBK	AdaBoostM1
Accuracy	0.89	0.90	0.25	0.47
Precision	0.899	0.907	0.250	0.472
Recall	0.892	0.905	0.250	0.477
F-Measure	0.893	0.906	0.250	0.449
ROC Area	0.976	0.961	0.498	0.801
Mean absolute error	0.0541	0.0706	0.375	0.2988
Execution Time Model Building	1860ms	20ms	20ms	25ms

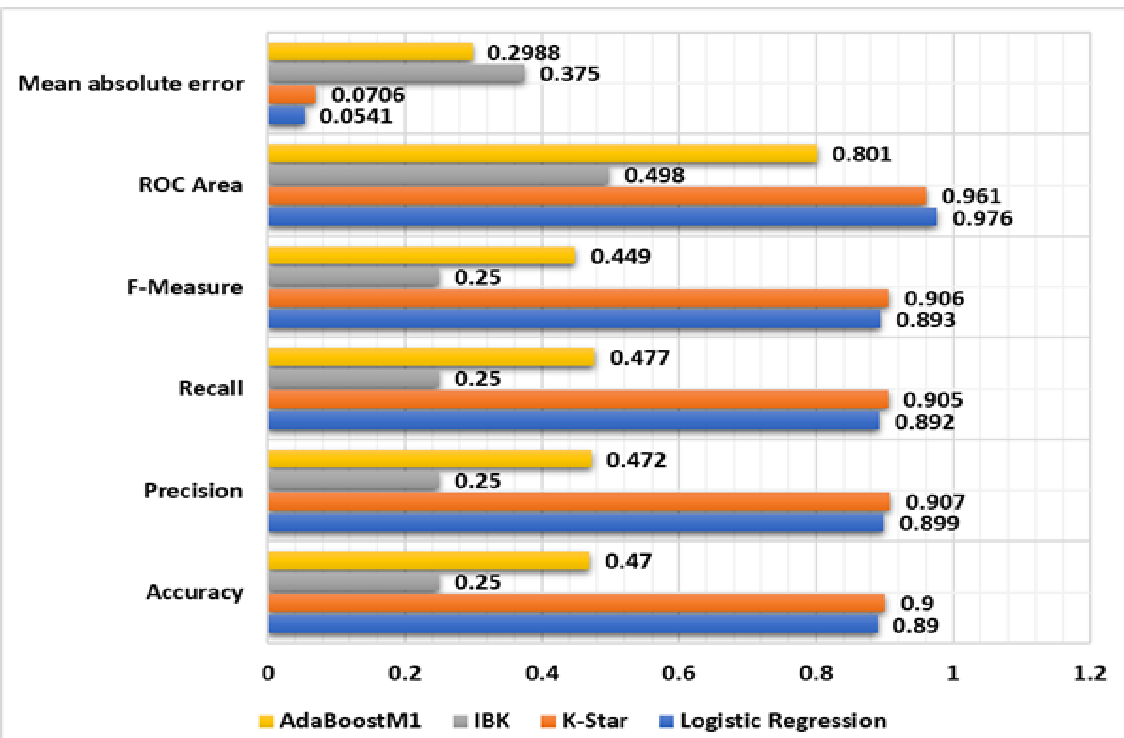

Figure 4.7: Performance measure & classification algorithms at 25-fold cross validation (Number of Task: 160 and Nodes: 4)

With the setup settings of cross validation 25 folds, 160 tasks, and 4 nodes, it is evident from the above figure that Logistic Regression and K-star are the best algorithms for task scheduling and resource allocation.

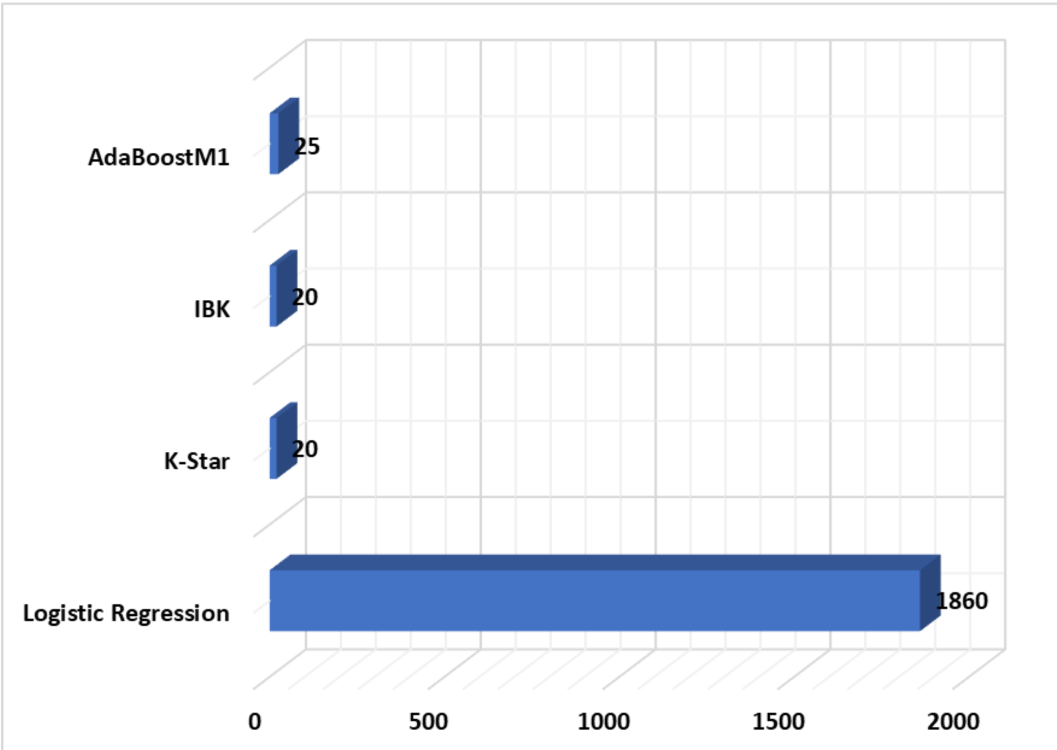

Figure 4.8: Average Execution Time (ms): 25 folds
(Number of Task: 160 and Nodes: 4)

K-Star and IBK are determined to be the best suitable algorithms based on average execution time when taking into account 160 tasks, 4 nodes, and cross validation 25 folds.

5. Conclusions:

Finally, it can be concluded that the null hypothesis (H_0 1) is being rejected which means there is significant difference between various classification algorithms based on performance measure accuracy as the accuracy level of K-Star, IBK, Logistic Regression etc. is different from each other. Based on overall accuracy measure it can be suggested that K-Star is the most appropriate classification algorithm with highest overall accuracy of 90.74% as compared to others. The second-best algorithm is found to be Logistic Regression with overall accuracy level of 86.74%.

References:

1. Lindong Liu, Deyu Qi, Naqin Zhou, Yilin Wu (2018). "A Task Scheduling Algorithm Based on Classification Mining in Fog Computing Environment", Wireless Communications and Mobile Computing, vol. 2018, Article ID 2102348, 11 pages, 2018. https://doi.org/10.1155/2018/2102348.
2. F. Saqib, A. Dutta, J. Plusquellic, P. Ortiz, and M. S. Pattichis, “Pipelined decision tree classification accelerator implementation in FPGA (DT-CAIF),” Institute of Electrical and Electronics Engineers. Transactions on Computers, vol. 64, no. 1, pp. 280–285, 2015.

3. R. Bruni and G. Bianchi, "Effective Classification Using a Small raining Set Based on iscretization and Statistical Analysis," IEEE Transactions On knowledge and data engineering, vol. 27, no. 9, pp. 2349–2361, 2015.
4. M. Xiao, Y. Yin, Y. Zhou, and S. Pan (2017), "Research on improvement of apriori algorithm based on marked transaction compression," in Proceedings of the 2nd IEEE Advanced Information Technology, Electronic and Automation Control Conference, (IAEAC '17), pp. 1067–1071, China, March 2017.
5. J. Yang, H. Huang, and X. Jin (2017), "Mining web access sequence with improved apriori algorithm," in Proceedings of the 20th IEEE International Conference on Computational Science and Engineering and 15th IEEE/IFIP International Conference on Embedded and Ubiquitous Computing, CSE and EUC 2017, pp. 780–784, China, July 2017.
6. S. Zhang, Z. Du, and J. T. L. Wang (2015), "New techniques for mining frequent patterns in unordered trees," IEEE Transactions on Cybernetics, vol. 45, no. 6, pp. 1113–1125.