

A Novel Ensemble Approach to Enhance the Performance of Web Server Logs Classification

Mohammed Hamed Ahmed Elhebir¹ and Ajith Abraham²

¹ Faculty of Mathematical and Computer Sciences, University of Gezira ,
P.O. Box 20, Wad Medani, Sudan
elhibr@uofg.edu.sd

² Machine Intelligence Research Labs (MIR Labs), Scientific Network for Innovation and
Research Excellence, P.O. Box 2259, WA, USA
ajith.abraham@ieee.org

Abstract: The World Wide Web (WWW) is growing in both the volume of traffic and the complexity of website, it has become very important to classify this web traffic and the usage of the web site according to predetermined attributes. Web Usage Mining (WUM) is the process of extracting knowledge from the accessed data by the web users. Classifying web users' sessions provides valuable information for web designers to respond to their individual needs in time. The main objective of this paper is to classify users' sessions. However, most of classification algorithms obtained good performance for specific problems, but they are not robust enough for all kinds of problems. Combination of multiple classifiers can be considered as a general solution method for pattern discovery. It has been shown that the combination of classifiers obtains better results compared to a single classifier provided that its components are independent or they have diverse outputs. This paper compares the accuracy of ensemble models, which take advantage of groups of learners to yield better results. The Base classifiers that have been used in this approach are: decision tree algorithm, k-Nearest Neighbor, Naive Bayesian and BayesNet. Stacking and Voting are used as Meta classifiers. The performance of our approach is measured and compared using Sudan University of Science and Technology (SUST) web log data with session based timing. Different comparative analysis and evaluation were done using various metrics, such as Error Rate, ROC curves, Confusion Matrix, F- measure and the Matthews correlation coefficient. The results show that these ensemble machine learning models using voting meta classifier can significantly improve users sessions classification. It can achieve high accuracy in comparison with the outcomes of the all base and meta classifiers proposed.

Keywords: Web Usage Mining, Base Classifiers, Meta Base Classifiers, Ensemble Methods, Voting.

I. Introduction

The World Wide Web (WWW) is rapidly emerging as an important communication means of information related to a wide range of topics (e.g., education, business, Government). It has created an environment of abundant consumer choices, where organizations must give importance to improve customer loyalty. The navigation patterns of users are generally gathered by the web servers and stored in server

access logs. Analysis of server access log data provides information to restructure a web site to increase effectiveness, better management of work group communication, and to target ads to specific users. Web usage mining involves with the application of data mining methods to discover user access patterns from web data, to better serve the needs of web-based applications. Three different tasks of usage mining are data pre-processing, pattern discovery and pattern analysis are extraction of hidden predictive information from large databases [1]. Pattern discovery uses statistical and machine-learning techniques to build models that predict the behavior of the data. One of the most pattern discovery techniques used to extract knowledge from preprocessed data is classification.

Conventionally an individual classifier, such as K-Nearest Neighbor (KNN), Decision Tree (J48), Naive Bayes (NB) or BayesNet (BN) is trained on web log data set. Depending on the distribution of the patterns, it is possible that not all the patterns are learned well by an individual classifier.

A classifier performs poorly on the test set under such scenarios. One of the most attractive topics in supervised machine learning is learning how to combine the predictions of multiple classifiers. This approach is known as ensembles of classifiers in the supervised learning area. The motivation for doing this derives from the opportunity to obtain higher prediction accuracy, while treating classifiers as black boxes, i.e. without considering the details of their functionality. Meta-learning is a process of learning from learners (classifiers); the inputs of the meta-learner are the outputs of the base-classifiers (the basic classifiers). The goal of meta-learning ensemble is to induce a meta-model that combines base-classifier predictions into a single prediction. In order to create such ensemble, both the base-classifier and the meta-learner (meta-classifier) need to be trained. Since the meta-classifier(s) training requires an already trained base-classifier, these must be trained first.

After the base-classifiers are trained, they are used to produce outputs (classifications), from which the Meta level dataset is made. This dataset will be used for training the

meta-classifier(s). In the prediction phase, when the ensemble is already trained, the base classifiers output their predictions to the meta-classifier(s) that combines them into a final prediction (classification). In this paper our experiments were conducted using SUST web log data set. Firstly; we considered and compared the performance of four algorithms namely J48, K-NN, NB and BN. Secondly; we carried out a thorough investigation comparing the performance of various base classifiers. The meta-classifiers used were: Stacking, and Voting under Test mode: 10-fold cross-validation. Thirdly; we used an ensemble method constructed based on meta-classifiers. The rest of this paper is organized as follows: Section 2 presents Classification Model; Section 3 describes the proposed methodology; Section 4 presents the experimental results and Section 5 gives the main conclusions of this study.

II. Classification

Given a training data set, the classification model was used to categorize the given training data set into attributes and the attributes were referred to as class. In our web log data time stamp, users, etc. were considered as attributes or class. Classification can be performed using different techniques. Our goal was to predict the target class based on our source data (web log data). Our model takes into consideration the category type of classification in which the target attribute has only two possible variations: forenoon and afternoon.

A. Base Classifiers

Base classifiers refer to individual classifiers used to construct the ensemble classifiers. J48, k-NN, NB and BN classifiers are some of the commonly used base classifiers. However, the proposed technique is a very general approach and its performance may further improve depending on the choice and/or the number of classifiers as well as the use of more complex features.

1) Decision Tree

Decision tree is one of the most popular approaches for both classification and predictions. It is the predictive machine-learning model that classifies the required information from the data. Each internal node of a tree is considered as attributes and branches between the nodes are possible values [2]. Building algorithms may initially build the tree and then prune it for more effective classification. With pruning technique, portions of the tree may be removed or combined to reduce the overall size of the tree. The time and space complexity of constructing a decision tree depends on the size of the data set, the number of attributes in the data set, and the shape of the resulting tree [3]. Decision tree classifier has limitations as it is computationally expensive because at each node, each candidate splitting field must be sorted before its best split can be found [4].

2) K-Nearest Neighbor

Nearest Neighbor (also known as Collaborative Filtering or Instance-based Learning) is a useful data mining technique that allows using the past data instances, with known output

values, to predict an unknown output value of a new data instance. Hence, at this point, this description should sound similar to both regression and classification. Many researchers have found that the k nearest neighbors (KNN) algorithm achieves very good performance in their experiments on different data sets [5]. The general principle is to find the k training samples to determine the k nearest neighbors based on a distance measure. Next, the majority of k nearest neighbors decides the category of the next instance.

3) Naive Bayes

A Naive Bayes (NB) classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong independence assumptions. It can handle an arbitrary number of independent variables whether continuous or categorical [6]. The final classification is done by calculating the posterior probability of the object by multiplying the prior probability and likelihood. Based on the posterior probability, it takes the decision. The performance of Naive Bayes depends on the reality of data set [7].

4) BayesNet

BayesNet (BN) is based on the Bayes' theorem. So, conditional probability on each node is calculated and formed a Bayesian Network. Bayesian Network is a directed acyclic graph. In BN, it is assumed that all attributes are nominal and there are no missing values. Different types of algorithms are used to estimate the conditional probability such as Genetic Search, Hill Climbing, Simulated Annealing, Tabu Search, Repeated Hill Climbing and K2[8]. The output of the BN can be visualized in terms of graph. Figure 1 shows the visualized graph of the BN for a SUST web data set. Visualize graph is formed by using the children attribute of the web data set. In this graph, each node represents the probability distribution table within it.

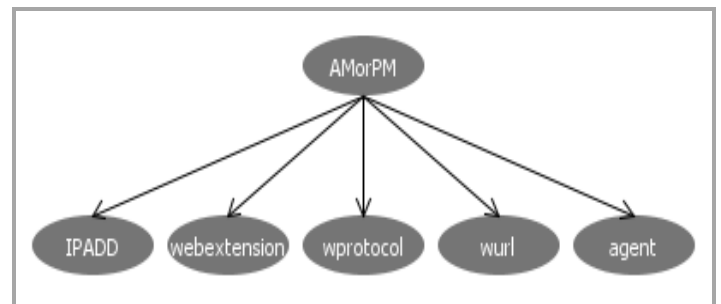


Figure 1. Visualize Graph of the BayesNet for a web data set

A new neural network architecture referred to as BAYESNET (Bayesian network) is capable of learning the probability density functions (PDFs) of individual pattern classes from a collection of learning samples, and designed for pattern classification based on the Bayesian decision rule. Bayes nets are often used as classifier to predict the probability of a target class label given features [9].

B. Meta Classifiers

Meta-learning means learning from the classifiers produced by the inducers and from the classifications of these classifiers on

training data. The following sections describe the most well-known Meta combining methods: Stacking and Voting.

1) *Stacking*

The first method that we employ for classifier combination is stacking, where the rule-based classifier is applied on the output produced by the based Classifier. Stacked generalization (or stacking) [10] is a different way of combining multiple models that introduces the concept of a Meta learner. Stacking procedure as follows:

1. Split the training set into two disjoint sets.
2. Train several base learners on the first part.
3. Test the base learners on the second part.
4. Using the predictions from 3) as the inputs, and the correct responses as the outputs, train a higher-level learner.

2) *Voting*

In the voting framework for combining classifiers, the predictions of the base-level classifiers are combined according to a static voting scheme, which does not change with training data set [11]. Voting does use a simple combination scheme of the base-classifier predictions to derive the final ensemble prediction. There are several types of voting schemes, which differ by the number of votes required for an ensemble prediction. Alternately, often a more powerful voting technique is to use a sum of each classifier’s probability distribution for the classes and predict the class with the highest value.

III. Methodology and Tool

A. *Data Set*

The data were collected from SUST web server log from 00:00:00 Nov 7, 2008 through 23:59:59 Aug 10, 2009. The total number of records was 23200 after removing unwanted data from the web log data.

B. *Classification Model*

In order to gauge the performance of ensemble techniques in the web usage mining, we set up classification accuracy tests to compare ensembles against base classifier. Here we first compare the performance of base and Meta classifiers on training set. Then select the best classifier, we combine those classifiers to generate ensembles using the best Meta classifier method. If ensemble techniques were useful in this domain, then we would expect a higher level of classification accuracy. If classification accuracy does not increase, then the added complexity and computational overhead of using an ensemble of classifiers will outweigh the benefit.

Classification was defined as the automated process of assigning a class label and mapping a user-based on the browsing history. The data were classified according to the predefined attributes. In this paper we consider four algorithms namely; J48, KNN, NB and BN. Combination of Multiple Classifiers (CMC) can be considered as a general solution method for the session classification. The inputs of the CMC are results of separate classifiers and output of the

CMC is their combined decisions [12,13]. Since the generalization ability of an ensemble could be significantly better than a single classifier, combinational methods have been a hot topic during the past years [14]. By combining classifiers, we intended to increase the performance of classification. There are several ways of combining classifiers. This work was done using voting majority method, which is the simplest way to find the best classifier as shown in Figure 2.

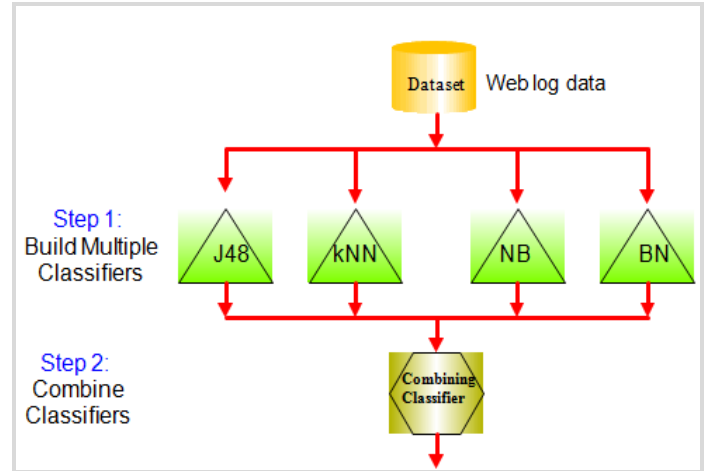


Figure 2. Majority Vote

C. *Performance Measures*

The performance of the classifiers is evaluated using the 10-fold cross-validation. In this paper we compared different classifiers, based on the measures of performance evaluation. According to Confusion matrix for two possible outcomes P (Positive) and N (Negative), as shown in Figure 3, many concepts often used:

		Actual		Total
		P	N	
Predicted	P	True Positive (TP)	False Positive (FP)	P
	N	False Negative (FN)	True Negative (TN)	N
Total		P	N	

Figure 3. Confusion matrix for two possible outcomes

i- **Precision:** Means the positive predictive value in information retrieved, which can be defined as:

$$Precision = \frac{TP}{TP + Fp} \quad (1)$$

ii- **Recall:** Proportion of actual positives which are predicted positive.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

iii- **Accuracy:** The Accuracy of a classifier on a given set is the percentage of test set tuples that are correctly classified by the classifier. Technically it can be defined as:

$$Accuracy = \frac{TP + TN}{P + N} \quad (3)$$

iv- **F-Measure**: It is another performance measure, needed because the accuracy determined using equation 3 may not be an adequate performance measure when the number of negative cases is much greater than the number of positive cases. F-Measure is defined in equation 4.

$$F - Measure = \frac{2(precision * recall)}{(precision + recall)} \quad (4)$$

v- **MCC**: The Matthews correlation coefficient is used in machine-learning as a measure of the quality of binary (two-class) classifications. MCC between the actual and predicted.

$$MCC = \frac{(TP * TN - FP * FN)}{\sqrt{(TP + FP)(TP + FN) + (TP + FP)(TN + FN)}}$$

vi- **ROC graphs**: It is another way besides Confusion matrices to evaluate the performance of classifiers. A ROC graph is a plot with the false positive rate on the X axis and the true positive rate on the Y axis. The point (0, 1) is the perfect classifier: it classifies all positive cases and negative cases correctly.

D. WEKA Data Mining Software

In this paper we used WEKA (Waikato Environment for Knowledge Analysis) software as the tool. WEKA includes several machine learning algorithms for data mining tasks. The algorithms can either be called from the users own Java code or be applied directly to the ready dataset. WEKA contains general-purpose environment tools for data preprocessing, regression, classification, association rules, clustering, feature selection and visualization [15].

IV. Experimental Results

A log file data with approximately 23242 entries was classified according to the predefined attributes, such as the pages visited by each user categorized into two sessions namely; forenoon (form 00:00:00 to 11:59:59) and afternoon (form 12:00:00 to 23:59:59). Figure 4 explains the number of entries classified into forenoon and afternoon.

We compared the performance of Decision Tree Classifier (J48), K-Nearest Neighbor Classifier (KNN), Naive Bayesian Classifier (NB), K-Nearest Neighbor Classifier (KNN) and BayesNet classifier (BN).

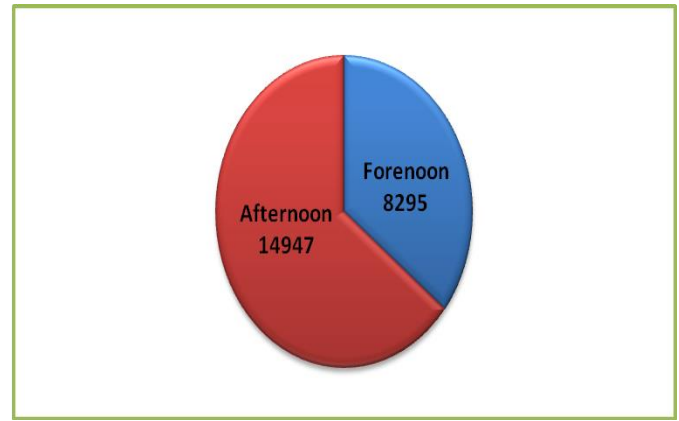


Figure 4. Users Count in each session

The results were displayed in form of tables. The comparison of accuracy, time and kappa statistic is presented in Table1. Table 2 shows the result based on recall, precision, f- measure, MCC, ROC Area and Error Rate. Meanwhile, Table 3 shows the mean absolute error (MAE) and the root relative squared error (RMSE). Figure 5 shows the obtained accuracy using different classification techniques. Figure 6 shows the performance metrics on balance-scale. The result inferred is that BayesNet classifier outperformed the others: base and meta classifiers with MAE = 0.3218 and 73.4274 % correctly classified. The Stacking meta classifiers had the same results with Voting, but it will take longer time to build model.

Table 4 shows the classifier performance using Ensemble Model of Meta Voting Classifiers combining with KNN, NB and BN classifiers. Voting combining two classifiers named 2 classifiers with vote. Voting combining three classifiers named 3 classifiers with vote. Table 5 shows the mean absolute errors (MAE) and root mean squared error (RMSE) of the ensemble of different classifiers.

Table 1. Comparison of different classifiers using Accuracy, time and kappa statistic for individual Base and Meta Classifiers. Best results are shown in **bold**.

Algorithm	Correctly Classified Instances (%) Value)	Incorrectly Classified Instances (%) Value)	Time Taken to build model (in seconds)	Kappa Statistic
J48	14947 (64.3103 %)	8295 (35.6897%)	5.14	0
KNN	16192 (69.667 %)	7050 (30.333 %)	0.04	0.2661
NB	16895 (72.6917 %)	6347 (27.3083 %)	0.09	0.4038
BN	17066 (73.4274 %)	6176 (26.5726 %)	0.04	0.4379
Stacking	14947 (64.3103 %)	8295 (35.6897%)	0.16	0
Voting	14947 (64.3103 %)	8295 (35.6897%)	0.01	0

Table 2. The classification performance of each Base and Meta Classifier in term of Recall, Precision, F- measure, MCC, Roc Area and Error Rate. Best results are shown in **bold**.

Parameters Algorithm	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Error Rate
J48	0.643	0.643	0.414	0.643	0.503	0.000	0.500	0.541	0.357
KNN	0.697	0.457	0.685	0.697	0.670	0.289	0.723	0.741	0.303
NB	0.727	0.324	0.726	0.727	0.727	0.404	0.799	0.810	0.273
BN	0.734	0.283	0.744	0.734	0.737	0.440	0.814	0.825	0.266
Meta Classifiers	0.643	0.643	0.414	0.643	0.503	0.000	0.500	0.541	0.357

Table 3. The mean absolute errors (MAE) and root mean squared error (RMSE) for each Base and Meta Classifier.

Base and Meta Classifier	MAE	RMSE
J48	0.459	0.4791
KNN	0.373	0.4432
NB	0.3373	0.4137
BN	0.3218	0.4106
Meta Classifiers	0.459	0.4791

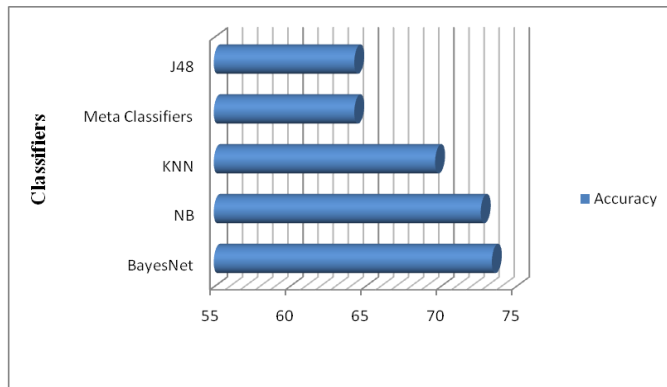


Figure 5. Comparison between Accuracy using different classification techniques.

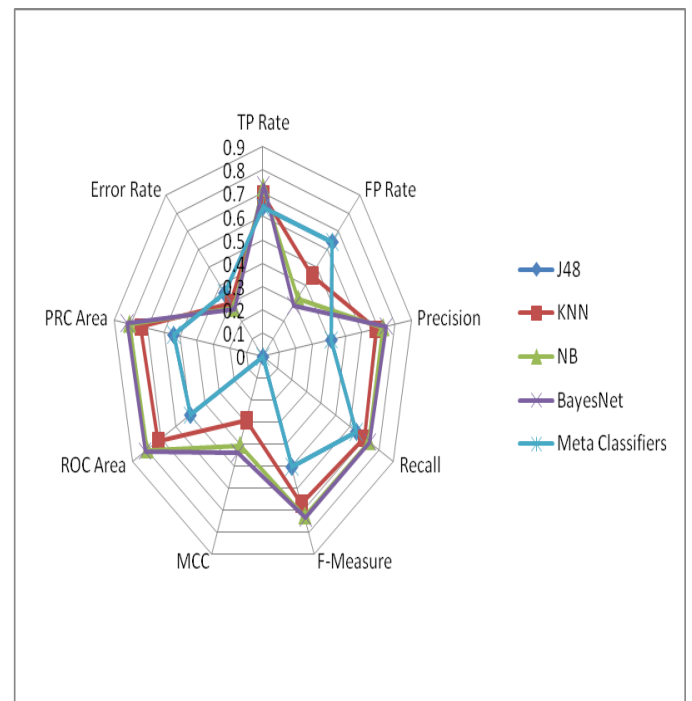


Figure 6. Performance metrics on balance-scale

Table 4. Comparison of the ensemble of different classifiers using the accuracy, time, kappa statistic. The best results are shown in **bold**.

Ensemble	Correctly Classified Instances (% Value)	Incorrectly Classified Instances (% Value)	Time Taken to build model (in seconds)	Kappa Statistic
KNN and NB with vote	16939 (72.881%)	6303 (27.119%)	0.04	0.3648
KNN and BN with Vote	17133 (73.7157%)	6109 (26.2843 %)	0.03	0.4217
NB and BN with vote	17036 (73.2983 %)	6206 (26.7017 %)	0.08	0.427
3 classifiers with vote	17114 (73.6339 %)	6128 (26.3661%)	0.07	0.4212

Table 5. MAE and RMSE of the ensemble of different classifiers.

Ensemble	MAE	RMSE
KNN and NB with vote	0.3552	0.415
KNN and BN with vote	0.3474	0.409
NB and BN with vote	0.3295	0.4109
3 classifiers with vote	0.344	0.4078

It was inferred from Tables 4 and 5, that the ensemble,3 classifiers with vote had the least RMES than ensemble 2 classifiers with vote, but will take longer time to build model. It was inferred from Table 1 and Table 4, that ensemble of **KNN and BN with Vote** had the best correctly classified than all individual Base and Meta Classifiers.

Table 6 shows the classification performance of each Ensemble model in term of recall, precision, f- measure, MCC and Roc Area for Forenoon and Afternoon class.

Table 7 shows the overall Ensembles, Base and Meta classifiers performance ranked by accuracy and error rate. It was inferred from Table 7 that ensemble KNN and BN with Vote classifier had the highest accuracy. The Base classifiers J48 and Meta classifiers had the lowest accuracy and greater error rate.

Table 6. The classification performance of each Ensemble model in term of Recall, Precision, F- measure , MCC and Roc Area for Forenoon and Afternoon class.

Parameters Ensemble	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
KNN and NB with Vote	0.463	0.124	0.675	0.463	0.549	0.378	0.798	0.690	Forenoon
	0.876	0.537	0.746	0.876	0.806	0.378	0.798	0.873	Afternoon
KNN and BN with Vote	0.610	0.192	0.638	0.610	0.623	0.422	0.811	0.705	Forenoon
	0.808	0.390	0.789	0.808	0.798	0.422	0.811	0.883	Afternoon
NB and BN with Vote	0.660	0.227	0.618	0.660	0.638	0.428	0.808	0.706	Forenoon
	0.773	0.340	0.804	0.773	0.788	0.428	0.808	0.882	Afternoon
3 classifiers with Vote	0.613	0.195	0.635	0.613	0.624	0.421	0.812	0.707	Forenoon
	0.805	0.387	0.789	0.805	0.797	0.421	0.812	0.885	Afternoon

Table7. Overall Ensembles, Base and Meta classifiers performance ranked by: accuracy and error rate.

Models	Accuracy	Error Rate
KNN and BN with vote	73.7157	0.263
3 classifiers with vote	73.6339	0.264
BN	73.4274	0.266
NB and BN with vote	73.2983	0.267
KNN and NB with vote	72.881	0.271
NB	72.6917	0.273
KNN	69.667	0.303
Meta Classifiers	64.3103	0.357
J48	64.3103	0.357

V. Discussions

In this work, we evaluated the performance in terms of classification accuracy of J48, KNN, NB, BN, Stacking and Vote Meta classifiers using various accuracy measures on log file dataset like TP rate, FP rate, Precision, Recall, F-measure and ROC Area.

- It was observed from results that an error rate of KNN and BN with Vote classifier was the lowest i.e. 0.263 and it will take shorter time to build model (0.03 seconds) in comparison with the others classifier, which was the most desirable.
- Accuracy of **KNN and BN with Vote** classifier was the highest i.e. 73.7157% in comparison with the others classifier, which was highly required. This investigation suggests that, the **KNN and BN with Vote** classifier is the optimum ensemble since it gives more classification accuracy for class session in web log file dataset having two values forenoon and afternoon.

- J48 was slightly bad algorithm. Thus we found that J48 was bad algorithm in most of performance measures.
- **KNN and BN with Vote** classifier had the highest accuracy, followed by the **three** classifiers together with **Voting**, followed by **BN**, followed by **NB**, followed by **NB and BN with voting**, followed by **KNN and NB with voting**, followed by **NB**, followed by **KNN**, followed by **Meta Classifiers**, followed by **J48**.

VI. Conclusions

Classification techniques arrange information in various classes depending on predefined attributes. There are different methods used to classify users' session. One of these is to classify them into "forenoon" and "afternoon". Performance evaluation between the classifiers was calculated. The result shows that ensemble learning-techniques can increase classification accuracy in the domain of web usage mining. The ensemble KNN and BN with vote classifier typically had the highest classification accuracy for SUST web log file dataset having two values "forenoon" and "afternoon".

References

- [1] Arvind Kumar Sharma, Dr. P.C. Gupta, "Exploration of Efficient Methodologies for the Improvement in Web Mining Techniques: A Survey", *International Journal of Research in IT & Management*, Vol 1, Issue 3, pp.85-95, July 2011.
- [2] Jameela, A., and P. Revathy. "COMPARISON OF DECISION AND RANDOM TREE ALGORITHMS ON A WEB LOG DATA FOR FINDING FREQUENT PATTERNS.", *International Journal of Research in Engineering and Technology*, Volume: 03 Special Issue: 07, pp. 155 -161, May-2014.
- [3] Tani, Fauzia Yasmeen, Dewan Md Farid, and Mohammad Zahidur Rahman. "Ensemble of Decision Tree Classifiers for Mining Web Data Streams." *International Journal of Applied Information Systems*, Volume 1– No.2, pp.30 – 36, January 2012.
- [4] Supreet Dhillon, and Kamaljit Kaur. "Comparative Study of Classification Algorithms for Web Usage Mining", *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 4, Issue 7, pp. 137-140, July 2014.
- [5] Baoli, Li, Chen Yuzhong, and Yu Shiwen. "A comparative study on automatic categorization methods for Chinese search engine." *Proceedings of the Eighth Joint International Computer Conference, Hangzhou: Zhejiang University Press*, pp. 117-120, 2002.
- [6] D. K. Tiwary, "A Comparative Study of Classification Algorithms for credit card approval using WEKA," *GALAXY International Interdisciplinary Research Journal*, vol. 2, no. 3, pp. 165–174, 2014.
- [7] S. K. Sarangi and V. Jaglan, "Performance Comparison of Machine Learning Algorithms on Integration of Clustering and Classification Techniques," *International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)*, pp. 251–257, 2013.
- [8] V. Vaithyanathan, K. Rajeswari, et al. "Comparison of different classification techniques using different datasets." *International Journal of Advances in Engineering & Technology*, Vol. 6, Issue 2, pp. 764-768, 2013.
- [9] T. Roos, H. Wettig, P. Grünwald, P. Myllymäki, and H. Tirri. On discriminative bayesian network classifiers and logistic regression. *Mach. Learn.*, 59(3):267–296, 2005.
- [10] Wolpert, D. (1992). Stacked generalization. *Neural Networks*, 5:2, 241–260.
- [11] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.
- [12] B. Minaei-Bidgoli, G. Kortemeyer and W.F. Punch, *Optimizing Classification Ensembles via a Genetic Algorithm for a Web-based Educational System*, (SSPR /SPR 2004), *Lecture Notes in Computer Science (LNCS)*, Volume 3138, Springer-Verlag, ISBN: 3-540-22570-6, pp. 397-406, 2004.
- [13] A. Saberi., M. Vahidi, B. Minaei-Bidgoli, *Learn to Detect Phishing Scams Using Learning and Ensemble Methods*, *IEEE/WIC/ACM International Conference on Intelligent Agent Technology, Workshops (IAT 07)*, pp. 311-314, Silicon Valley, USA, November 2-5, 2007.
- [15] T.G. Dietterich, Ensemble learning, in *The Handbook of Brain Theory and Neural Networks*, 2nd edition, M.A. Arbib, Ed. Cambridge, MA: MIT Press, 2002.
- [16] David, Satish Kumar, Amr TM Saeb, and Khalid Al Rubeaan. "Comparative Analysis of Data Mining Tools and Classification Techniques using WEKA in Medical Bioinformatics." *Computer Engineering and Intelligent Systems*, Vol.4, No.13, pp. 28-38, 2013.