RESEARCH ARTICLE OPEN ACCESS

Performance Evaluation of feature selection methods for Mobile devices

A. Chaudhary[1], S. Kolhe[2], Rajkamal[3]
[1,3]Devi Ahilya Vishwa Vidyalaya, Indore, Madhya Pradesh , India
[2]Directorate of Soybean Research, Indore, India

Abstract
Machine Learning deals with programming computers that learn from experience. The field of Machine learning is a popular research area in Computer Science. These techniques are helpful in different fields of Computer Science, Mobile Computing, Bioinformatics, Digital Forensic, Agriculture and Text Classification. Machine learning classification algorithms are used in Pattern Recognition, Text Categorization, Mobile message classification, Mobile Image tagging applications, Mobile music interaction, Mobile learning. An optimized Naïve Bayes classifier is used for this work .In this work performance evaluation of three feature selection methods with optimized Naïve Bayes is performed on mobile device. Correlation based method, Gain Ratio method and Information Gain method methods were used in this work. The three methods are evaluated for performance measures of Accuracy, Precision, True Positive Rate, F- Measure, Recall, Mathew's Correlation Coefficient and Receiver Operating Characteristics.

I. INTRODUCTION

The stream of Machine Learning field has evolved from the field of *Artificial Intelligence*, in which machines are trained to mimic intelligent skills of humans. Machine learning is programming computers to optimize a performance criterion using example data or past experience [1]. This field incorporates theory of statistics in designing mathematical models as the basic aim is making inference from a data sample. Applications of machine learning include pattern recognition, face recognition, medical diagnosis, crop disease detection, speech recognition, mobile text categorization, mobile music categorization, mobile learning, mobile automatic image tagging, internet traffic flow monitoring, mobile user location identification, Biometrics, digital forensics. Basically there are three types of machine learning namely supervised, unsupervised and reinforcement learning. In supervised learning a supervisor or guide is available to guide the entire learning process. Machine learns from different examples provided as training dataset. Regression and classification problems come under this category. In case of unsupervised learning there is no supervisor available and aim here is to find regularities in the input data. In this case there is a structure to the input space such that certain patterns occur more often than others and aim is to draw inference on the basis of input patterns. Data Mining problems come under this category. Reinforcement learning is when learning is associated with some scalar reward or punishment. In this type of learning aim is to maximize the scalar rewards. Problems related to game playing, Robot path planning, traffic control systems can be designed as reinforcement learning problems.

II. LITERATURE SURVEY

Machine learning techniques were used to represent the behavior of children suffering with autism communicating with a humanoid robot; a comparison was made between a static model and a dynamic model using hand-coded features in [2]. A higher accuracy (above 80%) was achieved in predicting child vocalizations. Additionally directions for future approaches for modeling the behavior of children with autism were suggested in [2]. Machine learning techniques of Bayesian learning and Neural Networks were used for modeling the response time of service-oriented systems in [3]. The results showed that Bayesian learning offered better accuracy but have less sensitivity to limited data set size [3]. Bayesian models were suggested more suitable for changing environments and need frequent response time model reconstructions. Neural Networks were suggested to achieve faster model evaluation time and support management routines which demand intensive response time predictions. Learning techniques were also useful in the prediction of condition numbers of sparse matrices in [4]. Condition number of a matrix offered a significant measure in numerical analysis and linear algebra [4]. Support Vector Machines and Modified K Nearest Neighbor techniques were used to estimate the condition number of a given sparse matrix. The experimental results proved that modified K Nearest Neighbor performed much better than Support Vector Machines on the chosen data set. Support Vector Machines, Artificial Neural Networks, K Nearest Neighbor were applied for estimating Function Points of software in [5]. The Experiments performed in [5] showed that Artificial Neural Networks and Support Vector Machines are efficient models for function point prediction. A framework for evaluating Machine Learning based approaches for Call Admission Control was presented in [6]. A comparison of performance of two major machine learning approaches Neural Networks and Bayesian Networks for QoS prediction was done in [6]. The training data size for Bayesian Network model was relatively smaller than the Neural Network model. A comparison of machine learning methods Decision trees, Flexible Neural Tree and Particle Swarm Optimization for intrusion detection on network traffic was performed in [7]. The reported results showed that Decision Tree had better accuracy in classification than other methods. Machine learning techniques of random forest and lasso regularization were used to predict software anomalies in [8]. Machine learning techniques are also helpful in transportation. Support Vector Machines were used for the short-term prediction of travel time in [9]. A Comparison between Artificial Neural Networks and Support Vector Machines was performed and it was concluded that Support Vector Machines performs better for the short-term prediction of travel time that was when the amount of training data is less, or when the training data had more variations as compared to the testing data [9]. It was also found that the influence of the amount of training data used was more on the Artificial Neural Networks method than on the Support Vector Machines method. Bayesian learning model and Multilayer Perceptron were applied to dynamic video adapting in [10]. The results showed that Multilayer Perceptron was superior to Bayesian learning model. The estimated bit rate error for Multilayer Perceptron was of 0.33% whereas Bayesian learning model presented 11%. Reinforcement learning, Swarm intelligence, Heuristics based learning techniques were applied to Wireless Sensor Networks and Mobile Ad-hoc Networks in [11]. Black-box optimization algorithms based on machine learning techniques such as Genetic algorithms and particle swarm optimization were used in the design of Multi-Hop Broadcast Protocols for VANET in [12]. New traffic identification techniques for monitoring the performance of eHealth systems for home and mobile users employing machine learning based methods was suggested in [13]. Different techniques for traffic identification that combined the results from multiple machine learning classifiers were suggested in [13]. A flexible Machine Learning approach based on non-parametric Gaussian Process regression for learning user-specific touch input models to increase touch accuracy on mobile devices was suggested in [14].

III. FEATURE SELECTION METHODS

In order to reduce the complexity of algorithms many machine learning feature selection methods are used. Feature selection methods outcast important features and eliminate unimportant, redundant or noise features to decrease the dimensionality of the feature subspace. It improves efficiency, classification accuracy of the models designed by learning algorithms. There are two methods for reducing dimensionality namely feature selection and feature extraction. The aim of feature selection algorithms is to find features that give best classification and discard the other irrelevant features. Feature extraction methods find a new set of dimensions that are combinations of original dimensions. There are many feature selection methods.

Correlation Based Feature Selection method

This feature selection method uses heuristic for evaluating the value or merit of a subset of features.

This heuristic works upon the usefulness of individual features for anticipating the class label at the same time the level of inter correlation among them [15]. This method uses hypothesis ‘‘A good feature subset is one that contains features highly correlated with (predictive of) the class, yet uncorrelated with (not predictive of) each other’’. In this method, unimportant features are ignored when their correlation with the class is weak. Redundant features are also not desirable once they are highly correlated with one or more of the other features. The subset evaluation function used in this method is given by

$$Mer_s = \frac{k\overline{r_{cf}}}{\sqrt{k+(k-1)\overline{r_{ff}}}} \quad (1)$$

where k represents the number of features within a given subset s, r_{cf} is the feature–class correlation mean, r_{ff} is the average feature–feature inter-correlation and Mer_s represents the merit of S.

Information Gain

Information Gain is an important measure used for ranking features. It works upon the strength of information gained for classification as long as the feature is considered. This method of feature selection measures the quantity of impurity in a group. A common measure of identifying impurity in a group is Entropy. This method is popular method for machine learning based classification. Information Gain is computed by the feature's influence on decreasing overall entropy [16,19]. It calculates the change in information entropy given a Feature as

$$Information - Gain(C, Feature_i) = H(C) - H(C|Feature_i) \quad (2)$$

where

$$H(C) = -\sum_{c \epsilon C} p(c) log_2\, p(c) \quad (3)$$

$$H(C|feature_i) = -\sum_{f \epsilon Feature} p(f) \sum_{c \epsilon C} p(c|a) \log_2 p\,(c|a) \quad (4)$$

using this simple metric, an ordered set of features can be obtained depending upon the information they provide for classification.

Gain Ratio
Gain Ratio is somewhat similar to Information Gain. This metric measures the gain in information for classification related to entropy of the given feature. In other words it evaluates the merits of an attribute by calculating the gain ratio with corresponding to the class.

$$Gain\ Ratio(C, Feature_i) = \frac{H(C) - H(C|Feature_i)}{H(Feature_i)} \quad (5)$$

where H(C) represents the entropy of class C, $H(C|Feature_i)$ represents entropy of class C given $Feature_i$ and $H(Feature_i)$ is the entropy for $Feature_i$.

IV PERFORMANCE MEASURES FOR MACHINE LEARNING ALGORITHMS

The performance of any Machine learning algorithm is determined by some simple measures. Any classification is correct if it can be judged by calculating the number of correctly identified class samples
(true positives), the number of correctly identified samples that are not members of the class (true negatives) and samples that either were incorrectly allocated to the class (false positives) or that were not identified as class samples (false negatives) [17]. These four components makeup a confusion matrix for binary classification

$$Confusion\ Matrix = \begin{bmatrix} t_P & f_n \\ f_p & t_n \end{bmatrix} \quad (6)$$

Accuracy: It is also termed as classification accuracy. It is the simplest measure in order to evaluate a classifier. It is defined as the degree of correct predictions of a model. It is also measured in percentage.

Precision: It is the number of accurately classified positive instances with respect to the number of instances that exist in the system as positive. Precision for true positive (t_p) and false positive (f_p) is given by

$$Precision = \frac{t_p}{t_p + f_p} \quad (7)$$

Recall: It is the number of accurately classified positive instances divided by the number of positive instances in the data. Recall is a measure of sensitivity. Precision for true positive (t_p) and false positive (f_p) is given by

$$Recall = \frac{t_p}{t_p + f_n} \quad (8)$$

F-Measure: It is also termed as FScore. It is a metric for accuracy of test. It makes use of both precision and recall to compute this score. F-Measure is calculated by

$$FMeasure = 2.\frac{prec\ .\ rec}{prec + rec} \quad (9)$$

where $prec$ represents precision and rec represents recall

Mathew's Correlation coefficient (MCC): It is used in machine learning for measuring the quality of binary classifications. It considers true and false positives and negatives. It is regarded as an optimal measure that can be used even for classes of varying sizes. MCC is in turn a coefficient of correlation between the observed and predicted binary classifications.MCC is computed using t_p, t_n, f_p, f_n as

$$MCC = t_p.\frac{t_p.\ t_n - f_p.f_n}{\sqrt{(t_p + f_p)(t_p + f_n)(t_n + f_p)(t_n + f_n)}} \quad (10)$$

Receiver Operating Characteristics: This is a curve known as ROC curve or ROC space. In order to trace this curve only the true positive rate and false positive rate are required. It represents tradeoffs between true positives and false positives.

V. CLASSIFICATION ALGORITHM

Naïve Bayes

Machine learning classification problems deal with allocating a class or category to an instance or observation [19]. The classifier classifies on the basis of a training set that consists of many correctly classified examples. These are known instance and class pairs. Every instance is related by a number of features and the classifier maps this set of features into a particular class. Naïve Bayes is a simple classification algorithm. For a given observation it estimates the probability of every class and chooses the class with maximum probability [19]. It assumes that all features are independent. Practically the assumption that all features are independent may be true or false. It gives performance comparable to other methods. Naive Bayes classifier is basically a Bayesian learning model where the class has no parents and every attribute has the class as its sole parent. Hence Bayesian models have principles from graph theory, probability theory, computer science and statistics. An optimized Naïve Bayes classifier that performed better than Naïve Bayes was used in this work with Indian crop variety dataset [20]. This dataset contained 27 features with 1200 instances. There were 75 crop varieties of India. Feature selection methods were applied to rank crop features. Feature extraction method was applied for extracting relevant features and discarding irrelevant features. Feature space was divided into three feature subsets. Every time six new features were added to the selected features. Ten runs were performed by including the features from best ranked feature subset, average ranked feature subset and below average ranked feature subset. Experiments were performed using [18] on mobile device in which several new metrics were added that did not exist in the current distribution. Experiments were validated using cross validation of 10 folds in which the entire dataset was fragmented into 10 equal sized subsets and classifier was trained on 9 subsets and tested on remaining subset.

VI. EXPERIMENTAL RESULTS

The results of optimized Naïve Bayes classifier for the three feature selection schemes namely Correlation based, Gain Ratio based and Information Gain based methods are presented below for True Positive rate, Precision, F-Measure, Recall, MCC, ROC and Classification accuracy.

TABLE I: TRUE POSITIVE RATE OF NAÏVE BAYES FOR THREE FEATURE SELECTION METHODS

Features	True Positive Rate with Correlation method	True Positive Rate with Gain Ratio method	True Positive Rate with Information Gain method
3	0.5	0.667	0.528
9	0.5	0.583	0.528
15	0.694	0.639	0.556
21	0.722	0.722	0.528
27	0.722	0.722	0.722

TABLE II: PRECISION OF NAÏVE BAYES FOR THREE FEATURE SELECTION METHODS

Features	Precision with Correlation method	Precision with Gain Ratio method	Precision with Information Gain method
3	0.478	0.664	0.497
9	0.483	0.592	0.499
15	0.721	0.672	0.555
21	0.741	0.722	0.526
27	0.72	0.72	0.72

TABLE III: RECALL OF NAÏVE BAYES FOR THREE FEATURE SELECTION METHODS

Features	Recall with Correlation method	Recall with Gain Ratio method	Recall with Information Gain method
3	0.5	0.667	0.528
9	0.5	0.583	0.528
15	0.694	0.639	0.556
21	0.722	0.722	0.528
27	0.722	0.722	0.722

TABLE IV: F-MEASURE OF NAÏVE BAYES FOR THREE FEATURE SELECTION METHODS

Features	F-Measure with Correlation method	F-Measure with Gain Ratio method	F-Measure with Information Gain method
3	0.48	0.664	0.5
9	0.489	0.586	0.512
15	0.694	0.643	0.553
21	0.727	0.721	0.525
27	0.72	0.72	0.72

TABLE V: MCC OF NAÏVE BAYES FOR THREE FEATURE SELECTION METHODS

Features	MCC with Correlation method	MCC with Gain Ratio method	MCC with Information Gain method
3	0.324	0.555	0.357
9	0.33	0.451	0.363
15	0.606	0.535	0.413
21	0.64	0.631	0.377
27	0.631	0.631	0.631

TABLE VI: ROC OF NAÏVE BAYES FOR THREE FEATURE SELECTION METHODS

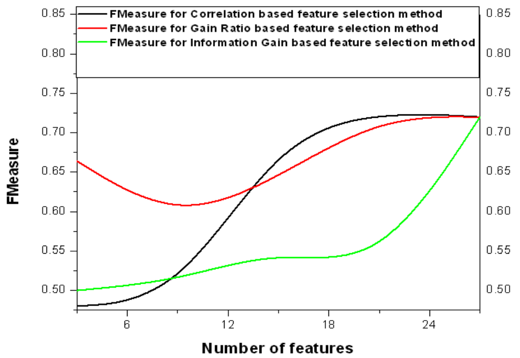
Features	ROC with Correlation method	ROC with Gain Ratio method	ROC with Information Gain method
3	0.751	0.863	0.767
9	0.737	0.853	0.756
15	0.848	0.878	0.778
21	0.89	0.896	0.799
27	0.92	0.92	0.92

TABLE VII: CLASSIFICATION ACCURACY OF OPTIMIZED NAÏVE BAYES FOR THREE FEATURE SELECTION METHODS

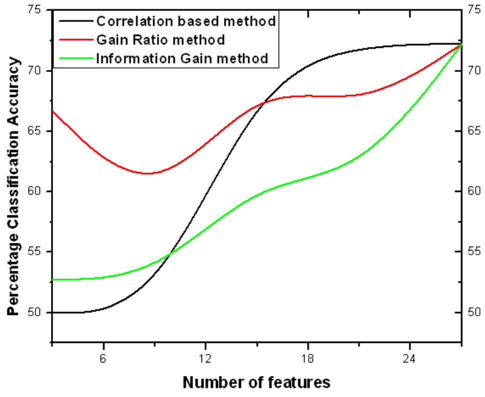
Features	Percentage Classification Accuracy with Correlation method	Percentage Classification Accuracy with Gain Ratio method	Percentage Classification Accuracy with Information Gain method
3	50	66.66	52.77
9	50	58.33	52.77
15	69.44	69.44	61.11
21	72.22	66.66	61.11
27	72.22	72.22	72.22

It is clear from the above tables that Gain Ratio method of feature selection outperforms other feature selection methods for every performance parameter. Correlation based feature selection method also performs better than Information Gain method of feature selection for each performance parameter. The graphs below also show that Gain ratio method of feature selection performs better than the other feature selection methods.

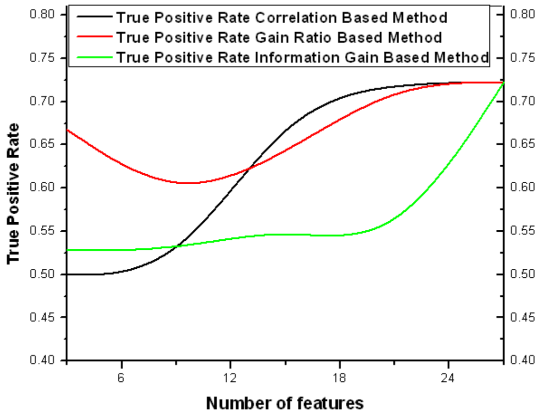
GRAPH I - F-MEASURE WITH FEATURE SELECTION METHODS



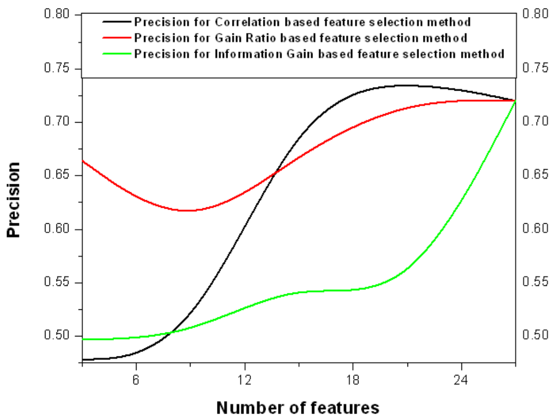
GRAPH II - PERCENTAGE CLASSIFICATION ACCURACY WITH FEATURE SELECTION METHODS



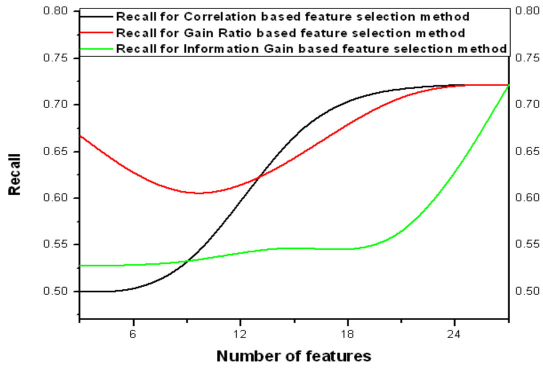
GRAPH III - TRUE POSITIVE RATE WITH FEATURE SELECTION METHODS



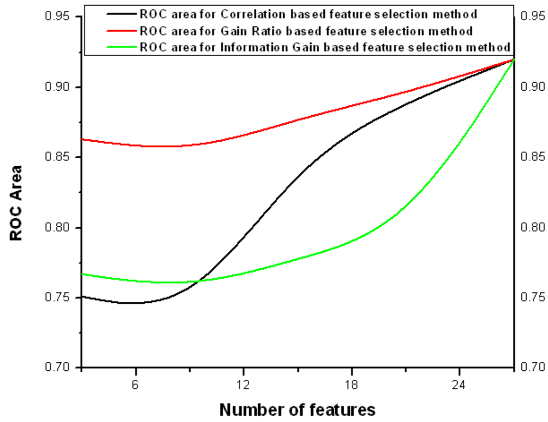
GRAPH IV PRECISION WITH FEATURE SELECTION METHODS



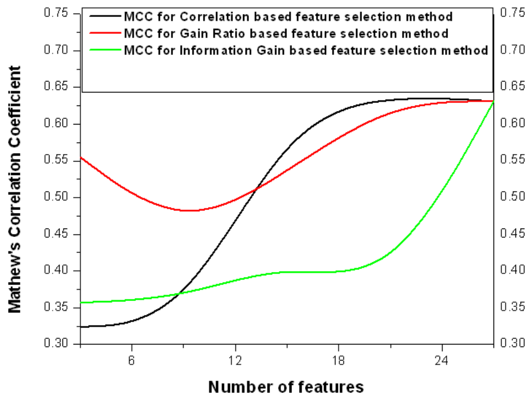
GRAPH V - RECALL WITH FEATURE SELECTION METHODS



GRAPH VI – ROC AREA WITH FEATURE SELECTION METHODS



GRAPHVII - MATHEW'S CORRELATION COEFFICIENT WITH FEATURE SELECTION METHODS



VII. CONCLUSION

The average percentage Classification Accuracy of optimized Naïve Bayes classifier with Gain Ratio feature selection was 66.66%. Correlation based feature selection yielded accuracy of 62.77% and Information Gain method had accuracy of 59.99%. Average True Positive Rate for Gain Ratio feature selection was observed as 0.666 which was better than the other two methods. Average Precision for Gain Ratio feature selection was 0.674 which was better than Correlation method that yielded precision as 0.628 and that of Information Gain which resulted in 0.559. Average Recall and F-Measure for Gain Ratio method was 0.666 which was comparably better than other two methods. Average Mathew's Correlation Coefficient and Receiver Operating Characteristics area with Gain Ratio feature selection was 0.560 and 0.882 which was better than other two methods. We conclude that Gain Ratio feature selection performs comparably better than the other two methods discussed in this work.

REFERENCES

[1] Ethem Alpaydin. Introduction to Machine Learning, Second Edition, MIT press, Cambridge, London, 2009.

[2] E. Short, D.F.Seifer, M. Matari, "A Comparison of Machine Learning Techniques for Modeling Human-Robot Interaction with Children with Autism", Human-Robot Interaction (HRI), 6th ACM/IEEE International conference, 251 – 252, Lausanne, 2011.

[3] R. Zhang, A.Bivens, "Comparing the Use of Bayesian Networks and Neural Networks in Response Time Modeling for Service-oriented Systems", In Proceedings of the 2007 workshop on Service-oriented computing performance: aspects, issues, and approaches, pp. 67 – 74, New York, USA, 2007.

[4] D. Han, J. Zhang, " A Comparison of Two Algorithms for Predicting the Condition Number", Sixth International Conference on Machine Learning and Applications, 13-15 Dec. pp. 223 – 228, Cincinnati, OH, 2007.

[5] G. Sikka, A. Kaur, M. Uddin, "Estimating Function points: Using Machine Learning and Regression Models", 2nd International Conference on Education Technology and Computer (ICETC), Shanghai, 2010.

[6] A. Bashar, G. Parr, S. McClean, B. Scotney, D. Nauck, "Machine Learning based Call Admission Control Approaches: A Comparative Study", IEEE International Conference on Network and Service Management (CNSM), pp. 431 – 434, Niagara Falls, ON, 2010.

[7] M. Bahrololum, E. Salahi, M. Khaleghi, "Machine Learning Techniques for feature Reduction in Intrusion Detection Systems: A Comparison", Fourth International Conference on Computer Sciences and Convergence Information Technology, pp. 1091 – 1095, Seoul, 2009.

[8] J. Alonso, L. Belanche, D.R. Avresky, "Predicting Software Anomalies using Machine Learning Techniques", IEEE International Symposium on Network Computing and Applications (NCA), pp. 163 – 170, Cambridge, MA, 2011.

[9] L. Vanajakshi, L. Rillet, "Support Vector Machine Technique for the Short Term Prediction of Travel Time", IEEE International Symposium on Intelligent Vehicles, pp. 600 – 605, Istanbul, 2007.

[10] R. Eisinger, R. Romero, R. Goularte, "Machine Learning Techniques Applied to Dynamic Video Adapting", IEEE Seventh International conference on Machine Learning and Applications(ICMLA '08), 2008.

[11] A. Forster, "Machine Learning Techniques Applied to Wireless Ad-Hoc Networks: Guide and Survey", International conference on Intelligent Sensors, Sensor Networks and Information, pp. 365 – 370, Melbourne, Queensland, 2007.

[12] M. Slavik, I. Mahgoub, "Applying Machine Learning to the Design of Multi-Hop Broadcast Protocols for VANET", Seventh International Conference on Wireless Communications and Mobile Computing (IWCMC), pp. 1742 – 1747, 2011.

[13] M. Grajzer, M. Koziuk, P. Szczechowiak, A. Pescap, "A Multi-Classification Approach for the Detection and Identification of eHealth Applications", Twenty First International conference on Computer Communications and Networks (ICCCN), pp. 1 – 6, Munich, 2012.

[14] D. Weir, S. Rogers, R. Murray-Smith, M. Lochtefeld, "A User-Specific Machine Learning Approach for Improving Touch Accuracy on Mobile Devices", UIST, Cambridge, Massachusetts, USA, 2012.

[15] M.A. Hall, "Correlation-Based Feature Subset Selection for Machine Learning", University of Waikato, Hamilton, 1999.

[16] A. Sharma, S. Dey, "Performance Investigation of Feature Selection Methods and Sentiment Lexicons for Sentiment Analysis features", Special Issue of International Journal of Computer Applications (0975 – 8887) on Advanced Computing and Communication Technologies for HPC Applications, 2012.

[17] M. Sokolova, G. Lapalme, "A systematic analysis of performance measures for classification tasks", Information Processing and Management, Volume. 45, pp 427–437, 2009.

[18] P. Liu, Y. Chen, W. Tang, Q. Yue, "Mobile WEKA as Data Mining Tool on Android.

Advances in Intelligent and Soft Computing Volume 139, pp 75-80, 2012.

[19] L. Silva, M. Koga, C. Cugnasca, A. Costa, "Comparative assessment of feature selection and classification techniques for visual inspection of pot plant seedling", Computers and Electronics in Agriculture, Volume 97, pp. 47–55, 2013.

[20] Directorate of Soybean research, India. (www.nrcsoya.nic.in).