

High Accurate and a Variant of k-fold Cross Validation Technique for Predicting the Decision Tree Classifier Accuracy



D. Mabuni, S. Aquter Babu

Abstract: In machine learning data usage is the most important criterion than the logic of the program. With very big and moderate sized datasets it is possible to obtain robust and high classification accuracies but not with small and very small sized datasets. In particular only large training datasets are potential datasets for producing robust decision tree classification results. The classification results obtained by using only one training and one testing dataset pair are not reliable. Cross validation technique uses many random folds of the same dataset for training and validation. In order to obtain reliable and statistically correct classification results there is a need to apply the same algorithm on different pairs of training and validation datasets. To overcome the problem of the usage of only a single training dataset and a single testing dataset the existing k-fold cross validation technique uses cross validation plan for obtaining increased decision tree classification accuracy results. In this paper a new cross validation technique called prime fold is proposed and it is experimentally tested thoroughly and then verified correctly using many bench mark UCI machine learning datasets. It is observed that the prime fold based decision tree classification accuracy results obtained after experimentation are far better than the existing techniques of finding decision tree classification accuracies.

Keywords : Decision tree classification accuracy, k-fold cross validation technique, machine learning, prediction.

I. INTRODUCTION

The k-fold cross validation is a bench mark technique for evaluating the performances of classification algorithms. Cross validation is a technique that is used for finding the optimal complexity of the model. It is one among the many techniques useful to fine-tune model complexity. When the data set is very large it is divided into k-distinct groups such that each group is again divided into one training set and one validation set. All the k-folds are trained and validated separately. When the dataset is small then the same dataset is spilt into different ways for getting k-groups of training and validation sets. Repeatedly dividing the same small data set into k-groups such that each group consists of well separated training dataset and validation dataset pair is called cross validation. Cross validation is a machine learning technique

used for evaluating machine learning model when the data set size is small. The bootstrap method is a basic building block for developing advanced machine learning algorithms such as Ada-Boost and XGBoost.

Bootstrap technique is a data re-sampling technique using random sampling with tuple replacement. Bootstrap aggregating is frequently used in ensemble based machine learning techniques.

If the sample dataset is very big then there is a possibility of getting a number of training/validation set pairs for k-fold cross validation and the typical value of k is generally 10 or 30. In k-fold cross validation given input dataset is divided randomly into k-partitions and again each partition is randomly divided into two equal parts such that one part is used for training and another part is used for validation. The k-fold cross validation is particularly useful when the dataset is small. It is a well known fact that repeatedly dividing the same small dataset into different pairs such that within the pair non overlapping is applied and there is a overlapping between the pairs such that each pair consists of one training set and validation set is called cross validation. In order to obtain robust error estimated training and validation sets dataset size must be large enough with minimum overlap between different k-fold groups. Given dataset is generally divided into pairs of training and validation sets for exclusively testing purpose. After completion of all the tests finally the entire single dataset is used only once for training and once for testing. In many cases one third training samples of the original dataset is kept separately for final test without overlapping with any one of the earlier sub-partitions of the original dataset.

That is final test set is not seen in any previous parts of the training datasets as well as any parts of the testing datasets. A classification algorithm uses a training dataset and generates a classifier model.

A single training and test dataset pair may not control randomness factors of the training dataset. Also note that training errors cannot be used for comparing two different algorithms because training dataset errors always less than the errors obtained with test set containing non overlapped instances with the corresponding training dataset. When a learning method is costly, generally it is trained and tested only once as a result of this the effectiveness of the classifier may not be possible to asses accurately and in other cases many runs of training and testing are performed.

The most desirable property or characteristic in any machine learning technique is that the validation dataset must be different from the training dataset and using only one run is not statistically sufficient.

Revised Manuscript Received on January 15, 2021.

* Correspondence Author

Dr. D. Mabuni*, Department of Computer Science, Dravidian University, Kuppam, India. Email: mabuni.d@gmail.com

Dr. S. Aquter Babu, Department of Computer Science, Dravidian University, Kuppam, India. Email: aquter@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

High Accurate and a Variant of k-fold Cross Validation Technique for Predicting the Decision Tree Classifier Accuracy

Definitely many runs are needed because of the reasons such as

1. Learning methods generally depend on many random factors that directly or indirectly affect generalization of the algorithm.
2. The sizes of the training datasets are small and sometimes very small and there is a provision of containing special and exceptional instances including outliers and noises.

Cross validation techniques and in particular k-fold cross validation technique is useful multitude ways

1. For comparing two different classification algorithms on the same dataset in the form expected error rates
2. For increasing the classification accuracy
3. For the utilization of all samples in the training dataset
4. To overcome the problem of presence of exceptional instances such as outliers and noise in the given training datasets
5. To increase the generalization capability of classification accuracy
6. To compare training times and space complexities of two different classification algorithms along the lines of same parameters
7. For tabulating cost sensitive learning details between two different classifiers
8. For providing instance and fast decision making in many fields and in particular in the medical field, and other research areas.
9. To fine tune the performance of the classification algorithm for obtaining highest data classification accuracy.
10. To create a base learner in the case of combining multiple base learners for the intent of increasing classification accuracy to the highest level
11. For validating the classifier or model

Receiver operating characteristics (ROC) curve is another technique that is frequently used for fine tuning a classifier.

A. K-Fold Cross Validation

The k-fold cross validation technique is a popular method for finding robust estimators and it is frequently used in machine learning applications. This technique is applied after dividing the dataset into k-groups such that each group consists of a single training dataset and a single testing dataset. Assume that the given dataset D is divided into D₁, D₂, D₃, D₄, ..., D_(k) sets. In the first time D₁ is used for data validation and D₂ + D₃ + D₄ + D₅ +, ..., D_(k) is used for training. Similarly in the second time D₂ is used for data validation and D₁ + D₃ + D₄ +, ..., D_k is used for training. This process is repeated for k-times and finally accuracy is computed by averaging all the individual k-results. The size of the training dataset increases as k increases and it causes to decrease the size of the validation datasets with increase robustness of the estimators. Also training costs increases as the k-value increases. When proportions of the class prior probabilities are same in each partition of the k-folds then it is called stratification.

Stratification is not possible in many real time applications especially when the data set size is very small. Leave-one-out is one special case of k-fold cross validation where one instance is used for validation and the remaining N-1

instances are used for training. This special case generally occurs in medical diagnosis data because getting labeled data is very difficult and sometimes very costly and causes to violate leave-one-out stratification property.

B. 5x2 Cross Validation

In this technique training dataset size is equal to the validation dataset size. In the first fold dataset is divided into two equal subsets called training set T₁ and validation set V₁. First fold pair is (T₁, V₁) and the second fold pair is (V₁, T₁).

TABLE-1 5x2 Cross Validation Datasets

Fold	Training Dataset	Validation Dataset
1	T ₁	V ₁
	V ₁	T ₁
2	T ₂	V ₂
	V ₂	T ₂
3	T ₃	V ₃
	V ₃	T ₃
4	T ₄	V ₄
	V ₄	T ₄
5	T ₅	V ₅
	V ₅	T ₅

Each fold consists of two pairs. The roles of first fold pair are reversed to get the second fold pair of training and validation sets. Same process is repeated in each fold. Totally there are five folds and in each fold there are two pairs such that T₁ U V₁ = T₂ U V₂ = T₃ U V₃ = T₄ U V₄ = T₅ U V₅ = D where D is the original dataset. Up to five folds this method gives good results with respect to statistical parameters. When the number of folds is taken more than five then the statistical significance of the results will decrease because of the sharing and overlap of the instances in the original dataset. There is a possibility of getting more training/validation sets. When the number folds are increased then validation error rates also increase automatically. When the numbers of folds are less than 5 then the sample size will become small as a result of this there will not be a scope for fitting the statistical distribution and the hypothesis testing.

C. Boot Strapping

Boot strapping is an alternative method for cross-validation for generating multiple samples from the given original sample dataset. New samples are generated from the original sample with replacement. Tuple overlapping percentage in bootstrapping is more than the tuple overlapping in cross validation. Bootstrapping is a potentially suitable technique for very small datasets. The overlapping percentage of bootstrapping is more than the overlapping percentage of k-fold cross validation. As the number of samples increases usage of tuples at least in one cross validation instance increases. All tuples are used in one or more fold groups of instances.

D. Bagging

In bagging samples are drawn with replacement. Bagging is simplified name for bootstrap aggregation.

Bagging is a variant of bootstrapping and it is voting based method and trained on different training datasets taken from slightly different samples.

One can use bagging for both classification and regression. It is particularly useful for learning unstable algorithms. An algorithm is said to be unstable if for small changes in the given training dataset there will be large number of changes in the corresponding output classifier model and both decision tree classification algorithm and multilayer perceptrons algorithm are examples for unstable algorithms.

E. Boosting

In the boosting technique complementary base learners are generated one after another in such a way that the subsequent learner is trained by taking all the mistakes of the previous learners. Also the main task or goal of boosting algorithm is to combine weak learning algorithms in order to obtain a strong learning algorithm. Boosting methods usage is increasing rapidly but the main disadvantage of boosting is that their training times are very high. For small training dataset it is not good because the sizes of sub-sequent samples decreases rapidly as learning proceeds. That is, initial sizes of datasets must be large enough for producing smooth and correct output models with less expected error rates. Boosting combines three base learners at the same time. Base learners should not be too weak. Boosting faces the problem of outliers and noise. Boosting is dependent on the data and base learners and it differs from problem to problem.

F. Ada-Boost

Ada-Boost is a variant of normal boosting technique. Its expansion is adaptive boosting. Same training dataset is used again and again. It means that the dataset need not be very large. Ada-Boost can have the ability to combine more than three base learners at the same time. In the literature many variants of Ada-Boosting were proposed. After completion of the training step Ada-Boost follows voting procedure. It can be used for both classification and regression. Now-a-days Ada-Boost algorithm is generally considered or take it for granted as one of the best machine learning algorithms and all of its functions are hundred percent automated and once the maximum number of base learners and the selected base learners are given then all the remaining tasks are automatically completed within small span of time interval.

II. LITERATURE SURVEY

Machine learning techniques are becoming popular with a tremendous speed both in research and in real world applications and it causes the attraction of many information technology professionals towards the machine learning fields and it's diversified and multitude related areas in the modern era. A classification result obtained by using a single training dataset and a single testing dataset is generally not better than the aggregated classification results of the many partitions of the same training dataset. In the research literature steps have been taken for getting high aggregated score based classification results. In many research finding outcomes it was clearly found that the dataset must be divided into many pairs of sub-datasets such that each pair consists of a single training and a single testing dataset for achieving high accurate classification results. Numerous techniques are available for increasing the classification accuracies and in

particular decision classification accuracy. Cross validation technique is used for fine tuning model complexity in machine learning. Once a classifier model is created using a training dataset and a classification algorithm then it must be required that the expected error rate of the model in future must be statistically correct and it must be in the pre specified expected error rate range with sufficient and enough confidence value say for example less than 3 percent expected error rate. Also cross validation techniques are required for comparing two different classification algorithms. In reality and indeed it is a real fact that a particular classifier may be high accurate on one dataset and may be less accurate on another dataset. In the real world no two things are identical and in a similar manner same classifier may not have same expected error rate on two different datasets. In machine learning literature it is revealed that no single learning algorithm always gives high accurate results in any domain. So, the alternative to this is experimentally applying many learners and then select the best one that performs the best outcome results with the separate validation set. The main point is that the learner performance must be fine-tuned for getting maximum possible accuracy on a validation dataset. One technique for improving the accuracy is by combining multiple base learners. Kaushika Pal and Biraj. V. Patel [1] have applied 5 machine learning techniques for document classification and then document classification accuracies are tested and compared using k-fold cross validation and holdout accuracy estimation methods. The experimental results have shown that support vector machines is very much better than the other document classification machine learning methods. Sanjay Yadav and [Sanyam Shukla](#) [2] said that k-fold cross validation technique is a potentially suitable method for finding data classification accuracy of small datasets with larger k values and also pointed out that the actual differences between k-fold cross validation and hold out methods. [Juan Diego Rodríguez](#) et al. [3] have analyzed statistical properties, bias and variance, of k-fold cross validation estimator and experimentally verified that the sensitive changes in the training datasets and sensitive changes in the k-fold cross validation methods. [Tzu-Tsung Wong](#) and [Nai Yu Yang](#) [4] proposed a statistical measure for the assumption of accuracy independence in k-fold accuracy and experimentally verified correctly that the assumption is perfectly valid.

Tzu Tsung Wong and Po Yang Ye [5] experimentally found the correlations between the folds of the k-fold cross validations and the experimental results have shown that the folds are highly correlated. Yanqiu zhang [6] have designed a new classification algorithm using SVM algorithm.

The new classification algorithm combines multiple weak SVM algorithms forming a strong and a new classifier in terms of classification accuracy.

Several comparison experiments were conducted with the new algorithm using datasets of varying sizes and the experimental results have shown that the proposed new prime fold algorithm is more accurate than the existing algorithms in terms of classification accuracy and execution speed.

High Accurate and a Variant of k-fold Cross Validation Technique for Predicting the Decision Tree Classifier Accuracy

Cesare Appilli [7] proposed a variant of k-fold cross validation called virtual k-fold cross validation with decreased training time with small set of models and the new method is successfully applied for linear models with hundred percent accuracy and for non linear models with some approximations.

Nneurhayathi [8] developed a new model for detecting the ground water flow wind tidal. The two algorithms namely hold-out and k-fold cross validation methods are used for accuracy determination. Experiments are conducted for ground water modeling and the results show that the k-fold cross validation is very much better than the hold-out methods. Ramasubbu Venkatesh and Charless row land [9] proposed a new model for developing biomedical genomic applications and it was successfully applied on many real world datasets and the results are more attractive in terms of accuracy and generalizability of the model. T.Q. Huynh, R Section [10] proposed a new data classification model of neural network and this model is applied on real time applications after performing many experiments on public datasets. The experimental results show that proposal model is better than the existing neural network and decision tree classification methods. Michal Vasinek [11] cross validation method is sometimes used for verifying predictive models. In this paper authors have predicted the stability of simple learning set of rules classifier using k-fold cross validation experiments are conducted and a set of rules that are passed are noted down as output and the passed rules have exhibited very low variance in terms prediction accuracy. Thine Swaran Gunasegaran [12] are used k-fold cross validation methods for generating folds or sections randomly. A fold is collection of training data subset and test data subset. The k-fold cross validation method finally computes aggregate accuracy of the individual fold accuracies, here many combinations are possible for selecting the training set and test set pairs. Different combinations of folds will produce different accuracies and there is a need to find a specific combination that yields optimal accuracies.

The study in this paper proposed a new procedure for finding the optimal folds in the given dataset for improving the predictive modeling accuracy. Proposed prime fold algorithm is a variant of k-fold cross validation and the experimental results reveal that the proposed algorithm is highly improved variant of k-fold cross validation technique.

In the case of k-fold cross validation all folds are not same and in general different folds will give different accuracies of classification results. One optimized solution for folds selection is obtained using genetic algorithm. There exist many algorithms for optimal selection of folds.

Every fold gives certain accuracy value for the selected model. There is a possibility of getting too many folds for the given dataset and getting of different folds can be formulated as an optimization problem. Authors have experimentally tried to find optimal folds of cross validation and experimentally verified that their results are very much better than the simple 10 fold cross validation.

Prediction error is the parameter frequently used to measure the performance of the learning algorithm. Cuixian Chen et al. [13] have carried out the sensitive analysis details with respect to changes in the training or test datasets. Experiments are conducted for finding different behaviors of cross-validation in order to find optimal prediction error estimates.

III. PROBLEM DEFINITION

The most and fundamental requirement in machine learning is that the dataset must be divided into many non overlapping pairs such that each pair consists of one training dataset and one test dataset. In general, when the dataset size is very large it is randomly divided into k equal sized partitions and then each partition is randomly divided into two subparts one subpart for training and another subpart for testing. Dividing given dataset, D, into k subparts is possible only when the dataset size is very large but in many real time situations the sizes of datasets are small and in particular in medical field areas sizes of datasets are very small. In such real applications scenarios the existing k-fold cross validation technique is useful for finding decision tree classification accuracy and in particular this method is useful for increasing the decision tree classification accuracy and many experimental results have shown that the k-fold cross validation is far better than the normal methods in terms of finding classification accuracies. In general, the typical k value in k-fold cross validation is 10 or 20 and it is also a well known fact that k must be very large and consequently training datasets must be very large for obtaining robust estimators. Though k-fold cross validation technique is a good technique still there is a need for further enhancements and also for finding more and more new robust classification techniques for increasing the classification accuracies. Cross-validation is widely employed to estimate the expected accuracy of a predictive algorithm.

IV. PROPOSED METHOD

A new cross validation technique called prime fold number based cross validation is proposed and folds or sections are selected based on prime numbers sequence. Number of folds is same as the list of prime numbers that are originally selected under consideration and usually it is 10 or 20 and sometimes it can be modified slightly according to the requirements real time scenarios. In each of the possible folds the sample dataset size is different and training and validation dataset sizes are also different. Overlapping between sections and at the same time overlapping within the section is reduced to the maximum extent. The proposed method is applied in finding the classification accuracy of the decision tree data classification. For larger datasets larger prime numbers are considered and for small datasets small prime numbers are considered in selecting fold sizes of validation datasets.

V. EXPERIMENTAL RESULTS

In computing decision tree classification accuracies the proposed prime fold method is applied on many UCI machine learning datasets and experimental results are tabulated. Keen observations of the experimental results have shown that the classification accuracies of the proposed method, prime fold cross validation is far better than the existing k-fold cross validation method based decision tree classification accuracies. Standard machine learning datasets are employed during experimentation and stratification is applied to increase the usage availability of the number of potential datasets in the experiments.

Direct method and k-fold cross validation method are the two existing methods that are used here for finding classification accuracy of many training datasets. In direct method, training dataset is used for model building and the test dataset is used for model testing or validation.

In k-fold cross validation many parts of the same dataset are used for training and then testing and then finally aggregate of all the independent classification accuracies are computed and used as the final and correct classification accuracies. In the proposed prime fold method many training datasets are employed and in each training dataset many parts are used for training and testing without stratification. After the complete experimentation with the many bench mark datasets it is observed that the proposed prime fold method has yielded high decision tree classification accuracies and it is considered to be the superior method than the existing methods.

TABLE-2 Decision Tree Classification Accuracy Comparisons

S. No	Dataset	Dataset Size	Classification Accuracy of Existing Direct Method	Classification Accuracy of Existing K-fold Cross Validation	Classification Accuracy of Proposed Prime Fold Cross Validation
1	Iris	150	98.0	97.3333	98.74006
2	Glass	214	85.5140	92.3809	87.3309
3	Iono	351	83.1908	83.1428	89.9919
4	Breast	569	98.2425	97.3214	99.4116
5	Vehicle	846	83.8061	79.8809	86.1437
6	Segment	1735	97.6190	96.5800	98.42
7	Page	5473	99.8355	99.5429	99.9546
8	Satellite	4435	89.8759	84.5352	90.8854

TABLE-3 Classifier Accuracies

Dataset	Existing	K-fold	Prime Fold
Iris	98.0	97.3333	98.74006
Glass	85.5140	92.3809	87.3309
Iono	83.1908	83.1428	89.9919
Breast	98.2425	97.3214	99.4116
Vehicle	83.8061	79.8809	86.1437
Segment	97.6190	96.5800	98.42
Page	99.8355	99.5429	99.9546
Satellite	89.8759	84.5352	90.8854

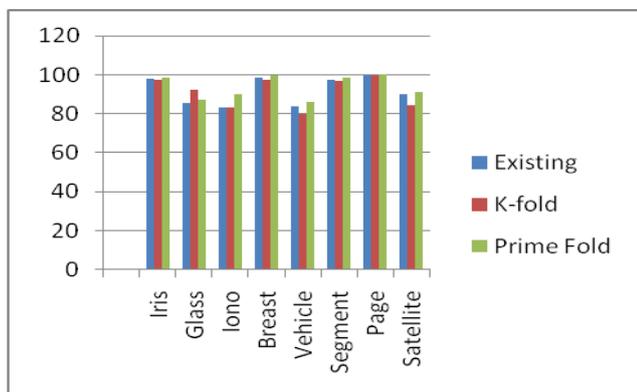


Fig-1 Classifier Prediction Comparisons

In fig-1 for different methods and for different training datasets classifier accuracies are given. Direct and k-fold cross validation methods are two existing methods useful for model accuracy prediction and in this paper the proposed prime fold method produces high prediction accuracies for all the training datasets except for glass dataset.

VI. CONCLUSION

Instead of dividing the dataset into a single training dataset and a single testing dataset sometimes it may be convenient and very much useful when it is possible to divide the given dataset into more number of groups of distinct and non overlapping and very large sized sub-datasets such that each group consists of non overlapping one training dataset and one testing dataset. But in many cases the sizes of datasets are limited and as a result of this it may not be possible for dividing given dataset into large number of non overlapping sub-datasets. To overcome this problem k-fold cross validation technique is available but still software industry is expecting more and more new state-of-the-art cross validation techniques. In this direction the present study has introduced a new cross validation technique which is a variant of k-fold cross validation method and at the same time it is better than the k-fold cross validation method. In the future for specific, advanced and reliable techniques will be investigated for inventing better, robust, state-of-the-art cross validation algorithms and ideas.

REFERENCES

1. Kaushika Pal and Biraj. V. Patel, "Data Classification with k-fold Cross Validation and Holdout Accuracy Estimation Methods with 5 Different Machine Learning Techniques", Publisher: IEEE, [2020 Fourth International Conference on Computing Methodologies and Communication \(ICCMC\)](#), Date of Conference: 11-13 March 2020, Date Added to IEEE Xplore: 23 April 2020, INSPEC Accession Number: 19569451, DOI: 10.1109/ICCMC48092.2020.ICCMC-00016.
2. Sanjay Yadav and Sanyam Shukla "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification", Published in: 2016 IEEE 6th International Conference on Advanced Computing (IACC), Date of Conference: 27-28 Feb. 2016, Date Added to IEEE Xplore: 18 August 2016.
3. Juan Diego Rodríguez, Aritz Pérez, Jose A. Lozano, "Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation", April 2010, IEEE Transactions on Pattern Analysis and Machine Intelligence 32(3):569 – 575.
4. Tzu-Tsung Wong, Nai Yu Yang, "Dependency Analysis of Accuracy Estimates in k-Fold Cross Validation", August 2017, IEEE Transactions on Knowledge and Data Engineering 29(11):1-1, DOI: 10.1109/TKDE.2017.2740926.
5. TZU TSUNG WONG AND PO YANG YE, "RELIABLE ACCURACY ESTIMATES FROM K-FOLD CROSS VALIDATION", [IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING](#) (VOLUME: 32, ISSUE: 8, AUG. 1 2020), PAGE(S): 1586 – 1594, DATE OF PUBLICATION: 25 APRIL 2019.
6. Yanqiu Zhang; Ming Ni; Chengwu Zhang; Shuang Liang; Sheng Fang; Ruijie Li; Zhouyu Tan, "Research and Application of AdaBoost Algorithm Based on SVM", Published in: 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAI), Date of Conference: 24-26 May 2019, Date Added to IEEE Xplore: 05 August 2019.
7. Cesare Alippi and Manuel Roveri, "Virtual k-fold cross validation: an effective method for accuracy assessment", Published in: [The 2010 International Joint Conference on Neural Networks \(IJCNN\)](#), Date of Conference: 18-23 July 2010, Date Added to IEEE Xplore: 14 October 2010.
8. Nurhayati; Indratmo Soekarno; Iwan K. Hadihardaja; M. Cahyono, "A study of hold-out and k-fold cross validation for accuracy of groundwater modeling in tidal lowland reclamation using extreme learning machine", Published in: [2014 2nd International Conference on Technology, Informatics, Management, Engineering & Environment](#), Date of Conference: 19-21 Aug. 2014, Date Added to IEEE Xplore: 19 January 2015.

High Accurate and a Variant of k-fold Cross Validation Technique for Predicting the Decision Tree Classifier Accuracy

9. [Ramasubbu Venkatesh](#); [Charles Rowland](#); [Hongjin Huang](#); [Olivia T. Abar](#); [John Sninsky](#), “Robust Model Selection Using Cross Validation: A Simple Iterative Technique for Developing Robust Gene Signatures in Biomedical Genomics Applications”, **Published in:** [2006 5th International Conference on Machine Learning and Applications \(ICMLA'06\)](#), **Date of Conference:** 14-16 Dec. 2006.
10. T.Q. Huynh; R. Setiono, “Effective neural network pruning using cross-validation”, Published in: Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005., Date of Conference: 31 July-4 Aug. 2005, Date Added to IEEE Xplore: 27 December 2005.
11. Michal Vasinek; Jan Plato; Vaclav Snasel, “Limitations on Low Variance k-Fold Cross Validation in Learning Set of Rules Inducers”, Published in: 2016 International Conference on Intelligent Networking and Collaborative Systems (INCoS), Date of Conference: 7-9 Sept. 2016, Date Added to IEEE Xplore: 27 October 2016.
12. Thineswaran Gunasegaran; Yu-N Cheah, “Evolutionary cross validation”, Published in: 2017 8th International Conference on Information Technology (ICIT), Date of Conference: 17-18 May 2017, Date Added to IEEE Xplore: 23 October 2017.
13. Cuixian Chen, Yishi wang, Yaw Chang, Karl Ricanek, “Sensitivity Analysis with Cross-Validation for Feature Selection and Manifold Learning”, International Symposium on Neural Networks, ISSN 2012: Advances in Neural Networks–ISNN 2012 pp 458-467, Springer.

AUTHORS PROFILE



Dr. D. Mabuni, completed M.Sc. (Computer Science), MCA, M.Phil. (Computer Science) and Ph.D. (Computer Science). Currently working as Assistant Professor in the Department of Computer Science at Dravidian University, Kuppam, Andhra Pradesh, India. My interested research areas are Data Mining, Databases, and User Interfaces.



Dr. S. Aquter Babu, completed MCA and Ph.D. (Computer Science). Currently working as Professor in the Department of Computer Science at Dravidian University, Kuppam, Andhra Pradesh, India. My interested research areas are Data Mining, Databases, and Natural Language Interfaces to databases.