CHAPTER 2

REVIEW OF LITERATURE

Review of Relief Algorithm

Relief is a feature selection algorithm for random selection of instances for feature weight calculation. The Relief algorithm adopts the random selection of instances for weight estimation. It uses the Monte Carlo Approaches for randomization selection of instances in the Relief.

Kira and Rendell proposed the Relief Algorithm. The statistical method is used in Relief instead of Heuristic search. Relief requires linear time in the number of given features and number of training instances regardless of the target concept to be learned. It selects the statistically relevant features (*Kira et al.,* 1992).

Relief-F is the extension of Relief algorithm. This Relief-F has enabled to work with noisy and incomplete data sets and to deal with multi-class problems (Kononenko, 1994), (*Abdolhossein Sarrafzadeh et al.,* 2012).

Relief-D is a deterministic version of Relief. It uses all instances and all near-hits and near-misses of each instance. This results in the equivalent of running Relief for an infinite amount of time (*John et al.,* 1994).

The Euclidean Based Feature Selection algorithm (EUBAFES) weights and selects features similarly to the Relief algorithm. It is also a distance-based approach that reinforces the similarities between instances that belong to the same class while deteriorating similarities between instances in different classes. A gradient descent approach is employed to optimize feature weights on this goal (*Scherf et al.,* 1997).

Principal component analysis and compression (information theory) play major role in feature selection by way of eliminating the features with less information for prediction (*Sa Wang et al.,* 2007). These approaches have adopted the feature selection technique for different areas to improve the model (*Liu et al,,* 2007).

Hua et al., reported that comparison of some famous feature selection method in the area of bioinformatics. The methods are information Gain, Gini index, t-test, Sequential Forward Selection. According to him, the Feature selection in the biological area is inevitable but quite challenging. Among the existing feature selection algorithms, the Relief and its variants are considered as successful one due to its simplicity and effectiveness (*Huanjing Wang et al.,* 2010). Relief algorithm was first proposed in (*Kira et al.,* 1992). After that lot of variants came to usage for feature selection. Every variant has its own merits and demerits depending on the nature of data sets. The key idea of Relief is to iteratively estimate feature weights according to their ability to discriminate between neighbouring models. Relief was extended to handle noisy and missing data and solve multi-class issues which the original Relief algorithm cannot deal with (*Robnik Sikonjam et al.,* 2003). This Relief algorithm was named as Relief F. Subsequently, to explore the framework of expectation maximization, Iterative-Relief is put forward in (Sun Yi Jun, 2007). Adaptive Relief is termed as A-Relief. This A-Relief algorithm offers effective feature subset for further identification (*Fab Wenbing et al.,* 2012). Likewise, many variants of Relief algorithms are available in the feature selection domains.

Lei Yu and Huan Liu introduced a feature selection algorithm called Fast Correlation-Based Filtering approach using the predominant correlation concept (*Lei Yu et al.,* 2003). In this algorithm, the feature subset was selected based on Symmetric Uncertainty and F-correlation measure. In this approach, redundant and irrelevant features were removed and it showed improvement in the classification accuracy.

Lei *Yu et al.*, identified that there was a need for explicit redundant analysis in feature selection (*Lei Yu et al.,* 2004). Hence, a new framework for efficient feature selection was defined through relevance and redundant analysis. This framework divided the relevant analysis and redundant analysis. A new feature selection algorithm was also implemented, and it was verified against various learning algorithms to extract the best feature set.

Xiubo Geng et al., investigated the existing feature selection algorithms for ranking models. They found that there was a striking difference between the ranking and classification. Hence, a feature selection algorithm was proposed. In this algorithm, two ranking models, Ranking Support Vector Machine and Rank Net were used to extract the best feature subset (*Xiubo Geng et al.*, 2007).

Chung-Jui Tu et al., proposed a feature selection algorithm using Particle Swarm Optimization (PSO) and Support Vector Machines (SVMs). The Particle Swarm Optimization was used to select the best feature subset. The Support Vector Machines with the one-versus-rest method was used to evaluate the fitness function of Particle Swarm Optimization. The algorithm was validated with various classification problems (*Chung-Jui et al.*, 2007).

Noelia Sanchez-Marono et al., studied the four filter methods for feature selection. In their study, they made a comparison among the four filter methods to find the best filter method. Based on the study, they proposed a hybrid filter algorithm using best filter methods (*Noelia Sanchez-Marono et al.,* 2007). To find the best filter method, a comparison was made among the four filter methods which are Relief, Correlation-Based Feature Selection, Fast Correlated Based Filter and INTERACT.

Antonio Arauzo-Azofra et al. defined a feature selection method, namely, Consistent-Based feature selection. It was a useful measure for various feature selection methods. Hence, the proposed method achieved similar accuracy result, than the wrapper approach and it also attained higher feature reduction (*Antonio Arauzo-Azofra et al.*, 2008).

Appavu *et al.* proposed a feature selection algorithm using association rule mining and Information gain. The Apriori algorithm was used to find the relevant attributes. Information gain was used to remove the irrelevant and redundant features in the dataset. The result of the algorithm showed that there was no improvement in the classification accuracy (*Appavu et al.,* 2009).

Huanjing Wang et al. introduced a feature selection method using the filter–based ranking techniques. The proposed technique was called Threshold Based Feature Selection (TBFS). Each attribute's value was normalized between 0, and 1 using the F-measure and the independent attribute was paired individually with the class attribute. This technique was useful to find the smaller subset of features, and it showed an improvement in the classification accuracy (*Huanjing Wand et al.,* 2010).

Athanasios et al. proposed a feature selection algorithm using the Correlation - Based filter approach called Relevance, Redundancy and Complementarily Trade-off. In this algorithm, the linear correlation coefficient was used to remove the irrelevant, redundant and noisy features. Gaussian distribution method was used to obtain the best feature subset *(Athanasios et al.,* 2010).

Yuxuan Sun et al., studied the existing Relief algorithm for feature selection using feature weight estimation. From the study, they found that Relief algorithm was unstable and that lead to poor accuracy of expected results. To overcome the problem, a feature selection algorithm was proposed based on Mean-Variance Model (*Yuxuan Sun et al.*, 2011).

The Mean-Variance Model was used to revise the feature weight estimation method according to the original Relief algorithm. In this algorithm, the mean and the variance of the discrimination among instances were considered as the criterion of feature weight estimation. The algorithm was validated through an experimental study, and the result indicated that the feature subsets generated by the proposed algorithm had a better performance.

Qinbao Song et al., developed a Clustering-Based Feature subset selection algorithm for high dimensional data. The Graph-theoretic method was used to divide the features into clusters. The features that were strongly related to the target class were selected as the best feature subsets. In this, they treated each cluster as a single feature. Hence, the dimensionality was drastically reduced. The algorithm was compared with various existing algorithms, and it showed a minimum improvement in the prediction accuracy and classification performance (*Qinbao Song et al*., 2011).

Debahuti Mishra and Barnali Sahu proposed a model for feature selection using Signal to Noise Ratio (SNR) ranking to enhance the predictive accuracy. In this algorithm, they proposed two approaches for selecting the best features. In the first approach, K-means clustering and SNR ranking were used to get the top ranked features. In the second approach, SNR ranking was used to obtain the best features. The two models were validated through different classifiers by conducting an experiment. They found that the performance of the learning algorithms decreased the classification accuracy (*Debahuti Mishra et al.,* 2011).

Tingquan Deng *et al.* introduced a notation called Knowledge Granularity. The knowledge granularity was used to find the relationship between conditional attributes and decision attributes. An evaluation function was used to measure the significance of conditional attributes. They developed an optimized algorithm for feature selection based on the evaluation function. The algorithm was validated, and it showed improvement in the classification accuracy (*Tingquan Deng et al.*, 2011).

Myo Khaing and Nang Saing Moon Kham studied the various feature selection algorithms using Multiple Correspondence Analysis (MCA) and found many disadvantages in the same. Hence, they proposed a feature selection algorithm called Modified - Multiple Correspondence Analysis (M-MCA). From the

experiment conducted, they claimed that the result of the experiment gave better performance compared to the existing algorithm using simple Multiple Correspondence Analysis (*Myo Khaing et al.,* 2011), (*Greenacre et al.,* 2006).

Boyang Li *et al.* designed a feature selection model based on correlation analysis and SVM ranking method. In this algorithm, the correlation-based clustering was used to group the feature into some clusters. Influence Qualities were calculated for each of the features in the cluster. Using the feature sensitivity in the Support Vector Machine, the best features were obtained. This model was tested with some real datasets, and they stated the result which showed improvement in the classification accuracy compared to the existing algorithms (*Boyang Li et al.,* 2011).

Danyang Cao *et al.* analyzed the discrimination feature selection algorithm using Information Theory, which did not consider the discriminate and continuous features of the datasets. From the analysis, they proposed a feature selection algorithm. In this algorithm, an entropy breakpoint concept was introduced. This algorithm was validated with various real-world datasets. The result of the experiment specified that the algorithm had a high computational complexity with very low prediction accuracy (*Danyang Cao et al.,* 2012).

Chinna Gopi *et al.* proposed an algorithm to solve the optimization problem in feature selection. This algorithm was used to find the best features using Greedy Search Method and Greedy Search Loss of ranking method. The algorithm was validated with the public dataset. This algorithm worked well in generated optimized feature subset but, the computational cost was higher than the existing algorithms (*Chinna Gopi et al.,* 2012).

Related Works on Feature Selection on Agriculture and Biological datasets

Data mining and its various methodologies are used for industrial, commercial, and scientific purposes (*Ebrahimi et al.,* 2010) (*Ehsan Bijanzadeh*

et al., 2010). Recently, agricultural and biological research studies have used various data mining techniques for analyzing large datasets and creating useful classification patterns in datasets. The novelty and advancement in feature selection on data mining technologies can bring more fruitful results in different discipline (*Hsiao et al*., 2006), (*Amiri Chayian*, 2010). Data mining tasks involve hundreds and thousands of attributes. The major portion of time in model building process involves examining the variables to be included in the model. Feature selection allows the features set to be reduced in size and it is creating a more manageable set of attributes for modeling (*Liu et al.,* 2008). Feature selection has been an active research area in pattern recognition, statistics, and data mining communities. The main idea of feature selection is to choose a subset of input variables by eliminating features with little or no predictive information (*Handl et al.,* 2006) (*Liu et al.,* 2008) (*Bijanzadeh et al*., 2010). When the feature space is complex, and the data distribution patterns are not uniform, the use of feature selection method allows analyzing the more complex data compared to other statistical techniques (*Drummond et al*., 2002) (*Gautam et al*., 2006).

Ehsan Bijanzadeh has reported that the supervised feature selection algorithm was applied to determine the most important features contributing to wheat grain yield. Four hundred and seventy-two fields from different parts of Iran which were different in 21 characteristics (features) were selected for feature selection analysis (*Ehsan Bijanzadeh et al*., 2010).

Selection of the wide range of features, including location, genotype, irrigation regime, fertilizers, soil textures, physiological attitudes and morphological characters, provided the opportunity for precise simultaneous study of a large number of factors in wheat grain yield topic by the hand off data mining. These features included culture type, location, soil texture, 1000 kernel weight, nitrogen supply, irrigation regime, biological yield and organic content of the

soil, the amount of rainfall, genotype, plant height, and spike number per unit area. Interestingly, growing season length and plant density were the second most important features for wheat grain yield. Based on the feature selection model, culture type, as dry land farming or irrigated, severely affected wheat grain yield. The soil pH had a marginal effect on wheat grain yield. The results of this investigation demonstrated that feature classification using feature selection algorithms might be a suitable option for determining the important features contributing to wheat grain yield, providing a comprehensive view about these traits. This is the first report in identifying the most important factors on wheat grain yield from many fields using feature selection model (*Suman et al.*, 2013).

Relief is considered as one of the most successful algorithms for assessing the quality of features due to its simplicity and effectiveness (*Dietterich*, 1997). The Relief algorithms are a family of attribute weighting algorithms that can efficiently identify associations between attributes and the class even if the attributes have nonlinear interactions without significant main effects (*Kira et al.*, 1992) (*Dietterich*, 1997). Relief was extended to handle noisy and missing data (Kononeko, 1994).

An iterative Relief (I-Relief) algorithm is used to alleviate the deficiencies of Relief by exploring the framework of the Expectation-Maximization algorithm. I-Relief was introduced to support to multiclass settings by using a new multiclass margin definition (Yijun Sun, 2007).

Abdolhossein Sarrafzadeh et al. studied the effects of reducing the number of features and selecting the most effective subset of features in the context of content-based image classification and retrieval of objects using the Relief-F algorithm. Their experimental result shows that employing Relief-F on Coil-20 image dataset improves the speed and accuracy (*Abdolhossein Sarrafzadeh et al.*, 2012).

Yan Wei *et al*., have briefed that in the feature set of complex products which are in high dimension, the set usually contains useful information, irrelevant information, and redundant information. However, the former is usually buried in the latter two. Therefore, the recognition of the most useful information in the original dataset, which is defined as the identification of Critical-To-Quality features, becomes a key process in the field of quality control. The traditional methods include Taguchi loss function and Decision tree, and the like. However, almost none of them can deal with the high dimensional quality feature set with both accuracy and easiness (*Yan Wei et al*., 2011).

Casey et al., have proposed the Speeded-up Robust Features (SURF) algorithm. They reported that Speeded-up Robust Feature's ability to detect interactions in this domain is significantly greater than that of ReliefF. Similarly Speeded Up Robust Features, in combination with the Total Unduplicated Research and Frequency (TURF) strategy significantly outperforms Total Unduplicated Research and Frequency alone for Single Nucleotide Polymorphisms (SNP) selection under an epitasis model (*Casey et al*., 2009).

Matthew Stokes *et al.*, have proposed and developed a new spatially weighted variation of Relief called Sigmoid Weighted Relief Star (SWRF*), and applied it to synthetic SNP data. When compared to Relief and SURF*, which are two algorithms that have been applied to SNP data for identifying interactions, SWRF* had significantly greater power. They reported that the new Relief algorithm called SWRF* that had greater ability to identify interacting genetic variants in synthetic data compared to existing Relief algorithms (*Mathew Stokes et al.*, 2012).

Fan Wenbing et al., have proposed an adaptive Relief (A-Relief) algorithm to alleviate the deficiencies of Relief by dividing the instance set adaptively. According to them, A-Relief has performed better in image datasets (*Fan Wenbing et al*., 2012).

Yuxuan Sun et al. devised a new strategy on the Relief algorithm. They briefed the defects of Relief algorithm. According to them, as Relief algorithm selects the instances randomly, the feature weight estimation is uncertain. So the randomicity and the uncertainty of the instances used for calculating the feature weight vector in the Relief algorithm, the results lead to poor evaluation accuracy. To overcome this issue, a novel feature selection algorithm based on Mean-Variance model is proposed by them. This algorithm takes both the mean and the variance of the discrimination among instances into account as the criterion of feature weight estimation, which makes the result more stable and accurate. Based on real seismic signals of ground targets, experiment results indicate that the subsets of feature generated by proposed algorithm have better performance (*Yuxuan Sun et al*., 2011).

Blessie E.C *et al.* proposed a new algorithm called Relief-Disc. It works based on Discretization. Discretization partitions feature into a finite set of adjacent intervals. Instead of using random sampling for selecting the instance, they have suggested taking instance from each interval which reduces the computational complexity and maintains the quality of features. Also, there was no need of user input for sample size parameter. Experimental results showed that the performance of the new algorithm is better when compared with the existing Relief algorithm. According to them, Relief-Disc performed better than Relief (*Blessie and Karthikeyan,* 2011).

Review of Simulated Annealing Algorithm

In the process of physical annealing, a solid is heated until all particles randomly arrange themselves forming the liquid state. A slow cooling process is then used to crystallize the liquid. This process is known as simulated annealing. Simulated Annealing is a stochastic computational technique that searches for global optimum solutions in optimization problems. The main goal here is to give the algorithm more time in the search space exploration by

accepting moves, which may degrade the solution quality, with some probability depending on a parameter called temperature. Simulated Annealing has been extensively applied to deterministic optimization problems, and the theoretical basis of the algorithm for this application has been known for some years. Many instances of practical and difficult problems were successfully solved by Simulated Annealing. The effectiveness of Simulated Annealing is attributed to the nature that it can explore the design space using a neighbourhood structure and escape from local minima by probabilistically a long uphill move controlled by a temperature parameter.

Kirkpatrick realized the similarity between the optimization of combinational optimization problems and the physical process of annealing. Simulated Annealing became one of the more popular optimization algorithms (Kirkpatrick, 1983). Sullivan and Jacobson studied generalized hill climbing algorithms and their performance. They extended necessary and sufficient convergence conditions for Simulated Annealing (*Sullivan et al.,* 2001).

Nader Azizi and Zolfaghari addressed changes in temperature based on the number of consecutive moves showing improvement by comparing two variations of the SA method in adaptive temperature control (*Nader Azizi et al.,* 2004).

Rosen and Harmonsky proposed a Simulated Annealing based simulation optimization method, which is an asynchronous, team-type heuristic. It improved the performance of Simulated Annealing for discrete variable simulation optimization with the conventional cooling schedule, the probability of transition decreases from the beginning of the search to the end (*Scott Rosen et al.,* 2005).

Ameur found a simple algorithm to compute the temperature in SA which is compatible with a given acceptance ratio of bad moves. He also provided a convex function and low temperatures and a concave function of high temperature based on a geometric schedule (Walid Ben-Ameur, 2004).

The first theoretical analysis of Simulated Annealing applied to solve discrete stochastic optimization problems was given by Gelfand and Mitter. They showed that if the noise in the estimated objective function values in the iteration has the normal distribution with zero mean and positive variance, then their procedure converges in probability to the set of optimal global solutions provided that the sequence be chosen properly (*Gelfand et al.,* 1991a).

In 1996, Walter Gutjahr and Pflug generalized a classical convergence result for the Simulated Annealing algorithm to the case where cost function observations are distributed by random noise (*Walter Gutjahr et al.,* 1996).

Saul Gelfand and Sanjoy Mitter examined the effect of using noisy or imprecise measurements of the energy differences on tracking the minimum energy state visited by the modified algorithms (*Saul Gelfand et al*., 1998).

Charon and Hudry suggested adding noise to the Simulated Annealing algorithm. Their approach adds random noise initially and then gradually reduces the noise to zero to perturb the solution space (*Charon et al.,* 1993).

Mehmoud H. Alrefaei and Sigrun Andradchir have presented a modified Simulated Annealing algorithm designed for solving discrete stochastic optimization problems (*Mahmoud Alrefaei et al.,* 1999).

In 2001 Charon and Hudry extended their noising method. The algorithm perturbs the solution space by adding random noise to the problem's objective function values. A stopping criterion is introduced in a precise way that gradually reduces the noise-rate (*Charon et al.,* 1993).

Prudius and Andradottir proposed two cooling schedule approach for controlling the probability of moving to seemingly inferior points and used the state with the highest estimated objective function value obtained from all the previous observations (*Prudius et al*., 2005).

Ling Wang and Liang Zhang proposed Simulated Annealing combined with hypothesis testing for stochastic discrete optimization problems and demonstrated the effectiveness of the proposed approach by the simulation results based on stochastic numeric optimization problems (*Ling Wang et al.*, 2005).

Review of Sequential Selection Algorithms

These algorithms are called sequential due to the iterative nature of the algorithms. The Sequential Feature Selection (SFS) algorithm starts with an empty set and adds one feature for the first step which gives the highest value for the objective function. From the second step onwards the remaining features are added individually to the current subset and the new subset is evaluated. The individual feature is permanently included in the subset if it gives the maximum classification accuracy. The process is repeated until the required numbers of features are added. This is a Naive sequential feature selection algorithm since the dependency between the features is not accounted for (*Pudil et al*., 1994) (Reunanen, 2003).

A Sequential Backward Selection (SBS) algorithm can also be constructed which is similar to sequential feature selection algorithm. It starts from the complete set of variables and removes one feature at a time whose removal gives the lowest decrease in predictor performance.

The Sequential Floating Forward Selection (SFFS) algorithm is more flexible than the Naive SFS because it introduces an additional backtracking step. The first step of the algorithm is the same as the SFS algorithm which adds one feature at a time based on the objective function. The SFFS algorithm adds another step which excludes one feature at a time from the subset obtained in the first step and evaluates the new subsets. If excluding a feature increases the value of the objective function then that feature is removed and goes back

to the first step with the new reduced subset or else the algorithm is repeated from the top. This process is repeated until the required numbers of features are added, or required performance is reached (*Pudil Novovicova et al.,* 1994) (Reunanen, 2003).

The Adaptive Sequential Forward Floating Selection (ASFFS) algorithm used a parameter r which would specify the number of features to be added in the inclusion phase which was calculated adaptively. The parameter o would be used in the exclusion phase to remove a maximum number of features if it increased the performance. The ASFFS attempted to obtain a less redundant subset than the SFFS algorithm. It can be noted that a statistical distance measure can also be used as the objective function for the search algorithms (*Langley,* 1994) (Blum and Langley, 1997) (*Pudil Novovicova et al.,* 1994) (*Pudil Novovicova et al.,* 1999). Theoretically, the ASFFS should produce a better subset than SFFS, but this is dependent on the objective function and the distribution of the data.

The Plus-L-Minus-r search method also tries to avoid nesting. In the Plus-L-Minus-r search, in each cycle L variables were added and r variables were removed until the desired subset was achieved. The parameters L and r have to be chosen arbitrarily.

Nakarlyakul and Casasent, improve the SFFS algorithm by adding an extra step after the backtracking step in the normal SFFS in which a weak feature is replaced with a new better feature to form the current subset (*Pudil Novovicova et al*., 1999) (*Nakarlyakul et al.,* 2009) (*Steams*, 1976).

Literature Review of Classifiers

The Instance based k-Nearest-Neighbor algorithm is a k-nearest-neighbor classifier that uses the same distance metric. The number of nearest neighbors can be specified explicitly in the object editor or determined automatically

using leave one out cross validation focus to an upper limit given by the specified value. The distance function is used as a parameter of the search method, and the Euclidean distance is used to measure the distance. Other options include Chebyshev, Manhattan, and Minkowski distances (*Kavitha et al.,* 2010).

The Naive Bayes classifier is a straightforward probabilistic classifier stand on applying Bayes theorem with strong Naive independence assumptions. A more expressive term for the underlying probability model would be "independent feature model." An inclusive comparison with other classification algorithms in 2006 showed that Bayes classification is outperformed by other approaches, such as boosted trees or random forests (*Caruna et al.,* 2006), (*Manikandan et al.,* 2014).

The J48 algorithm builds the decision tree from labeled training data set using information gain, and it examines the same that results from choosing an attribute for splitting the data. The measure to compare the difference of impurity degrees is called information gain. The attribute with highest normalized information gain is used to make the decision. Then the algorithm recurs on smaller subsets. The splitting procedure stops if all instances in a subset belong to the same class. Then the leaf node is created in a decision tree telling to choose that class (*Trilok et al.,* 2013) (*Nurul Amin et al.*, 2015).

Multilayer Perceptron classifiers are universal function approximators, and they can be used to create mathematical models by regression analysis (Cybeako, 1989) (*Nurul Amin et al.*, 2015).

Multilayer Perceptrons are popular machine learning solution, and it finds applications in the fields such as speech recognition, image recognition, and machine translation software (*Wasseman et al.,* 1988).

Pablo Bermejo *et al.* presented a proposal that is based on the combination of the Naïve Bayes classifier with incremental wrapper Feature Subset

Selection (FSS) algorithms. The advantage of this approach is analyzed both theoretically and experimentally, and the results show a striking speedup for the embedded FSS process (*Pablo Bermejo et al.,* 2014).

Li-Min Wang *et al*., proposed a novel algorithm, Self-adaptive NBTree, which induces a hybrid of the decision tree and Naive Bayes. The Naive Bayes node helps to solve overgeneralization and overspecialization problems. The experimental results on a variety of natural domains indicate that Self-adaptive NBTree has clear advantages on the generalization ability (*Li-Min Wang et al.,* 2006).

Chan et al., compared numerically to the conventional preprocessing approaches such as data elimination, averaging, imputation to treat missing values. The efficiencies were confirmed by the classification accuracies through BayesNet, Lazy Kstar, Decision table and Part method classifiers (*Chan et al*., 2013).

Kavitha et al., presented the classifying methods ID3, J48, Naive Bayes and OneR Their result shows that ID3 and J48 method carry the highest accuracy and sensitivity with seven and fourteen attributes. The Naive Bayes holds the highest degree of the specification for all three dimensionalities (*Kavitha et al.,* 2010).

Himadri Chauhan et al. presented the comparison of different classification techniques to detect and classify intrusions into normal and abnormal behaviors. They used the J48, Naive Bayes, JRip, and OneR algorithms (*Himadri Chauhan et al*., 2014).

Anshul Goyal et al. proposed a performance evaluation of Naïve Bayes and J48 classification algorithms. The experimental results shown in the study are about classification accuracy and cost analysis. J48 gives more classification accuracy for class gender in Bank dataset having two values Male and Female. The result in the study on these datasets also shows that the efficiency and accuracy of J48 and Naive Bayes are good (*Anshul Goyal et al*., 2012).

Kaushik Raviya et al., presented the comparison of three classification techniques which are K-nearest neighbor, a Bayesian network, and Decision tree respectively. There is a direct relationship between execution time in building the tree model and the volume of data records, and also there is an indirect relationship between execution time in building the model and attribute size of the data sets (*Kaushik Raviya et al*., 2013) (*Pallavi Kulkarni et al.*, 2014).

Harish et al., presented various text representation schemes and compared different classifiers used to classify text documents to the predefined classes. The existing methods are compared and contrasted based on various parameters namely criteria used for classification, algorithms and classification time complexities. There is no single representation scheme, and classifier can be recommended as a general model for any application. Different algorithms perform differently depending on data collections. None of them appears globally superior to the other. However, to the certain extent Support Vector Machine with term Weighted Support Vector Machine representation scheme performs well in many text classification tasks (*Harish et al*., 2010).

Aurangzeb Khan et al. proposed the important techniques and methodologies that are employed in text documents classification, while at the same time making awareness of some of the interesting challenges that remain to be solved, focused mainly on text representation and machine learning techniques. Several algorithms or combination of algorithms as hybrid approaches were proposed for the automatic classification of documents (*Aurangzeb Khan et al*., 2010).

Review of Evaluation Measures

Powers and David described the systematic analysis of performance measures for classification tasks regarding Precision, Recall and F-measure (*Powers et al.*, 2011).

Recall or Sensitivity (as it is called in Psychology) is the proportion of Real Positive cases that are correctly Predicted Positive. This measures the

coverage of the Positive real cases by the +P (Predicted Positive) rule. It is desirable feature is that it reflects how many of the relevant cases the +P rule picks up. It tends not to be very highly valued in information retrieval on the assumptions that there are many relevant documents that it does not matter which subset one finds, that one cannot know anything about the relevance of documents that aren't returned. Recall tends to be neglected or averaged away in Machine Learning and Computational Linguistics (where the focus is on how confident we can be in the rule or classifier). However, in a Computational Linguistics / Machine Translation context Recall has been shown to have a major weight in predicting the success of Word Alignment (*Fraser et al.*, 2007).

Perry, Kent and Berry stated that, in information retrieval contexts, precision and recall are defined in terms of a set of retrieved documents (e.g. the list of documents produced by a web search engine for a query) and a set of relevant documents (e.g. the list of all documents on the internet that are relevant to a certain topic) (*Perry et al.,* 1955).