FEATURE SELECTION APPROACH USING FOR FILTERING ON PREDICTION ALGORITHM

D.Richard, Dr.R.Mala, Priya
Research Scholar, Asst.Professor, Student
Dept of Computer Science,
Bharathiar University, Coimbatore, India

Abstract : Data mining refers to extracting or "mining" knowledge from large amount of data. Feature selection is very much useful to choose a subset of features from data set containing more than 100 to 1000 attributes by eliminating irrelevant features to improve predictive information and filtering helps to create mining models that use subsets of data potentially viable attributes. Machine learning processing these information were used to make choice support systems. These algorithms are CART algorithm, J48 decision tree algorithm and bagging algorithm for heart attack prediction

KEYWORDS: **Feature selection, Filter method, Wrapper method, and Embedded Method.**

I. INTRODUCTION

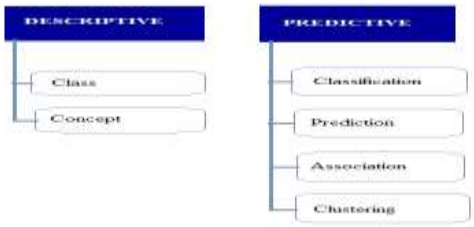
The heart force due to insufficient blood flood to a part of the heart is called a heart attack [1].early on analysis, in which clinical methods such as electrocardiography (ECG) [1] and blood tests are usually used, is a very important step in reducing sudden deaths from heart attacks. ECG is the process of the recording the electrical activity of the heart over a period of time. With the help of ECG signals anomalies in the heart can be detected. The blood as a display for a possible attack. In current years, values have been also checked for the analysis of heart attacks. Since the desire to achieve better results in health care services has increased the importance of computer-aided systems [1], they have begun to be used in addition to clinical methods. Therefore, data such as patient information, medical diagnostics, and medical images were started to be recorded [1]. Later, machine learning methods processing these data were used to build decision support systems. Some examples of these methods are CART algorithm, J48 decision tree algorithm and bagging algorithm for heart attack prediction. Data mining an interdisciplinary subfield of computer science which is the computational process of discovering patterns in large data sets by intersection of methods such as artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Feature selection is the most promising field of research in data mining in which most impressive achievements have been reported. Data based model filtering helps to create mining models that use subsets of data potentially viable attribute or feature from the feature space in mining structure of filtering gives to flexibility design data mining structures and data source for highly relevant or higher informative and also Eliminating irrelevant, irrelevant, noisy and redundant features

II.RESEARCH METHODOLOGY

Feature selection is a process that selects a subset of original features. The optimality of a feature subset is measured by an evaluation criterion. The feature selection process consists of four basic steps as shown in Figure 1, namely, subset generation, subset evaluation, stopping criterion, and result validation [5]. Subset generation is a search procedure [6] that produces candidate feature subsets for evaluation based on a certain search strategy. Each candidate subset is evaluated and compared with the previous best one according to a certain evaluation criterion. If the new subset turns out to be better, it replaces the previous best subset. The process of subset generation and evaluation is repeated until a given stopping criterion is satisfied.

2.1Pre-Process Method Diagram

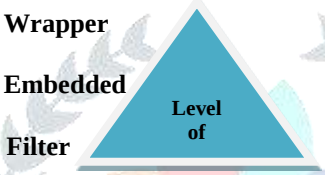
Fig 1.To view of Pre-Process Method Diagram



2.2 Feature Selection Method Diagram

Feature selection otherwise known as variable selection, attribute selection or variable subset selection. Defined as the process of selecting a subset of relevant features from the feature space for use in model construction. Reducing the features from the features space to a manageable size for processing and analysis. It is not cardinality reduction-arbitrary or predefined cutoff on the number of features

Fig 2.To view of Method Diagram



2.2. 1 Wrapper method

The feature selection using classifier and error Rate used as a measure Sub set feature selection to performance good for particular type model of computationally intensive

2.2.2 Embedded method

The takes-all group of techniques which perform feature selection and part of the model construction process.

2.2.3 Filter method

Independent of the classifier of information, correlation, distance between inter and intra class computationally fast and accuracy. Filtering method refers to the process of defining, detecting and correcting errors in given data, in order to minimize the impact of errors in input data on succeeding analysis.

IV. FEATURE SELECTION APPROACH

The Search of exhaustive, heuristic and random search for data set and evaluation of individual data set and subset data set evaluation whenever an algorithm/method for estimation of data is developed new need emerge for estimation.

4.1 General Characteristics of Filter Approach

The distances between classes for statistical dependencies between features of high number of feature can be selected above the thresh hold value of required to choose the feature selection

V. RESULT AND ANALYSIS

This paper provides the clear insight to different feature selection method. Data based model filtering helps you create mining models that use subsets of data in mining structure. Filtering gives to flexibility and design data mining structures and data source.

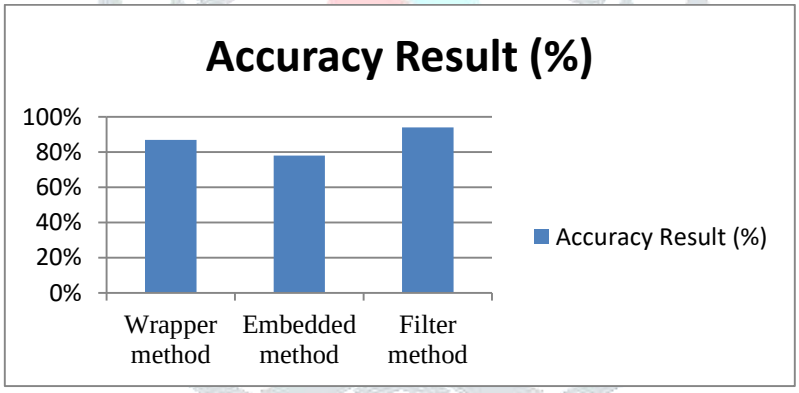
Filtering refers to the process of defining, detecting and correcting errors in given data, in order to minimize the impact of errors in input data on succeeding analysis

Fig 1.To view of Filter classifier using attributes

Table 1.Comparison of classification accuracy Result

Feature Selection	Accuracy Result (%)
Wrapper method	87%
Embedded method	78%
Filter method	94%

Fig 2 .Comparison of classification accuracy chart



VI. CONCLUSION

Feature selection has been a reported as ever green research topic with practical significance in many areas such as statistics, pattern recognition, machine learning, and data mining, web mining, text mining, image Processing, and gene microarrays analysis. These feature selection algorithms are very well useful to build simpler and more comprehensible models, improving data mining tasks performance and accuracy, and helps to understand predominant data.

REFERENCES

[1] Hidayet TAKCI.2018. Improvement of heart attack prediction by the feature selection methods, Turkish Journal of Electrical Engineering & Computer Sciences.221-240

[2] Han & Kamber. 2006. Data Mining Concepts and Techniques. Morgan Kaufmann Publishers.4 (20)

[3] Huan Liu and Lei Yu.2005. Toward Integrating Feature Selection Algorithms for classification and clustering. IEEE Transactionson Knowledge and Data Engineering, Volume17, Issue4, Pages: 491-502.

[4] M. Dash, K. Choi, P. Scheuermann, and H. Liu.2002 Feature selection for clustering – a filter solution. In Proceedings of the Second International Conference on Data Mining, pages 115–122.

[5] E. Xing, M. Jordan, and R. Karp. 2001.Feature selection for high-dimensional genomic microarray data.In Proceedings of the Eighteenth International Conference on Machine Learning, pages 601–608.

[6] M. Dash and H. Liu. 1997.Feature selection for classification. Intelligent Data Analysis: An International Journal, 1(3):131–156.

[7] P. Langley. 1994. Selection of relevant features in machine learning. In Proceedings of the AAAI Fall Symposium on Relevance, pages 140–144.