An Efficient Cancer Classification Using Mid value K-means and Naïve Bayes.

Abstract:

In the field of bioinformatics, cancer classification is important for cancer cell diagnosis . The main goal is to use machine learning to find the smallest set of genes possible.The proposed approach employs a flexible Naive Bayes classifier at first. However, due to major computational issues, the Naive Bayes is not appropriate for the classification of large datasets. Km-Naive Bayes (Naive Bayes combined with mid k-means clustering) is a quick algorithm designed to speed up both training and prediction of Naive Bayes classifiers by using the cluster centers obtained from the k-means clustering. The new techniques namely weighted Mid K-means-Naive Bayes is Implemented to improve accuracy and reduce misclassification. The algorithm was evaluated with different classifier which were applied on the same database. They classification accuracy has 88.8%

Introduction:

Cancer is a disease characterised by an uncontrolled, irregular growth that destroys and invades healthy body tissues nearby or elsewhere in the body. it is With the help of feature selection, most challenges to reduce the size of the data inside the microarray expression indeed of convenience in diagnosis of small genes during a specific type of cancer. The most challenging aspect of using this expression data is its high dimensionality, which requires a large number of samples for certain genes. This can be solved by using data preprocessing techniques such as feature extraction or feature selection to reduce the dataset size and improve accuracy. The high-density DNA microarray calculates thousands of genes at the same time, so the organic results are more accurate. phenomenon profiles are used for the cancer classification. This classification is completed so as to predict the sort of cancer person has in order that accurate treatments are often given to the person on time . Naïve Bayes is successfully

applied to the cancer identification problems. thanks to its generalization ability and has found success in many

applications. This paper presents a replacement technique for

cancer Classification Using Weighted k-means Naive Bayes

for predicting cancer cells in living organism by the technique

of ANOVA (Analysis Of Variance). Weighted K-means

(SWKM) to define weights of varied features so as to urge

un-class clusters provided a number of the objects has class

labels . Such clustering solutions are

often used for classifying unknown objects on cancer classification. from the quality optimization method point of view ELM for classification and Naïve Bayes are equivalent but ELM has

ELM for classification tends to achieve better generalisation efficiency than conventional Nave Bayes due to its unique separability function, which is analysed in theory and further checked by simulation results.

Dr : Sara El-Metwally

Name:Yosra Emad khattab