

PREDICTION OF ORAL CANCER BY NAIVE BAYESIAN

VIJAY ANAND V, VINOTH M, GANESH KUMAR S, SAMUEL PAULSON SITE, VIT University,
Vellore – 632014, Tamil Nadu, India
vijayanandv89@gmail.com

Abstract

Among the deadliest disease in the world cancer is one among them. Oral cancer is one of them. A survey among the developing countries of South Asia says that, one third of the person has a chance of having cancer. In India Oral Cancer is most widely spread. In this project we use two algorithms such as Naive Bayesian and Support Vector Machine. We compare results of both algorithms to show which algorithm is best. Here cancer patient database is built and data mining techniques is applied on it for analysing data. Keywords: IPPSCD – Intelligent Prognosis Prediction System for Cancer Disease; SVM – Support Vector Machine; LIF – Laser Induced Fluorescence; CCD – Charge Coupled Device; HPLC – High-Performance Liquid Chromatography.

INTRODUCTION

Oral cancer occurred when tumours is formed in oral cavity and two major item which is related to oral cancer is tobacco and alcohol usage. Our aim is to compare two algorithms i.e., Naive Bayesian and SVM to predict cancer at early stage. Our main goal is to build a model for cancer prediction using its symptoms and algorithm. We use attributes like doctor name, patient name, and patient habits like alcohol, smoking, ulcer status. There are many classification algorithm but we use only naive Bayesians and SVM algorithm because Naive Bayesian algorithm is simple among classification algorithm and SVM algorithm provides accurate among classification algorithm.

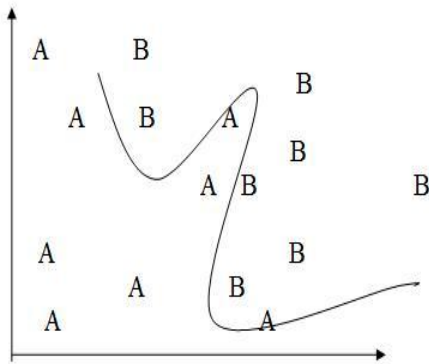
METHODS USED SUPPORT VECTOR MACHINE

SVM is mainly learning method used

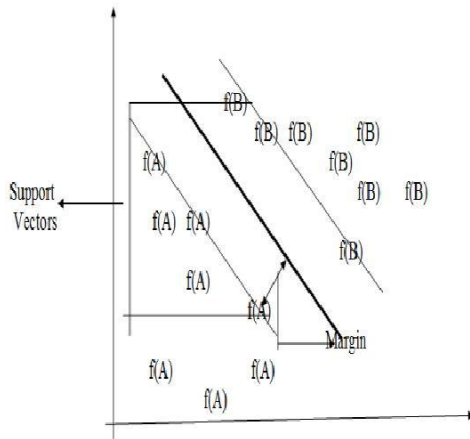
for classifying and for regression analysis. In SVM the user gives the input as a set and it predicts for each input, which makes support vector machine a non probabilistic in nature.

In other words, SVM algorithm predicts whether given input falls in which category. SVM model is used to represent example as point in space and it maps example to its category and divides it as much as possible. SVM is used in constructing hyper plane. Hyper plane may be dimension of infinite space which is used for classifying and regression analysis. Simultaneously best separation is achieved through hyper plane that has distance to nearest data points, in general maximum the margin minimum the generalization error occurs for the classifier.

SVM has the kernel function. They use kernel to transform problems. When kernel function is fixed, it uses kernel equation to arrange data instance. First limitation is user – chose, second is size and speed, in testing and training. Third one is efficiency lacking in problem handling.



ORIGINAL SPACE



NEW SPACE

Basic idea behind Support Vector Machine, We can clearly see from the above diagram, data instances which were not linearly separable in the original space have become linearly separable in new space, this is because of the application of kernel (transforms the position of data). When new instance occurs in the original domain, same kernel is applied and its position in the new domain is found.

NAIVE BAYESIAN THEOREM

variables from data sample through close observation of each sample independently.

Naive Bayesian theorem consists of conditional probability. It uses all the attributes contained in the data and analyses it individually. It works in the assumption that one attribute is independent to other attributes in the sample. Naive Bayesian theorem is used in finding the probability of occurring event for given probability of another that has occurred.

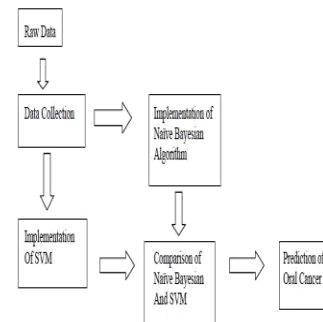
$$\text{Probability}(B \text{ given } A) = \text{Probability}(A \text{ and } B) / \text{Probability}(A)$$

For given A, calculate the probability of B, it counts no of times where A as occurred.

In this study, we are getting attributes such as doctor name, patient name, and patient habits like alcohol, smoking, ulcer status and prediction of cancer. When new patient comes the patient data is compared to the collected data attributes in the database to predict whether cancer is going to occur or not.

EXPERIMENTS

Naive Bayesian is the simplest in classification algorithms but in some cases it outperforms other complex algorithms. It uses



Raw data of patient are collected which comprises of doctor name, patient name, and health status of the patient. Patient health

comprises of attributes such as smoking, alcohol addiction and many more the Bayesian algorithm is applied on the collected data and its output is output is retrieved. Likewise SVM is applied on the same data and its output is retrieved. Both the outputs of the algorithm are compared to achieve good accurate result.

CONCLUSION

Naive Bayesian theorem is successfully implemented. The comparison between Naive Bayesian theorem and SVM is done easy and simple. By implementing both the theorem we can said that Bayesian theorem is best for prediction of oral cancer.

REFERENCES

- [1] Mehmed K : “Data Mining : concepts,models,Methods And Algorithms”,New Jersey:John Wiley,2003
- [2] Tang, Z.H., MacLenman, J.;”Dta Mining with QL Server 2005”, Indianapolis: Wiley, 2005.
- [3] Ho, T.J.:”Data Mining and Data Warehousing”, Prentice Hall, 2005.
- [4] Margaret H.Dunham.:”Dta Minig Introductory and Advanced Topics”, Pearson Education, 2006.
- [5] A B M Shawkat Ali, Salef A.Waimi.:”Data Mining: Method And techniques”, Indian Edition, 2009.