ABSTRACT

Communication system plays an important role in transmitting a message from source to destination. But this process of communication is affected by noise. To reduce the effect of noise and develop a strong system of communication, many researchers have contributed to improve this process. Development of information theory is a milestone in this direction. It gave the solution to many problems of communication system, but got boost after the remarkable work of C. E. Shannon. His work revolves around compression, lossy compression and channel capacity. Shannon model transmits the message from the source in encoded form and decodes it at destination. Later, researchers worked on the information measure introduced by Shannon and generalized the same to different forms as per the requirements of the problem under consideration.

The subject of this thesis is a theoretical study of information theoretic measures. It includes a detailed analysis of various generalizations of information related measures and their properties and their applications in different fields like; mechanics, finance, data mining etc. In the present work, we have applied information theoretic measures in one of its area of application i.e. data mining. We have also studied extensions of weighted residual information measure and its application in classification of data using algorithms of data mining. In addition, for more improvement, methods of classification are also compared and the result is analyzed.

Initially, among various information measures, we have studied about Shannon entropy, its characterization, generalization and as well as its properties. The role of entropy in different forms with their respective areas of applications and importance is analyzed. The characterized forms of information theoretic measures has a unique set of properties that are satisfied by a particular measure and are applied according to the nature of data to be dealt with, like for continuous values the concept of differential en-

tropy is used, while for discrete values the concept of Shannon entropy is appropriate.

With the increasing number of data ware houses, it becomes necessary to find the potentially valuable information which can be helpful in finding the solution of many related problems like prediction of future behavior of data, retainment of potential customers etc. Solutions to such problems can be obtained by data mining, with help of its various tasks like analysis, classification, prediction etc. Classification is one of important task of data mining which can be done by using various methods. ID3 algorithm, a well known classification algorithm, induce decision tree with the help of information gain. The role of information theoretic measures in ID3 algorithm is to find the amount of information gain, which is used as the splitting criteria for the data under consideration. We have applied ID3 algorithm to the data extracted from the census 2011 of India having four different attributes with some classes, to be classified into States and Union Territories. Decision tree was developed using Renyi entropy for particular values of its parameter and a set of rules were developed, on the basis of decision tree, which help in deciding the region for implementation of a new project.

In continuation with the above concept, we apply Varma entropy based ID3 algorithm for particular values of its two parameters to analyze health services in States and Union Territories of India based upon the data up to 2015. For the same data, with discrete values without intervals of an attribute, we applied C4.5 algorithm for the development of decision tree and these two resultant decision trees, obtained from using Varma entropy and C4.5 algorithm, are compared on the basis of rules induced. The study of information theoretic measures is further extended towards development of weighted generalized residual information measure of order α and type β, a generalized form of Varma entropy, which also characterize various distribution function.