Abstract

Information regarding forest land is highly required for developing ecosystem management. This paper provides an analysis related to classification and prediction estimation using machine learning techniques. The approach is to predict the forest cover type using the cartographic variables like aspect, slope, soil type, wilderness area etc. Various Data mining techniques such as decision trees, random forest, regression trees, and gradient boosting machines are used for prediction of the forest cover type. Using these machine learning methods models have been developed and tested for accuracy ranging from 19.4% to 74.8%. Kaggle dataset which is the standard benchmarking dataset, is taken for comparison studies. The comparisons of these models are done to identify a better model for predicting the forest cover type with better accuracy. For performance comparison, metrics like accuracy and error rate are used. An important aspect of the study is the use of different performance measures to evaluate the learning methods.
References

2. Jiawei Han and Micheline Kamber, “Data Mining Concepts and Techniques”, second edition Morgan Kaufmann publisher.  
4. Decision Trees for Business Intelligence and Data Mining: Using SAS Enterprise Miner.  

Index Terms

Computer Science

Information Sciences

Keywords

Machine learning, classification and regression, decision trees, random forest, gradient boosting machines.