

Introduction

In the forest cover dataset the goal is to predict the type of cover that a certain has based on the observation of the environment. The dataset provides multiple features like soil type, area elevation, distance from water as we described in the first report. We have 7 types of cover type and each row in the data frame is associated with only a single type, thus the problem is known as multi-class classification. The Kaggle competition website states that the model submitted is simply evaluated on the accuracy to classify the cover type on the test data, the value is returned as a percentage so 100% is perfect accuracy and obviously 0% the model is useless.

Data Analysis

Interesting Features

Data Processing Findings

Conclusion

References