

## Problem Statement: Data Analytics

Natural resource managers responsible for developing ecosystem management strategies require basic descriptive information including inventory data for forested land to support their decision making processes. However managers do not have this type of data for inholdings or neighbouring lands that are outside their immediate jurisdiction. One method for obtaining this information is through the use of predictive models.

Given is the attribute name, attribute type, the measurement unit and a brief description. The forest cover type is a classification problem. The order of this listing corresponds to order of numerals along the rows of database.

Different factors that are taken as follows:-

Index, Elevation in metres, Aspects in degree azimuth, Slope in degrees, Horizontal distance to hydrology i.e Horizontal distance to nearest surface water features, Vertical distance to hydrology i.e Vertical distance to nearest surface water features, Horizontal distance to roadways, Hillshade\_9A.M i.e Hillshade index at 9 AM, summer solstice, Hillshade\_Noon,. Rawah and Neota are wilderness areas that are encoded as one if present otherwise zero.

Candidates are required to classify whether the forest cover belongs to category one or category two. Category one shows forest cover of Spruce /Fir type. Category zero shows forest cover of Lodepole Pine.

Your model performance will be evaluated on the basis of your prediction of classification of forest cover for test data which contains similar data points as train except for forest cover. Your submission need to be in format as shown in "submission.prakriti.csv".

We at our end have actual forest cover for test data set against which your prediction will be evaluated. Submission will be scored on the basis of AUC-ROC score.

## Rules & Regulations:

- 1) Only teams whose members have registered online would be able to participate.
- 2) Each team can contain a maximum of Four (4) participants.
- 3) Students from different colleges can form a single team.
- 4) For the preliminary round, each team has to prepare an abstract of the solution based on the problem statement (abstract should not be more than 6 pages).
- 5) The abstract submission has to be a word/pdf document which should be mailed to prakritikgp@gmail.com. The abstract should contain team Name, Name of the team members, College and the subject of the mail as “Data Analytics submission”.
- 6) The whole analysis of the problem statement should be done by the participants themselves. Plagiarism will attract cancellation of participation.
- 7) Teams selected for the second round have to come IIT Kharagpur for the final presentation.
- 8) Teams presenting their analysis in the second round should come with a proper presentation along with their word document submitted in the first round.
- 9) The decision of judges is final and binding. Team Prakriti reserves the rights to disqualify any team in case of violation of any of the above rules.