

PROPOSAL

Mohammed A. Saeedi

Project Proposal

msaeedi@mrda.gov.sa

10/1/2021

First Stage the dataset:

- My dataset cover type type from [here](#).
- I will describe the dataset and the classification task.

Second stage data exploration:

- I will display number of instances, number of attributes and number of classes.
- For each class label, I will display the code of the class label and the name of that class.
- I will summarize the class distribution using a suitable graph.
- I will display a statistical summary for all the attributes.

Third stage data preprocessing:

- I will check data quality issue with suitable strategies to deal with any issue (if exists).
- I will convert the multiclass classification problem into a binary classification problem.
- I will use a features selection technique to select those features in my data that contribute most to the prediction.
- I will divide my dataset into training, validation and testing datasets.

Fourth stage classification:

- I will build classification models , three different learning algorithms to generate three classification models.
- For each classification model I will find the most accurate classifier.

Fifth stage evaluation:

- I will evaluate my classification models on the validation and the testing datasets.
- For each classification model, I will print out a confusion matrix for the validation and testing datasets.
- I will use the following evaluation measures to evaluate the performance of the generated classification models accuracy, error rate and F –measure.
- I will compare between the performances of all the classification models using suitable chart.