

IE615: Data Analytics for Operations Research

Assignment 2

Arun R
163190013

1) Classification Problem

Implementation Details:

- The columns with string entries (last two columns) are manually converted into integer entries
- The first column (serial number of employees) is not considered for learning.
- CART is used for constructing trees in Random Forests.
- Error is the ratio of number of misclassified data to total number of data.

Result:

HRA Dataset	Training Error	Test Error
CART	0	0.020170
C4.5	0	0.020170
Random Forests	0.007556 N=50	0.018503

