

Eclerx Recruitment Assignment: Data Science

About Data:

You have a file named as 'Opportunity.csv' which contains all the Deals information extracted from Salesforce instance. There are both closed deals and open deals in the data. Any deal which is closed can be of 2 types – Won or Lost

- Where "isWon"= true then the Deal is a win
- Where "isWon"=false then Deal is a Loss
- Where "isClosed"=true then Deal is Closed
- Where "isClosed"= false then Deal is open

Assignment

The assignment is to predict the Open Opportunity (Deals) outcome based on the historical data of closed opportunities.

–Out of the closed deals data, please keep 75% of the data as training dataset and rest 25% as the test dataset. Also predict the output for open deals.

–Use the model built to score the open deals.

–We will later have a discussion on why you choose one algorithm over the other, e.g. (logistic regression over Decision Tree)

- **Send back** the solution with the following files:

- Input (training dataset)
- Validation Dataset (after scoring)
- Open Deals Dataset (after scoring)
- Model Code(R/SAS/Python/Alteryx or any other tool of choice)