**Case Study:-**

Insurance Company has asked you to develop a predictive classifier model to determine fraudulent claims. The firm has provided you with a sample of 3000 historical example of claims described in terms of:

* Injured body part;
* The nature and cause of injury; and,
* Adjustor notes taken by insurance employees when in contact with the claimants or their employers.

After a lengthy process, each of the claims has been verified and annotated with the flags indicating:

* If the injury involved a vehicle (whether or not stated in the claim);
* Whether or not fraudulent claims have been detected and the applicant eventually sued.

**Requirements:-**

* Determine appropriate attributes as labels and predictor and explain the reason for selection.
* Determine the best classifier algorithm most k-NN, Decision Tree and Gradient Booster Tree.
* Build the model and tune the performance in terms of confusion matrix and kappa.
* Analyse ROC/AUC curves (true positives vs false positives) and find appropriate threshold parameter.

**Data:**



**Note to the candidate:**

* Feel free to use R, SAS, Python or Rapid Miner
* Please document and provide us with snapshot of your approach and recommendations in word document along with any visuals
* Provide rational for selecting the statistical approach, any observations and your recommendation
* Submit the outcomes at least 2 days before your interview.
* 30 mins of the interview will include discussion on your workings

**Further Clarification/Queries:**

For any further clarifications/queries, please contact Swaranjit Singh ([Swaranjit.Singh@anz.com](mailto:Swaranjit.Singh@anz.com)) AND Manmohan Toshniwal ([Manmohan.Toshniwal3@anz.com](mailto:Manmohan.Toshniwal3@anz.com))