**Problem Statement:**

Each year customers of an insurance company is impacted by Catastrophe (CAT) events. These are typically weather related (e.g. cyclone, bushfire or earthquakes) and cause millions of dollars in damage. QBE has reinsurance policies to mitigate the impact of these claims. It is therefore important that the insurer accurately identifies, and records claims associated with CAT events.

**Objective:**

When a new claim is lodged, amongst their other tasks, the claims officer should identify if the claim is associated with a CAT event and then apply the relevant CAT code to that claim. As with all process there is a chance for leakage to occur. The objective is to identify claims that are related to a CAT event and have potentially missed by the claims officer. Use CAT\_FLAG as the target variable in the data.

**Guidelines:**

* The data will be split into **70-20-10.** 90% data along with target variable and **10% Test data without target** is shared. Variable descriptions are also shared at the below link:
* **Submission** **docket to be uploaded on sharepoint site by 8:00 AM IST On Sunday, 10th June**:
  + **Email Subject Line: DSC | <Employee ID> | Train AUC | Val AUC**. Example: DSC | F00400 | Train: 84.36% | Val: 83.45%
  + **Scoring code in Python or R –** Filename: <Employee ID>\_Scoring Code\_<programming language>. Eg: F00400\_Scoring Code\_Py or F00400\_Scoring Code\_R
  + **Scored 10% test dataset –** Filename: <Employee ID>\_Test.csv. The dataset should be a csv file with only two columns: R\_UNIQUEID and <Employee ID\_Score>
  + **Approach note –** Very short – just models used, individual model results (AUC in Train – 70% and Val – 20%) and final model result – Filename: <Employee ID>\_Approach Note.docx
* Only **one submission** is allowed per email ID
* The submissions should be shared with Himanshu Sharma ([himanshu.sharma@fractalanalytics.com](mailto:himanshu.sharma@fractalanalytics.com)); Bibhas Dey ([Bibhas.dey@fractalanalytics.com](mailto:Bibhas.dey@fractalanalytics.com)) and Neelima Naidu ([neelima.naidu@fractalanalytics.com](mailto:neelima.naidu@fractalanalytics.com)) before deadline
* We will have a **Temp shared excel** in the sharepoint site, and you can update your Train and Val AUC to see where you are standing before making submissions. This will become useful only if everyone taking part update the excel
* **No restriction** on team size, etc.
* There is **scope for text mining and Deep Learning** in the problem statement
* **NO QUESTIONS TO BE ASKED** to the core team
* Submissions not meeting any of these guidelines (not following naming conventions, doing multiple submissions, etc.) will be disqualified
* The top 3 entries will have to convert their scoring code to R by Sunday EOD IST if the initial code is developed in Python. They will be contacted by the core team if needed for this on Sunday
* Any ideas are welcome by the participants if they couldn’t try something due to time crunch. They can mention this to the core team in the email during submission

Reward:

* Top 10 entries based on the AUC on a hold-out 10% sample will be published to Fractal All
* The core team will also be building models in parallel. Top 3 entries will be announced only if the **Test AUC is better than the benchmark models** **developed** by the core team
* Reward of vouchers worth INR 35,000 for First, INR 15,000 for Second and INR 10,000 for Third