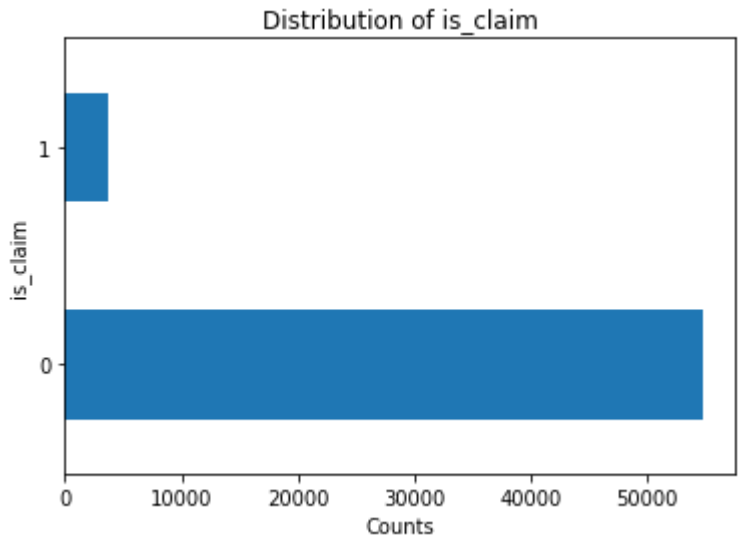
Approach

Team Name: **Chinmay_ee**
Members Name: **Chinmayee Bharati Dalai and Chinmay Kumar Das**

Catboost is all you need!!

- This problem statement deals with the classification with a lot of categorical features.
- As we checked, there were no null values in any features, which saved us much time which otherwise would have taken for the null values handling.
- And this is a case of an unbalanced dataset so we treated this differently and used some specific parameters of different classification models.



Feature Engineering :

- These are the columns that are boolean in nature so we will replace the value of ‘Yes’ with 1 and ‘No’ with 0.

['is_esc', 'is_adjustable_steering', 'is_tpms', 'is_parking_sensors', 'is_parking_camera',
'is_front_fog_lights', 'is_rear_window_wiper', 'is_rear_window_washer',
'is_rear_window_defogger', 'is_brake_assist',
'is_power_door_locks','is_central_locking','is_power_steering',
'is_driver_seat_height_adjustable', 'is_day_night_rear_view_mirror','is_ecw', 'is_speed_alert']

- Dummy columns :
  For the mentioned columns we will do one-hot encoding as these are independent values inside it.
  ['area_cluster', 'segment', 'model', 'fuel_type', 'rear_brakes_type', 'engine_type']

- Ordinal columns
  These are columns we replace the value in terms of there priority value
  Columns : ['transmission_type', 'steering_type']
  Ex: {'Manual': 1, 'Automatic': 2}
  {'Manual': 1, 'Power': 2, 'Electric': 3}

- Compound Features
  ['max_torque', 'max_power']: These features are split into 4 numerical features.

Modeling :

- We tried with following models
  - Random Forest
  - CatBoost
  - XGBoost

- As it is an unbalanced dataset the model with default parameters all gave 0 as output initially.
- As datasets have a lot of categorical value we thought of using Catboost.
- So tried Hyperparameter tuning with the Cateboost model, seems like default parameters work best with **scale_pos_weight as 10.**

Could have Tried:

- Ensemble of different parameter models optimized with f1 score and cross-validation.

Link to the notebook:

https://github.com/chinmayeeb96/Analytics_Vidhya_Dataverse_Hackathon

Github Ids: https://github.com/chinmayeeb96
https://github.com/chinmaydas96