Dataverse Hack

Online 13-11-2022 12:00 AM to 14-11-2022 12:00 AM

1963 Registered | **190** Number of Teams | **Win Prizes worth 1.5 lakh+ ($1800+)** Prizes

ENDS IN

15 HOURS 6 MIN 25 SEC

Registered

About Problem Statement Solution Checker My Submissions Leaderboard Team Discuss

Insurance Claim Prediction

Predict whether the policyholder will file a claim in the next 6 months or not.

Problem Statement

CarIns is a startup that provides insurance for cars. It is one of the best car insurance brands known for the highest claim settlement ratio. It was launched back in Oct 2020 and acquired its initial policyholders by providing a hassle-free claim process, instant policy issuance, and claim settlements at minimum coverages.

As it's a fast growing startup, the company would like to optimize the cost of the insurance by identifying the policyholders who are more likely to claim in the next 6 months.

Now the company would like to use Data Science to identify the policyholders whose chances of filing a claim are high in the next 6 months. The company challenges the Data Science community to build a high-performance algorithm to predict if the policyholder will file a claim in the next 6 months or not based on the set of car and policy features.

About the Dataset

You are provided with information on policyholders containing the attributes like policy tenure, age of the car, age of the car owner, population density of the city, make and model of the car, power, engine type, etc and the target variable indicating whether the policyholder files a claim in the next 6 months or not.

Data Dictionary

You are provided with 3 files - train.csv, test.csv, and sample_submission.csv

Train and Test Set

The train and test set contains information about different insurance policy holders. The train set includes the target variable *is_claim* whereas in the test set, you need to predict the target variable *is_claim.*

Variable	Description
policy_id	Unique identifier of the policyholder
policy_tenure	Time period of the policy
age_of_car	Normalized age of the car in years
age_of_policyholder	Normalized age of policyholder in years
area_cluster	Area cluster of the policyholder
population density	Population density of the city (Policyholder City)
make	Encoded Manufacturer/company of the car
segment	Segment of the car (A/ B1/ B2/ C1/ C2)
model	Encoded name of the car
fuel_type	Type of fuel used by the car
max_torque	Maximum Torque generated by the car (Nm@rpm)
max_power	Maximum Power generated by the car (bhp@rpm)
engine_type	Type of engine used in the car
airbags	Number of airbags installed in the car
is_esc	Boolean flag indicating whether Electronic Stability Control (ESC) is present in the car or not.
is_adjustable_steering	Boolean flag indicating whether the steering wheel of the car is adjustable or not.
is_tpms	Boolean flag indicating whether Tyre Pressure Monitoring System (TPMS) is present in the car or not.
is_parking_sensors	Boolean flag indicating whether parking sensors are present in the car or not.
is_parking_camera	Boolean flag indicating whether the parking camera is present in the car or not.
rear_brakes_type	Type of brakes used in the rear of the car
displacement	Engine displacement of the car (cc)
cylinder	Number of cylinders present in the engine of the car
transmission_type	Transmission type of the car
gear_box	Number of gears in the car
steering_type	Type of the power steering present in the car
turning_radius	The space a vehicle needs to make a certain turn (Meters)
length	Length of the car (Millimetre)
width	Width of the car (Millimetre)
height	Height of the car (Millimetre)
gross_weight	The maximum allowable weight of the fully-loaded car, including passengers, cargo and equipment (Kg)
is_front_fog_lights	Boolean flag indicating whether front fog lights are available in the car or not.
is_rear_window_wiper	Boolean flag indicating whether the rear window wiper is available in the car or not.
is_rear_window_washer	Boolean flag indicating whether the rear window washer is available in the car or not.
is_rear_window_defogger	Boolean flag indicating whether rear window defogger is available in the car or not.
is_brake_assist	Boolean flag indicating whether the brake assistance feature is available in the car or not.
is_power_door_lock	Boolean flag indicating whether a power door lock is available in the car or not.
is_central_locking	Boolean flag indicating whether the central locking feature is available in the car or not.
is_power_steering	Boolean flag indicating whether power steering is available in the car or not.
is_driver_seat_height_adjustable	Boolean flag indicating whether the height of the driver seat is adjustable or not.
is_day_night_rear_view_mirror	Boolean flag indicating whether day & night rearview mirror is present in the car or not.
is_ecw	Boolean flag indicating whether Engine Check Warning (ECW) is available in the car or not.
is_speed_alert	Boolean flag indicating whether the speed alert system is available in the car or not.
ncap_rating	Safety rating given by NCAP (out of 5)
is_claim	Outcome: Boolean flag indicating whether the policyholder file a claim in the next 6 months or not.

Submission File Format

The solution file should follow a format similar to that of the sample submission file**. sample_submission.csv** contains 2 variables - policy_id and is_claim.

Variable	Description

Support

policy_id	Unique identifier of the policyholder
is_claim	Outcome: Boolean flag indicating whether the policyholder file a claim in the next 6 months or not

Evaluation metric

The evaluation metric for this hackathon would be **F1 score**.

Public and Private Split

Test data is further divided into Public (**40**%) and Private (**60**%) data.

Your initial responses will be checked and scored on the Public data. The final rankings would be based on your private score which will be published once the competition is over

Submission Tutorials

1. All Submissions are to be done at the solution checker tab.
2. For a step-by-step view on how to make a submission check the below video

Guidelines for Final Submission

Please ensure that your final submission includes the following:

1. Solution file containing the predictions for the *policy_id* in the test set (Format is given in sample_submission.csv)
2. **A zipped file containing code & approach**
    a. **Code**: Clean code with comments on each part
    b. **Approach**: Please share your approach to solve the problem (doc/ppt/pdf format). It should cover the following topics:
        i. A brief on the approach used to solve the problem.
        ii. Which Data-preprocessing / Feature Engineering ideas really worked? How did you discover them?
        iii. What does your final model look like? How did you reach it?

Hackathon Rules and Conditions

1. The final standings would be based on the private leaderboard score.
2. Setting the final submission is recommended. Without a final submission, the submission corresponding to the best public score will be taken as the final submission
3. Use of external data is not allowed.
4. The submitted code file must be able to reproduce a similar score to that of the final submission file.
5. The code file pertaining to your final submission is mandatory while setting final submission.
6. Entries submitted after the contest is closed, will not be considered

7. Throughout the hackathon, you are expected to respect fellow hackers and act with high integrity.
8. Use of multiple Login IDs will lead to immediate disqualification
9. Analytics Vidhya holds the right to disqualify any participant at any stage of the competition if the participant(s) are deemed to be acting fraudulently.

Data

Test File | Train File | Sample Submissions