Car insurance

CAR INSURANCE

CAR INSURANCE CLAIMS CLASSIFICATION

BY:
AMAL ALTAMRAN,
AMIRAH ALOTAIBI,
ALANOUD ALHUSSAIN.

FINAL REPORT
SADAIA T5 BOOTCAMP

INTRODUCTION:

IN THE EVENT OF AN ACCIDENT TO THE INSURED MOTOR VEHICLE, THE COMPANY WILL REIMBURSE THE INSURED FOR THE ACCIDENTAL DAMAGES SUSTAINED BY THE INSURED VEHICLE, THIRD PARTY'S VEHICLE INVOLVED IN THE ACCIDENT, HIS PROPERTY AND BODILY INJURIES SUFFERED OR DEATH IN ACCORDANCE WITH THE TERMS AND CONDITIONS OF THE INSURANCE POLICY

DESIGN:

THE DATASETS WE USED IN THIS PROJECT IS TAKEN FROM KAGGLE WEBSITE.

DATA DESCRIPTION:

A MODEL PERFORMANCE DEPENDS HEAVILY ON THE DATA IT WAS TRAINED ON . TO ACQUIRE THE DATA, WE USED DATA FORM KAGGLE AND PREDEFINED DATA AS A DATA FRAME. MOREOVER, THE DATA SET CONTAINS MULTIPLE FIELDS WITH DIFFERENT DATA TYPES.

IN THE FEATURES WE HAVE 21 COLUMNS AND WE TAKE THE 6 MOST IMPORTANT FEATURES AS FOLLOWING:

1-AGE

2-DRIVING EXPERIENCE

3-VEHICLE OWNERSHIP

4-SPEEDING VIOLATIONS

5-DUIS (DRIVING UNDER THE INFLUENCE OF ALCOHOL)

6-PAST ACCIDENT

TARGET:

DID THE CUSTOMER HAS CLAIMED HIS/HER LOAN OR NOT.

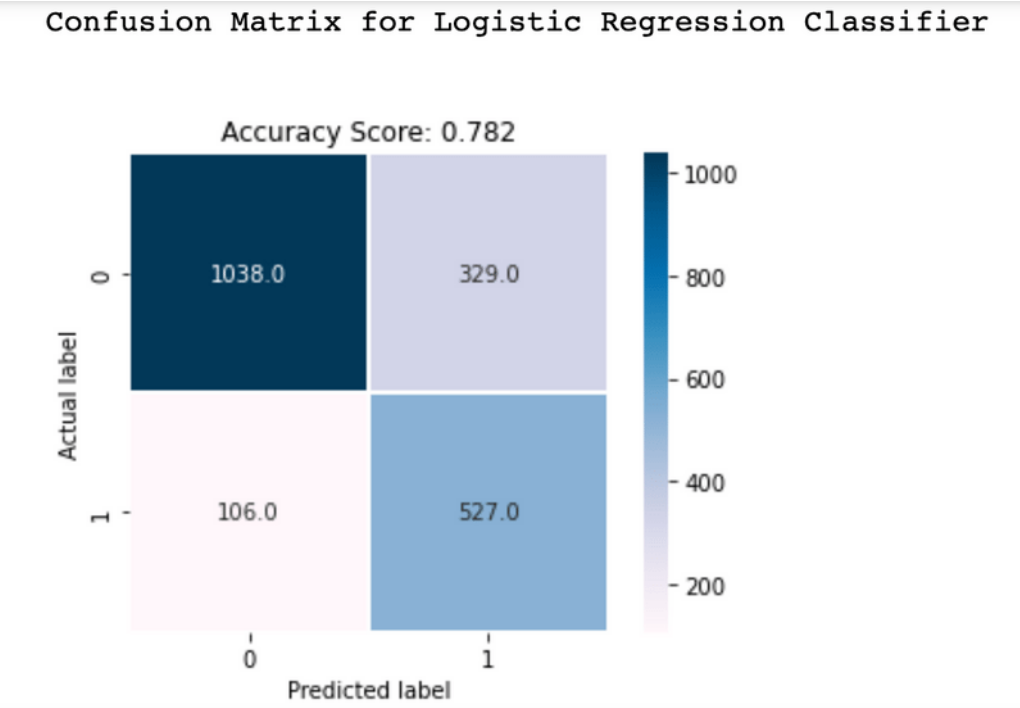
WORKFLOW:

AFTER LOADING THE DATA FROM KAGGLE WE START EXPLORING AND CLEANING THE DATA BY USING ALL FUNCTIONS LIKE INFO AND DESCRIBE.THEN CLEANING THE DATA BY REMOVING NULL VALUES, DUPLICATES AND REMOVING ANY IRRELEVANT DATA . AFTER THAT, PLOT THE GRAPHS USING MATPLOTLIB MODULE FROM PYTHON. FINALLY, FIND THE CORRELATION BETWEEN FEATURES AND FIT THE CLASSIFICATION MODEL IN ORDER TO FIND THE BEST MODEL THAT FITS OUR DATA.

WE USE 9 MODELS TO SEE WHAT THE BEST FOR OUR DATASET:

1.RANDOM FOREST
2.DECISION TREE
3.KNN
4.LOGISTIC REGRESSION
5.SVC
6.ADABOOST CLASSIFIER
7.XGB CLASSIFIER
8.GRADIENT BOOSTING CLASSIFIER
9.CATBOOST REGRESSOR

CONCLUSION



THIS PROJECT SHOWED THAT THE LOGISTIC REGRESSION IS THE BEST MODEL BECAUSE THE TRAINING ACCURACY 78%,TEST ACCURACY 78% AND F1 SCORE 70%.

FUTURE WORK WILL BE EXPLORING MORE FEATURES SUCH AS GENDER AND CREDIT SCORE.