Classifying car insurance claims

Albina Sitdikova
02/10/2021
Metis Bootcamp

Data

Dataset:

French Motor Claims Datasets

Target:

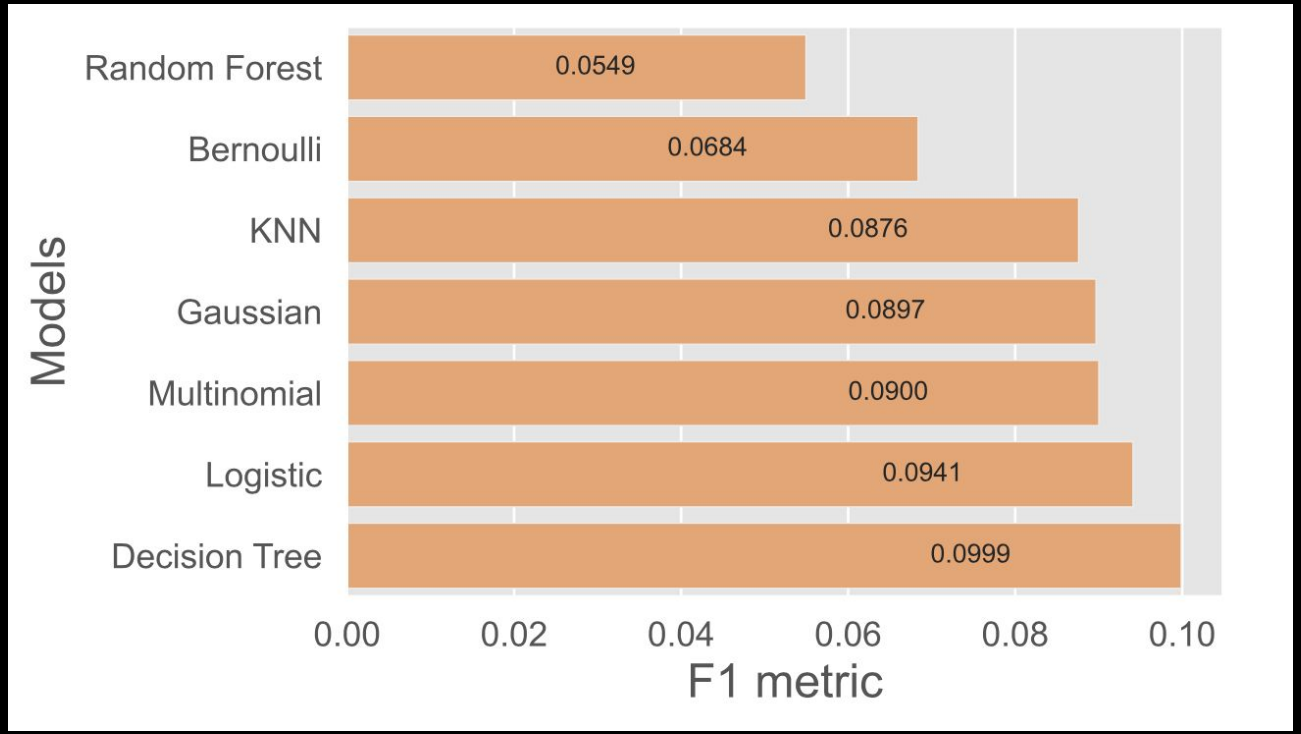
Claim for next year (yes/no)

target features

IDpol	Claim	Area	VehPower	VehAge	DrivAge	BonusMalus	VehBrand	VehGas	Density	Region
1.0	1	D	5	0	55	50	B12	Regular	1217	R82
3.0	1	D	5	0	55	50	B12	Regular	1217	R82
5.0	1	B	6	2	52	50	B12	Diesel	54	R22
10.0	1	B	7	0	46	50	B12	Diesel	76	R72
11.0	1	B	7	0	46	50	B12	Diesel	76	R72

Methodology

F1 metric for each model



Methodology

F1 score on Logistic Regression

F1 = 0.0941

Tuning hyperparameters

F1 = 0.0989

Adjusting threshold

F1 = 0.104

Feature importance

VehPower
VehAge
DrivAge
BonusMalus
Density

Area_A
Area_B
Area_C
Area_D
Area_E
Area_F

VehBrand_B1
VehBrand_B10
VehBrand_B11
VehBrand_B12
VehBrand_B13
VehBrand_B14
VehBrand_B2
VehBrand_B3
VehBrand_B4
VehBrand_B5
VehBrand_B6

VehGas_Diesel
VehGas_Regular

Region_R11
Region_R21
Region_R22
Region_R23
Region_R24
Region_R25
Region_R26
Region_R31
Region_R41
Region_R42
Region_R43
Region_R52
Region_R53
Region_R54
Region_R72
Region_R73
Region_R74
Region_R82
Region_R83
Region_R91
Region_R93
Region_R94

Conclusion

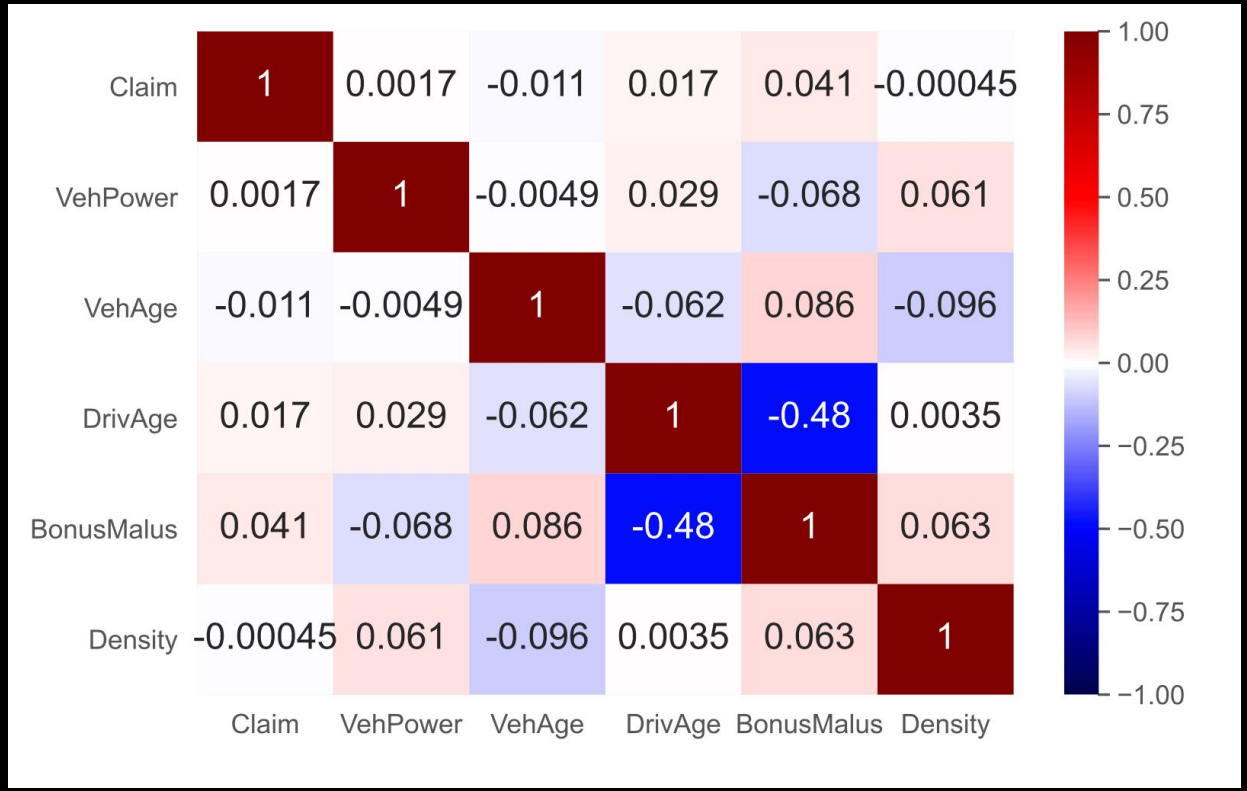
- To classify model was chosen F1 metric
- F1 = 0.104 after adjusting
- Metric doesn’t give sufficient result, analysis should be improve
-Feature importance was found

Feature work

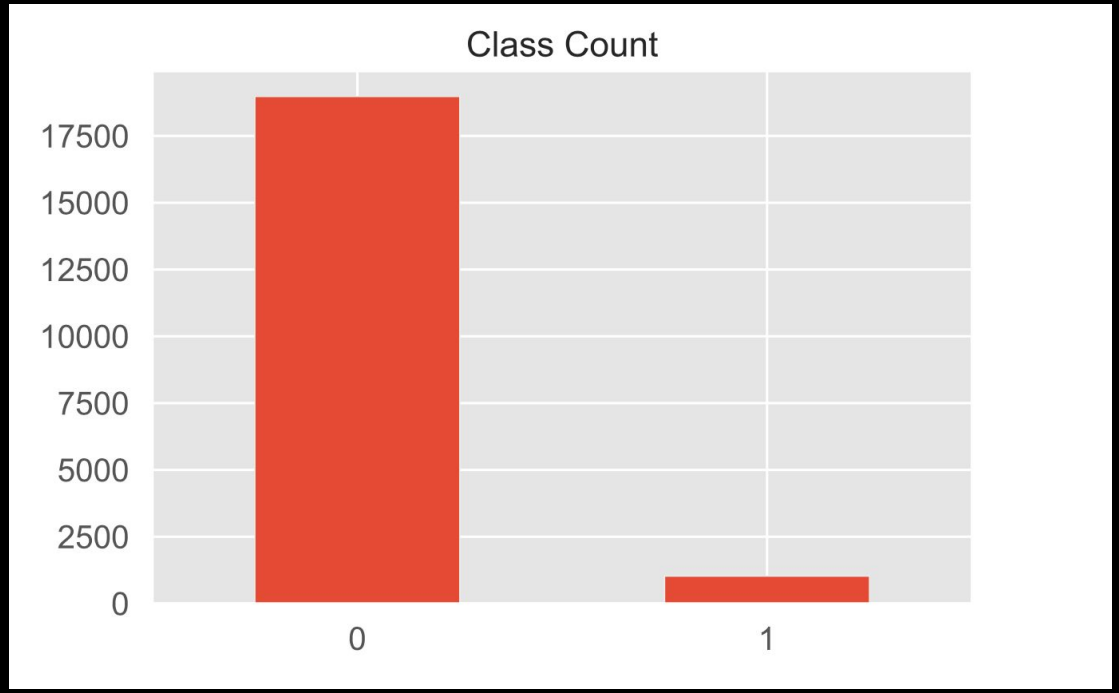
- Add more data
- Improve analysis by using other type of regression and different methodology
- Create an app that will recommend insurance company if they should give higher price for the client

Appendix

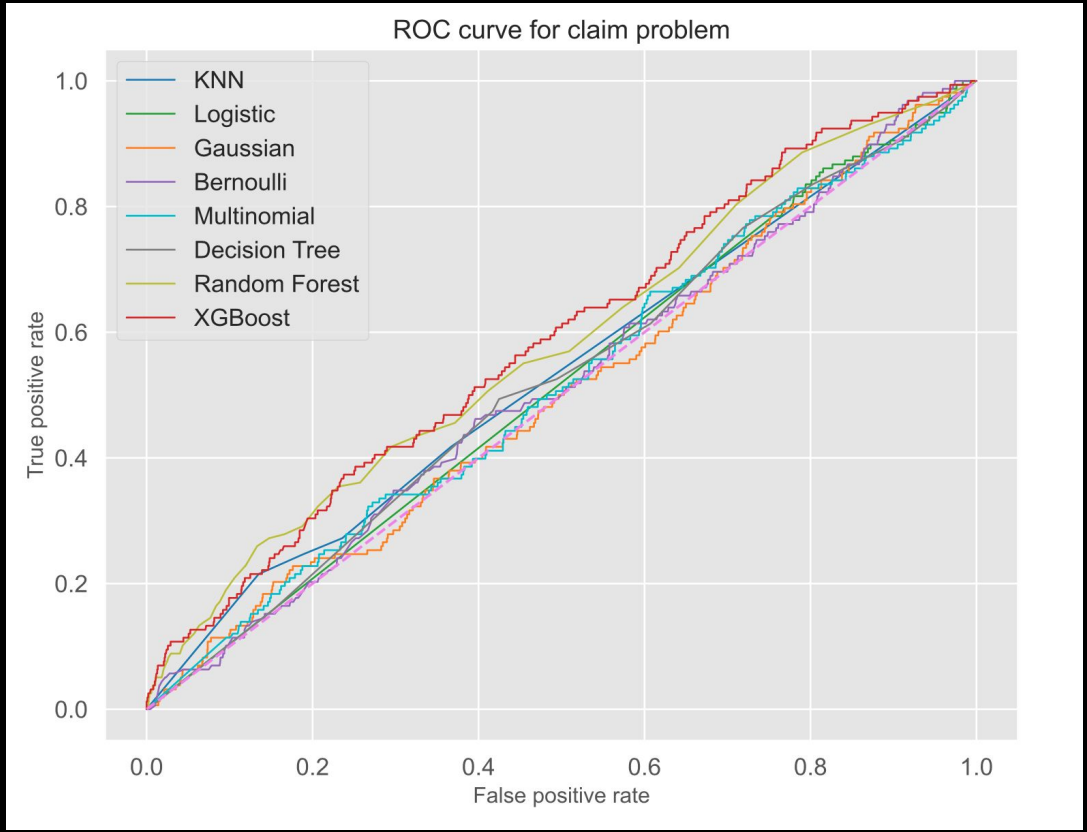
Features' Correlations



Class Imbalance



ROC AUC For All Models



Features' Importance