SUMMER INTERNSHIP

ON AI/ML CONCEPTS

HENOTIC TECHNOLOGY PVT LTD

Presented by:

M. YUVA GOPI ANJANI KUMAR

¾ CSE (**20B91A05G6**)

During : 7th July 2022 to 06th September 2022

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING,SRKREC

CONTENTS

- ❖ INTRODUCTION TO DATASET
- ❖ PROBLEM OF STATEMENT
- ❖ DATA SCIENCE PROJECT LIFE CYCLE
- ❖ DIFFERENT AIML MODELS
- ❖ FINAL RESULTS
- ❖ CONCLUSIONS & FUTURE WORK
- ❖ LIST OF CHARTS
- ❖ REFERENCES

ABSTRACT

The project is done on fraud detection of Vehicle insurance claims. The Vehicle insurance fraud involves conspiring to make false or exaggerated claims involving property damage or personal injuries following an accident. Some common examples include staged accidents where fraudsters deliberately "arrange" for accidents to occur; the use of phantom passengers where people who were not even at the scene of the accident claim to have suffered grievous injury, and make false personal injury claims where personal injuries are grossly exaggerated. While some insurance frauds are categorised as false claims, some frauds are classified as false policies. As a result, innocent people fall prey to fraud insurance providers and fraud insurance agents. Thus this project is to predict the trueness of the claim for vehicle insurance by providing whether it includes fraud or not ,by using popular AIML models.

INTRODUCTION TO DATASET

Dataset is : VEHICLE INSURANCE CLAIMS PROUD DETECTION

- ❖ This data set have **15421** records of **23** columns(attributes/variables).
- ❖ The target variable is " **Fraud_Found**" .
- ❖ And our target is to detect if a claim application is fraudulent or non-fraudulent.
- ❖ The remaining variables are { Month, WeekOfMonth, DayOfWeek, Make, AccidentArea, DayOfWeekClaimed, MonthClaimed, WeekOfMonthClaimed, Gender, MaritalStatus, Age, Fault, PolicyType, VehicleCategory, PolicyNumber, RepNumber, Deductible, DriverRating, PoliceReportFiled,WitnessPresent, AgentType, BasePolicy } (22 columns).

PROBLEM OF STATEMENT

❖ Predictive models are most effective when they are constructed using a company's own historical claims data since this allows the model to recognize the specific nature of a company's exposure as well as its claims practices.The construction of the model also involves input from the company throughout the process, as well as consideration of industry leading claims practices and benchmarks.

❖ And this problem includes two main solutions:

❏ Detection of vehicle insurance fraud earlier to avoid unknown losses to the vehicle insurance company.

❏ Providing all aspects that can prove the trueness of a vehicle insurance claim.

ROLE OF AIML IN VEHICLE INSURANCE

Machine Learning is a subset of artificial intelligence where computer algorithms are used to autonomously learn from data. Machine learning (ML) is getting more and more attention and is becoming increasingly popular in many other industries. Within the insurance industry, there is more application of ML regarding the claims.

TOP 7 BENEFITS OF AI FOR CAR INSURANCE

1. Increased efficiency

2. Reduced human overhead

3. Lower claims handling expenses

4. Valuable business insights

5. Improved user experience and service affordability

6. Reduced risk of human errors

7. Transparent decision-making

DATA SCIENCE PROJECT LIFE CYCLE

➔ **Data Pre-processing**

  ❖ Check the Duplicate and low variation data

  ❖ Identify and address the missing variables

  ❖ Handling of Outliers (Using Box-plots, Z-score , Interquartile Range)

  ❖ Categorical data and Encoding Techniques (Label Encoding , Binary Encoding)

  ❖ Feature Scaling (Min-Max Scaling)

➔ **Selection of Dependent and Independent variable**

  ❖ In this dataset,the target variable is **Fraud_Found**

➔ **Training Models**

➔ **Developing predicted values**

DIFFERENT AIML MODELS

Different Models Used in this AIML Classification are:

- ❖ LogisticRegression()
- ❖ DecisionTreeClassifier()
- ❖ RandomForestClassifier( )
- ❖ ExtraTreesClassifier()
- ❖ KNeighborsClassifier()
- ❖ SVC(probability=True)
- ❖ BaggingClassifier()
- ❖ GradientBoostingClassifier()
- ❖ LGBMClassifier()
- ❖ XGBClassifier()

We used our train dataset to build the above models and used our test data to check the accuracy and performance of our models.

FINAL RESULTS

|	Model Name	True +ve	False -ve	False +ve	True -ve	Accuracy	Precision	Recall	F1 Score	Specificity	MCC	ROC_AUC_Score	Balanced Accuracy
0	LogisticRegression()	0	285	0	4341	0.938		0	0	1		0.5	0.5
1	DecisionTreeClassifier()	58	227	269	4072	0.893	0.177	0.204	0.19	0.938	0.133	0.5707	0.571
2	RandomForestClassifier()	1	284	1	4340	0.938	0.5	0.004	0.007	1	0.038	0.5016	0.502
3	ExtraTreesClassifier()	2	283	2	4339	0.938	0.5	0.007	0.014	1	0.054	0.5032	0.503
4	KNeighborsClassifier()	5	280	23	4318	0.935	0.179	0.018	0.032	0.995	0.038	0.5061	0.506
5	SVC(probability=True)	0	285	0	4341	0.938		0	0	1		0.5	0.5
6	BaggingClassifier(n_estimators=100)	14	271	8	4333	0.94	0.636	0.049	0.091	0.998	0.165	0.5236	0.524
7	GradientBoostingClassifier(loss='deviance')	1	284	2	4339	0.938	0.333	0.004	0.007	1	0.029	0.5015	0.502
8	LGBMClassifier()	16	269	3	4338	0.941	0.842	0.056	0.105	0.999	0.208	0.5277	0.528
9	GaussianNB()	117	168	611	3730	0.832	0.161	0.411	0.231	0.859	0.178	0.6348	0.635

CONCLUSIONS & FUTURE WORK

The model results in the following order by considering the model accuracy, F1 score and ROC AUC score.

1) **Light GBM** with Simple Random Sampling
2) **Bagging Classifier** with Simple Random Sampling
3) **Gradient Boosting** with Simple Random Sampling

We recommend the model - **Light GBM** with Simple Random Sampling technique as a best fit for the given vehicle insurance claim fraud detection dataset. We considered Light GBM as it can lead to good predictions with this dataset.
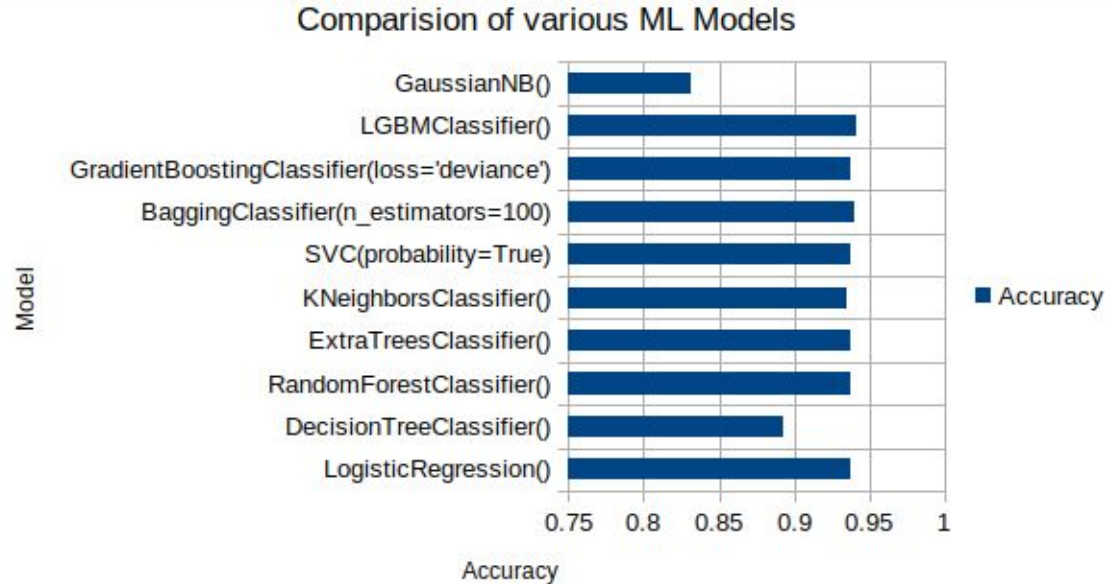
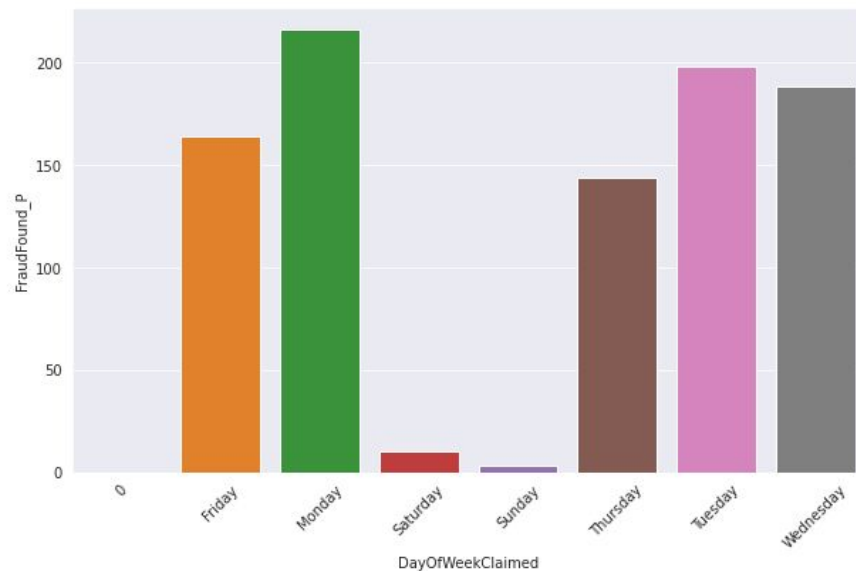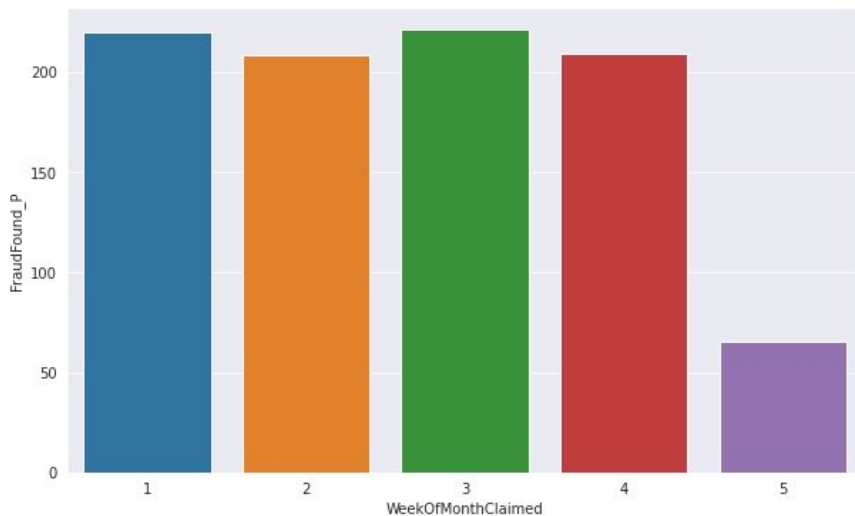# LIST OF CHATS (TOP2 MODELS)
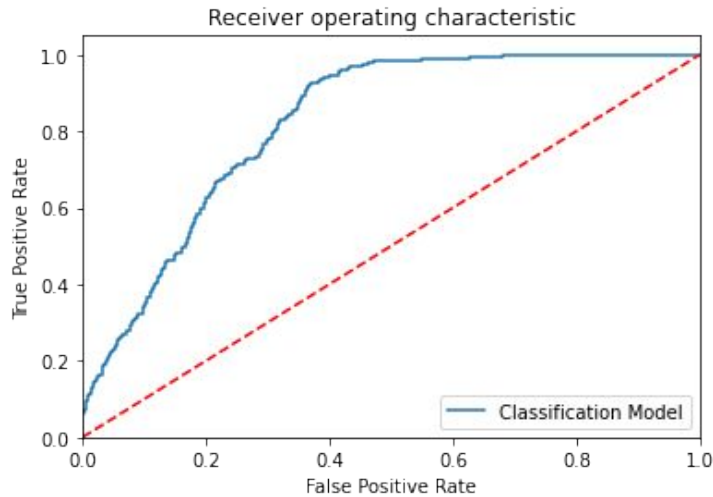
Light GBM Model
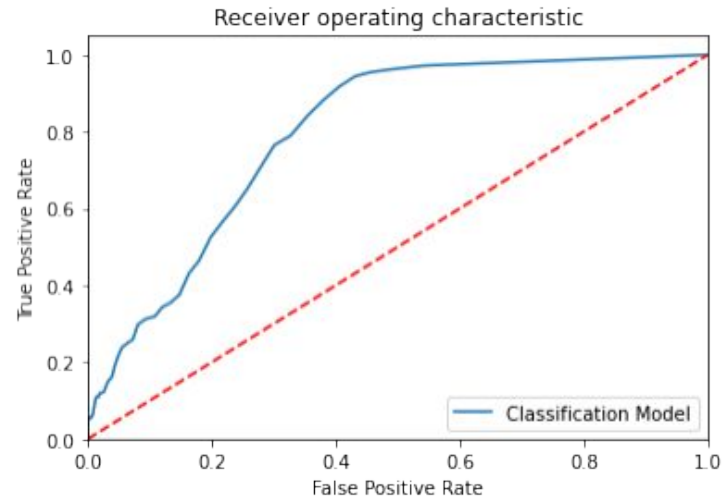
Bagging Classification Model

LIST OF CHATS

Comparision of various ML Models

LIST OF CHATS

REFERENCES

❖ The dataset taken for the project is from **Kaggle**.

   LINK: **https://www.kaggle.com/datasets/shivamb/vehicle-claim-fraud-detection**

❖ Some of the above information is taken from the websites like sklearn

   Author's name/Dataset Owner: Shivam Bansal

   Most of the above information Is Taken From The Websites

   ● **https://www.forbes.com/sites/davidteich/2021/03/02/machine-learning-can-help-the-insurance-industry-throughout-the-process-lifecycle/**

   ● **https://www.javatpoint.com/types-of-machine-learning**

   ● **https://www.iffcotokio.co.in/motor-insurance/car-insurance/the-dos-and-donts-of-filing-car-insurance-claims**