International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017),e-ISSN:2455-2585
International Conference on
Recent Explorations in Science, Engineering And Technology (ICRESET'19)
Volume-5, Special Issue-March, 2019.

VEHICLE THEFT ANALYSIS

Suchitra Priyadarshini K[1], Suvetha Sri V[2], Vijay Venkatesh R S[3], Mrs. Kamala V[4]
Assistant professor, Department of Computer Science and Engineering
KGISL Institute of Technology, Coimbatore

ABSTRACT:

The project "Vehicle Theft Analysis" is used to build a system that helps in analysing the theft claims and revert with premium changes based on various aspects of theft. More crimes are reported regarding vehicle thefts. The analysis performed by existing system focuses only regarding the motives behind the thefts and the methods implemented by them for performing the thefts. In our proposed system, initially the datasets are collected and the analysis is performed and the results are visualized. This enables better management and provides better understanding to the users. The datasets are updated in particular interval of time so that the results visualized are more accurate. Tableau is the technology used to visualize the data. It explores in real time and reacts to change faster.

Key Words: Theft Claims, Premium Changes, Visualized, Tableau, Analysing

I. PROBLEM DEFINITION

The analysis makes use of fixed datasets to produce results. The data may vary regularly. Hence day-today updated results are not visualized to the users. Our proposed system reduces the disadvantage by manually updating the database and producing updated visualized data to the users. There is a shortage of research especially on the reclamation of stolen motor vehicles. The system also seeks to expand the understanding of the role of insurance fraud.

II. SIGNIFICANCE

Vehicle Theft Analysis has a major significance such as follows:

- Quick data analysis
- Minimal time consumption
- Easy access and accuracy

III. LITERATURE REVIEW

In the existing system the analysis performed are not accurate. There are separate analysis performed for vehicle theft, insurance claims and for fraud analysis. The disadvantage of the existing application is there more missing data. The datasets are not updated continuously. Hence the analysis visualized to the users is not accurate. The analysis is published only in a documentation method and not in a visualized report. The analysis performed is not visualized to the users. In the existing system, validation of the data in the dataset is not made for obtaining accurate result. The dataset is permanent and no changes are made to the data. The time consumed for performing the analysis is also high. The data are collected by conducting survey or online. Here, the process of analysis is very slow. The system is not accurate because the datasets here are not updated. For example, in the existing system dataset uploaded will always be the same. Hence after some period of time when the data changes the result may have been outdated. So that is why the accuracy is affected in this system. The uploaded dataset circulates around only one or two constitutional problems. So when another transitional problem is met with, that is where the existing system faces backlog.

There are three categories in which the vehicle theft takes place. The three categories may also be stated as three purposes.

1) Theft for the purpose of recreating it.

2) Theft for the purpose of travel from one place to another.

3) Theft for the purpose of selling it.

Theft for the purpose of recreating it involves the re-modelling the bike in way that it cannot be identified as the bike that is stolen. So this gives an opportunity to reuse the vehicle without any risks involved. Theft for the place to travel from one place to another involves a little bit of risk while travelling in another region reduces the risk. Theft for the purpose of selling makes the person sell each and every part of the vehicle at different areas to avoid complete risk. These may be considered as datasets. In this existing system, the dataset may not contain all the categories. It contains one or two of the categories. The other one is not updated. So the accuracy is affected due to this

i. DRAWBACKS

- It takes very long time to complete the process
- The analysis is approximate but not accurate.
- The analyses for theft, insurance claims and fraud analysis are separate.

IV. VEHICLE THEFT ANALYSIS:

In our proposed system the analysis provided are more accurate. The datasets are requested from the insurance agencies to perform the analysis. The analysis for vehicle theft, insurance claims and fraud analysis are performed under single application. The datasets are updated with new data with regular intervals to obtain accurate result. The regions where the motor vehicle theft is high and agents under whom the false insurance claims are made is analysed and can be visualized using Tableau. The time consumed for performing the analysis is very less because it has every aspects of the theft that has been updated frequently. Unlike the existing system, our proposed system contains various datasets that can help the users to search the data in an accurate manner. This provides a clear data of vehicles that has undergone theft in specific area and it can be looked up to reduce the theft. And also this analysis takes data of the agency of insurance under which the vehicles has been stolen. This data gives an idea of any mismanagement of the agency involved in the theft of the vehicle. Likewise, many more aspects will be included as a dataset in regular intervals.

i. ADVANTAGES

- Data provided after analysis is accurate.
- All data are made available in single analysis
- The regions where the vehicle theft is high are analyzed and listed.
- Whereas in this application current updates are possible.

V. ALGORITHMS

- K-Means
- Linear Regression

i. K-Means ALGORITHM

K-Means is a clustering algorithm that is used for prediction in data mining. Cluster analysis is the way of grouping similar type of objects and separating the objects which are not similar into different clusters. Data Mining is process of storing large sets of pre-existing database to get desired information. The similar data can be easily identified with clustering process so that process of searching for data is made easy. The large drawback in K-Means algorithm is that it requires the number of clusters in the data to be initialized before prediction operation. The prediction process in K-Means algorithm is done using the following python code with pandas

```
import pandas as pd import numpy as np
from sklearn.cluster import KMeans
from sklearn.preprocessing import LabelEncoder from sklearn.preprocessing import MinMaxScalar import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline correct = 0
for I in range(len(x)):
predict_me= np.array(x[I].astype(float))
predict_me= predict_me.reshape(-1, len(predict_me)) prediction= kmeans.predict(predict_me)
if prediction[0] == y[I]:
correct += 1
print (correct/lex(x))
```

ii. LINEAR REGRESSION ALGORITHM

Linear regression is an algorithm which explains the abstraction between a dependent variable and an independent variable. If the regression algorithm deals with one independent variable, then it is a simple linear regression algorithm. If it deals with two or more independent variable, then it is multiple linear regression algorithms. The linear regression maybe applied in areas which require prediction, forecasting or error reduction. This algorithm also defines the relationship between the dependent and independent variables. The expression for linear regression is

$$Y = a+bX$$

where X is an independent variable and Y is the independent variable. The slope of line is b and a is the intercept.

```
plt.style.use('fivethirtyeight')
plt.scatter(reg.predict(X_train , reg.predict(X_train)- y_train, color = "green", s = 10, label = 'Train data')
plt.scatter(reg.predict(X_test), reg.predict(X_test) - y_test, color = "blue", s = 10, label = 'Test data')
plt.hlines(y = 0, xmin = 0, xmax = 50, linewidth = 2) plt.legend(loc = 'upper right')
plt.title("Residual errors") plt.show()
```

VI. VISUALIZATION

Data visualization in used in data analytics to represent the data in a graphical format or in a pictorial form such as pie chart. This helps the users to understand the patterns quickly and easily. Data visualization is used for following reasons

- Faster Result
- Communicate findings in constructive way
- Understand connections
- Embrace emerging trends
- Interact with data
- Create new discussion

i. Example:

	A	B	C	D	E
1	State	Rank	Make/Model	Model Year	Thefts
2	Alabama	1	Chevrolet Pickup (Full Size)	2005	499
3	Alabama	2	Ford Pickup (Full Size)	2006	357
4	Alabama	3	Toyota Camry	2014	205
5	Alabama	4	Nissan Altima	2014	191
6	Alabama	4	Chevrolet Impala	2004	191
7	Alabama	5	Honda Accord	1998	180
8	Alabama	6	GMC Pickup (Full Size)	1999	152
9	Alabama	7	Dodge Pickup (Full Size)	1998	138
10	Alabama	8	Ford Mustang	2002	122
11	Alabama	9	Ford Explorer	2002	119

Fig.1 Sample Dataset

This data represents the number of thefts took place in the state Alabama with number of thefts taken place including the Model of the vehicle. Now this particular data is visualized using data visualization. The below given code is used for visualizing the data to the users. The dataset for each state is created individually and visualization is performed.

```
import matplotlib.pyplot as plt import pandas as pd
file = r'E:\Pro\Alaska.xlsx' df = pd.read_excel(file) State = df["Make/Model"] data = df["Thefts"]
colors=['yellow','blue','red','green','violet','lightblue','purple','lightgreen', 'pink','brown']
plt.pie(data, labels=State, colors=colors, autopct='%1.1f%%', shadow=True, startangle=180) plt.title("Alabama")
plt.show()
```

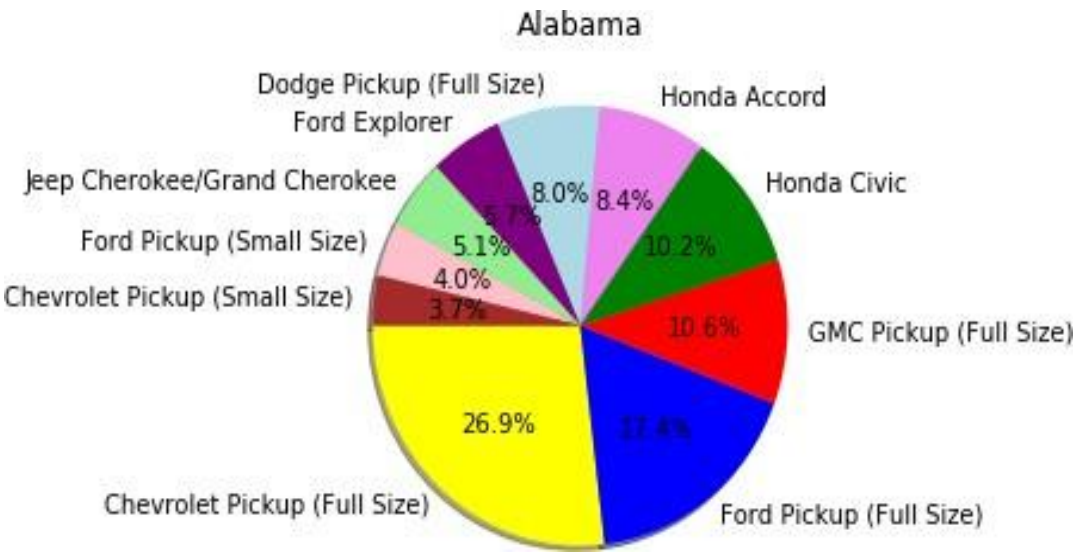

Fig.2 Visualization

This pie-chart shows us the visualized data of our provided data. The Model of the car is provided as column of the report for better understanding.

VII. ARCHITECTURAL DIAGRAM

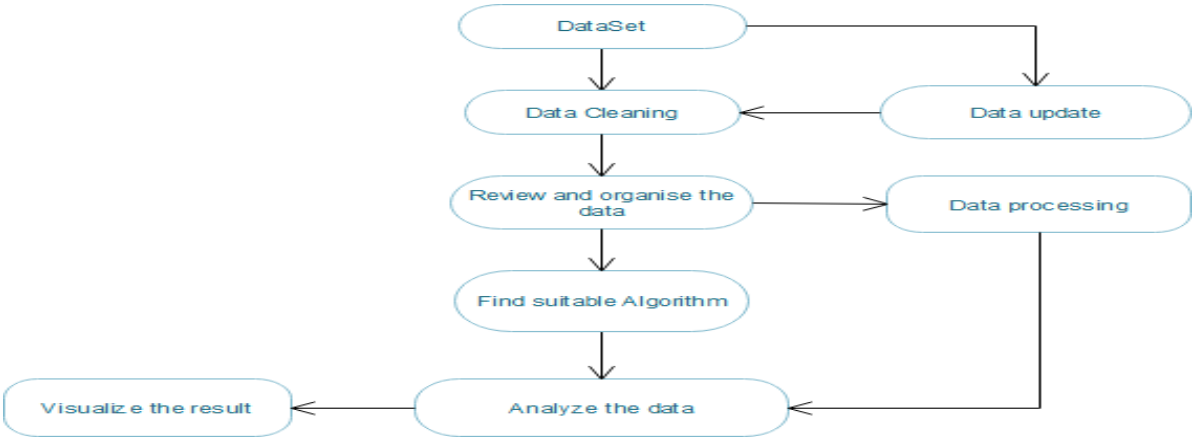

Fig.3 Architectural Diagram

Architectural diagram is used to understand the overall concept of our project. The developers analyse and visualize overall structure and requirements of their system using the architectural diagram. This also helps us to develop an idea about the project without any clarifications. The system requirements are also taken into considerations.

VIII. CONCLUSION

In vehicle theft analysis we have overcome major disadvantages of the existing system. The results visualized are more accurate. The Updating of the data in the dataset are made manual and it is regular. This process of updating data about all type of vehicle thefts consumes much lesser time. Regions where the theft is high are also analysed and visualized.

False theft cases can be analysed and reported in minimal time. Agents under whom the false insurance is claimed can also be analysed.

IX. REFERENCES

[1] Deborah Lamm Weisel ; William R. Smith ; G. David Garson ; Alexi Pavlichev ; Julie Wartell, "Motor Vehicle Theft: Crime and Spatial Analysis in a Non-Urban Region(2006) "

[2] Clarke, R.V. & Harris, P.M. (1992). Auto theft and its prevention. In Crime and Justice: A Review of Research, vol. 16, pp.1 –54. Chicago: University of Chicago Press.

[3] Field, S. Crime prevention and the costs of auto theft: An economic analysis. In Crime Prevention Studies; Clark, R.V., Ed.; Willan Publishing: Devon, UK, 1993; Volume 1, pp. 69–91

[4] Clarke, R.V.; Mayhew, P. Parking patterns and car theft risks: Policy-relevant findings from the British Crime Survey. In Crime Prevention Studies; Clarke, R.V., Ed.; Criminal Justice Press: Monsey, NY, USA, 1994; Volume 3, pp. 91–108.

[5] Webb, B. (1994). "Steering Column Locks and Motor Vehicle Theft: Evaluations from Three Countries." In

R.V. Clarke (ed.) Crime Prevention Studies Vol. 2, Monsey NY: Criminal Justice Press.

[6] Piza, E.L.; Carter, J.G. Predicting Initiator and Near Repeat Events in Spatiotemporal Crime Patterns: An Analysis of Residential Burglary and Motor Vehicle Theft. Just. Q. 2017, 1–29.

[7] De Melo, S.N.; Andresen, M.A.; Matias, L.F. Repeat and near-repeat victimization in Campinas, Brazil: New explanations from the Global South. Secur. J. 2018, 31, 364–380.

[8] Harlow, C.W. Motor Vehicle Theft; U.S. Department of Justice; Bureau of Justice Statistics: Washington, DC,

USA, 1998

[9] Roberts, A.; Block, S. Explaining Temporary and Permanent Motor Vehicle Theft Rates in the United States: A Crime-Specific Approach. J. Res. Crime Delinq. 2013, 50, 445–471

[10] Lu, Y. Getting away with the stolen vehicle: An investigation of journey-after-crime. Prof. Geogr. 2003, 55, 422–433.

[11] Harlow, C.W Motor Vehicle Theft; U.S. Department of Justice; Bureau of Justice Statistics: Washington, DC,USA, 1998.

[12] Light, R.; Nee, C.; Ingham, H Car Theft: The Offender's Perspective; Oxford University Press: London, UK, 1993.

[13] Bromley, R., & Thomas, C. (1997). Vehicle crime in the city center: Planning for secure parking. Town Planning Review, 68(2): 257-278.

[14] Tonry (ed.), Crime and Justice: A Review of Research (Volume 30), pp. 197-227. Chicago: University of Chicago Press.

[15] Krimmel, J.T. & Mele, M. (1998). Investigating stolen vehicle dump sites: An interrupted time series quasi-experiment. Policing: An International Journal of Police Strategies and Management, 21(3): 479-489.