



Car.ly - Vehicle Insurance Claim Fraud Detection

Machine Learning Strategies for Detecting
Vehicle Insurance Claim Fraud



Group No.: 30
Department of CS 4th Year
Project [PROJ-CS801]
University of Engineering and Management, Kolkata

Authors

Car.ly

Vehicle Insurance Claim Fraud Detection

- Abhishek Sharma
- Digbijoy Dasgupta
- Shreya Bose
- Ishita Pahari
- Udayan Misra
- Raktim Karmakar

Project Guide and Mentors:

- Prof. (Dr.) Sudipta Basu Pal
 - Prof. Piyali Chandra
-

Why this Project Work!?



Let's find out...

OBJECTIVE

- The goal is to develop and implement a fraud detection system that leverages advanced technologies such as data analytics, machine learning algorithms, and predictive modeling.
- Such a system can analyze large volumes of claims data to identify patterns and anomalies that indicate potential fraud.

Exploratory Data Analysis

- The original dataset(Kaggle) contains vehicle dataset - attribute, model, accident details etc along with policy details - policy type, tenure.
- The data first analyzed the missing values then the categorical and continuous data were separated. Then Exploratory data analysis was performed.

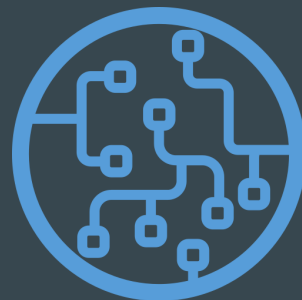
Month	# WeekOfMonth	DayOfWeek	Make	AccidentArea	
Jan	9%	Monday	17%	Pontiac	25%
May	9%	Friday	16%	Toyota	20%
Other (12642)	82%	Other (10359)	67%	Other (8462)	55%
Dec	5	Wednesday	Honda	Urban	Tue
Jan	3	Wednesday	Honda	Urban	Mon
Oct	5	Friday	Honda	Urban	Thu
Jun	2	Saturday	Toyota	Rural	Fri
Jan	5	Monday	Honda	Urban	Tue
Oct	4	Friday	Honda	Urban	Wed
Feb	1	Saturday	Honda	Urban	Mon
Nov	1	Friday	Honda	Urban	Tue
Dec	4	Saturday	Honda	Urban	Wed
Apr	3	Tuesday	Ford	Urban	Wed
Mar	2	Sunday	Mazda	Urban	Wed

Dataset used for Vehicle Insurance Fraud

Methods and Technologies

- ❖ Dataset Collection
 - Kaggle
- ❖ Algorithm Used to train the Model
 - Random Forest Classifier.
 - Support Vector Classifier.
 - K Nearest Neighbour.
 - Naive Bayes Classifier.
 - Logistic Regression.
 - Decision Tree.
 - XGBoost Classifier.

kaggle



Result Analysis

Gender and marital status of the accident victims

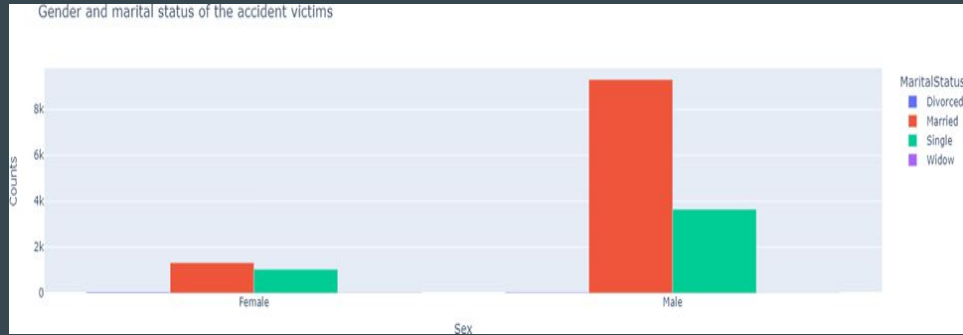


Fig: Gender and marital status of the accident victims

Regions where accidents occurred by year

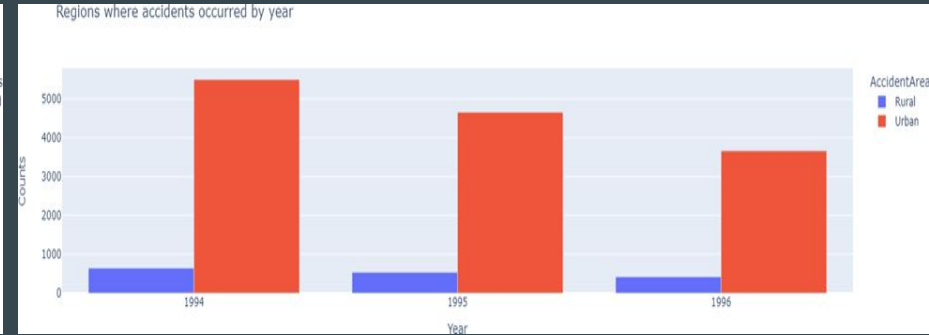


Fig : Regions where accidents occurred by year

Result Analysis

Age of the vehicle is also a cause of an accident. Different car brands have varying longevity of their vehicles, like, Toyota, Honda, and Subaru are generally considered to produce vehicles with longevity and proper maintenance. But luxury brands like BMW, Audi, and Mercedes-Benz may have higher maintenance costs and may not last as long as some other brands.

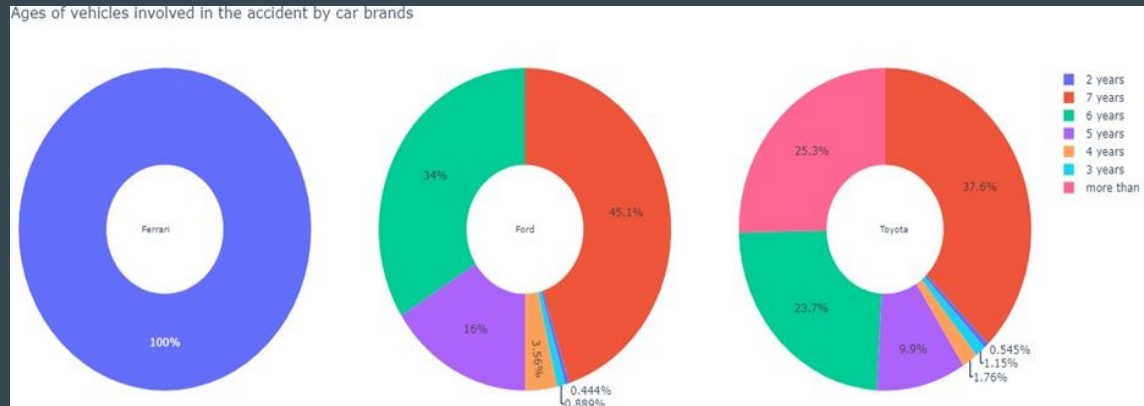
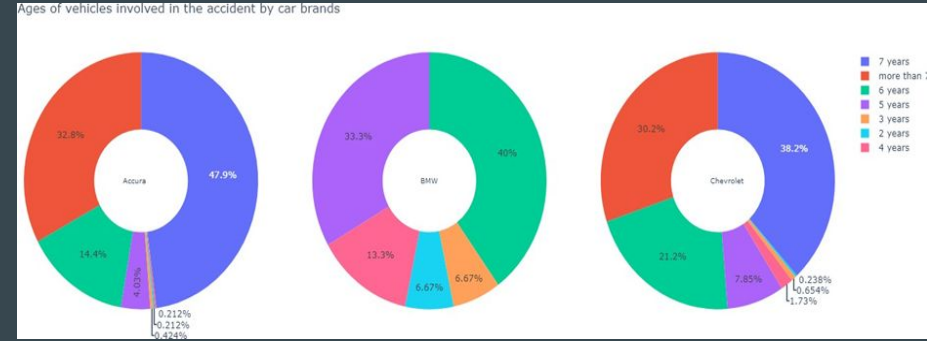


Fig : Ages of vehicles involved in the accident by car brands

Accuracy scores

In our model Random Forest had the highest accuracy out of all the others, followed by Support Vector and K Nearest Neighbour.

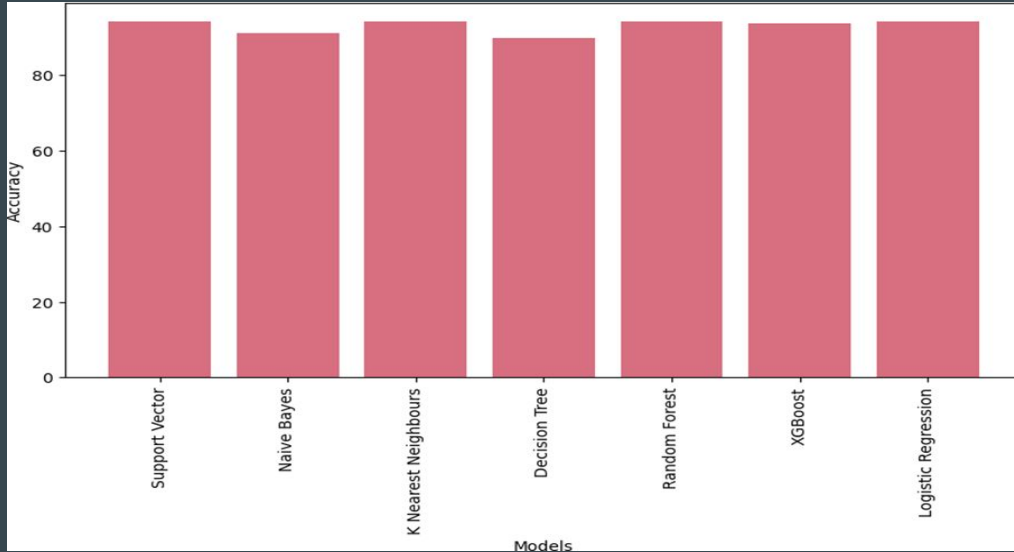


Fig : Accuracy scores of different models

Support Vector	94.0337224383916
Naive Bayes	90.9208819714656
KNN	94.0337224383916
Decision Tree	89.7751837440553
Random Forest	94.1634241245136
XGBoost	94.1634241245136
Logistic Regression	94.0337224383916

Why we are preferring Random Forest Classifier!

Well the answer is simple,

- Accuracy of this algorithm, is way better than any other algorithm.
- Easy to implement and easy to demonstrate.
- Last but not the least, this also shows the capability of Random Forest Classifier.

Hope, this will make you to support the Random Forest Classifier for this special framework!

Conclusion

- A Vehicle Insurance Claim Fraud Detection system can significantly reduce losses incurred by insurance companies due to fraudulent claims, thereby improving the company's profitability and financial stability.
 - By automating the process of detecting fraudulent claims and improving the efficiency of the claims process, a Vehicle Insurance Claim Fraud Detection system can enhance customer satisfaction by providing a faster, smoother, and more hassle-free claims experience.
-

References

- Hasan, R., & Islam, M. S. (2021). Vehicle insurance fraud detection using machine learning techniques: A systematic review. *IEEE Access*, 9, 26994-27011
- Zhang, X., Li, J., Li, Y., & Chen, X. (2021). A vehicle insurance claim fraud detection system based on deep learning and feature selection. *Journal of Intelligent & Fuzzy Systems*, 40(1), 707-718.
- Varga, M., & Lengyel, L. (2020). Insurance fraud detection with machine learning: A review. *Expert Systems with Applications*, 141, 112986
- Wang, Y., Huang, X., Li, J., Yang, C., & Shi, Q. (2019). An efficient fraud detection method for automobile insurance claims based on SVM and FCM. *Applied Soft Computing*, 82, 105545.
- Kim, Y. J., & Kim, D. K. (2017). A comparative study on fraud detection methods for automobile insurance. *Journal of Intelligent Information Systems*, 49(1), 73-93
- Zhang, T., Shi, J., & Zhou, X. (2016). An insurance fraud detection model for automobile claims based on Bayesian networks. *Journal of Intelligent & Fuzzy Systems*, 31(1), 281-290.
- Wu, M., Wu, T., Wang, Y., & Li, W. (2021). A fraud detection model for auto insurance claim based on XGBoost and SMOTE. *International Journal of Data Mining and Bioinformatics*, 26(1), 1-18
- Li, Q., Wu, J., & Zhang, J. (2020). A hybrid fraud detection model for automobile insurance based on feature selection and SVM. *Neural Computing and Applications*, 32(22), 16819-16831.
- Shi, J., Zhang, T., Xu, Y., & Zhou, X. (2018). Insurance fraud detection for automobile claims using clustering and decision trees. *Expert Systems with Applications*, 97, 72-82.
- Singh, A., Suman, S., Kumar, A., & Kumar, V. (2018). A novel approach for fraud detection in vehicle insurance using data mining. *Procedia Computer Science*, 132, 501-509
- Zhu, S., & Zhang, Y. (2017). An insurance fraud detection model for automobile claims based on improved decision tree algorithm. *Journal of Intelligent & Fuzzy Systems*, 32(4), 3037- 3045

Thank You!



Adieu.