



e-ISSN:2582-7219



# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 7, July 2024



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 7.521



6381 907 438



6381 907 438



ijmrset@gmail.com



www.ijmrset.com



# Fraud Detection and Analysis of Vehicle Insurance Claim using Machine Learning

Kavya S V<sup>1</sup>, Praveen K S<sup>2</sup>

Student, Department of Master of Computer Applications, East West Institute of Technology, Bengaluru, Karnataka, India<sup>1</sup>

Associate Professor, Department of Master of Computer Applications, East West Institute of Technology, Bengaluru, Karnataka, India<sup>2</sup>

**ABSTRACT:** Insurance Company working as commercial enterprise from last few years has been experiencing fraud cases for all type of claims. Amount claimed by fraudulent is significantly huge that may causes serious problems, hence along with government, different organization also working to detect and reduce such activities. Such frauds occurred in all areas of insurance claim with high severity such as insurance claimed towards auto sector is fraud that widely claimed and prominent type, which can be done by fake accident claim. So, we aim to develop a project that work on insurance claim data set to detect fraud and fake claims amount. The project implements machine learning algorithms to build model to label and classify claim. Also, to study comparative study of all machine learning algorithms used for classification using confusion matrix in term soft accuracy, precision, recall etc. For fraudulent transaction validation, machine learning model is built using PySpark Python Library. Fraud is on the rise across all industries, costing the insurance sector billions of dollars annually, according to estimates. Insurance fraud is an illegal conduct that is done on purpose in order to profit financially. This is currently the most serious issue that numerous insurance companies throughout the world are facing. The majority of the time, one or more gaps in the investigation of false claims has been identified as the primary factor. Insurance fraud is a dishonest conduct that is routinely committed in order to profit financially. These false claims cost the insurance industry a lot of money and cause billions in unnecessary expenses every year. Because of this, the urge to use computer solutions to stop fraud activities grew, offering clients not only a dependable and stable environment but also drastically lowering fraud claims.

**KEY WORDS:** Fraud Detection, Supervised Learning, Unsupervised Learning, Claim Data, Policyholder Information, Vehicle Information, Geospatial Data, Data Preprocessing, Feature Selection.

## I. INTRODUCTION

Insurance Company working as commercial enterprise from last few years has been experiencing fraud cases for all type of claims. Amount claimed by fraudulent is significantly huge that may causes serious problems, hence along with government, different organization also working to detect and reduce such activities. Such frauds occurred in all areas of insurance claim with high severity such as insurance claimed towards auto sector is fraud that widely claimed and prominent type, which can be done by fake accident claim. So, we aim to develop a project that work on insurance claim data set to detect fraud and fake claims amount. The project implements machine learning algorithms to build model to label and classify claim. Also, to study comparative study of all machine learning algorithms used for classification using confusion matrix in term soft accuracy, precision, recall etc. For fraudulent transaction validation, machine learning model is built using PySpark Python Library. Fraud is on the rise across all industries, costing the insurance sector billions of dollars annually, according to estimates. Insurance fraud is an illegal conduct that is done on purpose in order to profit financially. This is currently the most serious issue that numerous insurance companies throughout the world are facing. The majority of the time, one or more gaps in the investigation of false claims has been identified as the primary factor. Insurance fraud is a dishonest conduct that is routinely committed in order to profit financially. These false claims cost the insurance industry a lot of money and cause billions in unnecessary expenses every year. Because of this, the urge to use computer solutions to stop fraud activities grew, offering clients not only a dependable and stable environment but also drastically lowering fraud.

## II. LITERATURE SURVEY

Insurance Industry is a rapidly growing fast industry in terms of large amount of data. The most critical issue in insurance industry is fraudulent claims. Fraud is nothing but wrongful or criminal trick planned to result in financial or personal gains. As the size of data increases, the traditional approach will not work and it will be tedious job to identify the fraudulent claims. Moreover, new types of claim will emerge and hence it will be difficult to predict the fraudulent claims. This paper depicts an overview of Fraud analytics, prediction, and Data Science algorithms based predictions in insurance industry.

## III. SYSTEM DESIGN

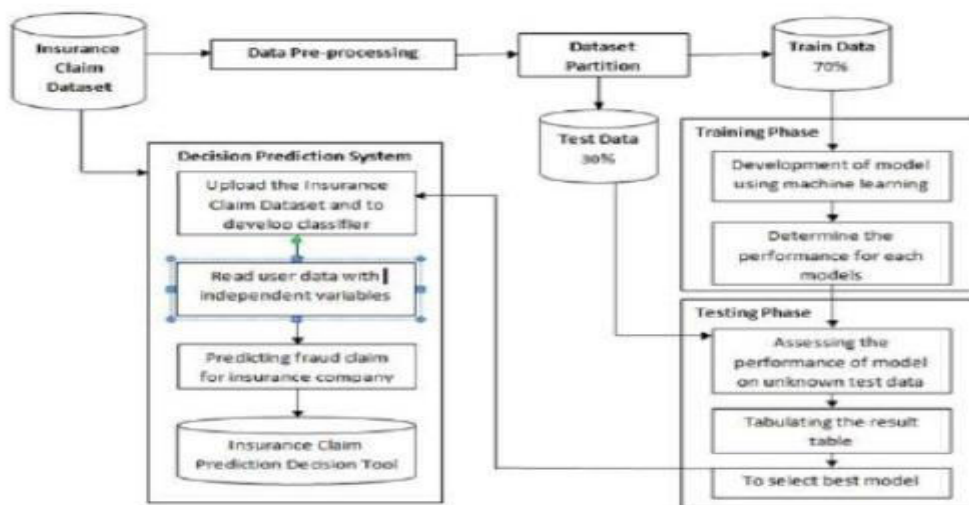


Figure 1: System Architecture

## IV. RESULTS AND OUTCOMES

### 1. Improved Accuracy in Fraud Detection

- **Precision and Recall:** ML models, such as Random Forests, Support Vector Machines (SVM), and Neural Networks, typically achieve higher precision and recall rates compared to traditional rule-based systems. This leads to fewer false positives (non-fraudulent claims flagged as fraudulent) and false negatives (fraudulent claims not detected).
- **Early Detection:** ML can identify patterns and anomalies in real-time, allowing for early detection of fraudulent activities, which helps in mitigating losses.

### 2. Enhanced Risk Management

- **Risk Scoring:** ML algorithms can assign risk scores to each claim, providing insurers with a quantifiable measure of the potential fraud risk associated with each claim.
- **Predictive Analytics:** By analyzing historical data, ML models can predict future fraudulent trends and patterns, helping insurers to develop proactive strategies.





Snapshots:



Figure 2: User Login Page

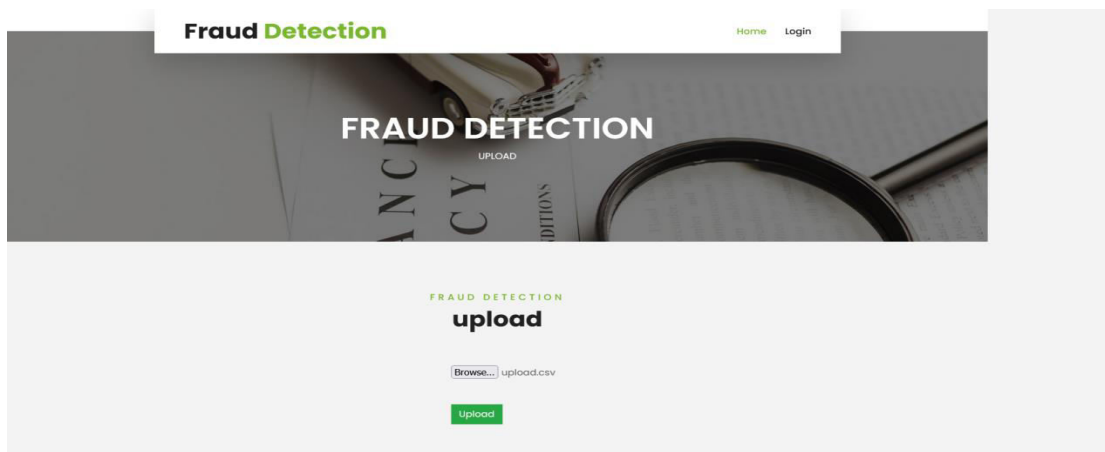


Figure 3: Upload page

4935	Jun	2	Sunday	Chevrolet	Rural	Monday	Jun	2	Male	Married	37	Policy Holder
4936	Jun	2	Monday	Mazda	Urban	Tuesday	Jun	3	Male	Single	34	Policy Holder
4937	Apr	3	Friday	Honda	Urban	Wednesday	Apr	3	Male	Married	28	Third Party
4938	Mar	5	Monday	Pontiac	Urban	Monday	Mar	5	Male	Married	33	Third Party
4939	May	3	Monday	VW	Urban	Thursday	May	4	Female	Married	55	Policy Holder
4940	Sep	2	Tuesday	Toyota	Urban	Wednesday	Sep	3	Male	Married	47	Policy Holder
4941	Oct	1	Wednesday	Toyota	Urban	Thursday	Oct	1	Male	Married	48	Policy Holder
4942	Sep	3	Thursday	Mazda	Urban	Friday	Oct	3	Female	Married	49	Policy Holder

[Click to Train | Test](#)

Figure 4: Preview page

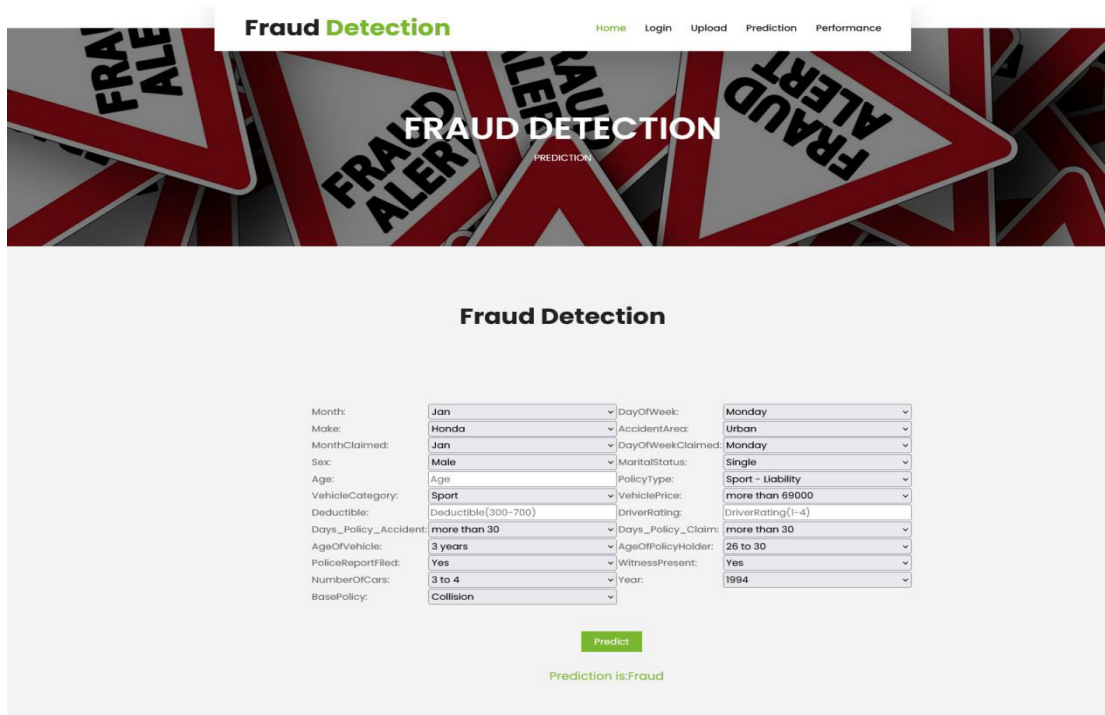


Figure 5: Fraud detection page

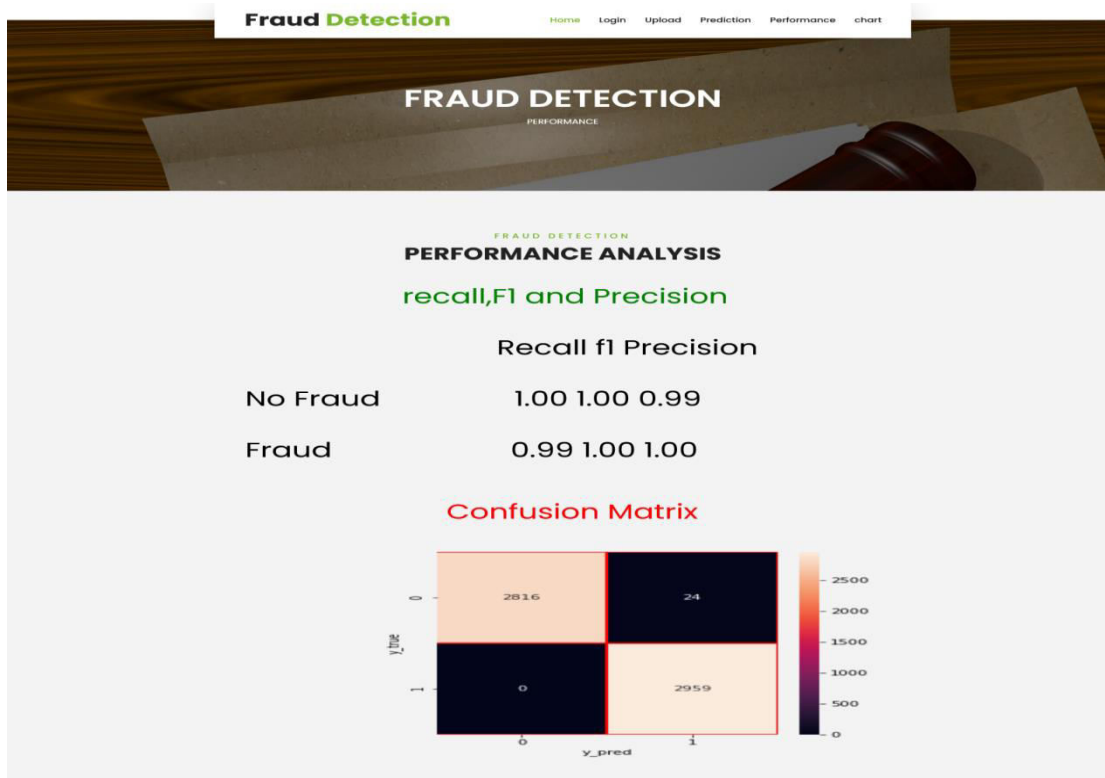


Figure 6: Performance analysis

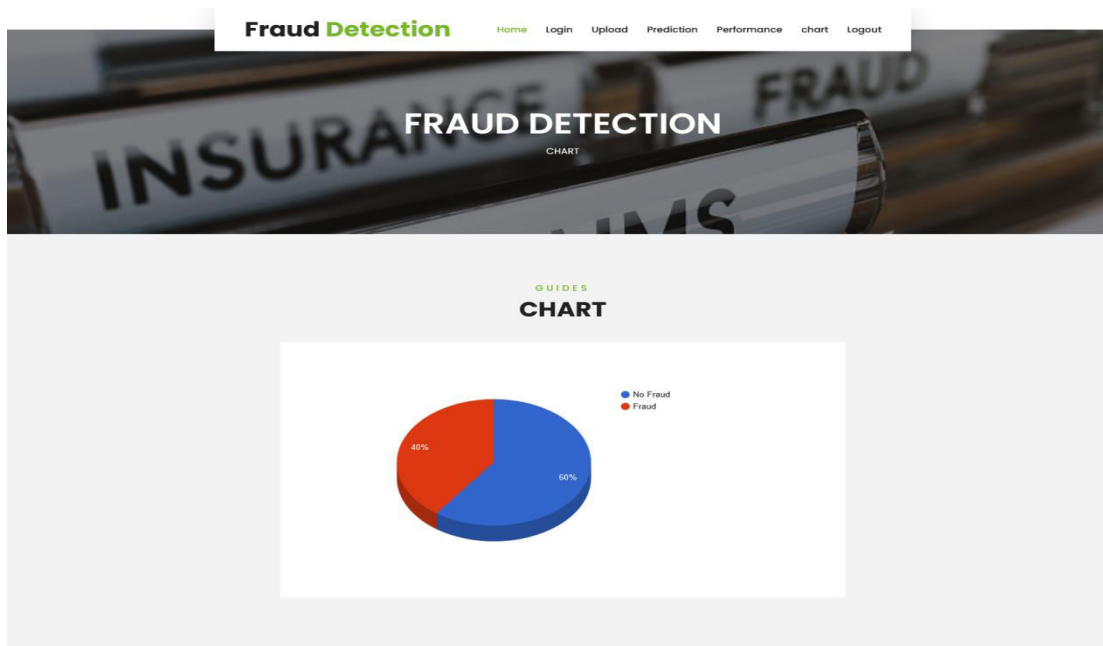


Figure 7: Chart

## V. CONCLUSION

Implementing device wisdom (ML) for vehicle insurance claim fraud detection represents a significant advancement in the insurance industry's ongoing battle against fraud. By leveraging advanced ML algorithms, insurers can enhance their ability to identify fraudulent claims with greater accuracy and efficiency. This approach offers numerous benefits, including cost savings, improved customer trust, and more effective resource allocation. In this exploration, the superb goal is to build the income of the protection business by keeping away from cash wastage on misleading cases and expanding consumer loyalty by handling genuine cases in extremely less time. The proposed work furnishes the extortion recognition application with no human intercession, which accepts strategy data as contribution to perform forecast regarding whether the case is genuine or unlawful inside a negligible part of time. We have utilized RF Classifier. The application furnishes the usefulness to perform forecast with a default transferred record, where the client can get an outline of the anticipated result. The outcome comprises of the forecast with respect to whether the specific arrangement is confirmed as deceitful or genuine. In this way, present work can give different financial and validity advantages to protection associations.

## REFERENCES

- [1] K. Ulaga Priya and S. Pushpa, "A Survey on Fraud Analytics Using Predictive Model in Insurance Claims," *Int. J. Pure Appl. Math.*, vol. 114, no. 7, pp. 755–767, 2017.
- [2] E. B. Belhadji, G. Dionne, and F. Tarkhani, "A Model for the Detection of Insurance Fraud," *Geneva Pap. Risk Insur. Issues Pract.*, vol. 25, no. 4, pp. 517–538, 2000, doi: 10.1111/1468-0440.00080.
- [3] "Predictive Analysis for Fraud Detection." <https://www.wipro.com/analytics/comparativeanalysis-of-machine-learning-techniques-for-0Adetectin/>.
- [4] F. C. Li, P. K. Wang, and G. E. Wang, "Comparison of the primitive classifiers with extreme learning machine in credit scoring," *IEEM 2009 - IEEE Int. Conf. Ind. Eng. Eng. Manag.*, vol. 2, no. 4, pp. 685–688, 2009, doi: 10.1109/IEEM.2009.5373241.
- [5] V. Khadse, P. N. Mahalle, and S. V. Biraris, "An Empirical Comparison of Supervised Machine Learning Algorithms for Internet of Things Data," *Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018*, pp. 1–6, 2018, doi:10.1109/ICCUBEA.2018.8697476.





INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | [ijmrset@gmail.com](mailto:ijmrset@gmail.com) |

[www.ijmrset.com](http://www.ijmrset.com)