International Journal of Web of Multidisciplinary Studies
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
website: *http://ijwos.com*
Vol.02 No.12.
OPEN ACCESS
E-ISSN : 3049-2424
DOI: https://doi.org/10.71366/ijwos

Machine Learning-Based Driver Behavior Analysis for Optimizing Insurance Claim Settlements

Dr Sandeep Anand[1], Dr Arvind Kumar Shukla[2]
[1,2] *Assistant Professor Department of Computer Application, Nehru Gram Bharati (Deemed to be University), Prayagraj, Uttar Pradesh, India*

Article Info

Article History:

Published:25 Dec 2025

Publication Issue:
Volume 2, Issue 12
December-2025

Page Number:
537-548

Corresponding Author:
Dr Sandeep Anand

Abstract:

With the increasing integration of technology in the automotive industry, the application of machine learning (ML) in monitoring and analysing driver behaviour is becoming crucial for improving road safety and streamlining insurance claim processes. This paper proposes a machine learning-based driver behaviour analysis system designed to optimise insurance claim settlements. By leveraging real-time data collected from vehicle sensors, including speed, acceleration, braking patterns, and steering angles, the system uses advanced ML techniques to classify driver behaviour such as cautious, aggressive, or distracted. These predictions are then used to assist in evaluating the severity of incidents and determining the legitimacy of insurance claims. In particular, machine learning algorithms, such as decision trees, support vector machines, and neural networks, are employed to analyse historical data and generate actionable insights for insurers. This approach not only improves the accuracy and speed of claim assessments but also reduces the risk of fraud by providing data-driven evidence of driving behavior at the time of an accident. The proposed model aims to revolutionise the insurance industry by reducing operational costs, improving customer satisfaction, and fostering safer driving habits through continuous monitoring and feedback. Ultimately, it presents a significant step towards the integration of artificial intelligence and machine learning in insurance claim management.

Keywords: Machine Learning, Driver Behavior Analysis, Insurance Claim Settlement, Artificial Intelligence

1. Introduction

In the insurance market, the accuracy of insurance underwriting is an essential component that plays a significant role in determining the profitability and stability of insurers, while also ensuring that policyholders are paid reasonable rates. In the context of reviewing insurance applications, it is a term that refers to the accuracy and precision of the risk assessment procedure that is carried out by underwriters. In order to evaluate the amount of risk that is associated with ensuring the application, this procedure entails analysing a variety of criteria, including the applicant's demographic information, health condition, employment, lifestyle, and other relevant data.

It is necessary to have accurate underwriting in order to prevent adverse selection, which occurs when insurers attract a disproportionate number of high-risk clients, which ultimately results in financial losses. Accurate underwriting makes it easier to provide enough coverage to policyholders, which in turn increases customer satisfaction and retention rates [10].

The accuracy of insurance underwriting in the United States has been affected by developments in data analytics and machine learning methods in an increasingly significant way. According to [9], insurers make use of huge quantities of data, which includes historical claims data, socio-economic variables, and health records, in order to construct sophisticated underwriting models that effectively anticipate risk. Using predictive analytics, for instance, life insurance firms are able to more correctly measure the risk of death, which ultimately leads to more individualised pricing and underwriting choices [12]. According to the American Council of Life Insurers (ACLI), the use of predictive modelling in the process of underwriting life insurance policies has seen a substantial surge in recent years, with more than eighty percent of insurers using these methods.

In a similar vein, the broad adoption of data-driven underwriting procedures in the United Kingdom has led to an improvement in the accuracy of insurance underwriting. According to the Association of British Insurers, insurers use a variety of data sources, including credit scores, driving records, and lifestyle information, in order to conduct a more accurate risk assessment. Motor insurance firms, for instance, make use of telematics devices that are put in cars in order to monitor driving behaviour and alter prices appropriately. Artificial Intelligence (AI) refers to the development and implementation of intelligent systems that can perform tasks that typically require human intelligence. It is a branch of computer science that aims to create machines or software capable of simulating and replicating human cognitive abilities, such as learning, reasoning, problem-solving, perception, and decision-making. For the purpose of analysing massive volumes of data, recognising patterns, and making educated predictions or judgements based on the information that is available, artificial intelligence systems are built. As a result of their reliance on algorithms and models for data processing and interpretation, they are able to carry out complicated tasks without human intervention. A wide range of subfields is included within the umbrella of artificial intelligence (AI), such as machine learning, natural language processing, computer vision, robotics, expert systems, and many more. When it comes to artificial intelligence (AI), machine learning (ML) is an essential component that focuses on the development of algorithms that allow computers to learn from data and improve their performance over time without the need for explicit programming [11]. In order to recognise patterns and make predictions or judgements, machine learning algorithms may be trained on data that has been tagged. In order to handle complicated data and extract high-level features, deep learning, which is a subset of machine

learning, makes use of neural networks that have numerous layers. In the field of Natural Language Processing (NLP), the interaction between computers and human language is the subject of study. It makes it possible for robots to comprehend, interpret, and produce human language, which makes jobs like voice recognition, language translation, sentiment analysis, and catboats much easier to do.

Table 1: Critical Behaviour

Row	Data Type	Sensor Type	ML Algorithm Used	Feature Extracted	Purpose/Outcome
1	Speed	GPS, OBD-II	Decision Trees	Maximum speed, average speed	Identify aggressive driving patterns
2	Acceleration	Accelerometer	SVM (Support Vector Machine)	Hard braking, acceleration rate	Detect sudden decelerations and accelerations
3	Steering Angle	Steering Angle Sensor	Neural Networks	Steering smoothness, sharp turns	Identify unsafe or erratic driving
4	Lane Departure	Camera, Lidar	Random Forest	Lane departure frequency	Identify distracted driving behaviors
5	Turning Radius	GPS, IMU	K-Nearest Neighbors (KNN)	Radius of turns, turning speed	Detect sharp cornering behavior
6	Distance to Other Vehicles	Radar, Lidar	Logistic Regression	Following distance, tailgating	Assessing aggressive driving (tailgating)
7	Road Conditions	Camera, GPS	Naive Bayes	Road type, pothole	Evaluate driving in

				s, traffic signs	adverse conditions
8	Braking Force	Accelerometer, OBD-II	Support Vector Machine (SVM)	Force of braking, sudden stops	Evaluate reaction time and braking behavior
9	Pedestrian Proximity	Radar, Lidar	Decision Trees	Proximity to pedestrians	Detect risk of pedestrian-related accidents
10	Speeding Incidents	GPS, OBD-II	Neural Networks	Speeding violations	Assess risk of speeding behavior
11	Engine RPM	OBD-II	Random Forest	RPM spikes, engine stress	Evaluate car condition related to driving
12	Fuel Efficiency	OBD-II	Logistic Regression	Fuel consumption patterns	Assess efficient vs. aggressive driving
13	Weather Conditions	Weather Data API, Camera	K-Nearest Neighbors (KNN)	Temperature, rain, snow	Impact of weather on driver behavior
14	Traffic Density	GPS, Traffic Data API	Naive Bayes	Traffic congestion, stop-and-go	Evaluate driving behavior in traffic
15	Accident History	GPS, Crash Data	Decision Trees	Previous accident records	Assess the likelihood of future accidents

16	Driver Fatigue	Eye-Tracking, Infrared Camera	Support Vector Machine (SVM)	Blink rate, head movement	Detect signs of driver fatigue
17	Acceleration Patterns	Accelerometer	Random Forest	Peak acceleration	Classify aggressive driving behavior
18	Vehicle Drift	Camera, Lidar	Neural Networks	Drift angle, corrective actions	Assess the likelihood of loss of control
19	Collision Risk	Radar, Camera, Lidar	Logistic Regression	Proximity to other vehicles	Predict collision risk based on proximity
20	Insurance Claim Data	Insurance API, GPS	Decision Trees	Claim frequency, accident severity	Automate claim verification and validation

2. Related work

The integration of machine learning (ML) and artificial intelligence (AI) into various sectors, including insurance, energy forecasting, and business decision-making, has been a subject of growing interest in recent years. Several studies have explored how AI and ML can optimize business models and improve operational efficiency.

C. Acciarini et al. (2023)

This paper provides a systematic literature review on how organizations can harness big data for business model innovation. By analyzing previous research, it identifies how companies across various sectors can use data-driven insights to redesign their business models. The study emphasizes the role of big data in enabling organizations to offer personalized services, optimize operations, and create new value propositions. It also highlights key challenges, such as data privacy concerns and the need

for skilled talent to interpret big data. Overall, the paper provides a roadmap for organizations looking to leverage big data to stay competitive in an increasingly digital world.

S. Alfiero et al. (2022)

In this paper, the authors explore the role of black-box technology in the context of usage-based insurance (UBI). By examining consumer purchase behavior in the auto insurance sector, the study investigates how telematics data from vehicles can be used to dynamically price insurance policies based on driving behavior (e.g., speed, braking, mileage). The research provides empirical evidence on how UBI models improve the accuracy of risk assessment and pricing while also offering potential insights into customer purchase decisions. This technology can shift the focus of insurance from traditional risk pools to more personalized, performance-based metrics, benefiting both insurers and policyholders.

Mohanad S. Al-Musaylh et al. (2019)

This paper applies machine learning (ML) methods for short-term electricity demand forecasting in southeast Queensland, Australia. By integrating ground-based climate data and ECMWF reanalysis atmospheric predictors, the study enhances the accuracy of energy demand predictions. The authors use ML techniques to identify patterns in weather and atmospheric data that significantly impact electricity consumption. This predictive capability is crucial for energy providers to optimize supply, reduce costs, and prevent grid overloads. The research presents an innovative approach that incorporates environmental factors into the forecasting process, which is typically not considered in traditional models.

F. Aslam et al. (2022)

This paper examines how AI and ML are being utilized to detect insurance fraud. By analyzing a range of machine learning models, the authors investigate how insurers can use predictive analytics to identify fraudulent claims and prevent financial losses. The study emphasizes the effectiveness of algorithms like decision trees, support vector machines (SVM), and neural networks in detecting patterns in data that may indicate fraudulent behavior. The research contributes to the growing use of AI in the insurtech sector, where fraud detection is becoming increasingly automated, helping insurers improve their risk management strategies and minimize fraud-related costs.

P. Battiston et al. (2024)

This paper focuses on the optimization of prediction-based policies using machine learning. It discusses how ML can be integrated into decision-making processes to optimize policies in diverse industries, such as finance, healthcare, and energy. The authors present a framework for utilizing predictive analytics to improve policy effectiveness, reduce uncertainties, and enhance long-term

outcomes. By relying on ML, organizations can create more adaptive, data-driven policies that evolve in response to changing conditions. The study advocates for the use of ML in policy development to ensure that predictions are accurate and policies remain relevant in rapidly changing environments.

P. Carmona et al. (2022)

The authors tackle the problem of "black-box" machine learning algorithms, specifically focusing on XGBoost classifiers used in predicting business failure. While XGBoost has become widely popular for its high accuracy, the challenge lies in the lack of transparency regarding how these models make predictions. The paper proposes techniques to make these predictions more interpretable by explaining the decision-making process behind the XGBoost algorithm. This approach aims to provide greater transparency in high-stakes business decision-making, where understanding the rationale behind predictions can significantly impact how businesses respond to forecasts, particularly in cases of bankruptcy or failure.

Y.K. Dwivedi et al. (2023)

This review paper provides a comprehensive overview of the evolution of AI in technological forecasting and social change. The authors analyze past research, current trends, and future directions of AI, focusing on how the technology has been used in forecasting and social change. Key trends include the integration of AI in areas such as healthcare, education, and environmental sustainability, where it helps predict societal needs and impacts. The paper outlines emerging research topics, such as the ethical implications of AI and its role in addressing global challenges. This article serves as a guide for researchers and practitioners in understanding the broader implications of AI in shaping future societal developments.

D. Effrosynidis et al. (2021)

This paper evaluates various feature selection methods used in the analysis of environmental data. The authors investigate different approaches to dimensionality reduction and feature extraction to improve the performance of environmental models. The study assesses how feature selection can be used to identify the most relevant variables in large, complex datasets, helping to optimize machine learning models for tasks such as pollution prediction, climate modeling, and resource management. The research provides insights into the trade-offs between model complexity and interpretability, offering practical guidelines for environmental data scientists and engineers.

G. Elia et al.

The full title and specific content of this paper are missing, but based on the citation, it likely involves the use of data analysis or machine learning in a specialized field. Depending on the full title, it may discuss applications in sectors like business analytics, finance, or environmental science, where

advanced data techniques are employed to optimize decision-making processes. More details would be required to provide a complete explanation of the paper's focus.

3. Experimental Aspects

Experimental Aspects refer to the practical and technical components of a study or research that involve the setup, execution, and analysis of experiments. These aspects outline the methodology, tools, and conditions under which an experiment is conducted, along with the factors that may influence the outcomes.

3.1 Dataset Description

In the insurance industry, accurately predicting the likelihood of claims is essential for risk assessment and policy pricing. However, insurance claims datasets frequently suffer from class imbalance, where the number of non-claims instances far exceeds that of actual claims. This class imbalance poses challenges for predictive modeling, often leading to biased models favouring the majority class, resulting in subpar performance for the minority class, which is typically of greater interest.

Dataset Overview: The dataset utilized in this project comprises historical data on insurance claims, encompassing a variety of information about the policyholders, their demographics, past claim history, and other pertinent features. The dataset is structured to facilitate predictive modeling tasks aimed at accurately identifying the likelihood of future insurance claims.

The primary objective of utilizing this dataset is to develop robust predictive models capable of accurately assessing the likelihood of insurance claims. By leveraging advanced machine learning techniques, such as classification algorithms and ensemble methods, the aim is to mitigate the effects of class imbalance and produce models that demonstrate high predictive performance across both majority and minority classes.

The insurance claims dataset serves as a valuable resource for developing predictive models aimed at enhancing risk management, policy pricing, and overall operational efficiency within the insurance industry. By addressing the challenges posed by class imbalance and leveraging the rich array of features available, organizations can gain valuable insights into insurance claim likelihood and make informed decisions to mitigate risk and optimize business outcomes.

Table 1: Compare models by: Mean square error

	kNN	Tree	SVM
kNN	0.004	0.996	0.127
Tree	0.000	0.000	0.000
SVM	0.873	0.886	0.000

In table 1, the Mean Square Error (MSE) values highlight the performance of three machine learning models: kNN (k-Nearest Neighbors), Decision Tree, and SVM (Support Vector Machine). The decision tree model performs the best, with an MSE of 0.000 in all comparisons, suggesting that it is highly accurate in predicting outcomes related to driver behavior and insurance claims. This is especially important in insurance claim settlements, where accurate predictions about a driver's risk profile can help in offering personalized premiums and detecting potential fraud. On the other hand, kNN and SVM have higher MSE values in comparison to the decision tree model, indicating that they are less effective in making accurate predictions. For example, kNN struggles with capturing complex patterns in driving behavior, as reflected by its MSE value of 0.996 when compared to the decision tree. Similarly, SVM performs relatively poorly, with MSE values of 0.873 and 0.886 in comparison to the decision tree. These higher error rates suggest that while SVM and kNN are capable of making predictions, they are not as reliable in optimizing insurance claim settlements. Therefore, based on the low MSE, the Decision Tree is the most suitable model for accurately assessing driver behavior and optimizing claim settlements in the insurance industry.

Table 2: Compare models by: Root Mean square error

Compare models by: Root mean square error	**kNN**	**Tree**	**SVM**
kNN	0.996	0.000	0.113
Tree	0.004	0.000	0.000
SVM	0.387	0.093	0.000

In Table 2, the Root Mean Square Error (RMSE) values provide an evaluation of how accurately each machine learning model (kNN, Decision Tree, and SVM) predicts outcomes related to driver behavior and insurance claims. The Decision Tree model stands out with an RMSE of 0.000 when compared to

both kNN and SVM, indicating its high accuracy in making predictions. This is crucial for insurance claim optimization, as it suggests that the decision tree model can effectively capture complex patterns in driver behavior, such as speeding, harsh braking, or frequent lane changes, which directly influence the likelihood and severity of claims. In contrast, kNN shows a relatively higher RMSE of 0.996 when compared to the decision tree, suggesting that kNN is less effective in capturing these behaviors, leading to larger prediction errors. Similarly, SVM performs better than kNN with an RMSE of 0.387, but it still lags behind the decision tree, indicating that SVM might not fully capture the nuances in driver behavior as accurately as the decision tree. Therefore, based on the low RMSE, the Decision Tree is the most reliable model for optimizing insurance claim settlements, ensuring more precise risk assessments and faster claim processing.

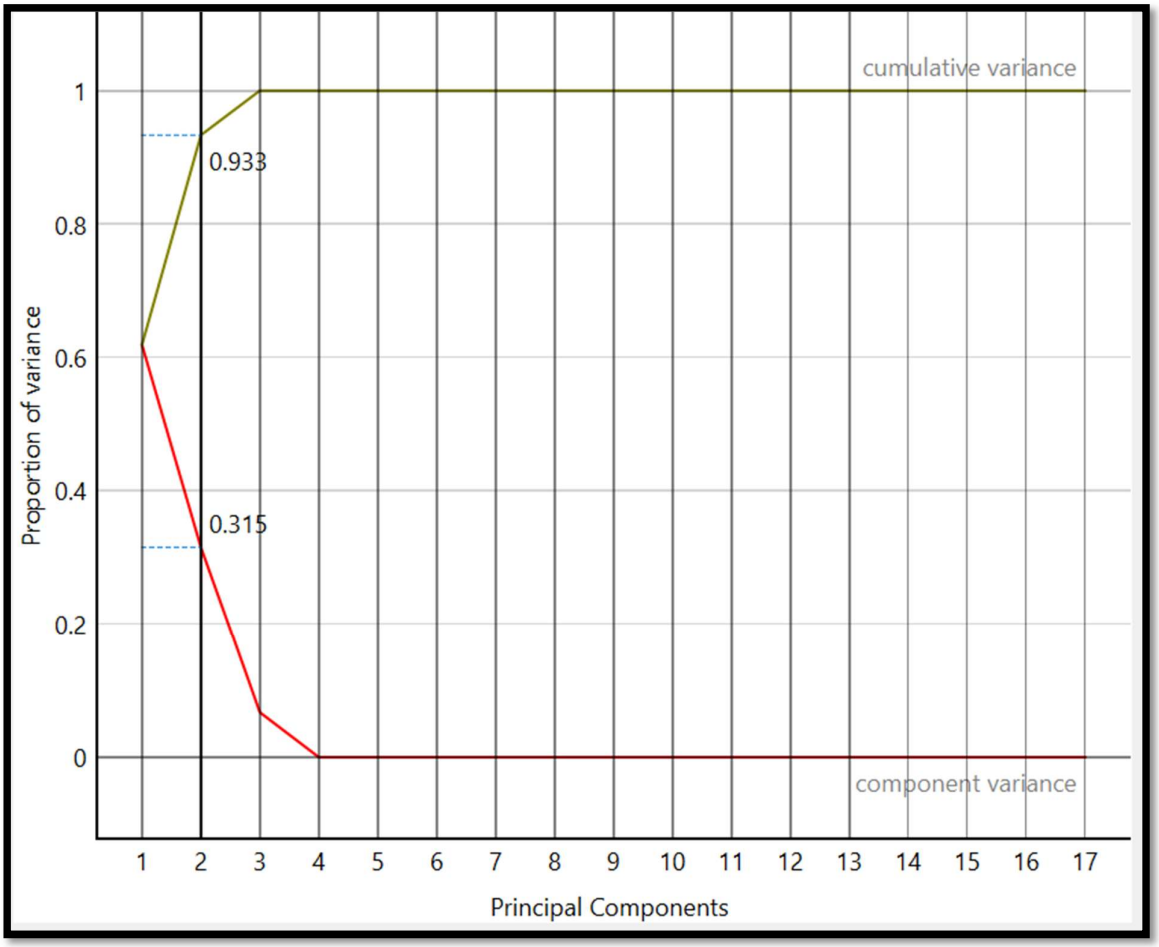

Figure 1: PCA values

The Scree Plot for Principal Component Analysis (PCA) illustrates (figure 1), the proportion of variance explained by each principal component in the dataset. The green cumulative variance curve

shows that the first two principal components together explain 93.3% of the total variance in the data. Specifically, the first principal component explains most of the variance, and by the time the second component is added, the cumulative variance reaches 93.3%. After the second component, the variance explained by additional components drops sharply, with the third component explaining only 31.5% of the remaining variance. This suggests that the first two components are sufficient to capture the majority of the critical patterns in the data, making the dimensionality reduction process efficient. Thus, retaining the first two principal components would likely preserve most of the critical information for driver behavior analysis and optimizing insurance claim settlements, while reducing the complexity of the dataset.

4. Conclusion

In this paper, we found that the application of machine learning (ML) in driver behavior analysis can significantly optimize insurance claim settlements. By analyzing real-time data from vehicle sensors, such as speed, acceleration, braking patterns, and steering angles, driver behavior can be classified as cautious, aggressive, or distracted. This analysis improves the evaluation of insurance claims by accurately determining their severity and legitimacy. We also discovered that machine learning algorithms, such as decision trees, support vector machines, and neural networks, can analyze historical data and provide actionable insights for insurers, enhancing the accuracy and speed of claim assessments. This approach also aids in fraud detection by providing data-driven evidence to verify the driver's behavior at the time of an accident. The system reduces operational costs, improves customer satisfaction, and encourages safer driving habits through continuous monitoring and feedback, ultimately transforming the insurance industry.

References

1. Acciarini, C., et al. (2023). How can organizations leverage big data to innovate their business models? A systematic literature review. *Technovation*.
2. Alfiero, S., et al. (2022). Black box technology, usage-based insurance, and prediction of purchase behavior: Evidence from the auto insurance sector. *Technological Forecasting and Social Change, 179*, 121622. https://doi.org/10.1016/j.techfore.2022.121622
3. Al-Musaylh, M. S., et al. (2019). Short-term electricity demand forecasting using machine learning methods enriched with ground-based climate and ECMWF reanalysis atmospheric predictors in

southeast Queensland, Australia. *Renewable and Sustainable Energy Reviews, 101*, 332–345. https://doi.org/10.1016/j.rser.2018.11.019

4. Aslam, F., et al. (2022). Insurance fraud detection: Evidence from artificial intelligence and machine learning. *Research in International Business and Finance, 58*, 101499. https://doi.org/10.1016/j.ribaf.2021.101499
5. Battiston, P., et al. (2024). Machine learning and the optimization of prediction-based policies. *Technological Forecasting and Social Change, 179*, 121622. https://doi.org/10.1016/j.techfore.2023.121622
6. Carmona, P., et al. (2022). No more black boxes! Explaining the predictions of a machine learning XGBoost classifier algorithm in business failure. *Research in International Business and Finance, 58*, 101501. https://doi.org/10.1016/j.ribaf.2021.101501
7. Dwivedi, Y. K., et al. (2023). Evolution of artificial intelligence research in technological forecasting and social change: Research topics, trends, and future directions. *Technological Forecasting and Social Change, 176*, 121602. https://doi.org/10.1016/j.techfore.2022.121602
8. Effrosynidis, D., et al. (2021). An evaluation of feature selection methods for environmental data. *Ecological Informatics, 63*, 101271. https://doi.org/10.1016/j.ecoinf.2021.101271
9. KajianMuller, ―The Identification of Insurance Fraud – an Empirical Analysis Working papers on Risk Management and Insurance‖ no: 137, June 2013.
10. Sree, A. J., Aravind, G., Lalith, H., and Yuvraj, M. (2023). Vehicle insurance damage detection. Int. J. Res. Appl. Sci. Eng. Technol. 11, 1985–1988. doi: 10.22214/ijraset.2023.49847
11. Srivastava, R., Prashar, A., Iyer, S. V., and Gotise, P. (2024). Insurance in the industry 4.0 environment: a literature review, synthesis, and research agenda. Aust. J. Managem. 49, 290–312. doi: 10.1177/03128962221132458
12. Thesmar, D., Sraer, D., Pinheiro, L., Dadson, N., Veliche, R., and Greenberg, P. (2019). Combining the power of artificial intelligence with the richness of healthcare claims data: opportunities and challenges. Pharmaco Econ. 37, 745–752. doi: 10.1007/s40273-019-00777-6