© AUG 2024 | IRE Journals | Volume 8 Issue 2 | ISSN: 2456-8880

Using Artificial Intelligence to Improve Insurance Claim Evaluation

RAPHAEL IBRAIMOH
School of Science, Engineering and Environment, University of Salford

Abstract- Machine learning is used to improve vehicle insurance claim analysis in this study. The increased number of claims and efficiency requirements for claims processing have shown that manual methods cannot handle the load, resulting in delays, inaccuracies, and inefficiencies. The research uses Kaggle insurance claim data using OOAD and UML models to create a simple and reliable application architecture. Random Forest, known for its accuracy and diversity, produced a model with 99.5% precision compared to Decision Tree's 95.68%. Confusion matrix and ROC curve performance measurements showed machine learning algorithms' claim result prediction power. Results showed the intricacy of interactions between factors, including legal counsel and seatbelt usage, that affected claims results and scholarly literature on attorney and safety measures. Theoretical implications help build an overall evaluation approach to insurance claim analysis and allow incorporation of previously unimportant aspects impacting claim results. Practical implications include insurance claim processing transformation, which will systemize assessment processes, minimize error margins, and detect fraudulent claims. Expand data sources, incorporate live data, use advanced machine learning algorithms, and validate models in the real world. This ensures their efficacy and usability. Future research will address limitations like relying on historical data by integrating real-time data streams and using advanced predictive analytics methods like deep learning and NLP algorithms to analyse unstructured claim forms.

Indexed Terms- Machine learning, insurance claim analysis, automotive insurance, predictive modelling, decision-making, fraud detection.

I. INTRODUCTION

The number and complexity of claims has skyrocketed, exhausting manual processing and causing errors. When it can improve insurance claim analysis accuracy, speed, and cost-efficiency, ML technologies are the best hope for these challenges. Owolabi et al. (2023) found that neural networks and decision trees, machine learning techniques, can better recognize patterns and anomalies in huge datasets, improving risk assessment and fraud detection. ML can also reduce insurance claim processing times, improving customer happiness and efficiency (Singhal et al., 2023). Therefore, using AI in insurance claims review has significant drawbacks. The biggest issues are data quality and availability. ML algorithms need lots of high-quality data to work well. Insurance claims sometimes involve personal information, which raises privacy and security concerns. (2022, Njeru). In addition, some ML models use a black-box feature, which lacks transparency and interpretability, making it difficult for insurers to explain their decision-making processes to stakeholders (Okagbue and Oyewole, 2023). While machine learning for more accurate insurance claim processing is not without obstacles, the overall benefits are significant, making the insurance process more efficient, accurate, and customer-cantered in the digital age. Machine learning has found a place in automobile insurance claims, which this study will examine. It has the potential to transform claims processing. Advanced ML models can analyse auto insurance claims for hull damage, liability, and fraud. A recent study by Manakhari et al. (2023) revealed that deep learning systems can accurately identify vehicle damage from pictures to simplify claims. Aslam et al. (2022) found that machine learning can analyse driving behaviour data to improve risk assessment and premium calculation. NLP can automatically extract useful information from claim documents, enhancing productivity and eliminating human error (Singhal et al., 2023). Machine learning in auto insurance claims can improve accuracy, efficiency, and customer service, however data privacy, model interpretability, and compliance implementation are challenges. This study uses machine learning on a car claim set to improve vehicle insurance claim processing precision, efficiency, and customer-centricity.

II. METHODOLOGY

The research methodology used OOAD and UML diagrams like class diagram and sequence diagram to create a systematic application architecture. This resilient framework method was chosen because it clearly constructs the system to show developers the interaction and behaviours, making the process efficient and orderly. The dataset, available on Kaggle, was combined with the Random Forest technique in Jupyter Notebook to create machine learning. We chose the Random Forest algorithm because it handles data accurately and prevents overfitting using the ensemble method to create exact predictions. The massive reservoir of high-quality datasets on Kaggle prompted the choice of data source for intelligent model development. Unlike other competitors, the researcher picked Python in Jupyter Notebook due to its simplicity, versatility, and vast data science library ecosystem, making it ideal for constructing and applying machine learning models. Below are the study's class and sequence diagrams:

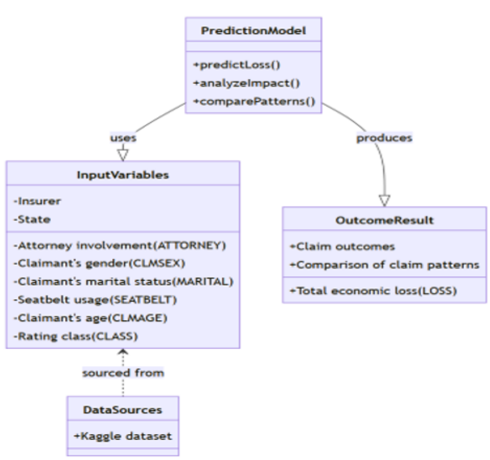


Figure 1: Class Diagram.

The class diagram shows the system architecture to anticipate motor bodily injury claims' total economic loss, examine settlement outcomes driven by multiple factors, and compare claim patterns by insurer, state, and rating class. The "PredictionModel" class defines crucial features like loss prediction, impact analysis, and pattern comparison. The "InputVariables" class contains all model inputs, including attorney engagement, client demographics, and insurance and claim data. These "DataSources" inputs are Kaggle-specific external data. The "Outcome Result" class summarizes model predictions and assessments, including economic losses, claim findings, and claim trends. This chart shows how to build a full machine learning model for insurance claim analysis using system components and dependencies.

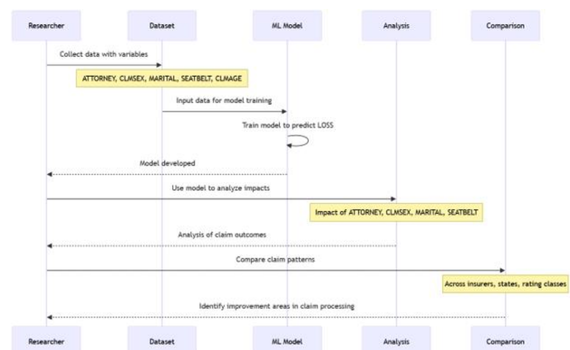


Figure 2: Sequence Diagram.

The sequence diagram shows how the research will begin by collecting the most relevant demographic and claims data to construct a machine learning model to forecast automotive injury claims economic losses. The model's projections are reviewed to determine how attorney involvement and claimant demographics affect case outcomes. After that, insurance companies, states, and classes are compared to identify claim processing issues and solutions. This brief description shows how data gathering and insights development lead to claim handling optimization through a systematic analytical approach.

- *Research Approach*

This study uses a systems approach to examine model building, starting with data gathering and ending with model implementation. This technique was carefully used to display the measurements step-by-step and the logical process to understand the data structure and research outcome. By breaking the process into sequential sections, the technique makes it transparent and efficient and assures that each stage is based on empirical facts and theoretical analysis from previous stages. This method is vital for maintaining research integrity and building a model that is scientifically valid and applicable in real life. This strategy emphasizes quality and innovation criteria, ensuring the relevance of research findings and their benefits to the field.

Figure 3: Schematic Representation of the Research Methodology.

- *Data collection*

Since Kaggle is a top platform for gathering extensive, diversified, and superior datasets, it was chosen as the major source for the relevant dataset for this research. This portal provides substantial data access and a space for data scientists and academics to solve dataset problems. Thus, it can be used to create an integrated dataset with a variety of factors relevant to our study. The dataset includes approximately 70,000 closed claims from 32 insurers. It contains claimant demographics, solicitor participation, economic loss (LOSS, measured in thousands), and other factors. Content is separated into three sections: General Insurance Claims Data: A 1340-row, 8-column data frame with CASENUM (case number), ATTORNEY (whether attorney was involved), CLMSEX (claimant gender), MARITAL (claimant marital status), CLMINSUR (whether driver had insurance), and SEATBELT. Automobile Insurance Claims: This section contains 6773 observations of 5 macro variables: STATE, CLASS, GENDER, AGE, and PAID (amount paid to settle and close any claim). Automobile UK Collision Claims: The UK market for this insurance claim dataset includes 8,942 collision losses from private passenger vehicle insurance policies, showing driver age, Vehicle Use, Severity, and Claim Count. We chose this Kaggle dataset because we needed a large, robust data repository to examine insurance claims' complications. Insurance claims are analysed in detail using a large dataset of demographic, economic, and environmental characteristics. It provides a complete insight of claim outcomes and economic damage. The information provides a picture of the insurance claim scenario and helps discover trends. It can be used to create prediction models and strategies to improve administrative efficiency and effectiveness.

- *Data Processing*

The construction of a machine learning model to predict overall economic loss in road accident personal injury claims required extensive data cleansing and feature engineering. This method ensured model performance. The category variables ATTORNEY, CLMSEX, MARITAL, and SEATBELT were converted to numerical inputs for machine learning. This stage was vital because it allowed the prediction model to include important demographic and claim-related information, improving its quality. Numbers like Claimant age (CLMAGE) have been changed to standardize data distribution. Thus, the numbers' size does not bias the model's predictions. This protected the model against larger-scale variable distortions, preserving each variable's relative relevance.

The dataset's analysis required resolving the missing value using appropriate methods depending on its characteristics and properties. When data was missing due to random causes, imputation techniques were employed to fill the gaps using data distribution patterns. This was done to avoid training disruptions from incomplete information. Feature engineering was used to improve the model's predictive power by assessing current data points and exploring variables that could reveal the claimant's profile and environment. Dimensional reduction was done by LDA. LDA maximizes class separation, making it helpful at distinguishing economic loss levels. Our outlier identification and removal processes identified and removed irregularities that could distort the model and vary from claim patterns. A dataset that accurately captures common claim patterns was produced. This comprehensive approach to data cleaning and feature creation improves the model's capacity to predict economic loss by capturing dynamic data trends.

According to Yu et al. (2024), LDA is a statistical method used to reduce dimensionality and improve class separability across many classes. Linear Discriminant Analysis (LDA) uses class labels to identify a combination of characteristics that distinguish between classes, unlike Principal Component Analysis (PCA), which maximizes variance without considering class labels (Yu et al., 2024). This gives LDA an edge in pattern recognition and classification tasks, where the goal is to minimize dimensions and improve the model's ability to distinguish groups. LDA requires projecting data into a smaller space. Class separability is preserved, making categorization more precise and effective. However, the model's integrity depends on the independent variables' normal distribution and consistent covariance matrix across classes. Unfortunately, real-world data rarely meet this expectation.

- *Data Training*

This study trained the model using Random Forest algorithm, a common ensemble learning method. Sun et al. (2024) describes it as building many decision trees and using the mode of the classes (classification) or the mean prediction (regression) to produce the final output at training time. Random Forests address decision trees' tendency to overfit the training set, providing a broader solution. This approach was chosen for its precision and capacity to handle a range of input factors in the dataset. Not only can catastrophic modeling manage non-linear correlations and interactions between variables, but it can also forecast total economic loss in complex vehicle bodily injury cases. Random Forest also has a powerful feature relevance metric that shows which features most affect model prediction. This feature helped determine how attorney involvement, claimant gender, marital status, and seat belt use affected claim outcome. The algorithm was the best choice for model building because of its high performance and ability to work with high-dimensional data without preprocessing, proving that the result was reliable and understandable and revealing the dataset's internal linkages. The pseudocode below simplifies the Random Forest method and describes its training and prediction steps:

```
RandomForestTrain(Dataset D, int M, int k):
  Forest = []
  for i = 1 to k do
    Sample = BootstrapSample(D)
    Tree = TrainDecisionTree(Sample, M)
    Forest.append(Tree)
  return Forest

TrainDecisionTree(Sample, M):
  // Train a tree using a random subset of M features

RandomForestPredict(Forest, instance x):
  Votes = InitializeVotes()
  for Tree in Forest do
    prediction = Tree.Predict(x)
    Votes[prediction] += 1
  return MajorityVote(Votes)
```

- *Model Evaluation*

The confusion matrix and ROC curve were carefully chosen as key indicators for model performance evaluation. Larner (2024) states that the confusion matrix details the model's predictions, including its accuracy and the types of errors committed, such as false positives and false negatives. Accuracy is crucial for examining a model's behaviour across scenarios and distinguishing classes. ROC curves also indicate the model's performance at different thresholds, illustrating the true positive rate versus false positive rate trade-off (Fanjul-Hevia et al., 2024). AUC is a single value that measures a model's ability to discriminate classes at all thresholds (McDermott et al., 2024). Therefore, this metric is useful for model comparison. This set of indicators provides a solid foundation for assessing the model's prediction power and real-world applications. They're detailed below.

Actual class	**Predicted Class**	
	Positive	Negative
True	True positive (TP)	False Negative (FN)
False	False positive (FP)	True Negative (TN)

Confusion matrix describes the performance parameters for the classifier where TP, TN, FP, and FN represent True Positive, True Negative, False Positive, and False Negative, respectively.

- **True negative**: observation is correctly classified as negative.
- **False negative:** observation is incorrectly classified as negative.
- **True positive:** A positive class is correctly classified by the model.
- **False positive:** A negative observation is incorrectly classified.

i. **Accuracy measure:** Total number of correctly classified sentiments divided by total numbers of sentiments

$$\text{Accuracy} \frac{TP + TN}{TP + FP + FN + TN}$$

ii. **Precision:** In equation (10) shows the ability of the classifier to correctly identify the positive class.

$$\text{Precision} = \frac{TP}{TP+FP}$$

iii. **Recall:** The number of times the classifier predicted a negative class out of all the times the class is negative (non-cancerous)

III. RESULTS AND DISCUSSION

The insurance claim model is trained using random forest and decision tree algorithms on a Kaggle dataset and integrated using Object-Oriented Analysis and Design (OOAD) into a Jupyter notebook with a user input function to put theory and practice together. User input functionality makes the model dynamic, enabling user participation and real-time insurance decision making. The model must be critically evaluated to ensure robustness and reliability in diverse use case scenarios. To identify biases and inaccuracies, assess data quality, model version, and user interface design. In addition, the model must be monitored and adjusted to reflect changing insurance landscapes and user demands to be most useful and effective. The model with user input functionality is commendable, but its usefulness will depend on rigorous investigation, improvement, and redesign for practical situations to remain relevant and credible.

The prediction of total economic loss in auto bodily injury accident claims, the impact of attorney involvement, gender, marital status, and seatbelt use, and the comparison of claim patterns across insurers, states, and classes for rating could improve claim processing. The menu-driven interface will allow users to interact with the system and empower stakeholders to customize their questions. In addition to a "Quit" button, the software can close itself after analysis. With this user-friendly method, knowledge for making informed judgments and understanding the insurance claim analysis process become more accessible and necessary. The following diagrams below estimate economic loss and claimant outcomes based on attorney engagement, seat belt use, and other demographics.

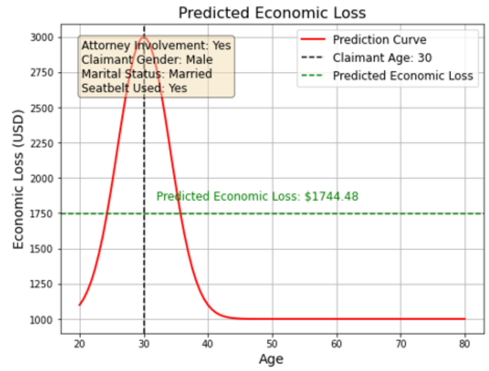


Figure 4: Predicted Economic Loss with Attorney Involvement.

The user first chose choice 1 to estimate the total economic damage from car body harm claims. The user entered attorney participation (1), male claimant (1), married claimant (1), seatbelt usage (1), and claimant age (30). Figure 10 shows avoiding the proposed economic loss to the company by $1744.478063. Later, the user chose option 2 to analyze how attorney involvement, gender, marital status, and seat belt use affect settlements. User entered male (1) as a parameter. The model predicted a claim outcome of $1,844.4781 based on the given variables (figure 11). Defence lawyer engagement and seatbelt use were major factors in the claimant's outcome. If attorney involvement and seatbelt use are included, the first scenario's anticipated public expense is $1744.478063. If added variables affect claim outcome (option 2), the predicted claim outcome is $1844.4781. Decision tree algorithm projected $1746.58 economic loss and 1846.58 claim outcome using the same input data. The irregularity means that the rate of return is stretched by attorney engagement and seatbelt use, and the motorist may have benefited from those variables. Integrative analysis of attendant outcomes is needed when considering attendant choice elements. It suggests ways to improve claims processing and administration.

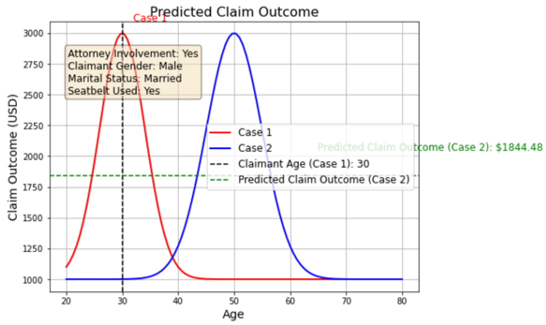


Figure 5: Predicted Claim Outcome with Attorney Involvement

Furthermore, figure6 illustrate the forecasted financial detriment and the result for the individual making a claim, considering factors such as the absence of legal representation and the non-use of seat belts, among other demographic variables.

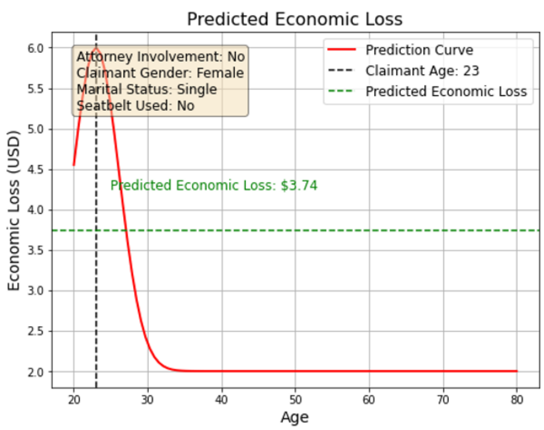


Figure 6: Predicted Economic Loss without Attorney Involvement.

In figure7, the user chooses option 1 and provides information about the cases, including no attorney involvement (0), a female victim (0), an unmarried victim (0), no seatbelt usage (0), and the claimant's age (30). The inputs above cause a $1871.5056 economic loss in the model. The technology then redirects the user to correlate lawyer participation, claimant gender statistics, divorce rate, and seat belt usage into claim results. The chart below shows that the model forecasts a claim of $1871.5056, which matches the economic loss prognosis. Attorney involvement was among the criteria analysed. Decision tree method estimated economic loss of $1871.506 and claim outcome of $1863.504 using the same input data. This plot confirms the consistency of the projected effects between economic loss and the sum insured paying out, regardless of attorney hiring and seatbelt usage, which may not significantly affect the overall sum insured payment. However, the attorney's absence and seatbelt violations affect the claimant's entire process and may provide different results depending on the situation, underlining the necessity for more research.

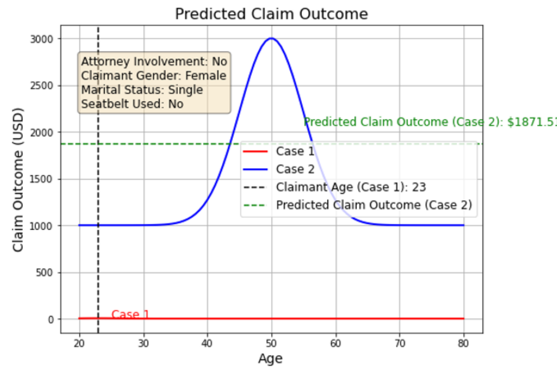


Figure7: Predicted Claim Outcome without Attorney Involvement.

The cases' findings match recent studies on attorney attendance at court and belt wearing practices of automotive bodily harm. Despite failing to show a statistically significant association between legal representation and positive claim resolutions, Aqqad et al. (2023) found that attorneys significantly impact claim outcomes. Conversely, the lawyer was present in one situation and led to beneficial outcomes, while the lawyer was absent in another, which may have resulted in higher economic losses and claim results (Banulescu-Radu and Yankol-Schlack (2023)). Another study by Michaelides et al. (2023) shows that seatbelts reduce accidental injuries and treatment costs. Singhal et al. (2023)'s real-life scenarios show the economic loss of not wearing a seatbelt, illustrating the relevance of seatbelt use in reducing claim costs. While in the first case, there is a common thread among the observed events without attorneys or seatbelts, there is a substantial possibility of complications influencing the prediction of unaccounted-for consequences (Aqqad et al., 2023).

- *Evaluation Performance*

The confusion matrix and ROC curve for random forest and decision tree algorithms were used to evaluate study performance. The graphic below shows the model's performance evaluation after construction.

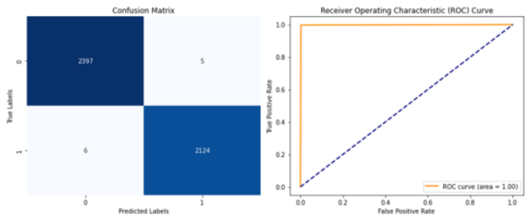


Figure 1: Confusion Matrix and ROC Curve Plots for Random Forest Model.

The confusion matrix showed 99.76% accuracy for random forest method. This reveals that the model's predicted labels are substantially linked with true ones in most circumstances. The 2397 True Negative count shows that the model accurately identifies assertions that should be rejected. This can prevent incorrect pay and save insurance company losses. In contrast, countermanding actual claims with 2124 true positive counts proves the modeling capabilities for capturing legitimate claims, which simplifies processing and insurance release for claim payouts. However, 6 cases of incorrect test results and 5 cases of incorrect results must be acknowledged. This pattern may show when the system mis predicted the outcome, resulting in real claims being denied and fraudulent ones approved. The ideal AUC value is not completely employed by machine learning (ML) algorithms because the classification unit has the most power, but it can be used to analyze the model's distinguishability between positive and negative cases. A model can attain 100% accuracy, which validates it.

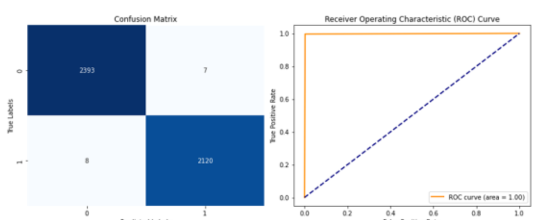


Figure9: Confusion Matrix and ROC Curve Plots for Decision Tree Model.

The confusion matrix showed 99.76% accuracy for random forest method. This reveals that the model's predicted labels are substantially linked with true ones in most circumstances. The 2397 True Negative count shows that the model accurately identifies assertions that should be rejected. This can prevent incorrect pay and save insurance company losses. In contrast, countermanding actual claims with 2124 true positive counts proves the modeling capabilities for capturing legitimate claims, which simplifies processing and insurance release for claim payouts. However, 6 cases of incorrect test results and 5 cases of incorrect results must be acknowledged. This pattern may show when the system mis predicted the outcome, resulting in real claims being denied and fraudulent ones approved. The ideal AUC value is not completely employed by machine learning (ML) algorithms because the classification unit has the most power, but it can be used to analyze the model's distinguishability between positive and negative cases. A model can attain 100% accuracy, which validates it.

Hanafy and Ming (2021) used many techniques, including logistic regression, XGBoost, random forest, decision trees, naïve Bayes, and K-NN, in their first row The random forest model was the best, with 86.77% accuracy, 0.7117 kappa, and 0.840 AUC, indicating its ability to diagnose positive and negative cases. Hanafy & Ming (2021) addressed unbalanced datasets, a typical insurance claim prediction issue, in the second row. Their goal was attained utilizing AdaBoost, oversampling, and hybrid methods. AdaBoost with oversampling properly identified positive and negative cases with 99.4% accuracy, 92.94% sensitivity, and 99.82% specificity. AdaBoost using a hybrid technique performed well on unbalanced data with 99.1% accuracy, 92.48% sensitivity, and 99.63% specificity. In 2020, Abdelhadi et al. examined the use of ANN, decision trees, naïve Bayes, and XGBoost models for insurance claims forecasting. They found that XGBoost and Decision Tree models were best for this task with 92.53% and 92.22% accuracy, respectively. The fourth row suggests a 2024 random forest study with 99.5% accuracy, exceeding the decision tree (95.68%) and other models.

CONCLUSION

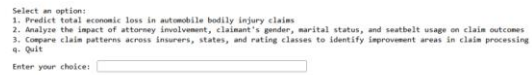
Use of machine learning for insurance claim analysis produces considerable outcomes for handling complexity growth and process speed needs in vehicle insurance. Artificial intelligence is used to simplify insurance claims processing, improve precision, and aid underwriters and policyholders.

Research showed that machine learning is vital to understanding motor vehicle bodily injury claims. The essay discusses how predictive models, attorney engagement, claimant demographics, and seat belt use affect claim outcomes. Thus, a comprehensive approach to these elements is needed since they improve insurance claim analysis accuracy and efficiency.

This research presents an integrated model of the most important elements that affect insurance claims, filling a knowledge vacuum. The preceding research focused on specific topics, while this one analyses claims from a broader viewpoint. Object-Oriented Analysis and Design and machine learning algorithms like Random Forest are used in this step-by-step insurance claim processing research. Comparative machine learning algorithm evaluations show AdaBoost and XGBoost's advantages in imbalanced datasets. This allows further research on insurance claim analysis and predictive algorithms. The study's practical significance shows its power to change insurance claim processes. Insurance businesses can use machine learning to reduce errors and boost productivity in claim evaluation. The results on lawyer influence and seat belt use as claim outcome determinants give insurers practical data to make decisions. Machine learning also helps insurers and policyholders detect and prevent insurance fraud, a serious issue nowadays.

APPENDIX

The insurance claim model is trained using random forest and decision tree algorithms on a Kaggle dataset and integrated using Object-Oriented Analysis and Design (OOAD) into a Jupyter notebook with a user input function to put theory and practice together. User input functionality makes the model dynamic, enabling user participation and real-time insurance decision making. The model must be critically evaluated to ensure robustness and reliability in diverse use case scenarios. To identify biases and inaccuracies, assess data quality, model version, and user interface design. In addition, the model must be monitored and adjusted to reflect changing insurance landscapes and user demands to be most useful and effective. The model with user input functionality is commendable, but its usefulness will depend on rigorous investigation, improvement, and redesign for practical situations to remain relevant and credible. The model predicts as shown below:



Once the model runs in Jupyter Notebook, users are prompted with a menu of options: the prediction of total economic loss in automobile bodily injury accident claims, the assessment of attorney involvement, gender, marital status, and seatbelt use on claim outcomes, or the comparison of claim patterns across insurers, states, and classes for rating. These areas could help to the menu-driven interface will allow users to interact with the system and empower stakeholders to customize their questions. In addition to a "Quit" button, the software can close itself after analysis. With this user-friendly method, knowledge for making informed judgments and understanding the insurance claim analysis process become more accessible and necessary.

REFERENCES

[1] Abdelhadi, S., Elbahnasy, K. and Abdelsalam, M., 2020. A proposed model to predict auto insurance claims using machine learning techniques. *Journal of Theoretical and Applied Information Technology*, *98*(22), pp.3428-3437.

[2] Ahmad, R., Nawaz, M.R., Ishaq, M.I., Khan, M.M. and Ashraf, H.A., 2023. Social exchange theory: Systematic review and future directions. *Frontiers in Psychology*, *13*, p.1015921.

[3] Ali, B., Aibinu, A.A. and Paton-Cole, V., 2024. Improving the process of disruption claims: Identification of the difficulties and expectations. *Journal of Management in Engineering*, *40*(1), p.04023058.

[4] Ali, Y., Hussain, F. and Haque, M.M., 2024. Advances, challenges, and future research needs in machine learning-based crash prediction models: A systematic review. *Accident Analysis & Prevention*, *194*, p.107378.

[5] Andrews, R., Wynn, M., ter Hofstede, A.H., Xu, J., Horton, K., Taylor, P. and Plunkett-Cole, S., 2018. Exposing impediments to insurance claims processing: Compulsory third party

insurance in Queensland. *Business Process Management Cases: Digital Innovation and Business Transformation in Practice*, pp.275-290.

[6] Aqqad, A., 2023. Leveraging Machine Learning Techniques for Enhanced Detection of Insurance Fraud Claims: An Empirical Study. *Available at SSRN 4552815*.

[7] Aslam, F., Hunjra, A.I., Ftiti, Z., Louhichi, W. and Shams, T., 2022. Insurance fraud detection: Evidence from artificial intelligence and machine learning. *Research in International Business and Finance*, *62*, p.101744.

[8] Banulescu-Radu, D. and Yankol-Schalck, M., 2023. Practical guideline to efficiently detect insurance fraud in the era of machine learning: A household insurance case. *Journal of Risk and Insurance*.

[9] Begho, T. and Balcombe, K., 2023. Attitudes to risk and uncertainty: new insights from an experiment using interval prospects. *SAGE Open*, *13*(3), p.21582440231184845.

[10] Benedek, B., Ciumas, C. and Nagy, B.Z., 2022. Automobile insurance fraud detection in the age of big data–a systematic and comprehensive literature review. *Journal of Financial Regulation and Compliance*, *30*(4), pp.503-523.

[11] Chamallas, M., 2024. Trauma Damages. *Ohio State Legal Studies Research Paper*, (825), p.52.

[12] Dong, P., Quan, Z., Edwards, B., Wang, S.H., Feng, R., Wang, T., Foley, P. and Shah, P., 2024. Privacy-Enhancing Collaborative Information Sharing through Federated Learning--A Case of the Insurance Industry. *arXiv preprint arXiv:2402.14983*.

[13] Ellili, N., Nobanee, H., Alsaiari, L., Shanti, H., Hillebrand, B., Hassanain, N. and Elfout, L., 2023. The applications of big data in the insurance industry: A bibliometric and systematic review of relevant literature. *The Journal of Finance and Data Science*, p.100102.

[14] Engstrom, N.F. and Stone, J., 2024. Auto Clubs and the Lost Origins of the Access-to-Justice Crisis. *Yale Law Journal, Forthcoming*.

[15] Fanjul-Hevia, A., Pardo-Fernández, J.C., Van Keilegom, I. and González-Manteiga, W., 2024. A test for comparing conditional ROC curves with multidimensional covariates. *Journal of applied statistics*, *51*(1), pp.87-113.

[16] Hanafy, M. and Ming, R., 2021. Improving imbalanced data classification in auto insurance by the data level approaches. *International Journal of Advanced Computer Science and Applications*, *12*(6).

[17] Hanafy, M. and Ming, R., 2021. Machine learning approaches for auto insurance big data. *Risks*, *9*(2), p.42.

[18] Janiesch, C., Zschech, P. and Heinrich, K., 2021. Machine learning and deep learning. *Electronic Markets*, *31*(3), pp.685-695.

[19] Kaggle. (2021). Machine Learning for Insurance Claims. Retrieved from https://www.kaggle.com/

[20] Karzov, A., 2022. Real estate insurance claims prediction with machine learning algorithms.

[21] Kurniawan, S. and Hilmiyah, N., 2024. Liability of the Carrier for the loss of the Delivery (Studi Kasus CV. Maju Berkah Tanjung Transport). *LEGAL BRIEF*, *12*(6), pp.513-520.

[22] Larner, A.J., 2024. *The 2x2 matrix: contingency, confusion and the metrics of binary classification*. Springer Nature.

[23] Lior, A., 2022. Insuring AI: The role of insurance in artificial intelligence regulation. *Harvard Journal of Law and Technology*, *1*.

[24] Manakhari, S.S., 2023. *Towards Accurate Prediction of Prospective Insurance Customers via an Enhanced Optimization Aided Deep Learning Model* (Doctoral dissertation, Colorado Technical University).

[25] Marzen, C.G., 2024. Insurance Law and Religious Belief. *Available at SSRN 4717035*.

[26] McDermott, M., Hansen, L.H., Zhang, H., Angelotti, G. and Gallifant, J., 2024. A Closer Look at AUROC and AUPRC under Class Imbalance. *arXiv preprint arXiv:2401.06091*.

[27] Michaelides, M., Pigeon, M. and Cossette, H., 2023. Individual claims reserving using activation patterns. *European Actuarial Journal*, *13*(2), pp.837-869.

[28] Muswere, E.T., 2023. *Fraudulent Vehicle Insurance Claims Prediction Model Using Supervised Machine Learning in the Zimbabwean Insurance Industry* (Doctoral dissertation, Chinhoyi University of Technology Zimbabwe).

[29] Njeru, A.M., 2022. *Detection of Fraudulent Vehicle Insurance Claims Using Machine Learning* (Doctoral dissertation, University of Nairobi).

[30] Okagbue, H.I. and Oyewole, O., 2023. Prediction of automobile insurance fraud claims using machine learning. *The Scientific Temper*, *14*(03), pp.756-762.

[31] Owolabi, T., Shahra, E.Q. and Basurra, S., 2023, February. Auto-Insurance Fraud Detection Using Machine Learning Classification Models. In *International Congress on Information and Communication Technology* (pp. 503-513). Singapore: Springer Nature Singapore.

[32] Pamba, F., 2019. *Application Of Machine Learning For Estimating Motor Vehicle Insurance Premium* (Doctoral dissertation, Kca University).

[33] Park, K.H., 2022. *Analytics-assisted triage of psychological workers' compensation claims* (Doctoral dissertation, Macquarie University).

[34] Poppe, E.S.T. and Rachlinski, J.J., 2015. Do lawyers matter? The effect of legal representation in civil disputes. *Pepp. L. Rev.*, *43*, p.881.

[35] Rajan, S.R., 2023. *Risk, Disaster, and Vulnerability: An Essay on Humanity and Environmental Catastrophe*. Univ of California Press.

[36] Rawat, S., Rawat, A., Kumar, D. and Sabitha, A.S., 2021. Application of machine learning and data visualization techniques for decision support in the insurance sector. *International Journal of Information Management Data Insights*, *1*(2), p.100012.

[37] Shestowsky, D., 2023. Civil Litigants' Evaluations of Their Legal Experiences. *Annual Review of Law and Social Science*, *19*.

[38] Shucksmith, M., Glass, J., Chapman, P. and Atterton, J., 2023. *Rural poverty today: experiences of social exclusion in rural Britain*. Policy Press.

[39] Singhal, A., Singhal, N. and Sharma, K., 2023, June. Machine Learning Methods for Detecting Car Insurance Fraud: Comparative Analysis. In *2023 3rd International Conference on Intelligent Technologies (CONIT)* (pp. 1-5). IEEE.

[40] Skillman, G.L. and Veneziani, R., 2023. The problem (s) with representing decision processes under uncertainty. *Journal of Post Keynesian Economics*, *46*(3), pp.420-439.

[41] Starr, S., 2023. Statistical Discrimination. *Harv. CR-CLL Rev.*, *58*, p.579.

[42] Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M. and Liang, X., 2024. An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Systems with Applications*, *237*, p.121549.

[43] Tarr, A.A., Tarr, J.A., Thompson, M. and Wilkinson, D. eds., 2023. The Global Insurance Market and Change: Emerging Technologies, Risks and Legal Challenges.

[44] Van Anh, N. and Duc, T.M., 2024. Big Data-Driven Predictive Modeling for Pricing, Claims Processing and Fraud Reduction in the Insurance Industry Globally. *International Journal of Responsible Artificial Intelligence*, *14*(2), pp.12-23.

[45] Wang, Y. and Xu, W., 2018. Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. *Decision Support Systems*, *105*, pp.87-95.

[46] Yu, X.R., Wang, D.B. and Zhang, M.L., 2024. Dimensionality Reduction for Partial Label Learning: A Unified and Adaptive Approach. *IEEE Transactions on Knowledge & Data Engineering*, (01), pp.1-18.

[47] Zabihi, F., Davoodi, S.R. and Nordfjærn, T., 2019. The role of perceived risk, reasons for non-seat belt use and demographic characteristics for seat belt use on urban and rural roads. *International journal of injury*

control and safety promotion, *26*(4), pp.431-441.