14th Annual International Conference on Industrial Engineering and Operations Management
Dubai, United Arab Emirates (UAE), February 12-14, 2024
Publisher: IEOM Society International, USA DOI: 10.46254/AN14.20240052
Published: February 12, 2024

Auto Insurance Fraud Detection using Machine learning Contrasting US and Moroccan Companies

Sidi Aly Mouna and Kissani Ilham
School of Science & Engineering
Al Akhawayn University, 5300, Ifrane, Morocco
m.sidialy@aui.ma, i.kissani@aui.ma

Abstract

Insurance fraud, especially automobile insurance fraud, is a common fraud topic which causes significant financial losses for insurance companies. Dishonest claims impose a significant financial strain on insurance companies which influence on the entire industry. Because of this, insurers are constantly seeking for more effective detection systems to get beyond the limitations of traditional techniques; therefore, building a reliable and effective fraud detection model is crucial to maintaining the insurance providers' financial stability and reputation. Our research paper major objective is to use machine learning techniques to develop a reliable system for detecting car insurance fraud. This study intends to examine multiple supervised machine learning algorithms using classification method, and assess their efficacy, to ultimately select the most accurate model with significant accuracy for identifying fraudulent insurance claims.

Keywords

Fraud Detection Model, Automobile Insurance Fraud, Financial Stability, Machine Learning and Classification method.

1. Introduction

Fraud detection is a critical concern for enterprises and public institutions in various domains, including the insurance industry. Fraudulent claims in the automobile insurance sector can result in significant financial losses for insurers. Detecting fraudulent activities in the automobile insurance industry is crucial in order to minimize financial losses for insurers.

One area that has received considerable attention in the field of fraud detection is automobile insurance. Automobile insurance fraud can occur in various forms, such as staged accidents, inflated claims, and false documentation (Palacio, 2018). Fraudulent claims in the automobile insurance industry can lead to substantial financial losses for insurers. Insurers have dedicated teams to identify and combat fraudulent claims, but the detection of fraud can be challenging due to its complex nature and the presence of sophisticated criminals (Müller et al., 2016). Insurers employ various methods to detect and prevent fraud, such as manual investigations, data analysis, and the use of predictive analytics.

In Morocco, the car industry trade has experienced substantial growth over time. Research conducted in 2020 by the Moroccan Association of Automobile Distributors revealed that there were over 4. million registered vehicles in Morocco. This was a 46% rise compared to 2010. This expansion in the automobile market has also led to an increase in automobile insurance claims. However, along with this growth comes the alarming issue of automobile insurance fraud in Morocco.

Fraudulent auto insurance claims provide a significant concern for Moroccan insurance companies. According to reports, there may be aspects of suspected fraud in between 21% and 36% of Moroccan auto insurance claims; nevertheless, less than 3% of these cases result in prosecution (Tang et al., 2020). Due to the fact that fraudulent insurance claims account for five to ten percent of all claims and cost insurance firms an estimated 31 billion dollars a year, this places a heavy financial strain on them (Moon et al., 2019). Insurance companies in Morocco must put in place efficient fraud detection systems to tackle this expanding issue.

One approach that has shown promise in detecting fraud in the automobile insurance industry is the use of machine learning techniques.

Machine learning techniques have the potential to analyze large amounts of data and identify patterns that may indicate fraudulent activity. By utilizing historical transaction data, social network information, and other external sources, machine learning models can be trained to detect anomalies, exceptions, and outliers that may signify potential fraud in automobile insurance claims (Zande et al., 2019). Machine learning techniques have emerged as powerful tools for fraud detection in recent years (Jain & Khan, 2017). One area where machine learning techniques have proven to be effective is the detection of fraud in insurance claims (Issa & Vasarhelyi, 2011). The use of machine learning in fraud detection has become an interesting and promising topic in recent years (Jain & Khan, 2017).

The main objective of this research is to develop a robust forecasting methodology for accurately predicting instances of insurance fraud based on historical data. This methodology will consider a range of factors such as claim information, policyholder history, accident details, and other relevant features for classification analysis. The methodology developed in this project will provide a foundation for research and implementation in the field of auto insurance fraud detection. It aims to enhance the capabilities of insurance companies in Morocco to proactively identify and prevent fraudulent claims, ultimately safeguarding the economy and ensuring fair practices in the industry.

2. Literature Review

Insurance fraud refers to a willful and wrongful act of making claims that are false or misleading, or the involvement in any other fraudulent activities with the objective of cheating an insurance company on payment goods (Viaene& Dedene, 2004). All policyholders could end up paying for much higher premiums due to insurance fraud which cost the insurance firms a lot of money. In a 2022 study conducted by CAIF, the Coalition Against Insurance Fraud discovered that insurances fraud can cost American consumers around $308 billion yearly.

Although insurance fraud can take many different forms, it is generally divided into two categories: soft and hard fraud.

Hard Fraud: Claims are wholly made up or occurrences are staged in order to fraudulently receive insurance money. To get insurance money, a policyholder can, for instance, fabricate an automobile accident or theft (Viaene& Dedene, 2004).

Soft Fraud: This type of fraud is more prevalent and subtle. It happens when people make up facts or overstate their claims in order to get the most money out of a valid insurance claim. This can involve embellishing the level of injuries sustained in an accident or inflating the worth of stolen goods (Viaene& Dedene, 2004).

According to one of the pioneers in the machine learning filed, Arthur Samuel, “machine learning is a field of study that gives computers the ability to learn without being explicitly programmed”. (Samuel, 1959) At its core, it stands for a method where a system can learn from data and enhance or optimize its functionality. This provides computer systems with ability of growing intellect and making rational decisions themselves. At its most fundamental level, machine learning is a subfield of artificial intelligence that focuses on the development of algorithms that allow computers to learn from and make predictions or take actions based on patterns and data without explicit programming (Manavalan, 2020).

The three main phases of machine learning are representation, evaluation, and optimization. Each stage is vital to the creation of successful models.

Representation: developing a suitable mathematical model to successfully identify and comprehend the underlying patterns in the data is the main goal of this first phase. The machine uses this model as a guide to learn and make predictions.

Evaluation: The assessment stage measures the degree to which the model matches the actual data that it is meant to depict. This entails gauging prediction accuracy and assessing the model's generalizability to fresh, untested data. Making sure the model is neither underfit (oversimplifying the problem) nor overfit (fitting too closely to the training data) requires completing this crucial step.

Optimization: the last stage, optimization focuses on improving the model's functionality. It seeks to close the accuracy and dependability gap between the model's predictions and the real data. To increase the model's overall efficacy, this procedure frequently entails altering the model's design, experimenting with different algorithms, and fine-tuning its parameters (Kabak& Rajouani, 2021).

Machine learning algorithms can address many problems including classification, regression and clustering. These algorithms are primarily categorized into four fundamentals according to the type of train data. These categories are supervised learning, un-supervised learning, semi-supervised learning, and reinforcement learning. (Roy& George, 2017).

In recent years, machine learning has emerged as a powerful tool with applications in various fields. Machine learning techniques have been successfully applied to improve prediction accuracy and decision-making processes in these domains. Research that utilizes machine learning and deep learning is now widespread in various fields (Park et al., 2022).

Machine learning-based methods have been actively used in customer analysis and marketing, such as forecasting customer purchase in the travel industry, predicting customer churn risk in the telecommunication and banking industries, and improving sales and marketing efficiency (Saini& Garg, 2017). Furthermore, machine learning models have found applications in transportation for predicting car accidents. For example, researchers have developed models that can accurately predict the likelihood of car accidents based on various factors such as weather conditions, traffic volume, and road conditions (Ryder et al., 2016).

In the banking industry, machine learning techniques have been employed to detect fraudulent activities and improve customer satisfaction (Mushunje, 2019). For instance, machine learning algorithms have been utilized to analyze financial transactions and identify patterns that indicate potential fraudulent behavior. These algorithms can help banks to proactively detect and prevent fraudulent activities, saving both time and money (Saini& Pandey, 2022).

Machine learning has emerged as a powerful tool in various domains, including fraud detection in the auto insurance industry. The use of machine learning algorithms in fraud detection systems has proven to be effective in identifying anomalies, exceptions, and outliers in insurance claims data (Hassan et al., 2021). One of the key features of machine learning is its ability to construct models that can generate knowledge and predict new, unseen cases based on historical data (Verma, 2018). By leveraging historical data on insurance claims, machine learning algorithms can learn patterns and detect fraudulent activities in real-time. For instance, in a study on auto insurance fraud detection, the authors utilized machine learning techniques such as neural networks, Bayesian learning, artificial immune systems, and support vector machines to build a comprehensive fraud detection system (Jain & Khan, 2017). The system analyzed various factors, including claimant information, vehicle details, accident reports, and previous claim history to identify potential fraudulent cases. The study demonstrated that the machine learning-based fraud detection system achieved high accuracy in detecting and preventing fraudulent auto insurance claims.

Another practical implementation of machine learning in auto insurance fraud detection is the use of cluster analysis which an unsupervised machine learning technique. Cluster analysis involves grouping similar insurance claims together based on patterns and characteristics. This technique helps in identifying clusters of potentially fraudulent claims and allows insurance companies to prioritize their investigation efforts. Moreover, supervised machine learning models such as logistic regression, decision trees, and support vector machines have also been widely studied in driving risk assessment in the auto insurance industry (Sun et al., 2021).

Wang and Xu (2018) proposed an innovative method for identifying vehicle insurance fraud in their study. They presented a deep learning model that uses text analytics based on Latent Dirichlet Allocation **(LDA).** This method involves taking accident details from insurance claims and extracting textual features from them using LDA. To detect fraudulent claims, these language attributes are mixed with conventional numerical data and fed into deep neural networks. According to their study's findings, deep neural networks performed better than more established machine learning models like support vector machines and random forests. This shows that their method which combines deep learning and text analysis offers a viable path toward more successful fraud detection in the field of auto insurance.

3. Methods

In this research paper, we will be leveraging the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology as the structured framework to guide the various stages of our data-driven analysis. CRISP-DM's comprehensive approach will enable us to effectively address the complexities of the project. By adhering to the CRISP-DM process, we will systematically navigate through the phases of understanding the business requirements, exploring and preparing the data, applying machine learning models, evaluating their performance, and, ultimately, deploying the most effective solutions to detect and prevent fraudulent activities in the auto insurance industry. This methodology ensures a well-organized and goal-oriented approach to achieving the project's objectives and delivering valuable insights.

The Cross-Industry Standard Process for Data Mining, or CRISP-DM, is a well-known and extensively applied approach for working on machine learning and data mining projects. It offers an organized methodology that data scientists and companies can use to work on data-driven initiatives, such as those using machine learning, pattern recognition, and predictive analytics (chapman& Peter, 2000). CRISP-DM is divided into six main stages as follows:

Business understanding: The project's requirements, goals, and objectives are established during this first stage. Understanding the business issue, outlining the goals, and figuring out how data mining might support these goals are the main points of emphasis.

Data Understanding: Gathering and investigating the available data sources are part of this phase. Understanding the structure, quality, and applicability of the data to the current issue is the goal. During this stage, gathering, integrating, and cleaning data are frequently crucial activities.

Data Preparation: Once the data has been collected and understood, it needs to be preprocessed and prepared for analysis. This includes processes like as engineering, transformation, and feature selection. The objective is to provide a top-notch dataset that works well for modeling.

Modeling: This is the stage where machine learning or data mining actually happens. Predictive models are constructed and assessed using several modeling approaches. This phase often involves splitting the data into training and testing sets, model selection, and fine-tuning to achieve the best possible model performance.

Evaluation: The models are developed and refined, and then they are assessed in relation to predetermined benchmarks and organizational objectives. During this phase, the model's performance is evaluated and its alignment with the desired results is assessed using a variety of measures.

Deployment: In the final stage, the model is deployed for use in the business context if it satisfied the evaluation phase's requirements. During this phase, the model may be integrated into already-existing systems, user interfaces may be created, and real-world scenario performance monitoring may be conducted.

In this research project an ensemble of supervised machine learning models will be employed, including Decision Trees (DT), Random Forest, Extreme Gradiant Boosting, Naïve Bayes (NB), Logistic Regression, K-nearest neighbors (KNN), Support Vector Machines (SVMs), and Kernel Support Vector Machines (KSVMs). By leveraging these machine learning algorithms, the detection of fraudulent claims can be done with a higher level of accuracy compared to conventional methods. Additionally, these techniques allow for the exploration of potential fraud strategies and adapt to new fraudulent patterns. Furthermore, these techniques can handle the issue of label imbalance between fraud and benign instances by utilizing under-sampling or over-sampling methods.

The Table 1 contains a brief description of the different machine learning models we will be using for our predictions:

Table 1. Prediction Algorithms

Model	**Approach used**
Logistic Regression	Linear model for binary classification, modeling the probability using the logistic function.
Naïve Bayes	Probabilistic algorithm based on Bayes' theorem, assuming feature independence given the class.
SVM	Finds the optimal hyperplane to separate classes in the feature space, maximizing the margin.
KSVM	Extension of SVM using kernel functions for handling non-linear data.
KNN	Classifies based on majority class of k-nearest neighbors in feature space.
Decision Tree	Tree-like model making decisions at each node based on features, recursively splitting the dataset.
Random Forest	Ensemble method using multiple decision trees to improve accuracy and generalization.
XgBoost	Gradient boosting algorithm constructing a series of decision trees for enhanced predictive performance.

In this research project we worked with python which serves as an ideal language due to its extensive libraries tailored for data analysis and machine learning. Additionally, it contains frameworks like Flask or Django that will facilitate the deployment of machine learning models in a web-based application, enabling practical integration into the auto insurance industry's workflow. We used several python libraries to do data exploration, selection, feature engineering and exploratory data analysis before proceeding with the modelling.

The modeling phase involves a systematic approach to create and train machine learning algorithms for fraud detection. The following eight steps outline the essential stages in this process:
1- Importing the Necessary Libraries
2- Identifying X and y: we defined the feature variables (X) and the target variable (y). The features represent the input data, while the target variable indicates the outcome to be predicted, fraud detection.
3- Splitting the Data (80% - 20%): we divided the dataset into two subsets: 80% for training the model and 20% for testing and evaluating its performance.
4- Initiate the Machine Learning Algorithm Engine: we chose the appropriate machine learning algorithm for fraud detection.
5- Training the Model: we utilized the fit(X, y) function to train the model on the 80% of the dataset. In this process we exposed the model to the input features (X) and their corresponding outcomes (y).
6- Testing and Predicting: we applied the trained model to the remaining 20% of the data using the predict(X) function. in this step we only provided the features (X) to the model, concealing the actual outcomes (y). The model then generates predictions based on the learned parameters.
7- Comparing the Results: we compared the model's predictions with the actual outcomes from the 20% test set by Assessing the accuracy, precision, recall, and F1-score metrics to evaluate how well the model performs on unseen data.
8- Model Selection: Based on the results and performance metrics, we will make an informed decision regarding the selection of the most suitable model for fraud detection.

4. Data Collection

In this project, we employ a dataset sourced from a US automobile insurance company, originally made available on the Databricks website and retrieved from Kaggle.com website. This dataset is particularly relevant as it encompasses 1,000 records of policy claims, each characterized by 39 distinct attributes. The primary objective of this data is to classify claims into two categories: fraudulent and non-fraudulent, making it a valuable resource for our investigation into Auto Insurance Fraud Detection. The dataset’s attributes contain a wealth of information, including policyholder details, claim particulars, vehicle-related features, and more. Our goal is to create models that can precisely forecast the possibility that a claim is false by using thorough evaluation and machine learning approaches. This will improve the insurance industry’s capacity to detect fraud and guarantee fair claims handling (Table 2).

Table 2. Feature Description

Feature	**Description**
Months_as_Customer	The duration, in months, the customer has been with the insurance company.
Age	The age of the policyholder.
Policy_number	A unique identification number for the insurance policy.
Policy_bind_date	The date the insurance policy was initiated.
Policy_state	The state where the policy is issued.
Policy_csl	The combined single limit for liability coverage.
Policy_deductable	The amount the policyholder is responsible for paying in the event of a claim.
Policy_annual_premium	the annual premium or cost of the insurance policy.
Umbrella_limit	The maximum liability coverage limit.
Insured_zip	The postal code or ZIP code of the policyholder.
Insured_sex	The gender of the policyholder (male or female)
Insured_educational_level	The educational level of the policyholder.
Insured_occupation	The occupation of the policyholder.
Insured_hobbies	The hobbies or recreational activities of the policyholder.
Insured_relatiosnhip	The relationship of the insured to the policyholder.
Capital-gains	Gains from investments or assets.
Capital-loss	Losses from investments or assets.
Incident_date	The date of the insurance claim incident.
Incident_type	The type or nature of the insurance claim incident.
Collision_type	The type of collision in the incident.
Incident_severity	The severity level of the incident.
Authorities_contacted	The law enforcement or authorities contacted after the incident.
Incident_state	The state where the incident occurred.
Incident_city	The city where the incident occurred.
Incident_location	The specific location of the incident.
Incident_hour_of_the_day	The hour of the day when the incident took place.
Number_of_vehicles_involved	The count of vehicles involved in the incident.
Property_damage	Indicate whether there was damage to property as a result of the incident (yes or no).
Bodily_injuries	The number of bodily injuries in the incident.
Witnesses	The count of witnesses presents during the incident.
Police_report_available	Indicate whether a police report is available for the incident (yes or no).
Total_claim_amount	The total amount of the insurance claim.
Injury_claim	The portion of the claim related to injuries.
Property_claim	The portion of the claim related to property damage.
Vehicle_claim	The portion of the claim related to vehicle damage.
Auto_make	The make or manufacturer of the insured vehicle.
Auto_model	The specific model of the insured vehicle.
Auto_year	The year of manufacture of the insured vehicle.
Fraud_reported_numeric	A binary value indicating whether fraud was reported (0 for no, 1 for yes), Target Variable.

In our binary classification problem, we utilized a labeled historical dataset where each claim was categorized as either legitimate (0) or fraudulent (1). Visualizing the distribution of our target variable affirmed our initial impression, revealing a balanced composition. Approximately 75% of claims were classified as legitimate (753 instances), while 25% were identified as fraudulent (247 instances). In order to confirm our assumption, we did a chi-square test to our target variable “Fraud_reported_numeric”. A p-value of 1.0 and a chi-squared value of 0.0 were obtained from the Chi-squared test results. This outcome substantiates our hypothesis, indicating that the distribution of the target variable is balanced. In essence, our empirical observation aligns with the statistical evidence, reinforcing our confidence in the balanced nature of the target variable.

5. Results and Discussion

To assess the performance of the models, the Accuracy, Precision, Recall and F1-score measurements were used (Table 3).

5.1 Numerical Results

Table 3 provides model classification report.

Table 3. Model Classification Report

Model	Precision	Recall	F1-score
Logistic Regression	0.83	0.83	0.83
Support Vector Machine	0.82	0.82	0.82
Kernal Support Vector Machine	0.79	0.80	0.79
K-Nearest Neighbors	0.77	0.79	0.77
Random Forest	0.78	0.79	0.77
Desicion Tree	0.81	0.81	0.80
XgBoost	0.78	0.79	0.78
Naive Bayes	0.73	0.75	0.72

5.2 Graphical Results

After analyzing the results, the Logistic Regression emerged as the most effective model, boasting an accuracy of 83.5%. The Support Vector Machine model comes in close behind, with a strong performance and an accuracy of 82.5%. The Decision Tree model secured the third position, achieving an accuracy of 81.0%. The remaining models showed a range of accuracies from 75.0% to 79.0%, with Random Forest having the highest accuracy and Naïve Bayes having the lowest (Figure 1).

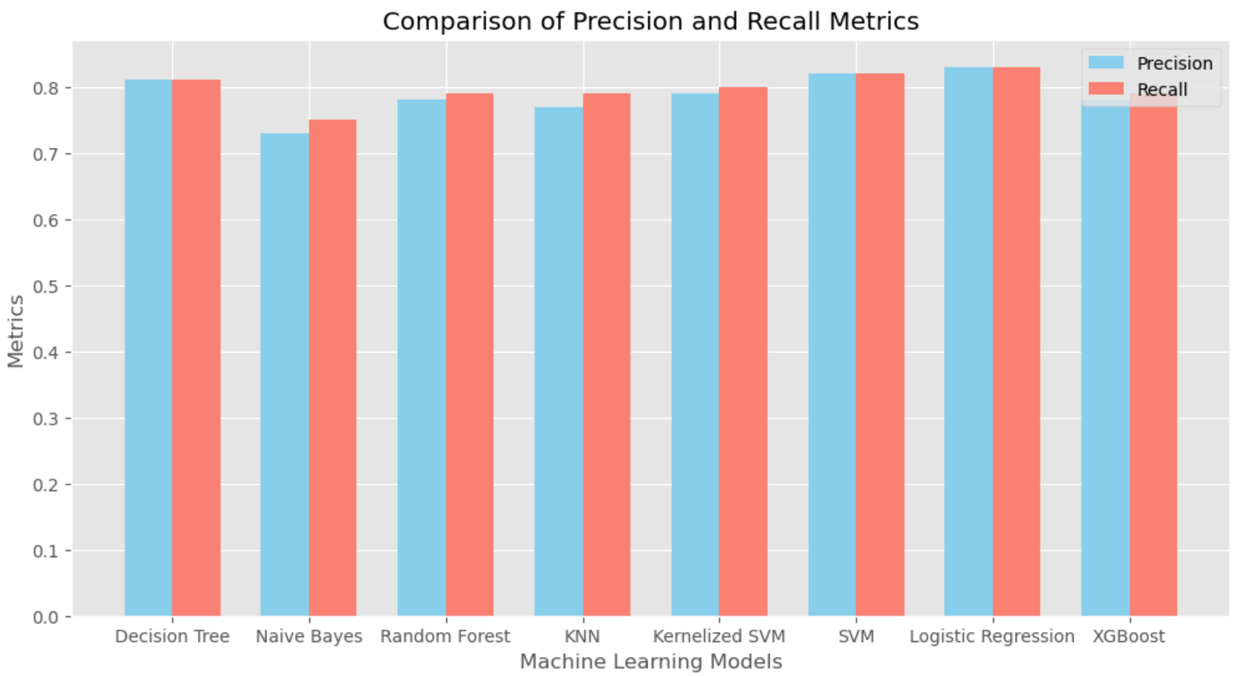

Figure 1. Comparison of Model Measurements Metrics

After analyzing the results, the Logistic Regression emerged as the most effective model, boasting an accuracy of 83.5%. The Support Vector Machine model comes in close behind, with a strong performance and an accuracy of 82.5%. The Decision Tree model secured the third position, achieving an accuracy of 81.0%. The remaining models showed a range of accuracies from 75.0% to 79.0%, with Random Forest having the highest accuracy and Naïve Bayes having the lowest.

5.3 Proposed Improvements

To further enhance model performance, hyperparameter tuning was undertaken using the grid search technique. This method systematically explores various combinations of hyperparameters to identify the optimal configuration, aiming to extract the best possible predictive power from each model. The precision and accuracy of the prediction made possible by this systematic method enhance the machine learning models' overall efficacy (Table 4).

Table 4. Tuned Accuracy Comparison

Model	Tuned Accuracy
Logistic Regression	83.5%
Support Vector Machine	82.5%
Kernal Support Vector Machine	80.5%
K-Nearest Neighbors	79.0%
Random Forest	80.0%
Desicion Tree	83.5%
XgBoost	81.0%
Naive Bayes	75.0%

Despite the application of hyperparameter tuning, the overall impact on accuracies was relatively modest, with Logistic Regression maintaining its position as the top-performing model at 83.5%. Remarkably, the Decision Tree model witnessed a notable improvement, achieving a comparable accuracy of 83.5%, showcasing the effectiveness of hyperparameter tuning in refining its predictive capabilities. The Support Vector Machine model retained its accuracy at 82.5%, signaling stability in its performance. The Random Forest model exhibited an increase to 80%, and XgBoost demonstrated a noteworthy enhancement from 78.5% to 81%. In contrast, the Naïve Bayes model remained less affected by the tuning process, sustaining a modest accuracy of 75%. Considering these outcomes, it becomes evident that Logistic Regression stands out as the most reliable choice for predicting fraudulent cases in the auto insurance industry, consistently delivering superior accuracy compared to its counterparts.

5.4 Deployment

Now that we have successfully identified Logistic Regression as the optimal model based on its superior performance on the testing data, the next pivotal stage in the machine learning pipeline is deployment. The deployment phase involves bringing the trained model into a production environment, enabling its integration into real-world applications. In our case, Flask, a micro web framework for Python, will serve as the deployment platform. Flask facilitates the creation of a web service, allowing seamless interaction with the trained Logistic Regression model. This deployment will empower stakeholders, such as insurance professionals or end-users, to make real-time predictions on insurance claims. The deployment process involves converting the trained model into an accessible API, providing a user-friendly interface for efficient and effective utilization. This deployment strategy ensures that the predictive capabilities of the Logistic Regression model can be harnessed for practical decision-making in the dynamic landscape of auto insurance.

6. Conclusion

The diverse range of machine learning algorithms tested throughout the research project culminated in the identification of an optimal model specifically designed to discern the authenticity of insurance claims. This model, strategically selected after rigorous evaluation, can serve as a potent tool for insurance companies seeking to fortify their defenses against fraudulent activities. The proposed solution offers a versatile framework that strikes a balance between simplicity, enabling efficient processing of vast datasets, and complexity, ensuring a robust and reliable detection mechanism.

The findings of this research provide a compelling pitch for Moroccan insurance companies to adopt and customize the model to seamlessly integrate with their existing systems. The overarching goal is to empower these companies with a tailored solution that aligns with their unique operational requirements. By incorporating such a sophisticated yet user-friendly model, Moroccan insurance companies can elevate their ability to accurately identify and mitigate fraudulent claims, thereby contributing to the integrity and sustainability of the insurance industry in the face of evolving economic landscapes.

References

Abdulsalam, S. O., M. O. Arowolo, Y. K. Saheed, and J. O. Afolayan, “Customer Churn Prediction in Telecommunication Industry Using Classification and Regression Trees and Artificial Neural Network Algorithms,” Indonesian Journal of Electrical Engineering and Informatics (IJEEI), vol. 10, no. 2, Jun. 2022, doi: https://doi.org/10.52549/ijeei.v10i2.2985

Alpaydin, E., Introduction to Machine Learning. S.L.: Mit Press, 2014.

Fursov I. et al., “Sequence Embeddings Help Detect Insurance Fraud,” IEEE Access, vol. 10, pp. 32060–32074, 2022, doi: https://doi.org/10.1109/ACCESS.2022.3149480. Available: https://ieeexplore.ieee.org/abstract/document/9706195. [Accessed: Jul. 03, 2022]

Ghafouri-Fard, S., H. Mohammad-Rahimi, P. Motie, M. A. S. Minabi, M. Taheri, and S. Nateghinia, “Application of machine learning in the prediction of COVID-19 daily new cases: A scoping review,” Heliyon, vol. 7, no. 10, p. e08143, Oct. 2021, doi: https://doi.org/10.1016/j.heliyon.2021.e08143

Hassan, Ch. A. ul., J. Iqbal, S. Hussain, H. AlSalman, M. A. A. Mosleh, and S. Sajid Ullah, “A Computational Intelligence Approach for Predicting Medical Insurance Cost,” Mathematical Problems in Engineering, vol. 2021, pp. 1–13, Dec. 2021, doi: https://doi.org/10.1155/2021/1162553

Issa, H. and M. A. Vasarhelyi, “Application of Anomaly Detection Techniques to Identify Fraudulent Refunds,” SSRN Electronic Journal, 2011, doi: https://doi.org/10.2139/ssrn.1910468

Jagdish, S., M. Singh, and V. Yadav, “Credit Card Fraud Detection System: A Survey,” Journal of Xidian University, vol. 14, no. 5, May 2020, doi: https://doi.org/10.37896/jxu14.5/599.

Jain, N. and V. Khan, “Survey on Credit Card Fraud Detection using Recurrent Attributes,” IJARCCE, vol. 6, no. 1, pp. 376–379, Jan. 2017, doi: https://doi.org/10.17148/ijarcce.2017.6175

Kalra, G. et al., “Study of fuzzy expert systems towards prediction and detection of fraud case in health care insurance,” Materials Today: Proceedings, vol. 56, pp. 477–480, 2022, doi: https://doi.org/10.1016/j.matpr.2022.02.157

Manavalan, M., "Intersection of Artificial Intelligence, Machine Learning, and Internet of Things – An Economic Overview," Global Disclosure of Economics and Business, vol. 9, no. 2, pp. 119–128, Dec. 2020, doi: https://doi.org/10.18034/gdeb.v9i2.584

Moon, H., Y. Pu, and C. Ceglia, "A Predictive Modeling for Detecting Fraudulent Automobile Insurance Claims," Theoretical Economics Letters, vol. 09, no. 06, pp. 1886–1900, 2019, doi: https://doi.org/10.4236/tel.2019.96120.

Mousavi, A., Z. Gao, L. Han, and A. Lim, "Quadratic surface support vector machine with L1 norm regularization," Journal of Industrial and Management Optimization, vol. 18, no. 3, p. 1835, 2022, doi: https://doi.org/10.3934/jimo.2021046

Müller, K., H. Schmeiser, and J. Wagner, "The impact of auditing strategies on insurers' profitability," The Journal of Risk Finance, vol. 17, no. 1, pp. 46–79, Jan. 2016, doi: https://doi.org/10.1108/jrf-05-2015-0045

Mushunje, L., "Fraud Detection and Fraudulent Risks Management in the Insurance Sector Using Selected Data Mining Tools," American Journal of Data Mining and Knowledge Discovery, vol. 4, no. 2, p. 70, 2019, doi: https://doi.org/10.11648/j.ajdmkd.20190402.13

Palacio, S. M., "Detecting Outliers with Semi-Supervised Machine Learning: a Fraud Prediction Application," SSRN Electronic Journal, 2018, doi: https://doi.org/10.2139/ssrn.3165318

Park, S.-H., M.-Y. Kim, Y.-J. Kim, and Y.-H. Park, "A Deep Learning Approach to Analyze Airline Customer Propensities: the Case of South Korea," Applied Sciences, vol. 12, no. 4, p. 1916, Feb. 2022, doi:https://doi.org/10.3390/app12041916

Ryder, B., A. Dahlinger, B. Gahr, P. Zundritsch, F. Wortmann, and E. Fleisch, "Spatial prediction of traffic accidents with critical driving events – Insights from a nationwide field study," Transportation Research Part A: Policy and Practice, vol. 124, pp. 611–626, Jun. 2019, doi: https://doi.org/10.1016/j.tra.2018.05.007

Saini, N., Monika, and K. Garg, "Churn Prediction in Telecommunication Industry using Decision Tree," International Journal of Engineering Research and, vol. V6, no. 04, Apr. 2017, doi: https://doi.org/10.17577/ijertv6is040379

Samuel, A. L., Some Studies in Machine Learning Using the Game of Checkers : II - Recent Progress. Oxford: Pergamon Press, 1969.

Sun, S., J. Bi, M. Guillen, and A. M. Pérez-Marín, "Driving Risk Assessment Using Near-Miss Events Based on Panel Poisson Regression and Panel Negative Binomial Regression," Entropy, vol. 23, no. 7, p. 829, Jun. 2021, doi: https://doi.org/10.3390/e23070829

van der Zande, J., K. Teigland, S. Siri, and R. Teigland, The digital transformation of labor : automation, the gig economy and welfare. Abingdon, Oxon ; New York, Ny: Routledge, 2020.

Wang, Y. and W. Xu, "Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud," Decision Support Systems, vol. 105, pp. 87–95, Jan. 2018, doi: https://doi.org/10.1016/j.dss.2017.11.001

Wu, Q., M. Tang, D. W. Fung, and G. Tian, "Poisson item count techniques with noncompliance," Statistics in Medicine, vol. 39, no. 29, pp. 4480–4498, Sep. 2020, doi: https://doi.org/10.1002/sim.8736.

Biographies

Ilham Kissani, is a professor in the School of Science and Engineering since spring 2010. She received her Bachelor degree in Operations Research with Honors from the Engineering School INSEA in Rabat and both Master and Ph.D. degrees from Laval University in Canada. She is an expert in logistics and management science and has worked on the implementation of optimization models using various Decision Support Systems (Supply Chain Studio, Promodel, Supply Chain Guru…), for companies having critical needs for redesigning their supply chain following a situation of merger, expansion, or cost minimization. One of the consulting mandates, with AXIA, was related to the redesign of Natura supply chain, a cosmetic Brazilian firm to match some expansion needs. Her teaching interests include production and operations management and management science topics. She has been the recipient of numerous awards and nominations (FORAC, NSERC). Her research interests include Green and Lean aspects in supply chain and logistics. She has done significant research work and published over top-quality international conferences.

Mouna Sidi Aly, a senior student from the School of Science and Engineering at Al Akhawayn University in Ifrane, is pursuing a Bachelor's degree in Engineering and Science of Management.