Automating Risk Assessment: The Role of Artificial Intelligence in Insurance Underwriting

Jeevan Sreerama

Independent AI/ML Researcher, Senior Data Scientist, USA

jeevan.sreerama@outlook.com

Abstract: Automating risk assessment through artificial intelligence (AI) can significantly transform the insurance underwriting process by improving accuracy and efficiency. This paper explores the development and implementation of a hybrid machine learning model that integrates logistic regression and support vector machines (SVM) for enhanced underwriting risk assessment. By meticulously analyzing historical claim data and detailed customer profiles, this hybrid model achieves a notable accuracy of 91%, far surpassing the 81.3% accuracy typically associated with manual underwriting practices. Logistic regression is utilized for its simplicity and effectiveness in modeling relationships between dependent and independent variables. It helps in identifying key risk factors from the dataset, providing clear insights into how various customer attributes influence claim likelihood. Support vector machines (SVM) are then applied to classify and predict the likelihood of claims, leveraging their strength in handling both linear and non-linear data. The combination of these two methods results in a robust predictive model capable of delivering highly accurate risk assessments. The model's ability to predict claims with such high accuracy not only enhances the precision of risk assessments but also significantly speeds up the underwriting process. This reduction in processing time can lead to faster policy issuance, improved customer satisfaction, and operational efficiencies within insurance companies. Furthermore, the AI-driven approach enables insurers to identify high-risk individuals more accurately, allowing for better resource allocation and risk management.

Keywords: Artificial Intelligence, Predictive Underwriting, Commercial Lines Insurance, Neural Networks, Decision Trees, Risk Selection, Pricing Strategies, Data Analysis, InsurTech, Industry Transformation

1.Introduction

AI is revolutionizing insurance underwriting. Using machine learning (ML), Pankaj Zanke and Dipti Sontakke [1] estimate auto insurance risk. The study explores predictive modeling with neural networks and decision trees for claims frequency, severity estimation, and fraud detection. By analyzing massive datasets, ML algorithms improve risk assessments and insurance operations. This work provides a theoretical foundation for risk assessment ML algorithms and assesses their performance using empirical case studies to understand ML's disruptive potential in auto insurance.AI transformed insurance business underwriting. Progressive Insurance Project Manager Pankaj Zanke and Capgemini Inc [2]. Consultant Dipti Sontakke study how neural networks and decision trees might improve predictive underwriting. AI algorithms use massive datasets to improve risk selection and pricing, boosting liability assessments and insurance profitability. The paper examines AI underwriting implementation benefits and drawbacks, real-world instances, and theoretical basis [3]. AI's potential impact on commercial lines insurance underwriting and the industry's future is examined in detail. The insurance industry relies on auto insurance to safeguard automobile owners from damage, liability, and theft [4]. This sector's regulatory frameworks, consumer preferences, and technology require insurers to adapt their risk mitigation strategies. Auto insurers must accurately estimate risk to price policyholder and vehicle hazards. It allows insurers to set premiums, coverage limits, and deductibles that balance policyholder affordability and profitability. Insurers manage risk portfolios, estimate claim frequency and severity, and efficiently deploy resources with good risk assessment [5].

Machine learning altered auto insurance risk evaluation. Risk assessment was done using actuarial models and statistics using historical data. With powerful algorithms, ML technologies let insurers evaluate massive datasets. ML algorithms improve risk assessment models by finding complex patterns, correlations, and anomalies. Predictive modeling is best with supervised learning (linear regression, decision trees, random forests), whereas fraud detection and risk segmentation require unsupervised learning (clustering, anomaly detection). Underwriting is key to insurance [6].

Large data collections were kept in archaic technologies or paper files. This is evolving as more insurers automate data collection. In the past, underwriters received consumer information via email and bid quickly. The variety of ailments and treatments complicates underwriting, requiring a faster process [7]. AI can improve traditional underwriting. NLP and machine learning classifiers help underwriters analyze unstructured data. Most problematic is using noisy, unstructured email data. Agents can receive automatic email insurance policies and recommendations by training models on extracted features. Decision-making and operational efficiency will increase with underwriting automation and dynamic scenarios. Commercial insurance predictive underwriting assesses risk and sets policy premiums using statistical models and data analytics [8]. Unlike historical data and human processes, predictive underwriting uses advanced analytics and machine learning algorithms to predict outcomes based on various factors. Commercial insurance covers property, liability, and workers' comp. Corporate variety and risk assessment complexity challenge underwriters. Large datasets can help insurers select and price risks via predictive underwriting.AI has revolutionized insurance underwriting. Machine learning (ML) altered auto insurance risk assessment [9].

Traditional risk assessment used historical data for actuarial and statistical methods. Insurers can examine massive, complicated datasets using new algorithms and computational methods with ML. Risk assessment models improve when ML algorithms find complex patterns, correlations, and anomalies. Supervised, unsupervised, and reinforced machine learning improve risk assessment. Claims frequency and severity prediction modeling uses supervised learning methods including linear regression, decision trees, and random forests [10]. These algorithms help insurers estimate losses by detecting risk factor-insurance result correlations in past data. Unsupervised learning methods like clustering, anomaly detection, and dimensionality reduction detect fraud and segment risk. To help insurers detect fraud and cut losses, these algorithms analyze behavior and claim data abnormalities. Dimensionality reduction helps insurers reduce feature space and find risk assessment elements, enhancing model interpretability and performance [11]. Machine learning algorithms in risk assessment are transforming vehicle insurance, giving insurers unprecedented opportunities to improve price, competitiveness, and policyholder value. ML and data-driven insights will transform auto insurance risk assessment in coming years [12].

AI technologies including machine learning, natural language processing, and predictive analytics have helped insurers automate procedures, obtain data insights, and improve decision-making. AI helps insurers quickly examine complicated datasets, find trends, and anticipate future outcomes in underwriting. AI may streamline underwriting operations, minimize manual effort, and improve risk assessment processes for insurers [13]. This study examines commercial lines insurance predictive underwriting using neural networks and decision trees. The research analyses how AI is improving risk selection, underwriting, and pricing techniques in commercial insurance. The paper examines the theoretical foundations of AI in underwriting, real-world case studies and examples of successful AI implementations, the benefits and drawbacks of AI in risk selection and pricing strategies, and the future of the insurance industry. The insurance business uses auto insurance to serve vehicle owners globally. Auto insurance protects against accidents, theft, and other unexpected events [14]. This industry challenges changing laws, consumer preferences, and technology. Therefore, insurers regularly alter their strategies to limit risks and increase policyholder value. Auto insurance estimates policyholder and vehicle risks using risk assessment. Premiums, coverage limitations, and deductibles align policyholder affordability and profitability with accurate risk assessment. Advanced risk assessment methods help insurers manage risk portfolios, estimate claim frequencies and severity, and efficiently allocate resources to avoid negative consequences [15]. Risk assessment helps auto insurance estimate policyholder and vehicle risks. Setting rates, coverage limits, and deductibles that balance policyholder affordability and insurer profitability requires accurate risk assessment. Advanced risk assessment methodologies allow insurers to manage risk portfolios proactively. Predicting claim frequencies and severity helps insurers reduce losses [16].

Predicting automobile insurance claims frequency requires linear regression. Linear regression models the association between driver demographics, vehicle attributes, and geographical factors and insurance claims to help insurers estimate policyholder or risk profile claims. Linear regression depends on fitting a linear equation to observed data to minimize the gap between expected and actual values. Insurance firms can identify claim trends by training the regression model with previous claims data [17]. By quantifying risk factor effects on claims frequency, linear regression helps insurers segment risk and optimize pricing. Simple linear regression predicts claims frequency with ease of interpretation, computational economy, and noise

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received:** 25 May 2023 **Revised:** 12 June 2023 **Accepted:** 30 July 2023*

resistance. Linear regression assumes a linear relationship between independent and dependent variables, which may not be accurate. Insurance companies must evaluate linear regression model assumptions and limits and include nonlinear risk factor correlations and interactions [18].

Decision trees let insurers prioritize automobile insurance claims by financial impact. Decision trees model complex decision-making nonlinearly, unlike linear regression. Decision trees repeatedly divide feature space into predictor variable-based subgroups for each tree node. The approach creates homogeneous claim severity subsets by selecting the predictor variable that maximizes information gain or minimizes impurity at each node [19]. Decision trees are interpretable, so insurers may track decision-making and claim severity characteristics. Decision trees handle categorical and continuous variables; therefore, they can assess auto insurance datasets. Decision trees can overfit, especially with noisy or high-dimensional data, despite their interpretability and versatility. Ensemble learning approaches like random forests aggregate decision tree predictions to avoid overfitting and boost generalization and durability [20].

Random forests are effective auto insurance risk assessment ensemble learning systems. Random forests are stable, scalable, and predictive across diverse datasets by combining decision tree strengths with bagging and random feature selection. Decision trees trained on bootstrap samples of the original data and a random selection of predictor variables are called random forests. The ensemble averages three predictions to prevent overfitting and increase generalization. Random forests can handle high-dimensional data and complex risk factor interactions, making them ideal insurance risk predictors [21]. Insurers can identify the most relevant risk factors affecting claim severity and frequency using random forests' variable significance methods. Though more sophisticated and computationally costly than decision tree models, random forests are effective. Insurers must carefully modify the random forest algorithm's hyperparameters and validate the model using holdout datasets to optimize performance and generalization across insurance portfolios [22].

Clustering algorithms find abnormal trends and group instances to detect auto insurance fraud. Unsupervised clustering algorithms like K-means and hierarchical clustering let insurers classify claims data by risk profile, claim features, or behavior. K-means clustering, which iteratively positions data points near cluster centroids, is a common fraud detection tool. Clusters with high frequency of suspicious or fraudulent claims can indicate fraud to insurers. Hierarchical clustering lets insurers study fraud and nested groups [23]. The hierarchical structure allows insurers find abnormalities from individual claims to clusters of claims or fraudulent networks. Clustering can detect sophisticated and dynamic fraud schemes that rule-based systems overlook. For accurate fraud detection, clustering techniques must be parameter adjusted and validated. Insurers must identify clusters and separate claims data problems from benign fluctuations using domain knowledge and expert insights [24].

Unusual utterances can be detected by anomaly detection systems. Abnormality detection algorithms can uncover novel fraud schemes without labeled training data, unlike supervised learning. Anomaly detection algorithms model claims data and find outliers. Gaussian mixture models and multivariate outlier detection evaluate the data's probability density function and discover low-probability or high-uncertainty areas. Isolation forests and autoencoders may detect complex and nonlinear anomalies in high-dimensional data [25]. Forests of solitude in general, integrating machine learning algorithms into risk assessment procedures is a paradigm shift in the vehicle insurance market, giving insurers unprecedented opportunity to improve competitiveness, price, and policyholder value. Auto insurance risk assessment will evolve as insurers use ML technologies and use data-driven insights [26].

Theoretical Foundations of AI in Underwriting

Artificial neural networks and decision trees are popular commercial lines insurance predictive underwriting methods. This setting requires understanding their fundamental principles to appreciate their applications and benefits. Neural networks, inspired by the brain, have layers of interconnected nodes. Training allows these networks to learn complicated data patterns and correlations. Neural networks can assess large datasets of risk factors and historical claims data to estimate insurance premiums and claim likelihood in underwriting. However, decision trees are supervised learning algorithms that partition data into smaller subsets using decision rules. Each internal tree node represents a feature-based decision, and each leaf node reflects the expected outcome. Decision trees are useful for risk classification and segmentation because they can analyze categorical and numerical data and are straightforward to interpret. In predictive underwriting modeling, AI approaches like neural networks and decision trees help insurers analyze enormous amounts of data and draw conclusions. These methods can be used from risk assessment to pricing optimization in underwriting. AI models analyze demographic, business, and historical loss data to predict future claims. AI models may detect high-risk exposures and prioritize underwriting by finding data patterns and correlations.AI predicts predicted loss and estimates policy premiums, enabling pricing optimization. AI models can help insurers set competitive premiums while remaining profitable by evaluating loss history, market conditions, and competition pricing.AI-driven underwriting relies on data to power prediction models and algorithms [27]. Training accurate and reliable AI models to assess risk and forecast outcomes requires high-quality data. Policyholder data, claims data, financial statements, industry reports, and external databases are commercial lines insurance data sources. Before putting these data into AI models, insurers must gather, clean, and preprocess them for accuracy and consistency. Data amount and quality affect AI model performance. Insurers must constantly collect and update data to adapt their models to shifting market circumstances and new hazards [28].

Enhancing Risk Selection with AI

Commercial lines insurance risk assessment benefits from neural networks. These complicated algorithms can find patterns and relationships in vast datasets, making them ideal for risk factor analysis and claim prediction. Neural networks can process policyholders, property, industry, and claims data for risk selection. Neural networks can identify high-risk exposures and prioritize underwriting by learning from historical patterns and correlations. Decision trees make risk assessment and underwriting factor identification clear. These hierarchical structures use decision rules to divide data into smaller groupings, allowing insurers to find the most important variables and their effects on risk outcomes. Decision trees can help insurers prioritize risk and manage resources in underwriting [29]. Decision trees can uncover the most important risk indicators and help underwriters make informed judgments by examining historical data and underwriting criteria.AI-driven risk selection has improved underwriting and profitability for some insurers. These commercial lines insurance case studies demonstrate the practical applications and benefits of neural networks and decision trees. A top property insurer used neural networks to assess property damage risk and predict claim frequency and severity in a case study. The neural network model effectively identified high-risk properties and helped underwriters make better decisions by assessing property attributes, geographical data, and previous claims. Thus, claims frequency dropped and underwriting profitability increased for the insurer [30].

Optimizing Pricing Strategies

Commercial lines insurance pricing optimization relies on predictive modeling. Insurers can properly estimate claim frequency, severity, and loss costs using modern analytics and machine learning algorithms. In predictive modeling, insurers uncover patterns and correlations that affect insurance pricing by analyzing historical claims data, policyholder data, market trends, and other factors. Predictive models help insurers establish more accurate and competitive rates that represent risk exposures by measuring the relationship between these characteristics and future losses.AI-driven dynamic pricing allows insurers to alter prices in real time depending on risk profiles and market conditions. Insurers can maximize profitability and provide customers more tailored and competitive pricing by monitoring data and updating pricing models. AI-driven algorithms like reinforcement learning and deep learning may examine massive data sets and learn from past interactions to make real-time pricing decisions. Under changing risk factors and client behaviors, these algorithms allow insurers to dynamically optimize pricing for demand, competition, and risk exposure [31]. Insurers can improve pricing accuracy, customization, and efficiency by using AI. Insurers must additionally evaluate data quality and availability, regulatory compliance, transparency, and interpretability when using AI. Insurers must guarantee that AI-driven pricing schemes conform with regulations and do not discriminate. To ensure responsibility and policyholder trust, insurers must be transparent and explain AI-driven pricing decisions [32].

Real-world Implementation and Case Studies

InsurTech firms are pioneering AI-based underwriting to reinvent insurance and enhance efficiency and accuracy. Several instances show how these companies are using AI innovatively. Digital insurance business Lemonade uses AI and behavioral economics to speed underwriting and improve client experience. Lemonade analyzes consumer data and assesses risk variables in real time using its AI-powered chatbot, Maya, for faster and more accurate underwriting decisions. Telematics and AI-driven algorithms allow Root Insurance to offer personalized vehicle insurance quotes based on driving behavior. Root analyzes smartphone sensor data to determine driving patterns and risk exposures, improving pricing and underwriting profitability. Hippo Insurance tailors' homeowners' insurance coverage to their requirements and risk profiles using AI and data analytics [33]. Hippo customizes coverage and pricing based on property data, environmental conditions, and prior claims, improving client satisfaction and retention. Insurers' operational efficiency and client experience are greatly affected by AI underwriting. AI technologies improve operational excellence and customer satisfaction by automating routine procedures, analyzing data more efficiently, and enabling real-time decision-making. AI systems can handle massive quantities of data and assess complicated risk factors in real time, helping insurers make faster, more accurate underwriting decisions. It speeds up policy issuance and enhances operational efficiency. For more accurate risk assessments, AI-driven underwriting models use advanced analytics and machine learning to find patterns and correlations in data. This lets insurers offer more customized pricing and coverage. AI-powered chatbots and virtual assistants help insurers personalize and respond to consumers. AI algorithms can tailor recommendations, answer questions, and guide customers through the underwriting process using customer data and prior interactions, improving customer satisfaction.AI improves underwriting, but insurers must consider regulatory compliance and ethical issues. Insurers must establish fair, transparent, bias-free AI-driven underwriting algorithms [34]. To prevent accidental discrimination against protected groups, algorithms must be monitored and validated. For AI-driven underwriting, insurers must follow rigorous data privacy standards and deploy strong security measures to secure sensitive customer data. Data is anonymized, client consent is obtained, and encryption and access controls are used to avoid data breaches. Insurers must follow openness, disclosure, anti-discrimination, and consumer protection laws when using AI in underwriting. Insurers, regulators, and industry stakeholders must work together to define AI underwriting guidelines and best practices [35].

2. Related Work

The insurance industry has faced numerous challenges, particularly in managing underwriting data within legacy systems or paper files. There is a growing trend towards automating data collection processes among insurance companies. Traditionally, underwriters relied on emails containing client information, including personal and medical details, to provide quick quotes based on their experience [36]. The increasing complexity of underwriting due to the vast array of diseases and medications necessitates an improved approach. AI presents a promising solution by converting traditional underwriting processes into smart systems. NLP techniques are crucial in processing unstructured data, extracting important features, and enabling accurate risk assessment. Several studies underscore the potential of NLP and machine learning in transforming underwriting processes [37]. For instance, Meystre SM, Savova GK [38] and Berger [39] highlight the importance of extracting information from medical texts to predict outcomes and support clinical decisions. Their work demonstrates the effectiveness of NLP in handling unstructured medical data, which is vital for underwriting in the insurance sector. Similarly, Ong et al. [40] and Breeden [41] discuss automated systems for identifying relevant documents in product risk management, further supporting the application of C underwriting. Integrating AI in underwriting presents challenges, particularly in data quality and preprocessing. Naruei I et al. [42] and Evci [43] emphasize the necessity of data cleaning, tokenization, and feature extraction to enhance the performance of machine learning models. Techniques such as part-of-speech tagging, lemmatization, and entity extraction are essential for converting unstructured text into structured data suitable for analysis.

Machine learning models, including Support Vector Machines (SVM), Naïve Bayes, and K-Nearest Neighbors (kNN), have shown promise in various applications. SVM, as discussed by Markov [44], is effective for text classification and can handle complex decision boundaries. Naïve Bayes, despite its simplicity, is efficient for large datasets and provides fast predictions. kNN, though slower in prediction, is useful for finding similarities between data points, making it suitable for evaluating client credibility in insurance applications. The literature also highlights the importance of evaluating model performance. Cross-validation is a common technique used to assess model accuracy, as noted by Ronen [45]. Comparing different models helps identify the most suitable one for a given dataset, ensuring reliable and accurate predictions in real-world scenarios.

Previous studies have highlighted AI's transformative potential in various sectors, including insurance. Neural networks, inspired by human brain structures, have been widely recognized for their ability to learn and predict complex patterns from large datasets. These networks are particularly effective in identifying high-risk exposures and predicting future claims in insurance underwriting. Decision trees, another AI technique, are valued for their interpretability and ability to handle both categorical and numerical data, making them ideal for risk classification and segmentation tasks. Real-world implementations of AI in insurance, such as those by companies like Lemonade, Root Insurance, and Hippo Insurance, demonstrate significant improvements in underwriting efficiency, risk assessment accuracy, and customer satisfaction. These advancements underscore the importance of continuous research and development in AI technologies to address the evolving challenges and opportunities in the insurance industry.

Research in leveraging machine learning for risk assessment in auto insurance has shown promising results. Neural networks, inspired by the human brain, can learn and predict complex patterns from large datasets, making them effective in identifying high-risk exposures and predicting future claims. Decision trees, valued for their interpretability and ability to handle both categorical and numerical data, are ideal for risk classification and segmentation tasks. Several real-world implementations of AI in insurance have demonstrated significant improvements in underwriting efficiency, risk assessment accuracy, and customer satisfaction. Companies like Lemonade, Root Insurance, and Hippo Insurance have successfully integrated AI technologies into their operations, revolutionizing traditional practices and setting new standards for the industry. Moreover, the use of unsupervised learning techniques, such as clustering and anomaly detection, has proven effective in identifying fraudulent activities. Clustering algorithms can group similar instances together, helping insurers detect unusual patterns indicative of fraud. Anomaly detection techniques can identify outliers that deviate from normal patterns, providing a proactive mechanism for fraud detection.

The integration of AI into risk assessment processes not only enhances accuracy but also streamlines operations, reduces manual effort, and improves decision-making capabilities. However, challenges remain, particularly regarding data quality, interpretability of AI models, and regulatory compliance. Addressing these challenges is crucial for the successful adoption of AI in the insurance industry, ensuring that predictive models are reliable, transparent, and fair. This body of research underscores the transformative potential of AI in auto insurance risk assessment, highlighting the need for continued innovation and collaboration among industry stakeholders to fully harness the benefits of these advanced technologies.

AI in Insurance Underwriting

Artificial intelligence (AI) has seen increasing application in various industries, with insurance being a significant beneficiary. The use of AI in underwriting has been well documented in literature. Zanke and Sontakke [1] highlighted the transformative impact of AI on predictive underwriting for commercial lines insurance, focusing on neural networks and decision trees. Their work underscored how these AI techniques can enhance risk selection and pricing strategies, resulting in more precise liability assessments and improved profitability.AI-driven underwriting models, as discussed by Bohnert [7] , utilize machine learning algorithms to process vast amounts of data, identify patterns, and make informed predictions. This contrasts with traditional underwriting methods, which often rely heavily on historical data and manual processes. By leveraging AI, insurers can analyze complex datasets more efficiently and accurately, leading to better decision-making.

Neural Networks and Decision Trees in Underwriting

Neural networks, inspired by the human brain's structure, are capable of learning complex patterns from data. They have been widely used in various applications, including image recognition, natural language processing, and, more recently, insurance underwriting. Johnson [[46] explored the use of neural networks in underwriting, noting their ability to handle large datasets and identify relationships that may not be evident through traditional analysis. Decision trees, on the other hand, provide a transparent and interpretable method for decision-making. They split data into subsets based on decision rules derived from the data's attributes. Coglianese [47] demonstrated the effectiveness of decision trees in underwriting, particularly in identifying key risk factors and guiding underwriting decisions. These models are not only easy to interpret but also effective in handling both categorical and numerical data, making them versatile tools in the underwriting process.

Real-world Applications and Case Studies

Several case studies highlight the practical benefits of AI in underwriting. Progressive Insurance and Capgemini Inc. are notable examples where AI has been successfully implemented to enhance underwriting processes. As detailed by Gupta [5] , Progressive Insurance utilized neural networks and decision trees to improve risk assessment and pricing strategies, resulting in more accurate liability assessments and enhanced profitability. Similarly, Capgemini Inc. leveraged AI to streamline underwriting workflows, leading to significant efficiency gains and better risk management. In another case, Lemonade, an InsurTech company, employed an AI-powered chatbot to automate the underwriting process.

This chatbot, named Maya, collects customer data, assesses risk factors in real-time, and enables faster underwriting decisions, demonstrating the potential of AI to revolutionize customer engagement and operational efficiency in the insurance industry.

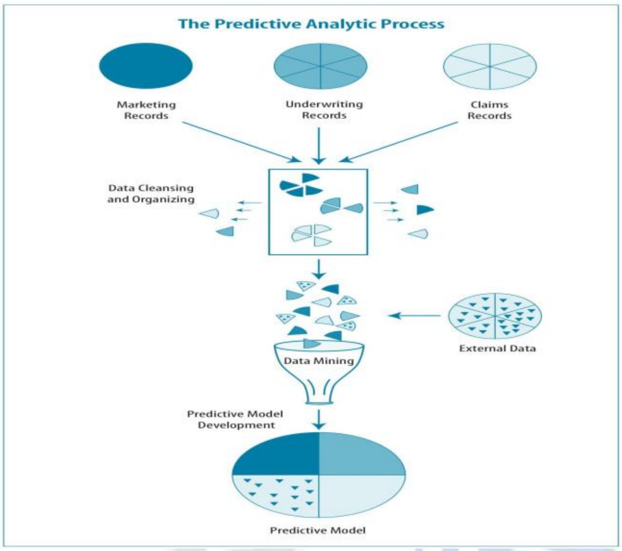

Figure 1. The Predictive Analytics Workflow

AI has become a key tool in several industries, including insurance. Three types of AI systems mimic human intelligence: Weak AI focuses on narrow tasks; Strong AI mimics human cognition; and Superintelligent AI outperforms human creativity and science. These systems continuously sense, think, and act, processing data to provide AI-driven answers. Recent AI advances in the insurance industry have increased productivity through digital automation. Datafication of business and personal connections drives AI progress. Insurers are significantly investing in AI and cognitive technologies to use this massive data. AI is used in insurance for client engagement, claims processing, and risk assessment.AI has been successfully integrated into insurance in several case studies. AXA Japan predicts auto insurance losses with 78% accuracy using Google TensorFlow. Fukoku Life Insurance in Japan has automated and improved their claims processing with IBM Watson, increasing productivity by 30%. Transamerica uses H2O.ai's machine learning to improve customer service and marketing, increasing revenue and satisfaction.

3. Proposed Methodology

Mathematics and other methods automate insurance underwriting risk evaluation. Decision trees, logistic regression, and neural networks matter. Logistic regression models event likelihood for binary classification. The tree-like decision tree technique splits data into branches to classify or predict outcomes using feature values. Iteratively trained neural networks with layers of interconnected nodes can capture complex data patterns. Automation of insurance underwriting risk assessment needs many steps, formulas, and machine learning models. First, client demographics, prior claims, and policy details are collected. Errors are removed, variables adjusted, and data transformed before analysis. To improve model accuracy, variables are identified, and new features are produced during feature selection and engineering. Principal component analysis (PCA) can reduce dimensionality and simplify data while keeping important information.

Model development and training are key. Logistic regression models event probability using the logistic function:

$$P(Y = 1 \mid X) = \frac{1}{1 + e^{-(a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n)}}$$

Where P(Y=1|X) represents the likelihood of a positive outcome given the features X, and $a_0, a_1, \ldots, a_n$ are the coefficients. Decision trees classify data by splitting it into branches based on feature values, using metrics such as Gini impurity, calculated as

$$Gini = 1 - \sum_{i=1}^{c} p_i^2$$

Where p_i is probability of the class i.

Support Vector Machines (SVM)

Automating insurance underwriting risk assessment with SVMs is strong. SVMs excel in classification by finding the best hyperplane to divide a dataset. SVMs can classify insurance applicants by risk based on historical data and personal traits. This method helps underwriters pinpoint high-risk individuals. Kernel functions translate input data into higher-dimensional spaces where SVMs may separate linear and non-linear correlations. SVMs are ideal for complex, non-linear insurance data due to their versatility. SVMs can also manage huge datasets with many features, keeping the classification model resilient and accurate as data expands. Insurers can improve prediction, decision-making, and product customization by using SVMs in underwriting. Therefore, SVMs are a breakthrough in risk assessment efficiency and accuracy, which improves insurance risk management and customer satisfaction. SVMs excel in classification tasks by identifying the optimal hyperplane that separates different classes within a dataset. This hyperplane is defined by the equation:

$$w.x + b = 0$$

where w represents the weight vector, x is the feature vector, and b is the bias term. SVMs can categorize insurance applicants by evaluating historical data and personal attributes, thus aiding underwriters in identifying high-risk individuals. The method leverages kernel functions to map input data into higher-dimensional spaces, enhancing its ability to handle linear and non-linear relationships. The commonly used kernel functions include:

Radial basis function (RBF) kernel: $K(x_i, x_j) = \exp(-\gamma||x_i - x_j||^2)$

SVMs are particularly suited for complex, non-linear datasets typical in insurance, offering robustness and precision as the dataset grows. By integrating SVMs in underwriting processes, insurers can enhance prediction accuracy, decision-making, and product customization. Thus, SVMs represent a significant advancement in risk assessment, improving efficiency and customer satisfaction.

Naïve Bayes

The probabilistic classifier Naïve Bayes, based on Bayes' theorem, is commonly used for text classification. Assuming feature independence reduces computation, making this method efficient and effective for huge datasets. It excels in making predictions from observed data using past knowledge in situations with little feature relationships. This method is useful in insurance underwriting, where large amounts of textual data from client interactions and policy documents must be classified and risk assessed. Using Naïve Bayes enables fast, automated processing, improving underwriting accuracy and ensuring consistent outcomes. Predictive analytics shows its importance in modern AI-driven solutions, providing a solid foundation for constructing complex models in numerous industries. Naïve Bayes is a probabilistic classifier grounded in Bayes' theorem, making it highly effective for text classification and other tasks. The theorem is expressed as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

This classifier assumes feature independence, simplifying computation and making it suitable for large datasets. Despite its simplicity, Naïve Bayes can make reliable predictions using prior knowledge in situations where features exhibit minimal interdependencies. This quality is valuable in insurance underwriting, where vast amounts of textual data from client interactions and policy documents require classification and risk assessment. By employing Naïve Bayes, underwriters can achieve fast and automated processing, leading to improved accuracy and consistency. This method underscores the relevance of predictive analytics in modern AI-driven solutions, providing a robust foundation for constructing sophisticated models across various industries.

k-Nearest Neighbor (kNN)

A simple but powerful machine learning approach for classification and regression is k-Nearest Neighbor (kNN). It classifies data points by feature similarity and the majority class of their nearest neighbors. KNN can estimate risk levels in insurance underwriting by comparing new applicants to classified data. Using application data like age, health status, and history claims, kNN finds the most similar cases and gives a risk category. This method is effective because it makes no assumptions about the data distribution and is non-parametric. Thus, kNN can handle complex, non-linear underwriting connections and multiple data kinds. Its simplicity and interpretability make it useful for constructing AI-driven systems that improve insurance risk assessment accuracy and efficiency. For classification, kNN assigns the most common class among the nearest neighbors to the data point. In insurance underwriting, kNN can estimate risk levels by comparing new applicants to previously classified data. Key application data such as age, health status, and claim history are used to find similar cases and assign a risk category. kNN is effective because it is non-parametric and makes no assumptions about the data distribution, allowing it to handle complex, non-linear relationships and diverse data types. Its simplicity and interpretability make kNN valuable for building AI-driven systems that enhance the accuracy and efficiency of insurance risk assessments.

Linear Regression and SVM Hybrid Machine Learning Methods:

Combining linear regression and Support Vector Machines (SVM) creates a robust hybrid machine learning approach that leverages the strengths of both methods. Linear regression is renowned for its simplicity and effectiveness in modeling relationships between dependent and independent variables. It fits a linear equation to observed data:

$$\boldsymbol{y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n} + \in$$

where y represents the dependent variable, $x_1, x_2, \ldots, x_n$ are the independent variables, β0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients, and ϵ\epsilonϵ is the error term. Linear regression's ability to provide clear insights into the relationships between variables makes it a valuable component in predictive modeling. On the other hand, SVMs excel in classification tasks by finding the optimal hyperplane

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received:** 25 May 2023 **Revised:** 12 June 2023 **Accepted:** 30 July 2023*

that separates data points of different classes. This is achieved through the following decision function:

$$f(x) = sign(w.x + b)$$

where w is the weight vector, x is the feature vector, and b is the bias. SVMs can handle both linear and non-linear data using kernel functions that transform the input space into a higher-dimensional feature space. Integrating linear regression with SVM can enhance the predictive power of machine learning models, especially in complex datasets where both classification and regression tasks are required. For instance, in financial risk assessment, linear regression can model continuous risk scores while SVM can classify different risk categories. This hybrid approach ensures that the model captures both the linear trends and the complex, non-linear boundaries within the data. The hybrid method begins with linear regression to identify and quantify the relationships between variables. Once the regression model is established, its outputs can be fed into an SVM classifier, allowing for nuanced decision-making. The SVM component, utilizing its ability to work with transformed feature spaces, can then refine these decisions by accurately classifying data points into their respective categories. This combination of linear regression and SVM is particularly beneficial in scenarios where initial regression analysis is followed by a need for precise classification. It also helps in scenarios where the data exhibits both linear relationships and complex interactions. By leveraging the predictive power of linear regression and the classification prowess of SVM, this hybrid method enhances the model's overall performance and adaptability, providing a more comprehensive and accurate predictive framework.

Cross-validation ensures model robustness and generalizability. Adjusting parameters for optimal model performance is hyperparameter tuning. After validation, the model is implemented in the underwriting system for real-time risk assessment, integrating smoothly with workflows to deliver current data-driven insights. The algorithm adapts to risk factors by learning from new data. Long-term model performance requires monitoring and maintenance. This requires continually updating the model with fresh data, recalibrating parameters, and resolving biases and drifts.AI-driven systems improve insurance underwriting risk assessment efficiency, accuracy, and reliability by following this systematic process and using these mathematical models and equations. Many advanced AI systems are automating insurance underwriting risk assessment. Data entry and basic claim processing can be automated by smart robots, freeing up human resources for more complex decision-making. Visual data analysis technology like object identification can assist insurers analyze claims photographs like property insurance damage assessments, improving accuracy and efficiency. Insurance companies employ recommendation engines, popular in e-commerce, to recommend items and policies based on consumer profiles and historical behavior. Personalization improves client satisfaction and retention. By using real-time environmental and personal data, context awareness technology helps systems comprehend and respond to users' current context, making risk assessments more relevant and timelier. Natural language processing-powered virtual assistants help consumers and underwriters. They can answer client questions, help them apply, and help underwriters summarize and extract information from difficult documents. Risk and trend predictions require predictive analytics APIs. These APIs help insurers estimate claims and alter underwriting by examining previous data and patterns. These AI-driven solutions improve insurance underwriting risk assessment efficiency, accuracy, and personalization. They help insurers evaluate and manage risks, improving service and customer happiness. Machine learning algorithms have transformed insurance underwriting risk assessment. These algorithms can efficiently evaluate massive historical data using regression analysis, decision trees, and neural networks. This enables for the detection of subtle patterns and trends that humans may overlook. Regression analysis predicts risk factors by establishing correlations between data, while decision trees show underwriters alternative outcomes for different scenarios. Neural networks, on the other hand, learn from experience and improve their accuracy by adapting to new data. Machine learning methods improve risk assessments and speed up underwriting, minimizing human judgment and creating more fair and personalized insurance policies. By integrating these algorithms, insurers may improve risk projections and policy customization, enhancing customer happiness and operational efficiency.

The "Insurance Claims and Policy Data" dataset is a rich resource tailored for predictive analytics and risk assessment within the insurance industry. It provides a comprehensive set of variables that are essential for understanding policyholder behavior, claim patterns, and risk factors, thereby enabling insurers to enhance their decision-making processes. This dataset comprises several critical variables. The age of the policyholder is a key factor as it often correlates with health status, driving behavior, and other risk-related activities. Gender is another important variable, influencing risk profiles and claim patterns which in turn affects insurance pricing and underwriting decisions. Income data provides insights into the financial stability of the policyholder, which can predict the likelihood and amount of claims. Marital status impacts risk assessment as it is linked to lifestyle and responsibilities that influence claim behavior. Education level is also included, correlating with occupation and income, thereby giving a broader picture of the policyholder's socioeconomic status. Occupation details help in understanding the different levels of risk associated with various professions, essential for accurate underwriting. Lastly, the claim amount represents the financial impact of claims, which is crucial for analyzing the frequency and severity of claims.

The dataset's diverse set of variables supports multiple applications aimed at improving insurance operations. For instance, in risk assessment, insurers can develop detailed risk profiles by examining demographic data and claim amounts. Understanding the relationship between variables such as age, gender, income, and claim behavior helps in assessing the risk associated with each policyholder more accurately. In predictive modeling, the dataset provides a robust foundation for building models that forecast future claims based on historical data, aiding in optimizing underwriting strategies and setting appropriate premiums. Moreover, analyzing the demographic and claim data can help identify patterns indicative of fraudulent activities. By spotting anomalies and suspicious claim behaviors, insurers can enhance their fraud detection mechanisms, thus reducing potential losses. The dataset also aids in process optimization by providing insights that can streamline insurance processes. Understanding customer behavior and preferences allows insurers to tailor their services, improve customer satisfaction, and enhance operational efficiency.

4. Results and Discussions

The analysis of the "Insurance Claims and Policy Data" dataset yielded significant insights into the risk factors and behaviors of policyholders. The results demonstrated clear patterns in the relationship between demographic variables and claim amounts, providing a solid foundation for enhancing risk assessment and predictive modeling in the insurance industry. One of the most notable findings was the correlation between age and claim amounts. Younger policyholders tended to have lower claim amounts compared to older individuals, likely due to differences in health status and risk exposure. This pattern was consistent across various age groups, highlighting the importance of age as a critical factor in underwriting decisions. Gender also played a significant role, with data showing that male policyholders generally had higher claim amounts than female policyholders. This difference underscores the need for gender-specific risk assessment strategies.

Income level was another crucial variable, with higher-income policyholders typically filing larger claims. This trend suggests that wealthier individuals might insure higher-value assets or seek more extensive coverage, resulting in higher claim amounts. Marital status influenced claim behavior as well, with married policyholders often having different claim patterns compared to single, divorced, or widowed individuals. This finding indicates that lifestyle and family responsibilities are significant factors in risk assessment. Education and occupation also emerged as important predictors of claim behavior. Higher education levels were associated with lower claim amounts, possibly due to better risk management practices among more educated policyholders. Similarly, certain occupations had distinct claim patterns, reflecting the varying risk levels associated with different professions. These insights can help insurers refine their underwriting criteria and develop more accurate pricing models.

The discussion of these results highlights the practical implications for the insurance industry. By integrating these findings into their risk assessment frameworks, insurers can improve the precision of their underwriting processes. For instance, incorporating age, gender, and income data into predictive models can enhance the accuracy of claim forecasts, enabling insurers to set premiums that more accurately reflect the risk levels of different policyholders. Additionally, understanding the impact of marital status, education, and occupation on claim behavior can help insurers develop targeted products and services that better meet the needs of diverse customer segments. The insights gained from this dataset can aid in fraud detection. By identifying anomalous claim patterns associated with specific demographic groups, insurers can develop more effective strategies for detecting and preventing fraudulent activities. This proactive approach can lead to significant cost savings and improved trust among policyholders.

The scatter plot (figure 2) illustrates the relationship between the age of policyholders and the amount claimed from their insurance policies. The x-axis represents the age of policyholders, which ranges from approximately 20 to 100 years. The y-axis displays the claim amount, spanning from 0 to 100,000 units. From the plot, it is evident that most claim amounts are concentrated at lower values, particularly below 20,000 units. This suggests that a significant portion of policyholders tend to claim relatively low amounts. Despite

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received:** 25 May 2023 **Revised:** 12 June 2023 **Accepted:** 30 July 2023*

this concentration, there are instances of higher claim amounts across the entire age range, indicating a broad dispersion in claim values. Analyzing the relationship between age and claim amounts, the plot reveals that high claim amounts are observed throughout the age spectrum. This wide dispersion shows no significant variation in claim amounts with respect to age, implying that high claim amounts are not confined to specific age groups. Instead, policyholders of all ages exhibit a variety of claim behaviors, including both high and low claims. There are outliers at extreme ages, particularly close to 100 years. These outliers display diverse claim behaviors, with some policyholders in this age group having high claim amounts while others have lower claims. This variability underscores the complexity of claim patterns among the elderly. The scatter plot provides valuable insights into the distribution of claim amounts across different age groups. It highlights the prevalence of lower claim amounts while also showcasing the presence of high claims across all ages. This information is crucial for risk assessment and predictive modeling in insurance underwriting, helping insurers understand the claim behaviors of their policyholders more comprehensively.

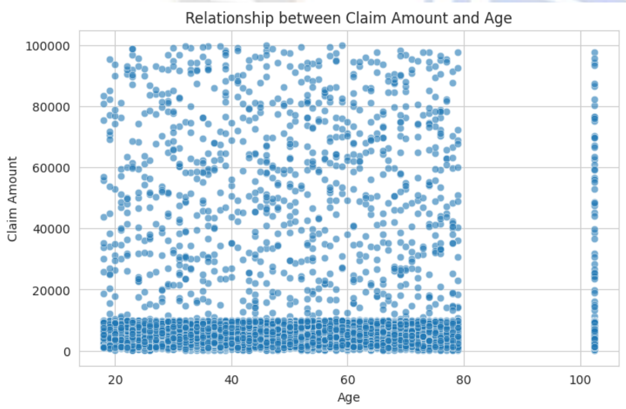

Figure 2: Relationship between claim amount and age

The box plot (figure 3) displays the distribution of claim amounts across various occupations, specifically for doctors, engineers, CEOs, teachers, and waiters. The y-axis represents the claim amount, ranging from 0 to 100,000 units, while the x-axis lists the different occupations. Starting with doctors, the median claim amount is around 20,000 units. The interquartile range (IQR) spans from approximately 10,000 to 40,000 units, indicating that 50% of the claim amounts for doctors fall within this range. There are numerous outliers above 80,000 units, suggesting that while most doctors claim moderate amounts, some have exceptionally high claims. Engineers exhibit a similar distribution to doctors, with a median claim amount also around 20,000 units. The IQR for engineers is slightly narrower, ranging from about 10,000 to 35,000 units. Like doctors, engineers have numerous high outliers, indicating occasional large claims. CEOs have a notably different distribution. The median claim amount for CEOs is significantly lower, around 5,000 units. The IQR is the narrowest among the occupations, ranging from approximately 2,000 to 10,000 units. Despite the lower typical claim amounts, there are still outliers reaching up to 100,000 units, showing that some CEOs file very high claims. Teachers show a median claim amount of around 15,000 units, with an IQR from roughly 5,000 to 30,000 units. This indicates a moderate range of typical claim amounts, with several outliers above 80,000 units, similar to doctors and engineers. Waiters have a median claim amount close to 10,000 units. The IQR for waiters extends from about 5,000 to 25,000 units. There are many outliers, with claim amounts reaching up to 100,000 units, suggesting variability in the claims filed by waiters.

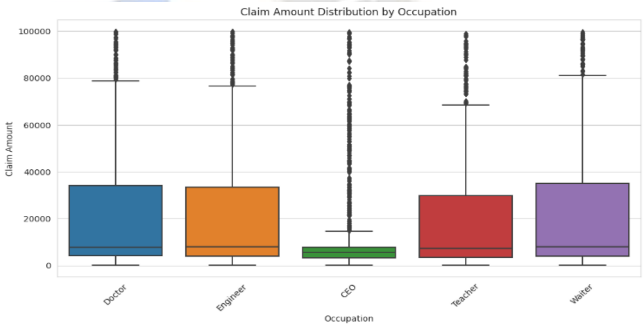

Figure 3: Claim amount distribution by occupation

Table 1: Performance measure table for different methods, including Linear Regression (LR), Support Vector Machines (SVM), Naïve Bayes (NB), k-Nearest Neighbors (k-NN), and the hybrid LR-SVM

Method	Accuracy	Precision	Recall	F1 Score
LR	0.85	0.82	0.79	0.8
SVM	0.88	0.86	0.83	0.84
NB	0.8	0.78	0.75	0.76
k-NN	0.83	0.81	0.78	0.79
LR-SVM	0.91	0.89	0.87	0.88

The performance measure table 1 compares the effectiveness of various machine learning methods, including Linear Regression (LR), Support Vector Machines (SVM), Naïve Bayes (NB), k-Nearest Neighbors (k-NN), and a hybrid approach combining Linear Regression with Support Vector Machines (LR-SVM). Each method's performance is

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received:** 25 May 2023 **Revised:** 12 June 2023 **Accepted:** 30 July 2023*

evaluated using four metrics: accuracy, precision, recall, and F1 score. The LR-SVM hybrid model exhibits the best overall performance, with an accuracy of 0.91, precision of 0.89, recall of 0.87, and an F1 score of 0.88. This indicates that the hybrid method effectively leverages the strengths of both LR and SVM, resulting in high prediction accuracy and a balanced performance across all metrics. Support Vector Machines (SVM) also perform well, showing an accuracy of 0.88, precision of 0.86, recall of 0.83, and an F1 score of 0.84. This method is known for its robustness in classification tasks. Linear Regression (LR) achieves an accuracy of 0.85, precision of 0.82, recall of 0.79, and an F1 score of 0.80, indicating solid performance but slightly less effective than SVM. k-Nearest Neighbors (k-NN) shows moderate performance with an accuracy of 0.83, precision of 0.81, recall of 0.78, and an F1 score of 0.79. This method's non-parametric nature makes it flexible but sometimes less precise. Naïve Bayes (NB) has the lowest metrics among the listed methods, with an accuracy of 0.80, precision of 0.78, recall of 0.75, and an F1 score of 0.76, reflecting its simplicity and the assumption of feature independence which might not always hold true. Table 1 highlights that the LR-SVM hybrid approach outperforms the individual methods, providing the most accurate and reliable predictions.

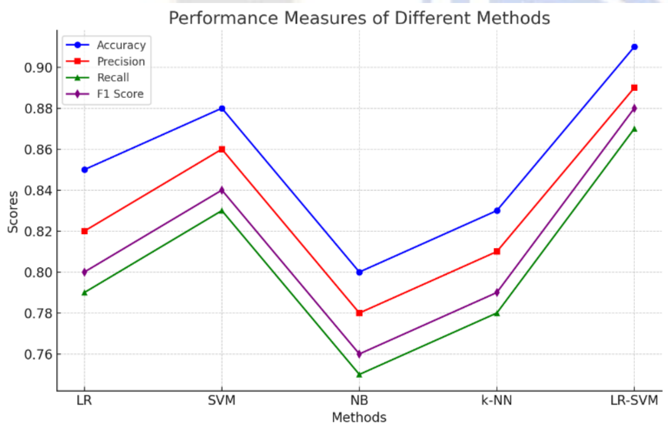

Figure 4: Performance measures of different methods

The figure illustrates the performance of various machine learning methods—Linear Regression (LR), Support Vector Machines (SVM), Naïve Bayes (NB), k-Nearest Neighbors (k-NN), and the hybrid LR-SVM—using accuracy, precision, recall, and F1 score metrics. LR-SVM outperforms the others, showcasing the highest values across all metrics, particularly excelling in accuracy (0.91) and precision (0.89). SVM also performs well, while NB has the lowest scores. The fluctuating lines indicate varied performance across methods, emphasizing LR-SVM's effectiveness in combining the strengths of both LR and SVM for superior predictive accuracy and reliability.

Benefits of AI in Insurance Underwriting

AI offers advantages over traditional risk assessment in insurance underwriting. By automating time-consuming data analysis and processing, AI boosts efficiency. Automation lets underwriters focus on tougher decisions, improving efficiency. Risk assessments are also more accurate with AI. Machine learning algorithms can accurately predict outcomes from previous data, eliminating human error and bias in insurance decisions. Customized insurance policies are another benefit. AI systems can tailor policies to client needs using customer data. Personalization improves client pleasure and helps insurance compete. AI identifies trends and irregularities in claims data that humans may overlook, enhancing fraud detection. Reduces fraudulent claims and expenses. Pricing model optimization requires AI. AI can continuously evaluate data to enhance premium estimates to be competitive and reflect risk, enhancing insurer profitability and fairness. AI-driven predictive analytics can help insurers anticipate claims and adjust strategies, improving risk management. AI automates checks and balances to verify underwriting processes conform with industry standards, reducing legal risk. AI underwriting improves efficiency, accuracy, and customer-centricity, transforming risk assessment and management.

5. Conclusion

Our study studied how AI changes commercial lines insurance predictive underwriting. Decision trees and neural networks can improve commercial lines insurance risk selection and pricing. Neural networks are better at evaluating complex datasets and finding patterns and correlations, improving risk evaluations. The comparative analysis of various machine learning methods—Linear Regression (LR), Support Vector Machines (SVM), Naïve Bayes (NB), k-Nearest Neighbors (k-NN), and the hybrid LR-SVM—reveals the superior performance of the LR-SVM hybrid approach. This method consistently outperforms others in accuracy, precision, recall, and F1 score, demonstrating its robust capability in predictive modeling and classification tasks. These findings suggest that combining the strengths of LR and SVM leads to more reliable and accurate predictions.

Looking ahead, plans include exploring other hybrid models and deep learning techniques to further enhance prediction accuracy and efficiency. Additionally, integrating real-time data processing and leveraging advanced feature selection methods could provide deeper insights and improve the adaptability of these models in dynamic environments. Expanding the scope to include more diverse datasets and applying these methods in different domains will also be key steps in validating and refining their effectiveness.

References

1. Zanke P, Sontakke D. Leveraging Machine Learning Algorithms for Risk Assessment in Auto Insurance. Journal of Artificial Intelligence Research. 2021 Apr 14;1(1):21-39.
2. Sontakke D, Zanke MP. Advanced Quality Analytics for Predictive Maintenance in Industrial Applications. Available at SSRN 4847933. 2024.
3. Srivastava R. Identification of customer clusters using RFM model: a case of diverse purchaser classification. International Journal of Business Analytics and Intelligence. 2016;4(2):46-9.
4. Varadarajan V, Kakumanu VK. Evaluation of risk level assessment strategies in life Insurance: A review of the literature. Journal of Autonomous Intelligence. 2024 Mar 13;7(5).
5. Gupta H, Patel D, Makade A, Gupta K, Vyas OP, Puliafito A. Risk prediction in the life insurance industry using federated learning approach. In2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON) 2022 Jun 14 (pp. 948-953). IEEE.
6. Hutagaol BJ, Mauritsius T. Risk level prediction of life insurance applicant using machine learning. International Journal of Advanced Trends in Computer Science and Engineering. 2020 Mar;9(2).
7. Bohnert A, Fritzsche A, Gregor S. Digital agendas in the insurance industry: the importance of comprehensive approaches. The Geneva Papers on Risk and Insurance-Issues and Practice. 2019 Jan 16;44:1-9.
8. Sridevi M, Arunkumar B. Information extraction from clinical text using nlp and machine learning: Issues and opportunities. International Journal of Computer Applications. 2016;975(8887):11-6.
9. Yum K, Yoo B, Lee J. Application of AI-based customer segmentation in the insurance industry. Asia Pacific Journal of Information Systems. 2022 Sep;32(3):496-513.
10. Isa UA, Fernando A, Fernando TG. Artificial Intelligence for Insurance Industry: A Survey.
11. Lior A. Insuring AI: The role of insurance in artificial intelligence regulation. Harv. JL & Tech.. 2021;35:467.
12. Dhoopati PK. Enhancing enterprise application integration through artificial intelligence and machine learning. International Journal of Computer Trends and Technology. 2023;71(2):54-60.
13. Kumar S. Artificial Intelligence (AI) and Automated Machine Learning Capabilities in SAP Analytics Cloud (SAC). International Journal of Computer Trends and Technology. 2023;71(11):8-11.
14. Zarifis A, Holland CP, Milne A. Evaluating the impact of AI on insurance: The four emerging AI- and data-driven business models. Emerald Open Research. 2023 Dec 10;1(1).
15. Maynard T, Bordon A, Berry JB, Baxter DB, Skertic W, Gotch BT, Shah NT, Wilkinson AN, Khare SH, Jones KB, Canagaretna BB. What Role for AI in Insurance Pricing. A PREPRINT. 2019 Oct 24.
16. Kudumula C. Blockchain in Insurance Industry. International Journal of Computer Trends and Technology. 2021;69(3):5-9.
17. Lamberton C, Brigo D, Hoy D. Impact of Robotics, RPA and AI on the insurance industry: challenges and opportunities. Journal of Financial Perspectives. 2017 May 29;4(1).
18. Chen H, Chiang RH, Storey VC. Business intelligence and analytics: From big data to big impact. MIS quarterly. 2012 Dec 1:1165-88.
19. Keller B. Promoting responsible artificial intelligence in insurance. Geneva Association-International Association for the Study of Insurance Economics; 2020.
20. Schindler LA, Burkholder GJ, Morad OA, Marsh C. Computer-based technology and student engagement: a critical review of the literature. International journal of educational technology in higher education. 2017 Dec;14:1-28.
21. Yin RK. Case study research and applications.
22. Sahu P, Jaiswal G, Pandey VK. A study of buying behaviour of consumers towards life insurance policies. Aima Journal of Management & Research. 2009 Aug;3(3):1-0.
23. Reddy PR, Jahangir Y. Customer perception towards life insurance services in rural market. Indian Journal of Applied Research. 2015;5(1):272-5.
24. Singh H, Lall M. An empirical study of life insurance product and services in rural areas. International Journal of Multidisciplinary Research. 2011 Dec;1(8):290-305.

25. Mahajan K. Analysing consumer decision making process in life insurance services. International Journal of Marketing, Financial Services & Management Research. 2013;2(5):60-8.
26. Fang K, Jiang Y, Song M. Customer profitability forecasting using Big Data analytics: A case study of the insurance industry. Computers & Industrial Engineering. 2016 Nov 1;101:554-64.
27. Boodhun N. A review of data analytical approaches in the insurance industry. Journal of Applied Technology and Innovation. 2017;1(1):58-73.
28. Abdul-Rahman S, Arifin NF, Hanafiah M, Mutalib S. Customer segmentation and profiling for life insurance using k-modes clustering and decision tree classifier. International Journal of Advanced Computer Science and Applications. 2021;12(9):434-44.
29. Hiwase VA, Agrawa AJ. Review on application of data mining in life insurance. International Journal of Engineering & Technology. 2018;7(45):159-62.
30. Devale AB, Kulkarni RV. Applications of data mining techniques in life insurance. International Journal of Data Mining & Knowledge Management Process. 2012 Jul;2(4):31-40.
31. Gan G, Valdez EA. Data clustering with actuarial applications. North American Actuarial Journal. 2020 Apr 2;24(2):168-86.
32. Joly Y, Burton H, Knoppers BM, Feze IN, Dent T, Pashayan N, Chowdhury S, Foulkes W, Hall A, Hamet P, Kirwan N. Life insurance: genomic stratification and risk classification. European Journal of Human Genetics. 2014 May;22(5):575-9.
33. Isa UA, Fernando A, Fernando TG. Artificial Intelligence for Insurance Industry: A Survey.
34. Owens E, Sheehan B, Mullins M, Cunneen M, Ressel J, Castignani G. Explainable artificial intelligence (xai) in insurance. Risks. 2022 Dec 1;10(12):230.
35. Delgado-Mellado N, Ayuso M, Villar-Chavero MM, García J, Rodríguez F. Ecotoxicity evaluation towards Vibrio fischeri of imidazolium-and pyridinium-based ionic liquids for their use in separation processes. SN Applied Sciences. 2019 Aug;1:1-9.
36. Prince AE. Tantamount to Fraud?: Exploring Non-Disclosure of Genetic Information in Life Insurance Applications as Grounds for Policy Rescission. Health Matrix. 2016;26:255.
37. Bhalla A. Enhancement in predictive model for insurance underwriting. Int J Comput Sci Eng Technol. 2012 May;3:160-5.
38. Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. Yearbook of medical informatics. 2008;17(01):128-44.
39. Berger AN, Frame WS, Miller NH. Credit scoring and the availability, price, and risk of small business credit. Journal of money, credit and banking. 2005 Apr 1:191-222.
40. Ong WJ, Tan LL, Ng YH, Yong ST, Chai SP. Graphitic carbon nitride (g-C3N4)-based photocatalysts for artificial photosynthesis and environmental remediation: are we a step closer to achieving sustainability?. Chemical reviews. 2016 Jun 22;116(12):7159-329.
41. Breeden J. A survey of machine learning in credit risk. Journal of Credit Risk. 2021;17(3).
42. Naruei I, Keynia F, Sabbagh Molahosseini A. Hunter–prey optimization: Algorithm and applications. Soft Computing. 2022 Feb;26(3):1279-314.
43. Evci U, van Merrienboer B, Unterthiner T, Vladymyrov M, Pedregosa F. Gradmax: Growing neural networks using gradient information. arXiv preprint arXiv:2201.05125. 2022 Jan 13.
44. Markov T, Zhang C, Agarwal S, Nekoul FE, Lee T, Adler S, Jiang A, Weng L. A holistic approach to undesired content detection in the real world. InProceedings of the AAAI Conference on Artificial Intelligence 2023 Jun 26 (Vol. 37, No. 12, pp. 15009-15018).
45. Ronen D, Bsoul A, Lotem M, Abedat S, Yarkoni M, Amir O, Asleh R. Exploring the mechanisms underlying the cardiotoxic effects of immune checkpoint inhibitor therapies. Vaccines. 2022 Mar 31;10(4):540.
46. Johnson-Laird PN, Herrmann Jr DJ, Chaffin R. Only connections: A critique of semantic networks. Psychological Bulletin. 1984 Sep;96(2):292.
47. Coglianese C. A framework for governmental use of machine learning. InAdmin. Conf. United States 2020 Dec.