

REVIEW ARTICLE

Evaluation of risk level assessment strategies in life Insurance: A review of the literature

Vijayakumar Varadarajan^{1,2}, Vijaya Kumar Kakumanu^{1,*}

¹ Research Division Swiss School Business and Management, 1213 Geneva, Switzerland

² International Division, Ajeenkya D Y Patil University, Pune 412105, India

* Corresponding author: Vijaya Kumar Kakumanu, vijaya@ssbm.ch

ABSTRACT

The viability of every insurance company depends on risk assessment of new life policy proposals. Machine learning techniques are increasingly shown to double case processing speed, reducing manual evaluation time. The underwriter evaluates the risk in several ways, including financial and medical evaluations and category classification based on customer data and other factors like previous insurance information, clinical history, and financial data. This research examines different academics' publications on risk prediction while offering a new insurance policy to an applicant. Multiple machine learning models developed by researchers have been extensively investigated. The researchers' model evaluation criteria were analyzed to understand and discover study gaps. The article additionally analyses how researchers found an accurate machine-learning model. This report also analyses various scholars' future work proposals to identify what could possibly be modified for further research. This study details the measures used by other academics to evaluate machine learning models. This study describes the criteria used by other scholars to evaluate machine learning models. The criteria used by investigators to assess the produced models were carefully evaluated to understand and spot any untapped potential for advancement. Researchers' methods for finding an accurate machine-learning model are also examined in this article. In addition, this study analyses several researchers' future work proposals to discover what may be changed for further research. Using previous academics' work, this review suggests ways to enhance insurance manual procedures.

Keywords: academics; criteria; evaluation; insurance; machine-learning; processing, risk prediction; researchers; speed; viability; underwriter

ARTICLE INFO

Received: 14 August 2023
Accepted: 1 December 2023
Available online: 13 March 2024

COPYRIGHT

Copyright © 2024 by author(s).
Journal of Autonomous Intelligence is published by Frontier Scientific Publishing. This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).
<https://creativecommons.org/licenses/by-nc/4.0/>

1. Introduction

The life cycle of the insurance proposal begins from the moment the representative submits the policy proposals to the insurance carrier. The new business process encompasses the whole life cycle of the case, from the time it is presented to the customer until the case is issued or brought to Inforce (active contract). The new business life cycle encompasses several stages, some of which are iterative and rely on data from third-party sources as well as internal systems (life insurance companies). The application proposal will be evaluated at the home office after it has been submitted. At this point, prescreening of the proposal will be done. The application will either be approved or refused once the case in subject assessment is completed by the underwriter to the best of their ability. However, because the review is a manual process, it may take longer than planned, resulting in consumer frustration and, in some scenarios, the carrier losing key business.

The literature study was conducted with the assistance of a wide range of resources, such as Google Scholar, Research Gate, Science Direct, IEEE, Springer, ACM, and many more. In addition, a large number of research articles authored by a variety of scholars over a period of fifteen to twenty years have been carefully reviewed. The following terms were used as search terms in order to locate publications that were pertinent to the topic at hand: insurance risk prediction, automation of underwriting, and the influence of machine learning and data mining on the insurance business.

Experts have identified and considered accelerating the digitization of the UW process by implementing AI in the system. Various automation approaches, such as building predictive models employing machine-learning technologies, are now available in the market. The applicant's traits that contribute to the risk level will be used to establish the risk level. Data analytics is revolutionizing the way insurance firms collect, classify, assess, and manage data, finally arriving at case decisions. Data analytics revolutionizes how insurance firms collect, classify, evaluate, and manage data. As a result, various industries in the insurance business, such as risk predictions, customer insights, handling of policy claims, and underwriting procedures, are fast expanding^[1]. The author has mainly focused on constructing a model that integrates cross-validation methodology on customer attributes to estimate the client's risk level and to provide decision-makers with an effective system that will allow them to create an excellent judgement on the issue of policies while simultaneously raising the level of client satisfaction.

This study on the subject of accelerating underwriting in the insurance industry offers significant insights into the potential advantages for community. Accelerating underwriting involves leveraging revolutionary technologies and data analytics to streamline and expedite the evaluation of insurance applications and risk assessment. This literature study not only helps insurance businesses provide better service to their customers, but it also contributes to the benefits to the community of applying AI, XAI, and data mining in the insurance industry. These potential benefits include such as Influenced judgment identification, promoting customer confidence, new product offerings, customer retention, Improved insurance availability, avoidance of fraudulent claims, improved coverage, insurance sector digitization, and a stable economy are only some of the other benefits.

This document summarizes the literature review and discusses the relevance of data mining technology to the insurance industry's automated underwriting needs. Both the insured and the insurer will profit from automating the underwriting process. The insurance company can issue the contract swiftly and satisfy the consumer by automating the process. This allows the insurance firm to keep an existing customer while also bringing in a new consumer. Furthermore, this technology significantly reduces human error and the misconceptions that underwriters may have about the information that applicants provide in their policy applications as well. The underwriting procedure usually takes a longer time to process the application form, typically a week. Using automated and data mining methods, this reaction time can be significantly lowered.

The underwriting phase typically takes weeks to complete. This response time can be reduced, however, considerably by automated and data mining technologies. In addition, rather than processing routine matters, underwriters may make better use of their time by focusing on the most intricate policies. A further issue with manual underwriting is that underwriters must be knowledgeable in a variety of topics connected to the underwriting activities, as well as the way they employ that knowledge to determine the client risk level, which is extremely tough in the real world. Underwriters can use AI techniques to gain a detailed understanding of patterns that exist in customer data and can use this expertise in policy-making decisions.

Because the evaluation is a tedious activity, it may consume longer than anticipated time and effort (brain draining), resulting in a loss of competitive advantage and eventually losing the business. In recent years, the proportion of people purchasing insurance coverage has risen dramatically. Additionally, the conventional method is mostly subjective, relies on individual UW personal experience, as well as a time-consuming

process. With the conventional risk assessment technique, there is a danger of overcharging low-risk consumers while undercharging customers with high-risk profiles^[2]. Accelerating the digitalization of the UW process by bringing intelligence into the process is one of the process enhancements recognized and acknowledged by industry professionals. Various automation approaches, such as constructing computational models leveraging machine-learning algorithms, are now accessible in the market. In the insurance sector, effective and reliable customer data collection is crucial. Data mining is used to assist industry specialists in their judgement process. ML algorithms rely heavily on client data that can be readily maintained through data analytics techniques^[3]. The major goal of this literature review is to explore the works of researchers in determining the ML model to enhance policy issue judgment accuracy. The principal focus of this study is to learn the applicant attributes that influence the level of risk, to ascertain the strategies followed by previous researchers to arrive at accurate risk prediction as well as to evaluate how AI technologies weigh the UW decision. This analysis attempts to explore the queries like the impact of personal and professional traits on risk level assessment and also cracks to identify a better ML algorithm by going through the previous investigations.

In previous studies, researchers focused on customer risk level prediction without considering the benefits of combining different strategies in their machine learning models. This study proposes various techniques for future study that include cross-validation, class balancing strategy, and hyperparameter tuning strategy to enhance the accuracy of risk assessment models in machine learning. In addition, the review also recommends employing more accurate performance metrics and also to apply system characteristics such as execution time, memory usage, etc.

Structure of the study

The remaining sections of this study are structured in the following manner: The chapter titled ‘Literature review’ provides an in-depth exploration of various aspects related to the underwriting process. It covers the manual underwriting process and its workflow, the advancements in automation of underwriting, the impact of data analytics on the insurance industry, the significance of data mining in the insurance sector, the role of KDD in facilitating risk prediction, Relevance of the Application of AI and the Role It Plays in the Insurance Industry, and a comprehensive overview of previous research conducted by other scholars in this field.

In the chapter titled ‘Conclusions’, the study portrays final statements and contributions in the field of risk level prediction in the insurance industry by the researchers that use the same dataset. In addition, the datasets used, algorithms implemented, methodologies followed, and the outcomes of experiments conducted by others have been compared in this section. Finally, the study highlights the areas that require further attention and development.

2. Literature review

2.1. Underwriting automation

Underwriting is a crucial division within an insurance company, responsible for evaluating the risk associated with policy proposals and analyzing the data provided by insurance agents regarding prospective customers^[4]. The insurance business involves offering individuals or businesses protection against potential risks they may face in the future. The underwriter will review each policy, and the risk associated will be assessed based on the firm’s established policies and rules^[5]. High-quality underwriting services are crucial for firms to retain a competitive edge, generate revenues, and preserve the trust of consumers. Typically, a number of machine learning techniques have been used to expedite process applications without compromising accuracy in evaluating risks^[6].

When it comes to the insurance business, one of the most difficult challenges is keeping current clients.

Maintaining existing relationships with clients is of greater importance than seeking out new ones. The general principle of Pareto can be applied here, that 80% of revenue comes from 20% of consumers^[7]. The conventional insuring procedure entails underwriters individually analyzing a digitized paper application and assessing the applicant’s risk rating. If the insurer has any reservations, he would ask for further evidence and then scrutinize the policy after the needed information is received. In certain circumstances, the underwriter will transfer the applicant’s proposal to another underwriter for opinion. The complete policy issuance procedure takes a long time. Automatic underwriting can significantly eliminate this sort of time waste. It takes a long time to have a policy issued. The digitized screening process can cut down on this sort of time waste significantly. Because policy issuing decisions are carried out manually, an internal review is carried out regularly to ensure the correctness of the outcomes. Furthermore, underwriters make choices based on a set of criteria and also make decisions based on their medical expertise and prior knowledge, resulting in discrepancies. Such challenges can be overcome by using machine learning techniques to automate the underwriting process.

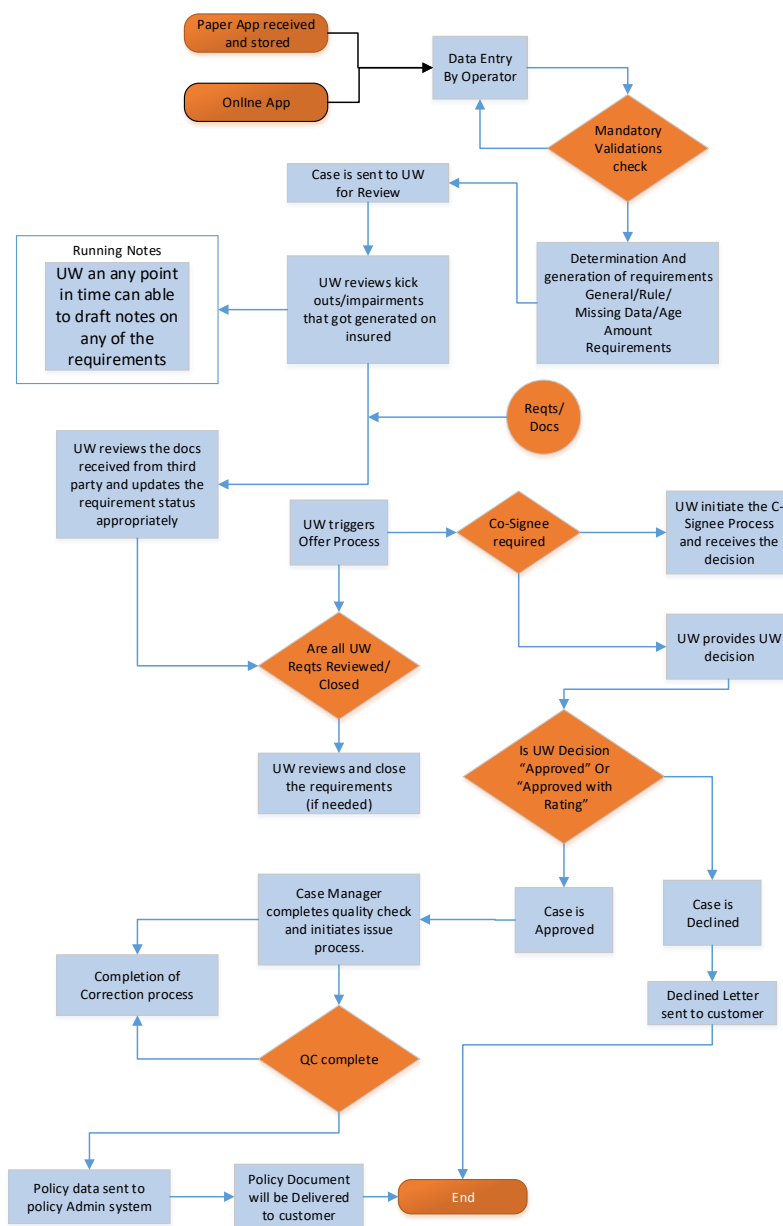


Figure 1. Underwriting workflow.

Figure 1 depicts the fundamental underwriting workflow. The initial undertaking is made by the sales

agent using Tele calling. If more evidence, like medical reports, is necessary, the insurance proposal will be sent to the assessor, who will underwrite the policy. If any extra information is necessary, the underwriter is likely to place queries. The underwriter determines whether to approve or reject the evidence after evaluating it. Furthermore, the underwriter will provide a rating to the insurance based on the risk classifications such as preferred, standard, substandard, and refused^[8].

The policy's underwriting approach has significantly improved since the introduction of data mining and machine learning technologies. The preliminary underwriting is done using machine learning approaches, and the applicant's chosen product type determines how the preliminary processing is carried out. Additionally, during the preliminary processing, requirements like missing medical requirements will be created. Apart from that, the rating category and insurance premium are determined. Following that, the proposal is forwarded to an assessor for additional consideration. The assessor has the authority to waive the prerequisites specified or establish additional criteria if necessary.

2.2. The influence of data analytics on the insurance industry

In today's world, data is a critical tool for any organization. With this data, every firm seeks to better its business. This information is even more critical for the insurance industry, as they rely largely on data statistics. Analytics allows firms to extract essential information from their data warehouse, allowing insurers to stay ahead of the competition with their counterparts^[3]. Sivarajah et al.^[9] provided a thorough overview of Big Data Analytics, and it is obvious from the research that it has greatly altered data analysis perspectives. Many companies are looking to Big Data Analytics to extract useful insights from large amounts of data. To enhance their businesses, an increasing number of companies are turning to business intelligence and data analytics. The analytics process is viewed as a tool for improving organizational efficiency, developing new revenue prospects, and winning market dominance over competitors.

Data analytics is crucial in every aspect of the insurance industry. Data analytics usage in the insurance industry will improve customer satisfaction and lead to considerable development in new possibilities. So, if data analytics is implemented appropriately, it has the potential to significantly influence how insurance firms operate. The following are the major sectors of the insurance industry that could really benefit from enhanced data analytics^[10].

2.2.1. Potential customer base

In the age of the internet, there is a huge volume of data available on the web in a chaotic format. In the insurance industry, data analytics may give detailed insights into client behaviour, which can aid in the discovery of new business prospects.

2.2.2. Organization image

Established consumer data studies suggest ways to improve service to customers, which boosts the company's brand recognition.

2.2.3. Avoidance of Fraudulent claims

Detecting fraud in insurance claims has always relied on traditional statistical approaches. Data mining and data analytics technology, however, may now be used to avoid these frauds. False claims made in the past are saved in the central repository of the insurance companies, and when evaluating the new claim, the insurer can compare the new claim pattern with the central database. This will aid in the reduction of fictitious instances.

2.2.4. Assessing risks and pricing premiums

Insurers are critical in assessing a consumer's risk level. Before offering a policy, they analyze the risk level based on many characteristics of the consumer. When it comes to assessing client profiles, data analytics

has a role. Furthermore, insurance premiums are generally determined by risk levels, which are based on a variety of factors including the customer's profession, family background, health information, age, and Body-mass index.

2.2.5. E-commerce portals for service and support

E-commerce portals for service and support: Customers have become adapted to online services, which are available in a variety of sectors, including the insurance industry. Customers could perhaps manage their insurance policies using the specialized customer website. Customers can thus make use of a variety of services, including acquiring a new policy and services offered post-policy issuance. Offering these services helps to keep ongoing expenses minimal while also keeping clients satisfied. Insurers may use data analytics to keep records of policy history and provide appropriate policy suggestions to customers.

The insurance sector is evolving as a result of the advent of data analytics. Data analysis, in particular, is one of the historical roots of insurance. For long underwriters have been using traditional rule sets to estimate premium rates and risk coverage. When selling policies, insurers collect huge amounts of data on their clients, which are then changed when the consumers file an insurance settlement. Underwriters utilize this altered policy data to estimate the amount of risk when the same consumer requests a new policy from any other insurance firm. Insurance firms have recently realized the relevance of business intelligence. They intend to employ data analytics to serve customers better, minimize risk assessment response time, swiftly handle insurance claims and avoid fraud while claiming coverage^[11]

Another key section of the insurance business where big data analytics is heavily exploited is customer profitability. This industry hasn't really taken off until big data from client databases, which contain information on past purchase habits, became accessible. Because insurers have access to every transaction information and claim the record of their clients, they may decide whether or not to continue them and whether or not they wish to preserve long-term ties with them^[12].

Bohnert et al.^[13] precise digital objectives are required to fully exploit digital technology's huge potential for long-term advancement in the insurance industry. Additionally, rather than going with the flow in their study and implementing digital technology in certain circumstances, the researchers found that businesses should think deeply about automation and acknowledge its influence on their whole business operations, both local and global.

The insurance sector is seeing an increase in the amount of data it collects. Data analysis techniques can be used to extract relevant information from the data collected. The key domains of insurance where data analytics may be employed include customer analysis, risk assessment, as well as fraud recognition. The primary purpose of using data analytics in insurance is to study and understand consumers, reduce risks, and gain a long-term competitive edge^[14]. In his research, Boodhun^[14] covered the three primary parts of the insurance sector where data analytics is applied, such as customer analysis, risk evaluation, and fraud detection.

2.2.6. Customer analysis

Customer profiling is the process of studying consumer behaviour in order for businesses to effectively estimate customers' demands. This is accomplished solely through data mining techniques focused on improving client networks and loyalty^[15]. The capacity to adequately categorize the appropriate policies that are sourced to customers might be provided using consumer analytics. Several researchers have presented various methods that have been extensively researched and applied to insurance data to develop modelling techniques, such as classification, Clustering, regression, and plenty more. In this study, authors Delafrooz and Farzanfar^[16] used a data mining approach to cluster end-user policy data to calculate their term worth. Consumer retention is yet another approach to defining and classifying customer behaviour and so estimating

their profitability for the business to enhance customer engagement initiatives^[16].

2.2.7. Risk evaluation

Insurance companies value risk assessment because issuing a policy to the customer means that the risk of loss is transferred from the policyholder. Insurance firms assess and evaluate their clients' risks to calculate the cost of coverage. Risk is usually calculated using real formulae., but with the introduction of data analytics, the subsequent study focuses on developing newer formulae with more precise criteria that impact policy holder's risk proportion^[17].

2.2.8. Detection of fraud

In the insurance industry, fraud and data manipulation is clearly a major challenge. Fraud is defined as the intentional use of certain data to create false claims. Insurance providers are overburdened by fraudulent activity, and they suffer significant losses when fake complaints are lodged. Because companies are currently seeing an enormous surge in transactional data that is produced regularly^[18], data analytics might assist in enhancing the manner in which insurance fraud is addressed. A variety of algorithms are available to identify fraud efficiently, hence strengthening the operational process to reduce avoidable damages.

The detection of fraud in insurance claims is yet another critical area where data analytics is extensively applied. The insurance industry is currently experiencing fraud risks while also dealing with client attrition issues. Mining and data analytics are increasingly being employed in the insurance industry, with the goal of managing data in a manner that stimulates insurance audits and prevents bogus claims. The number of fraudulent claims in the sector is believed to be 10%–15% of total claims, according to the TPA survey^[19]. As a result, insurance companies must devise creative strategies to keep their businesses healthy. It's a great move to use data analytics to help them make better judgments with their huge amount of data.

2.3. Data mining and its emphasis on the insurance industry

Machine learning identifies techniques that allow computers to learn autonomously without human involvement, whereas data mining collects knowledge from massive volumes of data^[20]. Data mining is an excellent tool for assisting insurers in focusing on accumulating valuable information about their client behaviours. To analyze consumer habits and patterns, it is required to describe and segment customers based on various characteristics. Furthermore, insurers could very easily model their consumers' data and contact them with plans that they would be interested in purchasing^[21]. The fast growth of life insurance is mainly determined by customer risk. Since the proportion of losses between many of the customers is so unequal, it's vital to examine their previous data to categorize them based on specific criteria that will aid insurers in making policy decisions. On a broad scale, data mining and machine learning technologies aid in the classification of clients.

On a bigger scale, the risk of the policyholder can have a significant influence on the profitability status of a life insurance firm. Data mining and machine learning provide a variety of ways and procedures to understand the true nature of individual behaviour, preferences, and patterns through data. As a result, it's critical not just to categorize the data but to develop a model to facilitate the categorization process much easier as well. The framework that has been built assists the life insurance sector in making decisions. Erroneous information could lead to misinterpretation, which could also prompt clients to migrate to other competitors, causing the insurance company to suffer losses^[22].

In his study, Ansari and Riasi^[23] underlines that relationship marketing aims to improve long-term customer relationships by providing better services and solutions to policyholders. To accomplish this, the customers can be segmented into small groups. It is considerably easier to detect and assess the characteristics of groups of clients than it is to investigate each client individually. Age and disability are significant predictors that should not be overlooked by substituting risk variables for particular types of insurance in the pursuit of

appropriate coverage^[24].

In the insurance industry, data mining can help companies gain a competitive advantage by speeding up decision-making. The insurance provider needs to understand the intricacies of selection and data mining processes to thrive in the life insurance industry. To assess new client bases, hold back current customer bases, accomplish customer categorization, and establish the association between product and policy choices, underwriters must have an extensive understanding of statistical patterns and probabilities. The clustering technique is used to gain potential customers by recommending an appropriate approach based on the user's attributes that are close to the cluster features^[25].

Another data mining technique that may be used to generate new products and services is classification^[14]. The association is another data mining function that focuses on retaining existing consumers. It explains the correlation between data and will, therefore, be exploited in market research by companies. The insurance carrier will create association rules to identify what additional policies have also been taken in conjunction with a specific policy. To assess the consumer's purpose in acquiring a policy, insurance companies conduct a correlation study between policy design and policy buying by the customer^[25]. Customers migrating to other competing firms is another critical issue that many insurance companies experience. Companies are exploring real-time data to reduce customer attrition^[26]. Data mining techniques aid these retention attempts. To thoroughly investigate the factors and determine the chance of customer switching, a range of diverse studies have indeed been carried out. Nevertheless, the measures that will be taken are the same for all of the strategies.

In his work, Sunita et al.^[27] went through the relevance and significance of data mining in the insurance sector in great depth. Furthermore, the author investigated data mining strategies along with consumer data issues, retaining customers, and organizational growth. By targeting critical areas of the insurance sector, such as establishing a new client base, maintaining customer retention, and identifying frauds, data mining can help insurers gain a competitive edge. By delving into the depths of customer data, data mining tools assist decision-makers and marketing executives in making the best judgments. The term "data clustering" encompasses the technique of grouping a collection of items into several distinct categories in order to ensure that the instances included within each class are more related to one another than they are to the data samples contained within any other class. Clustering methods may be broken down into two distinct types: partitional and hierarchical. An approach for partition clustering separates a dataset into one partition. A hierarchical clustering technique, on the other hand, creates a tree-like structure out of a dataset. Segmentation of data sets is one of the most commonly employed approaches for EDA and a significant part of the process of machine learning. Clustering data is often referred to as unsupervised learning due to the fact that it may be used for data that has not been labelled. As a direct consequence of this, the practice of data clustering has been implemented in a wide variety of scientific fields, such as engineering, computer science, and life science. Gan and Valdez^[28] reviewed several basic clustering methods and data clustering principles in their investigation. They have highlighted the use of various clustering techniques, distance measurements, and cluster validity. In order to showcase the practical benefits of data clustering in the field of actuarial science and insurance, they have utilized two scalable clustering algorithms, namely the truncated fuzzy c-means (TFCM) algorithm and the hierarchical k-means algorithm, to analyze VA contracts and develop predictive models. Based on the findings, it is evident that the hierarchical k-means algorithm demonstrates superior performance compared to the TFCM algorithm.

2.4. Role of KDD and data mining in risk assessment

The most important aspect of the insurance industry is risk assessment. Insurance premiums are measured by the degree of risk. Data mining technologies and approaches are increasingly being used to analyze risk levels. According to research conducted by LIMRA, the Association of Life Insurance and Market Research,

in 2015, “about 90% of financial services organizations are exploring the use of big data analytics to optimize membership processes.” At its most basic level, the insuring procedure includes data gathering, data processing for risk assessment, and making choices on policy issuance. With the introduction of predictive analytics, insurers will be able to make more accurate decisions based on the data. Predictive Modeling significantly decreases application turnaround time, saving time for both the company and the individuals. In addition, it reveals hidden patterns that exist in the data being analyzed, providing a comprehensive picture of potential risks. In other words, it enables insurers to take suitable safeguards and relate them to their learning perception^[29].

The firm evaluates policy proposals and assigns a risk rating to determine their premium. Typically, the policy includes a history of medical, employment, and other personal data. If firms misjudge risk, and if a corporation overestimates risk and costs more, it may lose the client to a competitor. A reliable risk predictor is needed to run the firm smoothly^[6]. The risk is passed from the policyholder to the insurance provider when the consumer is insured. Insurance carriers evaluate client losses to estimate the number of premiums to be paid^[17]. Traditionally, the risk was assessed using underwriting formulae. With the introduction of business intelligence tools, a much more precise investigation has been conducted on generating new rules to come up with alternative formulae to obtain the appropriate amount^[14].

Some of the most prominent approaches for identifying the risks associated with insurance businesses include regression analysis and classification strategies. In general, insurance companies pursue a product-based approach, focusing on providing a better product than their rivals. Furthermore, there are information gaps in the insurance industry. To anticipate a customer’s risk level, insurers need to gather a lot of information about customers. Big data analytics and machine learning models aid in the analysis of customer data, consequently enhancing the accuracy of the client’s risk level assessment^[30]. Insurance is typically regarded as a transfer of risk from an individual to an insurer. Because life insurance claims are usually sizable, it’s critical to assess the individual’s risk before offering insurance, particularly in the context of the firm’s economic position. In essence, the screening procedure, which evaluates the risk of the applicant, is crucial while offering policy^[31].

Life insurers’ access to clients’ genetic details has been a source of contention for many years. With the introduction of genome sequencing, the topic of genetic testing in the context of life insurance has become considerably more difficult, necessitating a review of existing rules and the exploration of new options. The researchers Joly et al.^[32] looked into the ethical, social, and legal implications of using genetic data in the insurance business to ascertain risk levels and insurance premiums when offering policies.

2.5. Significance of AI use and its influence on the insurance sector

Artificial intelligence (AI) is a rapidly growing technology, enterprise, and area of academic Research. Owing to its extensive acceptance and the development of new methods, AI solutions demonstrate remarkable effectiveness in several sectors, especially in Business and Finances. Nevertheless, the use of AI and machine learning in the field of Insurance is still in its early stages, with the majority of efforts focused on conducting preliminary tests. Many initiatives have yet to be expanded and launched in the industry. Life insurers are strengthening their digital infrastructure and relationship with customers to reduce the gap in coverage for those who are not covered or have inadequate coverage. The availability of extensive repository data sources and progress in AI can enhance the precision and openness of insuring in the life-insurance sector by using standardized indicators of mortality risk.

The writers Chan et al.^[33] conducted a study with the objective of gaining a deeper understanding by employing Artificial Intelligence (AI) methodology in the field of Insurance. More precisely, Research investigated the extent and marketplace reach of artificial intelligence (AI) in insurance offerings to address

persistent issues and enhance client appreciation. The model constructed aims to assess the correlation of artificial intelligence (AI) and their applications in the field of Insurance. The authors acquired realistic findings deemed very valuable for Insurance businesses in handling frustrated clients and other business challenges. Furthermore, the authors studied various uses of artificial intelligence (AI) in the insurance business, including facilitating innovative portfolio management, enhancing sales and marketing strategies, enhancing client delight, digitizing and enhancing the claims settlements, accelerating the underwriting procedures, improving risk prices, and combating insurance fraud. The writers concluded that AI could enhance client delight, increase profitability, mitigate fraud and streamline time and administrative challenges within the insurance business.

To balance the provision of cost-effective products with the management of the financial performance environment, life insurance firms use an underwriting procedure to evaluate the risk presented by individuals. Conventional insuring mainly relies on assessing a client's medical, demographic and other attributes^[34]. The authors used an AI approach to construct a mortality model. They explained the performance of the risk prediction model, which achieved remarkable accuracy and resulted in a significant per cent decrease in claims. Furthermore, the authors highlighted the need to adopt openness in order for the business to establish customer confidence and effectively address a constantly changing regulatory landscape that prioritizes computational choices.

Given the importance of Customer Relationship Management (CRM) in the insurance sector, it is anticipated that this business will be at the frontline of digitalization and extensively use cutting-edge Artificial Intelligence (AI) technology. Nevertheless, deploying artificial intelligence technologies in organizations does not necessarily ensure success. With the use of AI, businesses may refine their approaches to client segmentation, leading to more effective CRM practices. Yum et al.^[35] suggested a set of principles for AI-based client classification to standardize data mining projects in the insurance sector. This proposed guideline provides new insights for studies on AI-based technology implementation. The research primarily emphasized objectives such as promoting supplemental items to current clients, convincing consumers to modify their existing plans, and eventually acquiring new customers. The authors employed the measures of accuracy, recall, and f1-score in order to conduct an evaluation of the model that was developed.

Yum et al.^[35] conducted a study and found that the present client classification procedures are inaccurate and require even more investigation. They also noted that AI-based consumer classification has not been adequately explored. Additionally, they found that the conventional classification strategy fails to evaluate a significant portion of the data, leading to inaccurate customer categorize identification and customer dissatisfaction. Hence, the authors proposed a customer segmentation framework, CRISP-DM, specifically tailored for the insurance industry, with the aim of enhancing firms' effectiveness. In addition, as part of a future study, the authors recommended using real evidence from the insurance firm or carrying out case studies to get a deeper understanding of insurance consumers concerning their traits and preferences.

In the past few years, there has been an increase in the use of AI techniques in the insurance industry. However, researchers who work in the field of artificial intelligence (AI) sometimes face technical difficulties in keeping up with the ever-evolving literature and vast, complex, domain-specific expertise. Isa et al.^[36] aimed to unify the rapidly expanding research on AI in the insurance sector by conducting a thorough literature study. This survey would allow for the incorporation of a wide variety of AI approaches into all of the insurance sector's most significant responsibilities. The survey conducted by the authors explores the use of modern AI methods such as MLP, AdaBoost, SVM, Linear Regression, Naive Bayes, MLR, XGBoost, Random Forests, Logistic Regression, J48, Classification and Regression Trees (CART), REPTree, and Decision Trees in the insurance industry. The authors concluded that Overall, XGBoost performed better than the other AI methods evaluated. In addition, the authors found a few shortcomings in the earlier research, such as overlooking the

use of a deep learning algorithm to develop models, failing to make use of a variety of performance indicators, and not using the XAI approach to explain the models that were built.

Several AI systems, especially DL models, are seen as “black boxes” owing to their complexity and lack of transparency, prompting a requirement for XAI (Explainable Artificial Intelligence). While such algorithms are effective at making forecasts, they fail to explain their decision-making processes. Life insurance requires doing a risk analysis on each person to be insured and calculating the proper amount of coverage and cost for that coverage. The field of life insurance stands to gain much from using XAI. XAI is helpful in a number of areas in the context of the insurance sector, such as apparent underwriting, risk evaluation, influenced judgment identification, client training, identification of fraud, and settlement management. Furthermore, The confidentiality of data and its safety must be carefully considered while deploying XAI in the life insurance market because of the sensitive nature of the information being handled. In summary, XAI’s ability to explain underwriting and claims judgments, assist in the identification of fraud, and facilitate the development of products will increase openness, fairness, and consumer confidence in the industry.

Owens et al.^[37] analysis explored the explainability of existing AI implementations in insurance business operations and insurance research. According to the findings of their investigation, the researchers conclude that XAI procedures are prevalent in claims administration, policy underwriting, and pricing processes. XAI is a crucial AI advancement for the system’s environmental issues to be dependable, transparent, and righteous. The evaluation of these XAI focal points in the broader picture of the insurance sector offers an appropriate study into the unique benefits of XAI, showing experts in the field where specific attention should be devoted to the advancement of XAI. The authors suggest a revised interpretation of XAI based on a comprehensive evaluation of XAI research in the insurance industry. This new definition contributes to the body of knowledge on appropriate XAI definitions. The researchers recognize that it is tricky to define XAI as it applies to all circumstances and is unique to their respective fields. The authors suggested a definition specific to the insurance industry, like “XAI is the transfer of understanding to AI models’ end-users by highlighting key decision-pathways in the model and allowing for human interpretability at various stages of the model’s decision process. XAI involves outlining the relationship between model inputs and prediction while maintaining the model’s predictive accuracy throughout.” The study conducted by the authors presents a comprehensive review of the available research on the utilization of XAI in the field of insurance. Furthermore, the authors sought to fill the void left by the lack of extensive examination of definitions of XAI by proposing specific criteria applicable to the insurance industry. In addition, the study explores several other definitions of XAI and offers an alternative, more contemporary definition of XAI that is appropriate for the insurance industry.

The investigator Sushant^[38] focused on investigating how the two significant mainstays of artificial intelligence—machine learning (ML) and deep learning (DL)—help address problems in the insurance sector. Insurers are experiencing complicated and novel issues due to the characteristics of the data they generate. While constructing different prediction models, traditional statistical and machine-learning techniques may not be effective enough. NLP, DL, and RL have the potential to be effective. Both deep learning (DL) and machine learning (ML) serve as the backbones of artificial intelligence, enabling the field to advance toward its many goals in areas such as identifying potential customers, accelerated underwriting, locating new business, Customized services, client classification, Customer retention, insurance loss prediction, forecasting claims, etc. Information extracted from policy repository systems using a range of AI methods, including image processing, text analytics, natural language processing, natural language understanding, and natural language generation, may aid in product creation, marketing, and risk assessments. Improvements to both present and prospective services might be made with the use of speech recognition technology by better comprehending the difficulties experienced by customers.

Furthermore, the implementation of AI in the insurance industry would result in a wide range of business advantages. These include enhancing customer satisfaction and loyalty, improving interactions with clients, boosting revenue, enhancing customer intellectual ability, tailoring content to individual needs, increasing efficiency, and lowering operational expenses.

The insurance sector relies heavily on data, and AI is helping it make better use of all that information. Artificial intelligence can process, analyze, and assemble enormous data sets at a faster rate. The technique utilizes information to train machines to make predictions and take corrective measures. Artificial intelligence (AI) serves in areas such as customizing products to meet customer requirements, automating tedious server-side tasks, aiding in identifying and preventing fraud, accelerating the process of policy underwriting, etc. Insurers are leveraging digital platforms in an effort to broaden their customer base. Most insurance companies now use digital technologies like artificial intelligence and chatbots to improve their client experience. Several artificial intelligence (AI) technologies, including machine learning, RPA, text analytics, NLP, and audio/visual analysis, have the potential to alter the way services are delivered completely. Gupta^[39], in his study, investigates how AI has influenced the insurance business and outlines the numerous AI applications employed by insurance companies to provide services that consequently satisfy clients. AI has become increasingly prevalent across various industries, experiencing significant growth in recent years. It is bringing up numerous business prospects, significantly enhancing businesses across multiple sectors, such as Insurance. Insurers are using Chatbots, a kind of virtual assistant that uses natural language processing and sentiment analysis to understand their customers better and increase satisfaction.

Artificial intelligence (AI) is poised to revolutionize the insurance sector in many ways, including how services are underwritten, claims are handled, clients are reached, and fraud is uncovered. The goal of AI is to enable the delivery of revolutionary services to established user profiles. Its end goal is to promote consumer retention and confidence in the company via innovative services. Artificial intelligence helps insurers determine the optimal level of risk and create customized offerings for each client. Insurance companies will benefit on a long-term basis as a result of the lowered risk rate and preventative measures taken on time, which in turn provide chances for more business and cross-selling, which in turn will increase earnings.

2.6. Related research work

Data mining techniques have transformed the insurance sector. Insurers have gained the advantage of comprehensively analyzing client data with the development of data machine learning and data mining and may either intentionally or unintentionally sustain consumers by lowering turnaround time while issuing the policy. The core concept of ML is the idea that machines should possess the ability to learn and adapt through experience. Technology may help businesses realize the advantages of data gathering when used effectively^[40]. Machine learning algorithms have also crafted an impact on the insurance sector, including the development of claim risk prediction, financial analytics, fraud detection, new product offerings and reinsurance^[3].

The way insurance companies gather, analyze, edit, and handle data has changed as a result of data mining technologies. Data mining has made a significant contribution to risk prediction, customer inquiry, and claims processing. However, few life insurance firms still use classic actuarial formulae to forecast death rates and premiums. Method et al.^[41] Summarize in their research how an increase in data volume affects the risk level assessment process. The author employed dimensional reduction methods like principal component analysis and backward elimination strategy to estimate Random Forest, ANN, Random Tree, and Multiple Linear Regression since the considered dataset has a significant set of attributes. Furthermore, Method et al.^[41] considered RMSE and accuracy as crucial assessment metrics to evaluate predictive accuracy spanning various data mining methodologies^[41].

Machine learning models are proven to be quite powerful in standardizing the categorization issue, and their

performance will improve if a substantial training database is presented. Insurance is indeed dependent on estimates for the future. Insurance companies make premium decisions based on the applicant's historical features. However, as novel approaches become available, underwriters have begun to estimate premiums based on a broader range of applicant characteristics. Furthermore, these characteristics of applicants are used by automated models to determine risk levels. These risk ratings are used by underwriters to process the contract. Medical underwriting will also be skipped if this risk level is lower than the stated score, which will speed up the policy^[42].

Risk in life insurance may be caused by a variety of factors such as age, family history, medical history, living conditions, and kind of job. Furthermore, the risk factor differs from one individual to the next. As a result, before issuing the policy, a process is needed to determine the applicant's risk level. Manual processing might take quite some time to assess an applicant's risk level since it needs comprehending and evaluation of the applicant's varied facts. As a result, data mining technologies such as machine learning models may be used to alleviate the problem of policy processing slowness^[43]. The researchers used machine learning models such as Random Tree, Random Forest, and XGBoost to evaluate the models in their research. The weighted kappa coefficient was used to assess the models in their investigation. The Kappa coefficient value was used to determine the accuracy of all the algorithms that were investigated. The author finds that the XGBoost model is effective based on his observations. Furthermore, missing values have no effect on the XGBoost model, according to this research. This is due to the fact that the Gradient boost model learns despite missing values.

Risk assessments are conducted by life insurance firms to ascertain individual risk levels and assess client suitability as per the information given by customers through the application form. In the insurance industry, risk prediction is solely dependent on clients' past records. Furthermore, because history records would be unique to each claimant, it is much more probable that past data will be lost. As a result of its sparsity awareness, XGBoost is a much-needed algorithm to weigh the risk rating of the individual^[44]. Rusdah and Murfi^[44] explored by generating two sets of data, one by not using any imputation techniques and the other with the well-known imputation techniques Mean and KNN. In his research, Rusdah and Murfi^[44] used the XGBoost approach and investigated the methodology with and without data imputation, concluding that the accuracy of the two models is similar. Theoretically, similar work was also carried out by Mustika et al.^[43], in which the researcher proved that XGBoost gives better performance despite the fact that the data has significant missing values. A very essential component in assessing risk level is that it determines the insurance company's success. The manual life insurance application process has become outdated as a result of the emergence of data mining and machine learning technology. To label risk categorization and eligibility, underwriters are often given detailed demographic details of the applicant. This procedure usually takes around 30 days to complete. Customers have the option to approach alternative competitors in the market or, in the worst-case scenario, decide not to purchase the insurance at all.

Hutagaol and Mauritsius^[5] researched to explore the impact of automation on the risk assessment process for life insurance companies. They aimed to understand how this technology can assist in determining the level of risk associated with proposals from potential clients. They used the linear kernel SVM method, the Random Forest algorithm, and the Naive Bayes algorithm. The authors used Accuracy, Precision, and Recall to assess the efficiency of the models constructed. As a significant portion of the data points in the dataset belong to one class, the researcher converted the multiclass dataset to binary. Based on their research, they concluded that Random Forest provided the maximum accuracy^[5]. (Risk Level Prediction of Life Insurance Applicant using Machine Learning).

In Chen's work^[45], he briefly discusses how the consistency and reliability of the screening procedure affect the link between firm brand image and insurance company profitability. Fombrun and Van Riel^[46] state that "Corporate reputation is the overall representation of a company's previous actions and outcomes that

demonstrate the company's potential to produce valued returns to multiple shareholders." Hence, improving the turnaround time and the efficiency of the underwriting process, as well as the correctness of risk level evaluation, is crucial to retaining the customer base and attracting potential ones.

Typically, firms make their policy issuance decisions entirely on the information given by consumers. However, insurance businesses frequently encounter the issue of adverse selection, which is a widespread occurrence^[30]. It describes a scenario in which insurance firms lack sufficient information on an individual and offer a contract to someone whose risk levels are high. Insurance companies need to make an extra effort to refrain from such a negative choice since it will have a significant impact on their financial position. When providing a policy, insurers should consider all of the features of customers that impact the assessment process. The goal of this research is to establish a learning algorithm to solve the issue of negative choice that appropriately classifies the risk level of candidates.

The scholars Gopi and Govindarajula^[31], as part of their research methodology for data imputation, considered a computational approach for categorical variables and a median strategy for continuous variables. Furthermore, the authors used the strategy of categorizing risk levels to reduce dimensionality. In this study, the strongest model was chosen using the metric ROC index as the eligibility criteria. He proposed that the genetic element be considered when insuring insurance as a recommendation for future research. To assess application risk, companies utilize customer data. Depending on this evaluation, firms accept applications and compute premiums depending on applicant risk^[47]. Life insurance firms continue to use conventional underwriting rules to assess mortality rates and life insurance policy premiums. Furthermore, life insurance businesses have started adopting data analytics to boost their company, even though real research into how data mining might benefit the life insurance industry has yet to be undertaken. Data mining techniques have been studied to detect insurance business fraud^[48].

For years, life insurance firms have used conventional mortality tables and actuarial data to predict lifetime and create underwriting rule sets. Traditional methods, on the other hand, take time and are frequently expensive. As a result, it's critical to figure out how to make the underwriting process more efficient while still staying within budget^[49]. The authors of this study, Boodhun and Jayabalan^[49], used a prudential dataset and preprocessed it by removing characteristics having missing data percentages of more than 30. Furthermore, based on the study, their research concluded that missing data were MAR and used the appropriate imputation procedure, which included deleting category categories and considering numerical variables for imputation. Additionally, it followed a strategy that applied dimensional reduction methodology before applying ML algorithms to improve risk level assessments and considered MAE and RMSE as evaluation criteria for evaluating models. Furthermore, researchers compared models such as REPTree, ANN, Random Tree and MLR using feature extraction strategy, PCA, and CFS. They discovered that the CFS technique worked well with lower error rates for most models.

Boodhun and Jayabalan^[49] followed a strategy that applied dimensional reduction methodology before applying ML algorithms to improve risk level assessments and considered MAE and RMSE as evaluation criteria for evaluating models. **Tables 1** and **2** provide a comprehensive overview of the contributions made by academics to risk prediction for new business in the insurance industry.

Table 1. Consolidation of scholar's contribution.

Study	Insurance sector	Methodology	ML Models	Criteria	Outcome
Boodhun and Jayabalan ^[49]	Prudential Dataset 59,381(Rows) * 128 (Features) dataset	Dimensionality reduction approaches CFS, and PCA were used to improve model estimations.	Multiple Linear Regression REPTree, Random Tree Artificial Neural Network	MAE, RMSE, CFS, PCA	Overall, CFS outperformed the majority of the models investigated in this study. ANN, on the other hand, has performed better with PCA.
Mustika et al. ^[43]	Prudential Dataset 59,381(Rows) * 128 (Features) dataset	Different data imputation techniques were applied for DT and RF models and compared with XGBoost's optimized parameters to evaluate the final model.	XGBoost and its variants Decision tree Random Forest	Accuracy with Kappa coefficients	The Random Forest model worked well with the applied backward elimination strategy.
Method et al. ^[41]	Prudential Dataset 59,381(Rows) * 128 (Features) dataset	Employed feature elimination technique during the preprocessing stage to enhance model prediction accuracy.	Feature Elimination—PCA, Backward elimination NaiveBayes Logistic Regression Random Tree Random Forest	Accuracy RMSE	The Random Forest model worked well with the applied backward elimination strategy.
Bi et al. ^[2]	Prudential Dataset 59,381(Rows) * 128 (Features) dataset	During the preprocessing phase, several data imputation approaches and feature selection strategies were applied to improve model performance. In addition, the approach of transforming classification analysis into regression analysis was used.	Linear model SVM Random Forest XGBoost	Kappa Score, Accuracy	The choice of features has minimal influence on accuracy. Ensemble models can have a high level of accuracy. Since the classification issue was converted to regression, there has been a significant improvement in accuracy.
Rusdah and Murfi ^[44]	Prudential Dataset 59,381(Rows) * 128 (Features) dataset	Employed different data imputation techniques and evaluated the XGBoost model with and without imputation	XGBoost	Considered Confusion Matrix for evaluating the model.	XGBoost is required to manage missing values as there is a sparsely-aware finding algorithm to deal with missing values in Risk Prediction.
Biddle et al ^[50]	Australian insurance	transition-based and selection-based methods were used to minimize learning time and feature space during preprocessing to increase accuracy.	Logistic regression Xgboost Recursive feature elimination	receiver operating curve (roc)	reduced feature vector size saves learning time, increases accuracy, and prevents over-fitting. due to the necessity for a limited number of characteristics, xgboost is the most optimal model.

Table 2. Consolidation of scholar's contribution contd.

Study	Insurance Sector	Methodology	ML Models	Criteria	Outcome
Gopi and Govindarajula ^[31]	Prudential Dataset 59,381(Rows) * 128 (Features) dataset	Investigated using a statistical approach for categorical variables and a median technique for continuous variables while carrying out data imputation. In addition, the authors used the method of reducing dimensionality by classifying risk categories into three levels.	Neural Networks Ensemble model Decision Tree	Receiver Operating Curve (ROC)	The research demonstrated imputation methods used for missing values and dimensionality reduction techniques to help retain only the variables that explain the target variable. ANN model has the best AUC ROC Index.
Hutagaol and Mauritsius ^[5]	Prudential Dataset 59,381(Rows) * 128 (Features) dataset	the target variable will be changed to binary. Since the majority of the data is with response 8.	Support vector machine (SVM) Random Forest Naive Bayes.	confusion matrix, accuracy, precision, and recall	Random Forest performed well with the highest precision

Table 2. (Continued).

Study	Insurance Sector	Methodology	ML Models	Criteria	Outcome
Gupta et al. ^[6]	Prudential Dataset 59,381(Rows) * 128 (Features) dataset	The Dirichlet Procedure separated the data to create data allocations for distinct consumer groups. Training with a variable client distribution and count is simulated using the dataset divided across numerous clients. Also, the researchers compared the FL strategy to a centralized one that trains the whole dataset.	Keras Sequential Model XGBoost Decision Tree Random Forest	Quadratic Kappa Score	The stated FL method has been shown to be more effective than the standard centralized method.

The work done by Method et al.^[41] is conceptually built on the approach of using dimensionality reduction strategies during the preprocessing stage to enhance model power. Gupta et al.^[6] centred around safeguarding customer information privacy. To address this concern, Gupta et al.^[6] proposed a federated learning strategy that operates in a distributed manner, ensuring that actual data is not shared. The author used the Dirichlet Procedure to split the data in order to generate various data allocations across different customer groups. The dataset is split among many customers and used to simulate training with a variable client spread and client count. It has been suggested that the created models be evaluated using the measure known as the Quadratic Kappa Score^[6]. Gupta et al.^[6] followed the strategy of comparing the FL method with a centralized one that uses the whole dataset for training. The stated FL method has been shown to be more effective than the standard centralized method. The proposed FL learning methodology gives insurance firms the scope to collaborate and develop a robust predictive model without actually sharing their data^[6].

Rusdah and Murfi^[44] proposed a conceptually identical work on the effect of missing values on the accuracy of a machine learning model in risk prediction, in which he experimented and justified that the XGBoost algorithm works well despite missing values in the dataset. Biddle et al.^[50] suggested that applicants should comply with previous claims when determining the cost of underwriting decisions. Regression using feature extraction strategy, principal component Analysis and correlation-based feature selection technique, and discovered that the CFS technique worked well with lower error rates for most of the models.

3. Conclusions

As per the study carried out by researchers, Clustering, classification, and regression are the most extensively utilized methodologies for evaluating customer analytics to appropriately analyze the behaviour of customers, which assists insurers in forecasting client demands. Another purpose of studying customer behaviour has been to offer customers the desired services, thereby reducing turnover. According to the research studies, categorization is a generally used strategy in the insurance market for risk assessment and fraud identification that helps organizations enhance their revenue. In addition, research shows that segmenting clients into organized groups can help with customer engagement.

The research papers reviewed in this paper use many supervised machine learning models that focus on increasing applicant risk level assessment. Analogies with recognized insurance risk modelling techniques were used to assess the findings. Furthermore, the studies reviewed in the paper studied a variety of services offered in the insurance industry using artificial intelligence (AI) and XAI applications that include but are not limited to apparent underwriting, risk evaluation, influenced judgment identification, client training, explaining underwriting and claims judgments etc. A review of previous works conducted by the researchers was conducted systematically on the calibration and classification performance. Furthermore, for each study, data such as whether scholars disclosed the occurrence of incomplete information on evaluated findings and risk prediction model features were examined. And, if there was any missing data, the way it was handled was examined. In each article, model assessment metrics such as the AUC-ROC, accuracy, RMSE, and MAE,

confusion matrix were explored for each model.

Typically, academics have used accuracy as a criterion for determining an insurance applicant's risk level. There are various risk prediction models available for policy issuance, and statistical comparisons help to bring them closer together. The literature is definitely influenced by outcome selection and positive biases. This study reveals two new insights into insurance risk assessments using a customer dataset. The first is a combination of a class-balancing approach with cross-validation on the customer dataset, and the other is a combination of class balance with hyperparameter tweaking on the customer dataset, both of which have not received much attention in the literature. In addition, the researchers experiments did not take into consideration the performance of the system while they were assessing the performance of the model. It is suggested for future research to combine the above three techniques to identify the most suitable model for accurate risk assessment while also taking into consideration the system performance metrics in order to evaluate the performance of the model.

Acknowledgments

First and foremost, I am indebted to my mentor and research supervisor, Varadarajan Vijaya Kumar Prof. Dr Vijayakumar Varadarajan is currently an Adjunct Professor at the School of Computer Science and Engineering, University of New South Wales, Sydney, Australia, for guiding me throughout the study. I consider myself extremely fortunate to have him as my supervisor since he not only groomed me in data analytics but also paved the way for further research in this field.

Conflict of interest

The authors declare no conflict of interest.

References

1. Raj A, Joshi P. Changing face of the insurance industry. Available online: <https://www.infosys.com/industries/insurance/white-papers/documents/changing-face-insurance-industry.pdf> (accessed on 5 May 2023).
2. Bi F, Zhang H, Liu N, et al. Machine Learning Practice for Multi-Classification Problem in Life Insurance. Available online: https://www.mendeley.com/catalogue/5fae4282-e0d4-36a5-8c7a-1e660c77767c/?utm_source=desktop&utm_medium=1.19.4&utm_campaign=open_catalog&userDocumentId=%7B06e3337a-a69e-4a89-af0b-ed5295fc1bcd%7D (accessed on 10 Mar 2023).
3. Umamaheswari K, Janakiraman S. Role of Data mining in Insurance Industry. COMPUSOFT, An International Journal of Advanced Computer Technology. 2014, 3(6): 961–966.
4. Bhalla A. Enhancement in Predictive Model for Insurance Underwriting. International Journal of Computer Science & Engineering Technology. 2012, 3(5): 160–165.
5. 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, 9(2): 2213-2220. doi: 10.30534/ijatcse/2020/199922020
6. Gupta H, Patel D, Makade A, et al. Risk Prediction in the Life Insurance Industry Using Federated Learning Approach. 2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON). Published online June 14, 2022. doi: 10.1109/melecon53508.2022.9842869
7. Srivastava R. Identification of Customer Clusters using RFM Model: A Case of Diverse Purchaser Classification. 2002.
8. Bhalla A. Enhancement in Predictive Model for Insurance Underwriting. 2012, 3(5): 160–165.
9. Sivarajah U, Kamal MM, Irani Z, et al. Critical analysis of Big Data challenges and analytical methods. Journal of Business Research. 2017, 70: 263-286. doi: 10.1016/j.jbusres.2016.08.001
10. Shakeel F. Five Benefits of Data Analytics in Insurance Industry. Available online: <https://www.damcogroup.com/blogs/insurance-data-analytics-for-insurers/> (accessed on 22 December 2022).
11. Stonell K. How data analytics is changing the Insurance industry. Available online: <https://www.information-age.com/data-analytics-insurance-industry-123481252/> (accessed on 23 January 2023).
12. Fang K, Jiang Y, Song M. Customer profitability forecasting using Big Data analytics: A case study of the insurance industry. Computers & Industrial Engineering. 2016, 101: 554-564. doi: 10.1016/j.cie.2016.09.011

13. 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*. 2018, 44(1): 1-19. doi: 10.1057/s41288-018-0109-0
14. Boodhun N. A Review of Data Analytical Approaches in the Insurance Industry. *Journal of Applied Technology and Innovation*. 2017, 1(1): 58–73.
15. Zhou L, Chen QY. Customer segmentation, return and risk management: An empirical analysis based on BP neural network. *Journal of Chemical and Pharmaceutical Research*. 2014, 6(6): 698–703.
16. Delafrooz N, Farzanfar E. Determining the Customer Lifetime Value based on the Benefit Clustering in the Insurance Industry. *Indian Journal of Science and Technology*. 2016, 9(1). doi: 10.17485/ijst/2016/v9i1/72307
17. Polyakova M. Risk selection and heterogeneous preferences in health insurance markets with a public option. *Journal of Health Economics*. 2016, 49: 153-168. doi: 10.1016/j.jhealeco.2016.06.012
18. Bănărescu A. Detecting and Preventing Fraud with Data Analytics. *Procedia Economics and Finance*. 2015, 32: 1827-1836. doi: 10.1016/s2212-5671(15)01485-9
19. Lookman Sithic TBH. Survey of Insurance Fraud Detection using Data Mining Techniques. *International Journal of Innovative Technology and Exploring Engineering*. 2013, 2(3): 62–65.
20. Abdul-Rahman S, Arifin NFK, Hanafiah M, et al. 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). doi: 10.14569/ijacsa.2021.0120950
21. Desik PHA, Samarendra B, Soma P, Sundari N. Segmentation-Based Predictive Modeling Approach in Insurance Marketing Strategy. *The IUP Journal of Business Strategy*. 2016, 13(2): 35–45.
22. Hiwase VA, Agrawa AJ. Review On Application of Data Mining in Life Insurance. *International Journal of Engineering & Technology*. 2018, 7(4.5): 159. doi: 10.14419/ijet.v7i4.5.20035
23. Ansari A, Riasi A. Customer Clustering Using a Combination of Fuzzy C-Means and Genetic Algorithms. *International Journal of Business and Management*. 2016, 11(7): 59. doi: 10.5539/ijbm.v11n7p59
24. Européen GCA. Section 1-Executive Summary. Available online: https://www.mendeley.com/catalogue/5bcd0d75-2641-3cbb-8c4a-0ef9bf0b8d62/?utm_source=desktop (accessed on 17 August 2022).
25. Devale AB. Applications of Data Mining Techniques in Life Insurance. *International Journal of Data Mining & Knowledge Management Process*. 2012, 2(4): 31-40. doi: 10.5121/ijdkp.2012.2404
26. Goonetilleke TLO, Caldera HA. Mining Life Insurance Data for Customer Attrition Analysis. *Journal of Industrial and Intelligent Information*. 2013, 1(1): 52-58. doi: 10.12720/jiii.1.1.52-58
27. Sunita M, Patil D, Chavan S. Evaluation of Data Mining applications in Insurance Sector. pp. 44-47, 2017, Available online: http://ici2tm.sinhgad.edu/pcproc/ICI2TM2017_P/data/IC17011.pdf (accessed on 10 August 2022).
28. Gan G, Valdez EA. Data Clustering with Actuarial Applications. *North American Actuarial Journal*. 2019, 24(2): 168-186. doi: 10.1080/10920277.2019.1575242
29. Reddy Gaddam SK. Insurance Risk Assessment Using Predictive Analytics. Available online: <https://www.latentview.com/blog/insurance-risk-assessment-using-predictive-analytics/> (accessed on 26 July 2022).
30. Cappiello A. The Technological Disruption of Insurance Industry: A Review. *International Journal of Business and Social Science*. 2020, 11(1). doi: 10.30845/ijbss.v11n1a1
31. Gopi S, Govindarajula K. Classifying Risk in Life Insurance using Predictive Analytics. Available online: <https://mwsug.org/proceedings/2019/IN/MWSUG-2019-IN-083.pdf> (accessed on 26 July 2022).
32. Joly Y, Burton H, Knoppers BM, et al. Life insurance: genomic stratification and risk classification. *European Journal of Human Genetics*. 2013, 22(5): 575-579. doi: 10.1038/ejhg.2013.228
33. Chan L, Hogaboam L, Cao R. Artificial Intelligence in Insurance. *Applied Artificial Intelligence in Business*. Published online 2022: 189-199. doi: 10.1007/978-3-031-05740-3_12
34. Maier M, Carlotto H, Saperstein S, et al. Improving the Accuracy and Transparency of Underwriting with Artificial Intelligence to Transform the Life-Insurance Industry. *AI Magazine*. 2020, 41(3): 78-93. doi: 10.1609/aimag.v41i3.5320
35. Yum K, Yoo B, Lee J. Application of AI-based Customer Segmentation in the Insurance Industry. *Asia Pacific Journal of Information Systems*. 2022, 32(3): 496-513. doi: 10.14329/apjis.2022.32.3.496
36. Isa UA, Fernando A, Fernando T. Artificial Intelligence for Insurance Industry: A Survey. Published online June 22, 2022. doi: 10.21203/rs.3.rs-1766398/v1
37. Owens E, Sheehan B, Mullins M, et al. Explainable Artificial Intelligence (XAI) in Insurance. *Risks*. 2022, 10(12): 230. doi: 10.3390/risks10120230
38. Sushant K S. A Commentary on the Application of Artificial Intelligence in the Insurance Industry. *Trends in Artificial Intelligence*. 2020, 4(1). doi: 10.36959/643/305
39. Gupta R. Artificial Intelligence (AI) in Insurance: A Futuristic Approach. *Shodh Drishti*. 2020, 11(8): 6–10.
40. Das S, Dey A, Pal A, et al. Applications of Artificial Intelligence in Machine Learning: Review and Prospect. *International Journal of Computer Applications*. 2015, 115(9): 31-41. doi: 10.5120/20182-2402

41. Method R, Dwivedi SK, Mishra A, Kumar Gupta R. Risk Prediction Assessment in Life Insurance Company through Dimensionality. *International Journal of Scientific & Technology Research*. 2020, 9(1): 1528–1532.
42. Bhalla A. Enhancement in Predictive Model for Insurance Underwriting. *International Journal of Computer Science & Engineering Technology*. 2012, 3(5): 160–165.
43. Mustika WF, Murfi H, Widyaningsih Y. Analysis Accuracy of XGBoost Model for Multiclass Classification - A Case Study of Applicant Level Risk Prediction for Life Insurance. 2019 5th International Conference on Science in Information Technology (ICSITech). Published online October 2019. doi: 10.1109/icsitech46713.2019.8987474
44. Rusdah DA, Murfi H. XGBoost in handling missing values for life insurance risk prediction. *SN Applied Sciences*. 2020, 2(8). doi: 10.1007/s42452-020-3128-y
45. Chen TJ. Corporate Reputation and Financial Performance of Life Insurers. *The Geneva Papers on Risk and Insurance - Issues and Practice*. 2016, 41(3): 378-397. doi: 10.1057/gpp.2016.8
46. Hedengren D, Stratmann T. Is there adverse selection in life insurance markets? *Economic Inquiry*. 2015, 54(1): 450-463. doi: 10.1111/ecin.12212
47. Prince AER. Tantamount to Fraud?: Exploring Non-Disclosure of Genetic Information in Life Insurance Applications as Grounds for Policy Rescission Grounds for Policy Rescission. *Health Matrix: Journal of Law-Medicine*. 2016, 26(1).
48. Joudaki H, Rashidian A, Minaei-Bidgoli B, et al. Improving Fraud and Abuse Detection in General Physician Claims: A Data Mining Study. *International Journal of Health Policy and Management*. 2015, 5(3): 165-172. doi: 10.15171/ijhpm.2015.196
49. Boodhun N, Jayabalan M. Risk prediction in life insurance industry using supervised learning algorithms. *Complex & Intelligent Systems*. 2018, 4(2): 145-154. doi: 10.1007/s40747-018-0072-1
50. Biddle R, Liu S, Tilocca P, et al. Automated Underwriting in Life Insurance: Predictions and Optimisation. *Databases Theory and Applications*. Published online 2018: 135-146. doi: 10.1007/978-3-319-92013-9_11