INSURANCE UNDERWRITING AND TECHNOLOGY RELATIONSHIP: A BIBLIOMETRIC ANALYSIS

CAHYONO BUDY SANTOSO¹, HARCO LESLIE HENDRIC SPITS WARNARS², AHMAD NURUL FAJAR³, HARJANTO PRABOWO⁴

¹,²,³,⁴ Computer Science Department, Bina Nusantara University Program – Doctor of Computer Science, Bina Nusantara University, Jakarta, Indonesia

E-mail: ¹cahyono.santoso@binus.ac.id, ²spits.hendric@binus.ac.id, ³afajar@binus.edu, ⁴harprabowo@binus.edu

ABSTRACT

This study aims to determine the direction and objectives of current research on insurance underwriting and its relationship to the technology domain. We used a five-phase bibliometric analysis method. We extracted 4983 documents from the Scopus database for the period 1987 to 2021 and refined the keywords by adding inclusion criteria to generate 256 documents. We analyze the document metadata with features from the Scopus website, export it to a RIS-type data set, and then process it using VosViewer to visually map keywords for further analysis. As a result, we found publication trends related to this topic by time period, most influential publications, main topic trends, and potential topics for future research. The conclusions of research in this domain have increased publication interest from year to year, and the density diagram for the field of information technology is quite dark, which indicates that the level of novelty is still high for further research.

Keywords: Underwriting, Insurance, Bibliometric, Visualisation, Technology, VosViewer

1. INTRODUCTION

The term "insurance" may be defined differently from the point of view of different disciplines, but a more general definition is offered by the American Risk and Insurance Association as "the collection of losses by transferring the risk to an insurance company, which agrees to indemnify the insured person for losses." To provide other monetary benefits in exchange for the risk's occurrence, or to provide services related to the risk" [1]. The main goal of insurance companies is to protect their customers from economic stress in the event of unforeseen events by selling insurance policies and charging premiums. Insurance helps the economic development of a country in several ways, namely as a savings driver, financial intermediary, investment activity driver, financial market stabilizer, risk manager, and agent to allocate capital resources efficiently [2].

Underwriting is defined as an insurance function that is responsible for assessing and classifying the level of risk represented by the insured or proposed group and making decisions regarding the coverage of that risk [3]. Usually, insurance companies rely on age and policy history as tools for assessing risk, but technology has now become widespread. The existence of vast historical data, or Big Data, and new techniques such as machine learning and predictive modeling provide an unprecedented opportunity for the industry to transform the underwriting process. Today, more factors can be considered to support an insurance company's decision to make a risk assessment or underwriting process more personal and accurate [2].

Underwriting skills are learned through several months of training and knowledge exchange by senior underwriters. This task requires the underwriter to be sufficiently analytical, highly organized, and accurate to make a sound decision to approve or reject a risky application. Underwriting is sometimes required to change processes due to shifts in regulatory and compliance standards, investor requirements, and customer demands [4]. Manual underwriting tasks are very paper-based. Underwriters evaluate
scenarios by analyzing large amounts of dynamic information in the application. This can be a source of inconsistency, inaccuracy, and bias [5].

The term "InsurTech" refers to new insurance technology and developments that have an impact on the industry's regulatory practices. This is when the term "big data" comes into play. The way insurers handle vast amounts of data has changed as a result of Big Data. The volume of data that must be processed has expanded fast as a result of technological advancements, and this data might be unstructured, semi-structured, or structured [6]. For underwriting, several insurance companies now employ actuarial formulas. When compared to traditional data processing approaches, Big Data technology helps analyze data and extract significant information from it, which can lead to better strategic decision making and corporate development [7]. Fraud detection, risk assessment, claims analysis, underwriting analysis, marketing analysis, customer profile development, and product creation are all areas where data mining is used in the insurance sector [8].

Marine insurance was the first type of insurance utilized for insurance, making it one of the oldest industries on the world. The procedures utilized in this industry for policy registration and claim settlement have a lot of room for improvement as client data grows [9][10]. Artificial intelligence is the decade's buzzword, and it simply refers to the advancement of machines to replicate human intelligence [11] [12]. Machine Learning (ML) and predictive analytics are two types of artificial intelligence (AI) that help the insurance sector perform better. Almost all of the business needs of insurance businesses can be met with AI. AI improves risk selection and pricing techniques by speeding up underwriting processes [13][14]. Not only that, but customer data can yield a variety of insights, and policy personalization is facilitated by speedier underwriting and analysis [15][16]. Research related to the role of technology in insurance underwriting activities needs to be explored more deeply to obtain the latest research positions.

To fill this gap, the author aims to conduct research to determine the current state of the research position in the underwriting field, especially in the insurance sector, and to seek potential research directions that can bridge this problem. In particular, how the direction and role of information technology can be applied to improve underwriting performance or productivity in the insurance sector. For the method, we use the Bibliometric methodology. The use of the bibliometric method in written publications is applied to quantitative methods [17]. Bibliometric analysis identifies a collection of literature, usually linked as a data set published in a particular study domain [18].

2. RELATED WORK

There are no previous studies with Bibliometric methodology related to the topic of information technology systems in insurance underwriting. There is research related to insurance, especially about insurance technology developments that have an impact on the industry's regulatory practices. There is research related to insurance technology developed. This study is based on a bibliometric analysis of 561 papers retrieved from the Scopus database [19].

Other research explains how writings on Islamic insurance have evolved. From 2014 through 2021, the data was obtained from the Dimensions.ai database. The data was bibliometrically analyzed using the R biblioshiny software. A total of 500 articles from preprinted documents, journals, documents, books, and proceedings were collected. According to the findings, the number of articles on the subject of sharia insurance has increased. Not only in Malaysia, but also in other Muslim countries. The goal of this research is to benefit the development of Islamic insurance firms [20].

Other research provides a review of theoretical contributions and empirical studies on the external and internal mechanisms of corporate governance of insurance companies and their effects on performance and/or risk. Analysis of studies published between 1985 and 2019 via bibliometric tools, This research is able to describe the network of scientific collaborations (co-authors) and the relationships between the most frequently used terms, while also highlighting the groups of scholars and the most significant strands of research. In addition, this paper performs a meta-analysis of about thirty quantitative articles showing the relationship between the quantitative-qualitative characteristics of directors and the performance of insurance companies [21].
This research focuses on asking research questions: what is the direction, mapping, and trend of insurance underwriting technology using visual bibliometric analysis? From a bibliometric review, this study aims to study the visual mapping and research trends in the field of insurance underwriting technology on an international scale in the period 1987–2021. This research will explain the following research questions:

RQ 1: What is the trend of research on insurance underwriting technology based on the number of publications per year?

RQ 2: Which countries’ authors contribute the most to insurance underwriting technology research publications?

RQ 3: Who has contributed the most to insurance underwriting technology research publications?

RQ 4: What are the main topics in Insurance Underwriting Technology research?

RQ 5: What are the potential research topics in Insurance Underwriting and ICT in the future?

RQ 6: What are the important articles by citation size in insurance underwriting technology research?

RQ 7: What is the relationship between Insurance Underwriting and ICT?

3. METHODOLOGY

Pritchard [22] defined the term "bibliometrics" for the first time as "the application of mathematical and statistical methods to books and other communication media". The central theme of the research underlies the use of bibliometrics in the systematic literature analysis of related keywords, consisting of five steps [23] as shown in figure 1.

Figure 1: Keywords Network visualization

This bibliometric is used as a method for understanding global research trends within a field in Scopus publications or WoS publications [24]. First determine keywords, this research was conducted on March 10, 2022 using a combination of insurance underwriting keywords and also information technology searched for in the title, abstract and keyword section with a publication period until 2021 as:

( TITLE-ABS-KEY ( insurance ) AND TITLE-ABS-KEY ( underwriting OR "premium esti*" OR "underwrit*" OR "premium esti*" OR "premium calc*" OR "risk assessment" OR "risk prediction" OR "machine learning" OR "artificial intelligence" OR "big data" OR "internet of things" OR blockchain OR "mobile computing" OR "smartphone*" OR mobile application*" OR "knowledge base" OR "knowledge-base" OR "artificial intelligence" OR "big data" OR "internet of things" OR "mobile computing" OR "mobile application*" OR "knowledge base" OR "knowledge-base" OR "artificial intelligence" OR "big data" OR "internet of things" OR blockchain OR "mobile computing" OR "smartphone*" OR mobile application*" OR "recommendation system" OR "recommend system" OR "distributed ledger technology" OR blockchain )

The result from this search were 4983 documents. The second is database selection. In this study we used the Scopus database as a data source. According to Baas (2020) [25] "Scopus is one of the largest curated databases of abstracts and citations, with broad global and regional coverage of scientific journals, conference proceedings, and books, while ensuring only the highest quality data is indexed through content strict, selection and re-evaluation by an independent Content Selection and Advisory Board".

Third, namely the adjustment and refinement of the research criteria; From the results of the first search, the researchers added the following inclusion criteria: The subject is computer science, the types of documents are articles and conferences with English-language
Research on the insurance underwriting domain shows an increasing trend in the number of publications. From 1987 to 2021, a total of 256 articles have been produced. Publications seem to be increasing, in 2019 as many as 33 articles were produced.

4. RESULT AND DISCUSSION

4.1. Insurance Underwriting and Technology research trend based on number of publications per year (RQ1)

Figure 2: Underwriting Insurance and Technology Research Trend

4.2. Most productive country in Underwriting Insurance and technology research domain (RQ2)

Figure 3 explains that the countries that contributed to the Underwriting Insurance domain research were the United States with 89 publications, followed by China with 34 publications and India with 21 publications. Indonesia contributed three publications.

Figure 3: Productive country in insurance underwriting technology research

4.3. Most contributor Author in CBT and technology research domain (RQ3)

Figure 4 describes the top authors who contributed to the Insurance Underwriting research domain. Where Bonissone and Mukhopadhyay were the biggest contributors with 3 documents.
The next stage, we analyze the results with cluster analysis, we perform several steps. Frequently appearing keywords will be visualized for bibliometric analysis. This will make it easier for researchers to analyze topics in the insurance underwriting domain that have been studied for years (1987-2021). Cluster analysis is also applied to interpret the structure of intelligence with the VOSviewer Clustering method, which is a method that has the advantage of organizing objects into groups based on similarities or dissimilarities. The same cluster will contain keywords that are highly correlated with each other. The results of the research data in the previous step will be visualized in three different diagrams, namely (1) data network visualization; (2) overlay visualization; and (3) density visualization.

In Figure 5 about the Keywords Network visualization, it can be seen that the frequency of occurrence of keywords will reflect the size of the keyword node, where the more frequent the keyword, the greater the visualization of the node size. The relationship between the two keywords can be seen from the thickness of the connecting line in the two keywords, where the thicker the line, the closer the relationship [27]. From the results of the process into keywords in the previous stage, 827 keywords were obtained, the minimum occurrence of this keyword was 2 and produced 82 keywords that appeared according to the criteria, these keywords were then mapped to 10 cluster themes. We analyzed the keywords most and less. To investigate further as shown in table 1. For example, in Figure 5, the Machine Learning node has the largest node size visualization, because the keyword frequency is most often compared to other nodes. The machine learning node forms a network with other nodes in the green node cluster. The data structure of keywords in the green node cluster can be seen in table 1.
### Table 1: Cluster Analyze Result

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Most Frequent</th>
<th>Less Frequent</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red Cluster</td>
<td>Risk management (11 Cloud computing(4), cyber insurance(4))</td>
<td>Vulnerability(2)</td>
<td>Cloud computing(4), cyber insurance(4), cybersecurity(3), decision support system(3), hipaa(3), risk management(4), risk mitigation(2), security(3), stochastic optimization(2), strategic planning(2), threat(2), vulnerability(2)</td>
</tr>
<tr>
<td>Green Cluster</td>
<td>machine learning(19), data mining(8)</td>
<td>credit risk(2), data analysis(2)</td>
<td>Chatbot(2), classification(4), correlation(2), credit risk(2), data analysis(2), data mining(8), expert system(2), fraud detection(2), machine learning(19), prescription fraud(2), risk function</td>
</tr>
<tr>
<td>Blue Cluster</td>
<td>risk(8), logistic regression(6)</td>
<td>survival analysis(2)</td>
<td>Artificial neural networks(2), decision making(2), expert systems(2), fuzzy logic(2), internet of things(2), logistic regression(6), pricing(2), probability(3), risk(8), survival analysis(2)</td>
</tr>
<tr>
<td>Yellow Cluster</td>
<td>risk assessment(14), telematics(5)</td>
<td>驾驶行为(2), pay how you drive(2)</td>
<td>Deep learning(4), discrimination(2), driver behaviour(2), driving behaviour(2), pay how you drive(2), risk assessment(14), statistical analysis(2), telematics(5), usage-base insurance(4)</td>
</tr>
<tr>
<td>Purple Cluster</td>
<td>artificial intelligent(4)</td>
<td>automobile insurance(2)</td>
<td>Ahp(2), artificial intelligent(4), automobile insurance(2), hospital readmission(2), modelling(2), neural networks(3), risk prediction(3)</td>
</tr>
<tr>
<td>Light blue Cluster</td>
<td>insurance(22), Blockchain(5),</td>
<td>pay as you drive(2)</td>
<td>Blockchain(5), insurance(22), internet of thing(2), pay as you drive(2), privacy(3), radio frequency identification(2), ubiquitous computing(2)</td>
</tr>
<tr>
<td>Orange Cluster</td>
<td>big data(7), clustering(2)</td>
<td></td>
<td>Artificial neural network(2), big data(7), clustering(2), data science(3), natural language process(3), xgboost(2)</td>
</tr>
<tr>
<td>Brown Cluster</td>
<td>Cyber insurance(5)</td>
<td>security breach(2)</td>
<td>Cyber insurance(5), e-commerce(2), information security(2), is security(2), logit and probit model(2), security breach(2)</td>
</tr>
<tr>
<td>Light Purple Cluster</td>
<td>risk analysis(4)</td>
<td>decision making(2)</td>
<td>Decision support(2), decision making(2), insurance industry(4), risk analysis(4), risk measures(2), underwriting(3)</td>
</tr>
<tr>
<td>light red Cluster</td>
<td>Driving behaviour profiling(2), predictive model(2)</td>
<td></td>
<td>Driving behaviour profiling(2), intelligent transportation system(2), prediction models(2), predictive model(2)</td>
</tr>
</tbody>
</table>

In this paper, we only show 2 dominant keywords, for keywords that appear most often and rarely appear. The results of this table can be concluded which research still needs to be developed. In Figure 6, the overlay visualization is chosen as a tool to verify the latest trends in the academic field, this tool will classify items based on time units. Items will have a different color based on the year of issue [28]. Terms in the more recently emerging insurance underwriting research domain (average year of publication is 2020) will be brighter, in this case in green to yellow.
In Figure 7 regarding density visualization, the occurrence of keywords in numbers will be represented by color. A lighter color (yellow) would represent a lot of research that has been done, while a darker color would indicate the opposite. Thus, we can see that keywords with darker colors can be interpreted that the amount of research is still small, so the chances for his novelty contribution can be higher.
4.4. Main topics Research in the field of Insurance Underwriting Technology (RQ 4).

This study resulted in 10 clusters that describe the research themes studied in the Insurance Underwriting Technology domain. In each theme, there are dominant keywords that have a high frequency of occurrence, where this will represent a topic that is interconnected in one theme. For example, in the first cluster, the keyword that appears the most is risk management (11). Meanwhile, in the second cluster, the keywords that often appear are machine learning (19). The information in Table 4.1 related to the keywords that appear the most in each cluster is answering RQ 5 where this data is the main topic in research in the field of Insurance Underwriting technology.

4.5. Potential research topics in future Insurance Underwriting Technology fields (RQ 5)

Some keywords that rarely appear in the Cluster Table analysis, can be seen as potential research topics in the future. In this case, keywords that have a low frequency of occurrence in each cluster theme, for example in the green cluster are Chatbot (2), expert system (2) and the light purple cluster are decision making (2). A chatbot is a computer program that uses text or voice messages to communicate with humans. Artificial intelligence and natural language processing have been added to chatbots, making them intelligent computer applications capable of answering questions presented by humans [29]. While artificial intelligence includes the expert system. An expert system is a computer software that can mimic human behavior in a well-defined field of knowledge and act as a decision-maker[30]. Research using this technology has been carried out by Santoso [31] and Doultani [32], but it has not been explored in more depth and detail in its implementation and testing.

4.6. Significant articles by Citation size in Insurance Underwriting Technology research (RQ 6)

This study shows that Scopus journals have a significant effect on metrics that refer to citations. In Table 4, refers to the five articles that have the most citations in the field of Insurance Underwriting Technology. The importance of articles in this domain is indicated by the highest number of citations.

Where the article written by Bonissone entitled 'Evolutionary algorithms + domain knowledge = real-world evolutionary computation' has been cited as many as 126 pieces and is the most cited in this field.

Table 2: Top Five Significant Cited Articles

<table>
<thead>
<tr>
<th>Papers</th>
<th>Authors</th>
<th>Year</th>
<th>Cite</th>
</tr>
</thead>
<tbody>
<tr>
<td>[34]</td>
<td>Conforti, R</td>
<td>2015</td>
<td>97</td>
</tr>
<tr>
<td>[35]</td>
<td>Radanlie</td>
<td>2018</td>
<td>63</td>
</tr>
<tr>
<td>[36]</td>
<td>Baecke</td>
<td>2017</td>
<td>57</td>
</tr>
<tr>
<td>[37]</td>
<td>Carfora</td>
<td>2019</td>
<td>32</td>
</tr>
</tbody>
</table>

4.7. Relationship between Insurance Underwriting domain and Technology (RQ 7)

In the relationship between Insurance Underwriting and information technology, several keywords that have potential are Cloud computing(4), cyber insurance(4), Chatbot(2), data mining(8), expert system(2), machine learning(19), Artificial neural networks(2), expert systems(2), internet of things(2), Deep learning(4), telematics(5), artificial intelligence(4), neural networks(3), Blockchain(5), Artificial neural network(2), big data(7), data science(3) and natural language process(3).

In the visualization framework of Figure 7, it can be seen that the keywords related to technology, namely Chatbot, expert system, machine learning, and big data have a bright background color so that it can be interpreted that there has not been much research in this domain or in other words, this domain has a sufficient level of novelty tall.

5. CONCLUSION

This study uses a bibliometric approach to verify journal publications in the Scopus database.
Publications related to insurance and technology underwriting domains are increasing every year. This research has been carried out with an article publication time span of 30 years. The results obtained from this study were to collect 4983 publications related to Insurance Underwriting Technology from the Scopus database and after being refined based on predetermined inclusion criteria, it was reduced to 262 articles.

The results achieved show that research trends in the Insurance Underwriting domain are increasing, especially in 2019, where 33 documents were published. The most productive country in the insurance and technology research domain is the United States with the publication of 89 documents. The most contributing individual authors in this domain are Bonissone and Mukhopadhyay with the publication of 3 documents. The main topics in this domain can be divided into 10 clusters. The research potential in this domain can be seen from the lack of keywords in the cluster, especially those related to IT and technology. Most cited article Evolutionary algorithms + domain knowledge = real-world evolutionary computation' with 126 citations.

In addition, the data were analyzed using cluster analysis to reveal the main research topics in the insurance underwriting field based on the similarity or dissimilarity of keyword relationships in clusters that form a particular theme, otherwise a lower keyword frequency would indicate a potential topic for further research as a contribution from this research. The research potential found in this study can be used as a reference for research development, especially regarding technology in Underwriting.

Another contribution of this research is to fill the gap of bibliometric analysis that has not been carried out to analyze the literature of the Insurance Underwriting Technology domain and visualize the relationship between keywords based on their similarities so as to form a particular research theme since 1987.

The findings of this study are that there are still opportunities for novelty research in the use of Chatbot technology, expert systems, machine learning, and big data.

The limitation of this research is that the data source is taken only from the Scopus database due to limited access from the researcher, it is highly recommended for further research to expand the database from other sources such as Springer, IEE Plore or Web of Science (WOS).

REFERENCES


