ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 13 (2023) No. 4, pp. 277-298
DOI:10.33168/JSMS.2023.0417

Smart Insurance Model for Sea Cargo Business

Cahyono Budy Santoso, Harco Leslie Hendric Spits Warnars, Ahmad Nurul Fajar, Harjanto Prabowo

Computer Science Department, Bina Nusantara University Program – Doctor of Computer Science, Bina Nusantara University, Jakarta, Indonesia

cahyono.santoso@binus.ac.id; spits.hendric@binus.ac.id; afajar@binus.edu; harprabowo@binus.edu

Abstract. Additional factors may be considered to support an insurance company's decision to make a risk assessment or underwriting process more personal and accurate. This task requires the underwriter to be sufficiently analytical, well-organized, and accurate to make an informed decision to approve or deny a risky application. For this, it is necessary to have technology that can support the underwriting process or risk assessment. This research focuses on marine cargo insurance, which is related to underwriting technology features in determining premium rates and also related to the process of increasing engagement with customers, in this case in the customer communication model and customer recommendations related to insurance products. This study aims to increase the speed and accuracy of the underwriting process and improve customer service. Utilizing the Design Science Research approach, the model was developed in five stages. The findings of this research include a conceptual model and prototype for a smart insurance model for the sea cargo business. Based on the test results for the premium rate prediction feature using the linear regression model with an R2 value of 0.993 and an MAE of 0.0002727, As for the product recommendation features, the testing technique uses a lift value with a value above 1. As for testing the chatbot feature using the black box technique, the results varied according to the chat scenario. So it can be concluded that this model can meet the needs of the user.

Keywords: Underwriting, Technology, Marine Cargo, Insurance

1. Introduction

Insurance is a non-bank financial institution that plays an important role in protecting a business activity. One type of insurance business is general insurance which functions to bear the risk of loss from an insured object submitted by the insured by obtaining a reward in the form of a premium. According to Article 41 of Law No. 17 of 2008 on Shipping, shipping companies are obliged to ensure their responsibilities for the safety of passengers. Taking out sea freight insurance is also very important in the event of loss or damage to goods in transit and when accepting guarantees for transport delays and damage to third parties. Sea freight insurance is therefore not an option for companies that ship goods or products via export and import channels, but a statutory obligation. The principle applied in marine cargo insurance is the principle of indemnity, which is an indemnity agreement that is included in the insurance contract. Indonesia as a maritime country provides a great opportunity for insurance companies to provide marine cargo business line insurance services. Among the many products owned by general insurance, marine cargo insurance is one type of insurance product that guarantees losses due to damage or loss of transportation of goods while traveling by sea.

With the nature of the marine cargo insurance business, underwriters need a system to help determine the rate value appropriately in accordance with the risk category of the object of coverage. This rate determination is still a free market competition, and the determination of the rate value has not been definitively regulated by OJK (SE OJK, 2017). To determine the marine cargo insurance premium rate, you still rely on the underwriter's ability, knowledge, and experience from previous cases. The insured object is classified into one of a number of risk categories or rate classes (Yan & Bonissone, 2006) that will affect the premium rate. This process is also called the underwriting process, which is a complex decision-making task performed by a highly trained underwriter (Yan & Bonissone, 2006).

Some previous studies related to technology in the field of risk assessment or underwriting are as follows. (Joram, 2017) proposed a prototype application that can perform the underwriting process or risk assessment in life insurance. The application output can produce risk acceptance or rejection decisions and also determine the level of risk, so that it can be used as a basis for determining premium rate decisions. The algorithm used is a rule-based system based on the knowledge of expert underwriters. Another study (Dubey, 2018) proposed a decision-making system in underwriting that can perform an automatic email process containing insurance plans and recommendations. These results are based on historical data from previous emails. The algorithm used uses machine learning classification techniques by selecting the best model among Naive Bayes, Support Vector Machines, and K-Nearest Neighbors. (Fidelia, 2019) conducted research on underwriting in the field of motor insurance. This research aims to identify the classification of risk sources in motorized vehicles, identify variables for determining motor vehicle premiums, and establish machine learning models to estimate motor vehicle insurance premiums. The variables obtained in this research are related to experience, age, education, and history of the policyholder. The resulting premium rate determination model can be used as a reference for insurance companies when charging premiums dynamically. The algorithm used is to use the best model of machine learning decision trees and regression techniques. (Lephoto, 2014) conducted research for a decision-making system that supports medical underwriting applications. The output of this research is the architecture and implementation of the Medical Underwriting Program (MUP) system. The algorithm used is the Rule Base System.

Based on previous studies, it can be said that the database used is an expert knowledge base and historical database for producing risk assessment decisions or premium rate results. Rule-based system algorithms are used for expert-based data (Lephoto, 2014; Joram, 2017), while machine learning algorithms are used for data based on historical data (Dubey, 2018; Fidelia, 2019). To be able to further improve accuracy and reduce human intervention in making premium rate decisions, it requires a complete database of both expert knowledge and historical data experience in underwriting. For previous research related to the recommender system, see research (Ahmad, 2022; Haw et al., 2022; Chew et al., 2022).

One very important factor in increasing business value is customer involvement (Davenport, 2018). This research will discuss the right technology model to be able to communicate to customers in order to get the results of risk assessment and consultation in the process of inputting insurance objects. And also, customers get information on other insurance product offers that have been used by other customers. Based on the above, the researcher optimizes the prediction to determine the premium value using the best machine learning model (linear regression, support vector regression, random forest, and decision tree) based on 1 years of historical data from the marine cargo insurance premium rate dataset. Furthermore, the researcher will design an interactive communication model with customers related to the results of their risk assessment and also design system recommendations to get insurance product recommendations to customers.

The existing research and research focus are described below. The scope of research related to business value refers to research (Davenport, 2018).

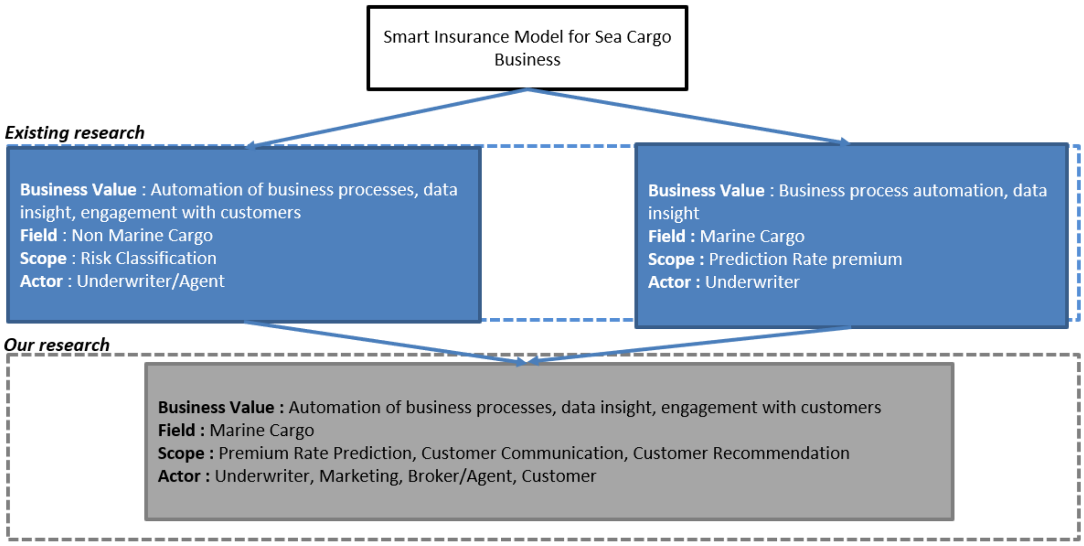

Fig. 1: Research Gap

The marine cargo insurance business is an international business that requires fast and precise product information services. Insurance product needs cannot be quickly obtained according to customer needs (Cebulsky, 2017). The underwriting productivity of the marine cargo business is currently still relatively lacking, because the time allocation is expended routine handling and similar issues. Risk assessment is still manual, so the decision-making process is less than optimal (Joram, 2017).

From the description of the problem formulation above, the Research Question is obtained as follows:

RQ 1 : What features are needed to build a smart insurance system model for marine cargo?

RQ 2 : How to build a smart insurance system model for marine cargo?

RQ 3 : How to test the smart insurance system model in the marine cargo business?

While the objectives of this research are

1. Obtain technology requirements and features to build the model. The main requirements needed are to be able to replace the role of the underwriter and be able to increase customer engagement with the role of Artificial Intelligence.
2. Modeling an intelligent insurance system in the sea cargo business based on literature review and expert interviews using Artificial Intelligence technology. The model is to get the best model of

premium rate prediction, customer communication, and recommendations with customers.

3. Get the best model testing results using Mean Square Error (MSE), R Square, and Mean Absolute Error (MAE) techniques. While the tool to test by making a prototype. This prototype is built by integrating between models using web technology. The prototype will be tested using white box and black box testing.

2. Literature Review

At this stage, the author conducts research using the Kitchenham method to obtain technological features for underwriting. The research source is scientific publication documents that have been indexed by Scopus from 2011 to 2021, and the research is carried out by applying the Literature Review study with the Kitchenham method to conduct a literature review. The result obtained is to determine what technological features are used and what objects are investigated.

The research database was indexed by Scopus and searched by title, keywords, and abstract. The background for determining the underwriting keyword is that it is the main activity in insurance companies. Likewise with insurance and premiums. For the research period, the last 10 years were selected. While the source of research data is based on journals and conferences. For the research subject, Computer Science, or Comp, was chosen because the scope of the research is related to the trend of computer technology. For the language in this research, English was chosen to make it easier to understand. The search keywords are presented as below:

TITLE-ABS-KEY ((insurance DAN premium) OR (underwriting)) AND (PUBYEAR > 2010) AND (PUBYEAR < 2022) DAN (LIMIT-TO (SRCTYPE , "j") OR LIMIT-TO (SRCTYPE , "p")) AND (LIMIT-TO (SUBJAREA , "COMP")) AND (LIMIT-TO (LANGUAGE , "English"))

Since 2011, 356 papers involving underwriting technology have been identified using abstracts, titles and keywords in this literature review. However, adding inclusion and exclusion filters resulted in 50 documents and after matching the results with keywords and research questions, researchers were able to find 24 studies that met our criteria.

Tabel 1: Technology Features in *Underwriting*

Technology	Description	Purpose	Object	Reference
Artificial Intelligent	involves the process of analyzing (big) data (using machine learning methods, for example) and generating automated decisions based on the data.	Assist Underwriting in risk assessment and determine premium rates based on data and knowledge.	Loan, Insurance (Medical, Vehicle, Credit, , Flight Delay, Life, Marine Cargo, Health, Travel)	(Dubey et al., 2018),(Mburu & Pamba,2019),(Doultani et al., 2021) , (Sachan et al., 2020),(Lephoto & Kogeda, 2014),(Lukas et al.,2019) (Maier et al.,2020),(Santoso et al.,2021),(Müller & Te, 2017)
Internet of things	Each connected element broadcasts and receives data through sensors.	Can dynamically determine premium based onIoT sensor data	Insurance (Vehicle, Life, Fire)	(Wengnoon & Limpiyakorn, 2014),(Lertpunyavuttikul et al.,2017),(Ramnathan et al.,2020)
Blockchain	A digital data storage system containing records linked through cryptography	Decentralized data storage that records transaction details in a secure, transparent, and efficient system	Vehicle Insurance	(Wan et al.,2018),(Singh et al., 2019)
Geographic Information	Computer-based risk mapping information	Create a premium rate map	Storm, Flood Insurance	(Lee, 2016)

Technology	Description	Purpose	Object	Reference
System	system			
Mobile devices with apps	*Smartphones* and tablets, along with their apps, have largely replaced desktop computers.	Can dynamically determine risk and premium based on smartphone app data	Vehicle Insurance	(Chakravarty et al.,2013),(Handel et al., 2014)
Global Positioning System	satellite-based navigation system	Analyzing driving risk patterns as a reference in determining risks and premiums.	Vehicle Insurance	(Ayuso et al.,2016)

Based on the conclusions of the authors in this study, the technological features used in the underwriting process are quite numerous and varied. These technologies include artificial intelligence, the Internet of Things, blockchain, geographic information systems, mobile devices with apps, and the Global Positioning System. Motor insurance is the subject of the most research related to technology utilization. In the future, it is expected that further research will be needed on objects other than motor insurance, for example, in the field of marine cargo. The recommendation system feature can also be used to provide recommendations for alternative insurance products. This is a research opportunity that can be further developed.

Henceforth, the author conducts research with keywords related to underwriting. From the table, it can be concluded that research on underwriting marine cargo business insurance has not been explored in depth. Other research related to underwriting is in the fields of life insurance (Biddle et all,2018; Mustika et al., 2019), motor vehicle insurance (Fidelia, 2021), medical insurance (Lephoto, 2014; Dubey, 2018), loan applications (Sachan, 2020), and aircraft delays (Lukas et al., 2019).

The focus of this research is not only related to underwriting, namely in determining the marine cargo insurance premium rate, but also in terms of the process of increasing engagement with customers, in this case in the customer communication model and customer recommendations related to insurance products.

3. Methodology

The research framework is a conceptual model that explains the stages of the flow of thinking so that a study can be carried out on the basis of a framework (variable or theory) that has been prepared (Kumar, 2014). The research framework is described in the flowchart in Figure 2. Broadly speaking, this research activity will be divided into 5 main stages based on the Design Science Research method (Lawrence et al., 2010; et.al 2007), namely the feature formulation process, model design process, model implementation process, model testing process, and conclusion.

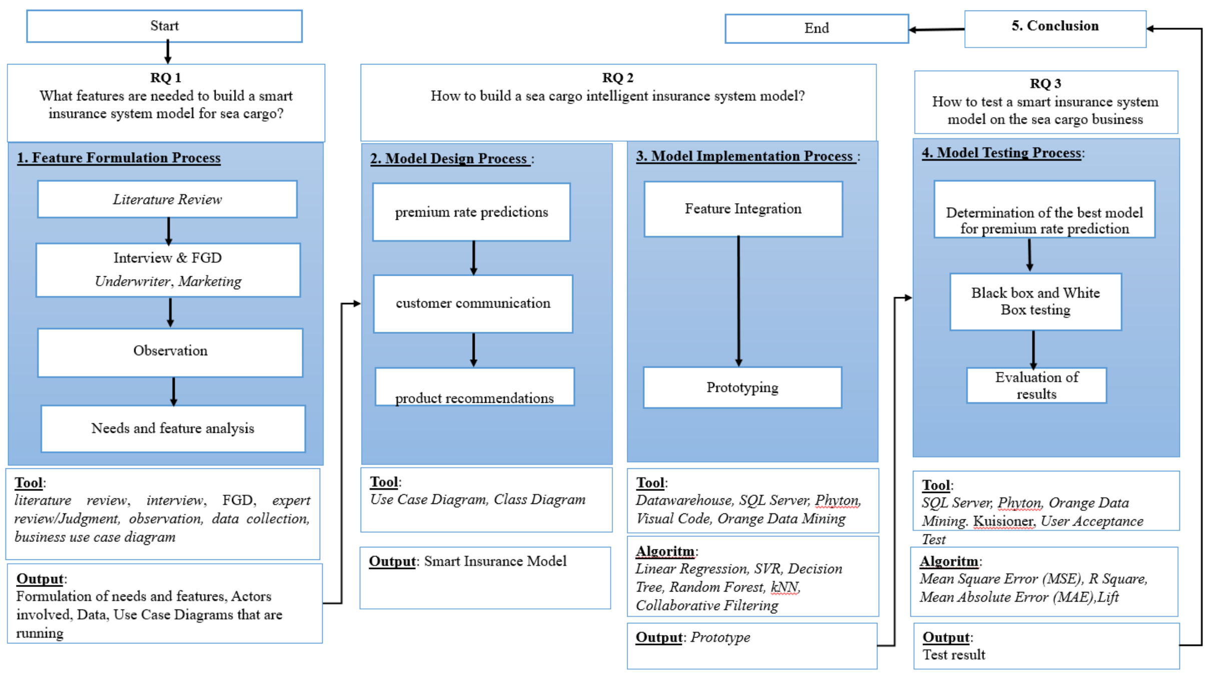

Fig.2: Research thinking framework

4. Results.

This research will be carried out in 4 stages: the feature formulation process, the model design process, the model implementation process, and the model testing process. The details of the stages are described below.

4.1. Feature Formulation Process

4.1.1. Literatur Review

For the literature review stage, it has been explained in the previous section

4.1.2. Interview

To obtain qualitative opinions or feedback from experts in the insurance industry, researchers use the qualitative research method using interviews. Through this interview, researchers will get enlightenment and feedback from experts who have experience in the insurance industry business process. The questions asked in the interview are outlined as follows, namely what problems are faced in the field of cargo underwriting, the parties involved in the cargo underwriting process, what factors affect the calculation of cargo underwriting, the reason for the need to automate cargo underwriting, and the technological features needed for underwriting automation or future expectations.

The selection of selected experts who will be resource persons for this research is based on the experience of experts who have long experience in the field of insurance, especially about underwriting and risk assessment. The expert is also familiar with related technologies. The criteria or standards for research resource persons are as in the table below.

Tabel 2: criteria for research resource persons

No	Description	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5
1	Still actively working in a national scale min insurance company	Yes	Yes	Yes	Yes	Yes
2	Understand business processes in insurance companies	Yes	Yes	Yes	Yes	Yes
3	Have a minimum position of manager	Yes	Yes	Yes	Yes	Yes
4	At least 5 years of experience in an insurance company.	Yes	Yes	Yes	Yes	Yes

At this stage, direct interviews were conducted with insurance company directors, marketing, underwriters, and data science experts at insurance companies in Jakarta. Interviews were conducted from March 5, 2022 to December 3, 2022 using zoom media.

4.1.3. Forum Group Discussion

Through this FGD, researchers will get enlightenment and input from experts who have experience in the insurance industry business process. Based on the standard of research sources that have been set previously, researchers get sources with a summary as shown in the table below. The FGD was conducted on December 3, 2022. For a summary of the curriculum vitae (CV) of the research sources, see table 3.

Table 3: Summary of CV of research resource persons

No.	Description	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5
1	Position	Director	Yes	Yes	Yes	Yes
2	Insurance experience	>10 years	>10 years	>10 years	>10 years	>10 years
3	Field	Underwriting	Underwriting	Underwriting	Marketing, Underwriting	Risk Analyst, Data Science
4	Degree	S1 ITS	S2 ITB	S2 Trisakti	S2 Undip	S2 ITB

4.1.4. Observation

Research began on September 6, 2021 using personal funds. The scope of this research observation is related to the technology model used to improve insurance performance, especially those related to underwriting. Based on these observations, the parties related to the research were found, namely underwriting, marketing, brokers, and customers. The media used for observation uses direct personal interview methods, telephone, email, and using zoom. For the data needed are such as underwriter work papers, reference data, running business flow, and expectations from each user.

4.1.5. Needs and Features Analysis

The course of activities on the system running in the insurance company begins with the process of registering the insured or broker as a potential customer in the company. Marketing will verify the integrity of the prospect's data and if complete, Marketing will onboard and register the prospect. Customers who have registered can go directly to the next stage. For the next step, the broker or client receives an insurance object form. This form contains detailed information about the object to be insured. For this research, the object data refers to business object data for ocean freight insurance. During the form filling process, brokers or clients can consult marketing to complete the form filling process. Once completed, the form will be sent to the marketing department. Marketing performs the process of checking and validating the completeness of the input data. If a form is not fully completed and the data needs to be confirmed, the form will be returned to the agent/principal for completion. When the form is complete, Marketing passes it on to the underwriter. Based on the form data and damage records for the object, the insurer provides risk assessment results for the object. If accepted, the premium rate is determined. In the meantime, if the result is rejected, a detailed justification for the rejection of the object of insurance is given. In addition, the customer receives further product recommendations.

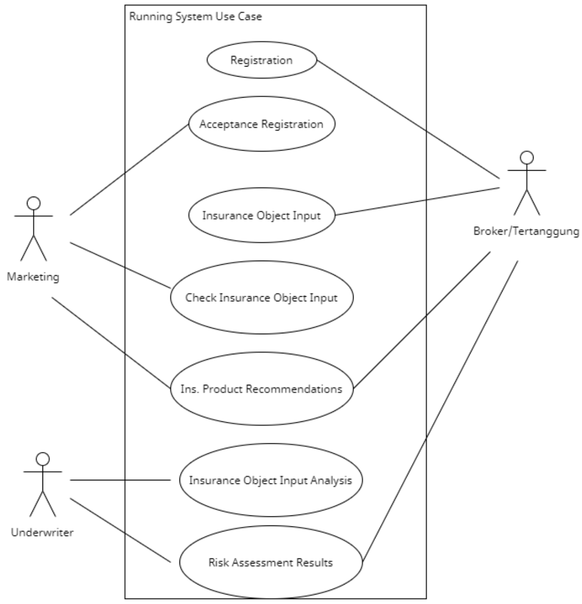

Fig.3: Use Case of the Running System

Based on the current system activities, the system requirements can be identified, namely: customer/broker registration process, process to fill out the insurance object entry form, process to create risk assessment results and process to make insurance product recommendations.

Based on the identification of these needs, the features and technology that will be used can be formulated, which are as follows:

- Functions of the online communication service with customers/brokers for customer registration and filling in input forms for insurance objects. The technology used can use chatbot and web.
- Underwriting automation feature for the need to automate underwriting risk assessment results. For technology used in Machine Learning & Rule Base System.
- Automatic recommendation feature for insurance product recommendation needs. For the technology used, a recommendation system can be used.

4.1.6. Requirements and Features Analysis

This section discusses the results of the research which is the purpose of Research Question 2 (RQ 2) to model an intelligent insurance system in the sea cargo business based on literature review and expert interviews using Artificial Intelligence technology. The model is to get the best model in determining premium rate predictions, customer communication, and product recommendations to customers.

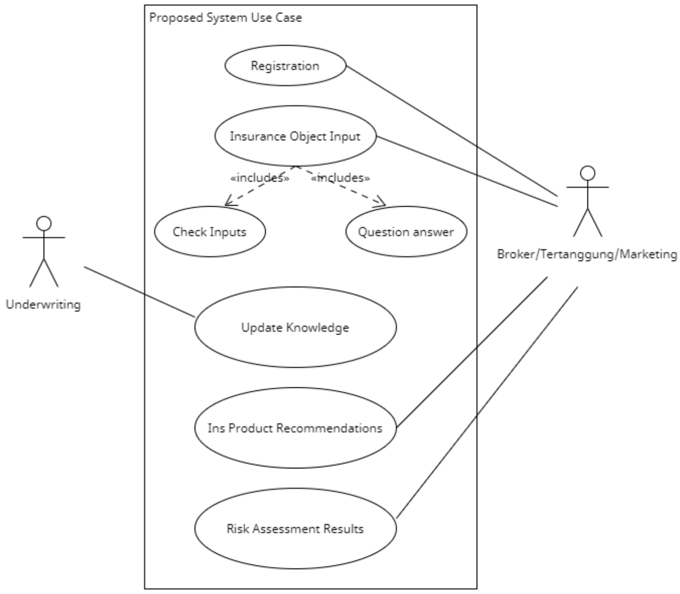

Fig.4: Use Case of the proposed System

The flow of activities in the proposed system begins at the stage of the insured/broker using a web-based application media to carry out the registration process and object insurance input. For things that need to be asked, you can use the chatbot media which is one of the features in the application. The application will check the completeness of the data entered. Clients/brokers can actively consult during the registration process and input of insurance objects using the application. This online communication can be done at any time. After the insurance object is inputted, the system will provide an automatic risk assessment in calculating the premium rate. In addition, the application can also provide recommendations for other insurance products automatically based on the history of purchase transactions by other clients who buy the same product.

4.1.6.1. Class Diagram

Based on the working paper and the data obtained, it is proposed to design the class diagram as below. User table has Order table and Chatbot table. The Order transaction table is formed from the CommodityType, RangeValue, ModeTransport, coverage, and MapPortFromPortTo reference tables. While the MapPortFromPortTo table refers to the Port table. The chatbot table has a ChatbotSource. For chatbot knowledge in the Chatbotsource table which contains FAQ (Frequenty Ask Question) data, namely data that is often asked during the determination process related to premium rates.

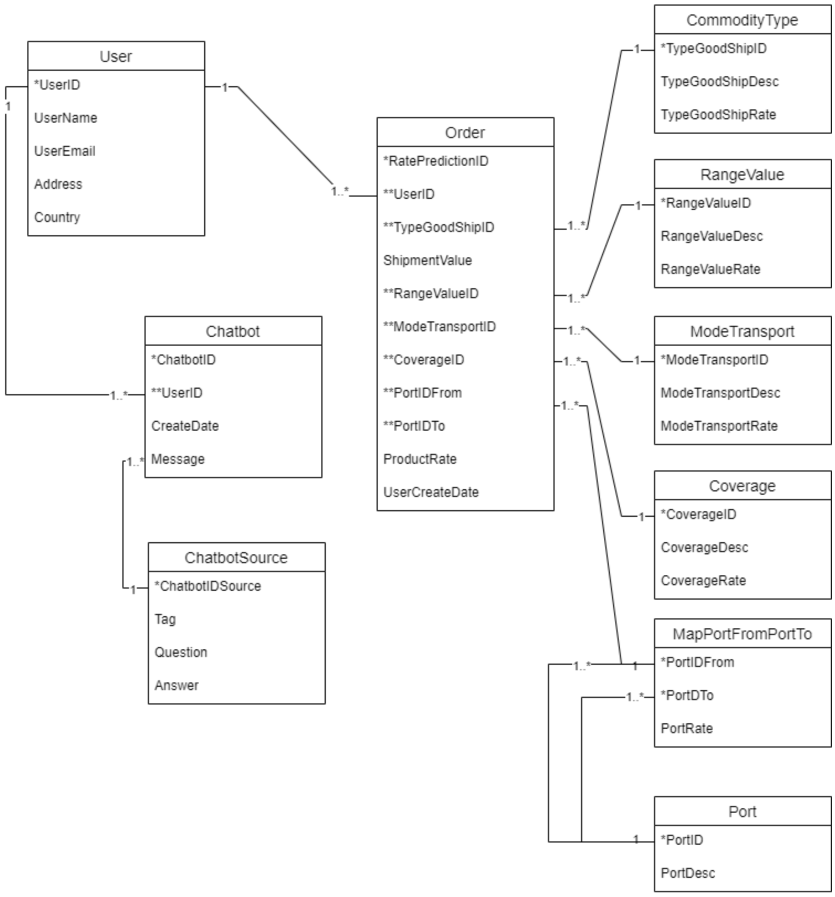

Fig.5: Class Diagram Model

4.1.6.2. Activity Diagram

For an explanation of the activity diagram image as below is as follows: The user logs in to the web application. If successful, it will enter the main page. Users can fill in the insurance object input parameters. If there are things that are asked, they can ask by clicking the chatbot button. On the chatbot page, users can perform the desired question and answer process. If there is a question that the chatbot does not know, the default answer will appear. When finished, the user can press the calculate button to calculate the premium rate. Click the product recommendation button, if the user wants to know other insurance product recommendations.

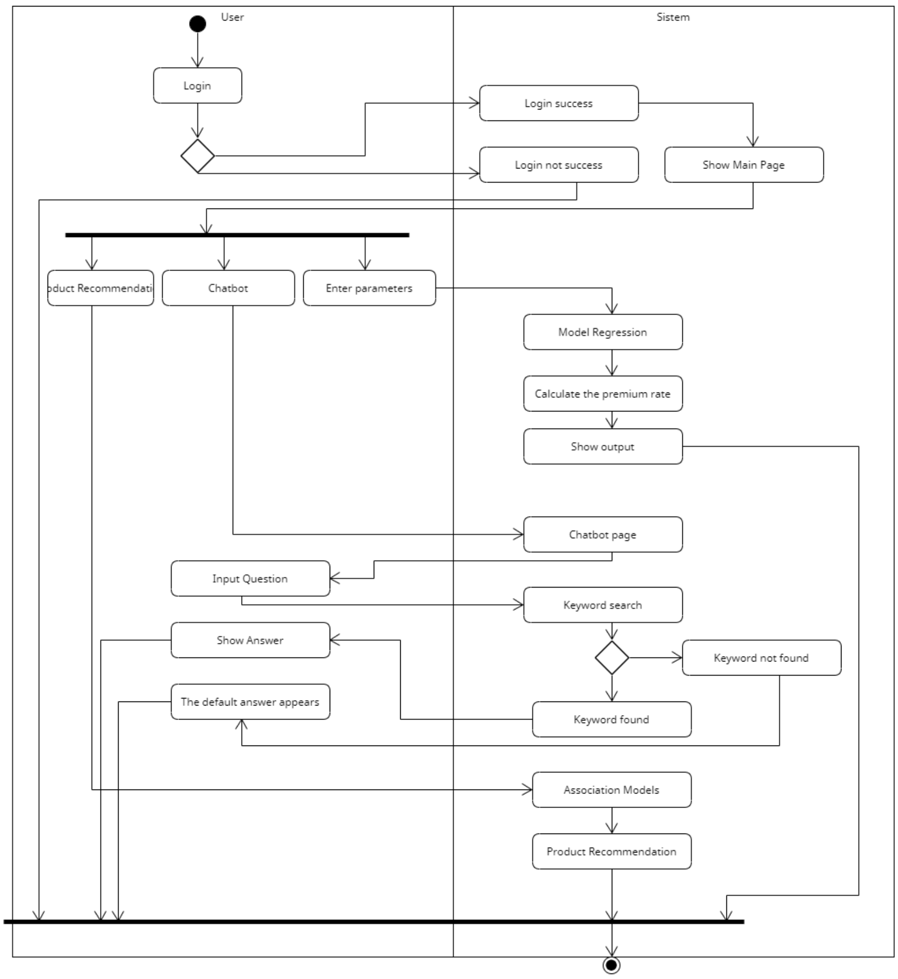

Fig. 6: Activity Diagram

4.2. Model Implementation Process

This section discusses the results of the research which is the purpose of Research Question 2 (RQ2), namely implementing the model by integrating existing features into a prototype. This prototype is built using web technology.

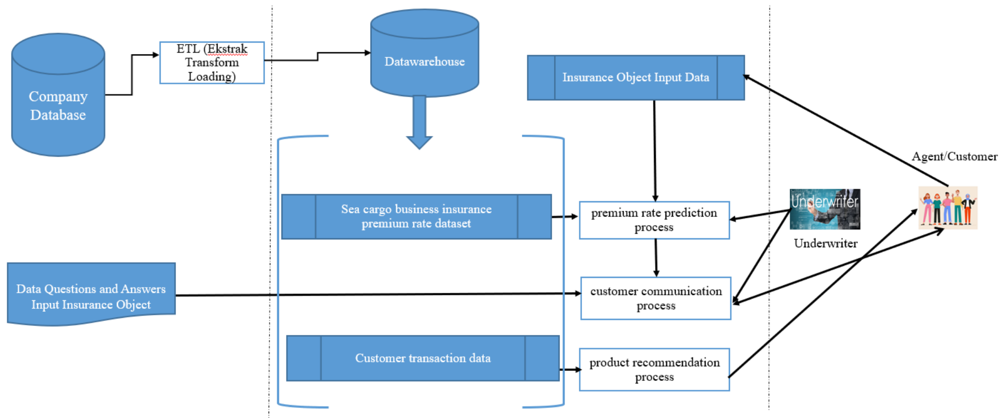

Fig.7: Proposed Smart Insurance System Model (source: researcher)

4.2.1 Premium Rate Prediction

This research uses primary data taken from the database of one of the Insurance Companies in Indonesia in 2021 with a total of 2216 data. The first stage is data pre-processing which includes theprocess of feature selection, data cleaning, transformation and data discretization. The next stage is feature selection which aims to select features that affect the determination of the rate value of the insurance company database based on expert judgment. The results of the initial feature selection by experts in marine cargo insurance consist of 12 features with details of 11 influencing features and 1classification feature (insurance rate value). These features are:

1. Obj_id : Object Id.
2. Pol_Nbr : Insurance Policy Number of each case
3. Sailing_Date : Date of Transportation
4. Voyage_From_Name : Port of origin of the shipment
5. Voyage_To_Name : Destination Place/Port of Shipment
6. Mode_Transport : Transportation Mode
7. Type_GoodShip_Code : Commodity Type
8. Value_Goods_Shipped : Shipping Value
9. Tsi_Cur_Id : Currency of interest amount
10. Tsi_Amt : Interest amount
11. Coverage_Code : Coverage Type
12. Rate_Premium : Final rate of transportation insurance

The next stage is to perform data cleaning on the obtained dataset, which aims to correct inconsistent data and identify outliers and noise. After transformation to numerical values using table mapping that has been determined by the underwriter. Columns with a categorical type will be converted to numerical using the mapping tables below.

Table 4: Mapping Table of Transportation Mode

Code	Description	Base Rate
Ba	Barge	0.002
Bu	Bulk	0.007
…	..	..
Tu	Tug	0.005

Table 5: Mapping Table of Commodity Type

Code	Description	Base Rate
DRB	Dry Bulk	0.109
GAS	Gas	0.005
..	..	..
UTC	Unitised Cargo	0.724

Table 6: Mapping Table of Coverage Type

Code	Description	BaseRate
ALR	All Risk	0.04
ICB	ICC B (or Equal)	0.03
ICC	ICC C (or Equal)	0.02
TLO	Total Loss Only	0.01

Table 7: Origin and Destination Port Mapping Table

Voyage_From_Name	Voyage_To_Name	BaseRate
SAMARINDA	INDIA	0.002
SURABAYA	KOREA, BUSAN	0.0013
..	…	…
JAKARTA, TJ. PRIOK	SINGAPORE	0.0011

In the Value_Goods_Shipped column, it is necessary to carry out a data binning process, namely grouping numerical data into several bins so that the data distribution is easier to understand, as shown in table 8.

Table 8: Item Value Range Limit Mapping Table

No.	Range	Base Rate
1	0 - 600.000	0.001
2	600.001- 5.000.000	0.001
..	..	..
11	196.000.001 - 300.000.000	0.033

At this stage, several features are reduced. For the reduction of features carried out, namely the Obj_id, Pol_Nbr, Sailing_Date, Tsi_Cur_Id, and Tsi_Amt columns. Through the data preprocessing stage, the final dataset results in 2216 data points with 7 features consisting of 6 factor features and 1 result feature as follows:

Voyage_From_Name	: Port of origin of the shipment
Voyage_To_Name	: Destination Place/Port of Shipment
Mode_Transport	: Transportation Mode
Type_GoodShip_Code	: Commodity Type
Value_Goods_Shipped	: Shipping Value
Coverage_Code	: Insurance Coverage Clause
Rate_Premium	: Final rate of transportation insurance

This next stage will be analyzed to find the best model in determining the premium rate. The tool used is *Data* Orange *Mining* software. Based on the calculation results, the comparison results are obtained as below

Table 9: Model Evaluation Comparison Table

Model	MSE	RMSE	MAE	R2
Tree	4.49E-05	0.006702	0.003617	0.986307
Random Forest	3.92E-05	0.006263	0.003439	0.98804
Neural Network	0.000520611	0.022817	0.017657	0.841279
Linear Regression	2.27E-05	0.004761	0.002727	0.993088
kNN	3.08E-05	0.005552	0.003076	0.990604
GradientBoosting	2.71E-05	0.005206	0.00294	0.991737

As seen in the table, the Linear Regression Model is the best model because it has R2 closest to the value of 1 and has the smallest MSE, RMSE, MAE values.

After obtaining the final dataset and the model used is *linear regression*, the next step is program creation. For making this program using *Python* programming language and the *tool* used is *Visual Studio Code*.

4.2.2. Customer Communication

This section discusses the results of the research, which is the purpose of Research Question 2 (RQ2), namely building a communication model with customers using chatbot features. Chatbots use artificial intelligence technology and programming languages to process input and responses. In implementing the chatbot, researchers will use Keras, the deep learning library, NLTK (Natural Language Processing Toolkit), and several other libraries.

For this chatbot architecture, refer to the research (Adamopoulou et al., 2020). The process starts with a user request, for example, "Hello?", to a chatbot that uses a web-based frontend application. Once the chatbot receives the user's request, Dialogue Management parses it to infer the user's intent and related information. Once the chatbot reaches the best interpretation it can, it can act directly on the new information, remembering whatever it has already understood and waiting to see what happens next, requiring more context information or asking for clarification. When the request is understood, action execution and information retrieval are performed. The chatbot performs the requested action or retrieves the desired data from its data source, which can be a database, known as the chatbot knowledge base. Once retrieved, the Response Generation Component prepares a human-like natural language response to the user based on the intent and context information returned from the user message analysis component. To create this *chatbot* program using *Python* programming language and the *tool* used is *VisualStudio Code*. The stages are as follows:

- Import library. Create a new python file and name it as train_chatbot and then we will import all the required modules. After that, we will read the JSON data file in the Python program
- Data Pre-processing, Models cannot take raw data. It has to go through a lot of processing to be easily understood by machines. For textual data, there are many preprocessing techniques available. The first technique is tokenizing, where it breaks the sentence into words. It tags eachpattern and adds the words in a list. Another technique is Lemmatization. It converts words intolemma form so that we can reduce all the canonical words. For example, the words play, playing,plays, played, etc. will all be replaced with play. In this way, we can reduce the total number ofwords in the vocabulary. So lemmatize each word and remove duplicate words.
- Creating Training and Testing Data For the Train model, each pattern will be converted and inputted into numbers. First, it will lemmatize each word of the pattern and create a list of zeroswith a length equal to the total number of words. Then assign a value of 1 only to the index thatcontains the word in the pattern.

4.2.3. Product Recommendations

This section discusses the results of the research which is the purpose of *Research Question 2 (RQ2)*, namely building a product recommendation model on the insurance product system by forming data patterns of insurance product sales transaction rules using the *Association Rule* and *FP-Growth* methods. The purpose of this insurance product recommendation is to be able to provide recommendations for insurance products based on historical data on insurance product sales transactions.

This section discusses the results of the research, namely building a system recommendation model using insurance product transaction patterns using the *Association Rule* method. This method uses several stages starting from literature studies, experimental methods, and through the *data mining* testing process. System recommendations for insurance products are made by forming rule patterns based on historical transaction data from buyers. The system design analysis process beginswith the process of retrieving historical transaction data of insurance users. The total data of insurance product sales transactions collected is 5,111 data. Based on these data, association rules will be madeto provide recommendations for recommended insurance products based on high frequency patterns.The stages are described as below

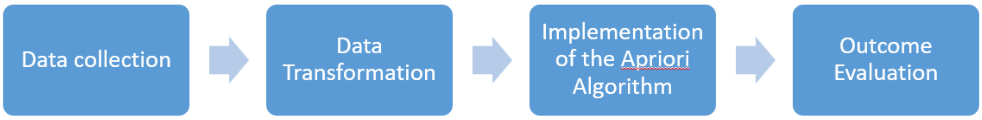

Fig.8: Recommendation System Development Process

To be able to design an insurance product recommendation system for users, a *dataset of* historical customer transactions is required. This data is in the form of types of insurance products used by customers to find high frequency patterns that will be used for product recommendations. The list of insurance products is in table 11. While the historical transaction data of insurance products per customer and per product collected as much as 6,480 data as shown in table 12. For the next data is processed into insurance product recap data per customer which amounts to 5,111 data as in table 13. This transaction data will be used for the construction of a more accurate insurance product recommendation system in making association method decisions.

Table 10: Insurance Product Master Data

Product Code	Product Description
A	AVIATION
B	BOND
C	MARINE CARGO
…	…
N	ONSHORE PSC

Table 11: Historical Sales Data per Insurance Product per Customer

No.	CustomerCode	BizCls	BizClsName	Total Policy
1	A000004500	L	LIABILITY	4
2	A000004500	O	OFFSHORE PSC	1
3	A000004500	V	MISCELLANEOUS	1
4	A000012200	F	FIRE	1
5	A000019400	E	ENGINEERING	2
...				
6479	Z000011600	F	FIRE	1
6480	Z000011700	F	FIRE	1

Table 12:Sales Transaction Data of Insurance Products per Customer

No.	Customer Code	Insurance Products
1	A000004500	L, O, V
2	A000012200	F
3	A000019400	E, F, H
4	A000030900	H
5	A000036400	L
6	A000036900	F, M, V
7	A000037100	F, M, V
...	...	...
5111	Z000011700	F

In this section, the process of building the *Frequent Itemsets* value will be carried out to find *itemset* rule patterns consisting of the most frequently used insurance product combinations. The formation of *high frequent itemsets* is determined based on the *support* value of each combination. The *support* value of an *itemset* can be calculated using equation (1) for each combination. In the formation of *frequent itemset*, *the minimum support* value is determined as a reference value to determine the pattern of the next iteration *itemset*. In this study, the *minimum support* value was determined to be 0.1%. This value is taken because the *support value* that will be generated from each combination is very small due to the comparison of the quantity of the *item* with the total transactions. In order to produce association rules that are not too few and more accurate, the *minimum support* value for this method is set at 0.1%. Each combination of *itemsets is* calculated fora combination of 1 *itemset*. If the value of *support* > *minimum support*, then the combination of *itemsets* will continue to be a combination of 2, 3, and 4 *itemsets*.

$$Support(A) = \frac{\text{Lots of transactions(A)}}{\text{Overall transaction total}} \times 100\% \quad (1)$$

Iteration 4: Combination of 4 Itemsets

Table 13: Combination of 4 Itemsets

Products	Quantity	Support (%)
C, E, F, L	12	0.2348
C, E, F, O	7	0.137
C, E, F, V	10	0.1957
C, E, L, O	6	0.1174
C, E, L, V	10	0.1957
...	...	...
F, H, L, V	9	0.1761
F, L, O, V	11	0.2152

Based on the iteration results according to tables 13, it is found that with a *minimumsupport* value of 0.1% transaction data can be iterated up to four times and produce a combination of*items* as many as 4. This combination states that there are several transactions there are purchases of 4 types of insurance products at once. This type of *item* combination will be continued to create a recommendation system based on the *support* value of each itemset. From the results of the *high frequent dataset* analysis, it was found that the types of insurance products most often used by customers include product F, namely fire insurance products, then L. These two *itemsets are* also purchased a lot at once in one transaction. Thus, the *output* obtained from this analysis process is to successfully find patterns of purchasing insurance products based on purchases in one transaction. This *high frequent dataset* pattern will be the new *dataset* to be implemented into system recommendations using the apriori algorithm.

After the results of the search for *high frequent datasets* are found. The dataset will be reprocessed using the Apriori algorithm to build a recommendation pattern for the insurance premium system. Each *frequent dataset* will be re-evaluated by creating *association rules*. Association rules consist of premises and conclusions that are built based on *high frequent dataset* patterns.

$$Confidence \equiv P(B|A) = \frac{\text{Many transactions contain (A dan B)}}{\text{Total transactions contain (A)}} \times 100\% \quad (2)$$

$$Lift = \frac{\text{Support (A dan B)}}{\text{Support(A)} \times \text{Support(B)}} \quad (3)$$

The Apriori algorithm is a number of steps used to find the most frequent *itemset* patterns in the *database.* (Lin et al., 2000). At the stage of implementing this algorithm, the combination of *itemsets* that have been obtained in tables 14 will be paired with other combinations. So that the rule {A} → {B} with {A} is a combination of *antecendent* or premise *itemsets* and {B} is a combinationof *consequent* or conclusion *itemsets*. In the iteration, the size of the *minimum support is* 0.1% or 0.001, the minimum *confidence is* 70%, and the *minimum lift is* 1. When the rules formed do not meet the requirements, the association rules will not appear in the *output*. The *dataset is* processed again and gets the results in the form of 77 association rulesthat meet the minimum limit.

Table 14: Iteration of Association Rules Using Apriori Algorithm

No.	*Antecendent (Premise)*	*Consequent (Conclusion)*	*Support*	*Confidence*	*Lift*
1	H, L, V	F	0.002	1	2.115
2	F, H, V	L	0.002	1	7.617
3	E, O, V	F	0.002	1	2.115
4	E, L, O, V	F	0.001	1	2.115
...	...	...	...	...	...
77	C, F, O	L	0.002	0.923	7.031

Based on the results of the iteration in table 14, 77 association rules are obtained. This rule has a minimum *support of* 0.1% or 0.001. When the *lift* value > 1 represents that the two variables are highly dependent on each other, so the purchase of product {B} will depend on the purchase of {A}. From these 77 rules, we can build a recommendation system for purchasing the next insurance product if you have purchased combination of insurance products. For example, for rule number 3, when there is a purchase of E, O, and V in one transaction, there is a 100% probability of buying F in the next transaction. If re-evaluated from the *lift* value, the rule has a value of 2,115 which represents that the F *itemset* depends on the purchase of E , O, and V.

4.3. Model Implementation

The design process includes the formation of a user interface based on the analysis process given in Figure 9. The underwriter or customer must first select the commodity type so that the system can determine what techniques will be used to calculate the predicted rate value, then fill in the new case data to be predictedon the form on the left. Then, the underwriter can press the "Calculate Rate" button so that the system processes the new case data and generates a prediction rate.

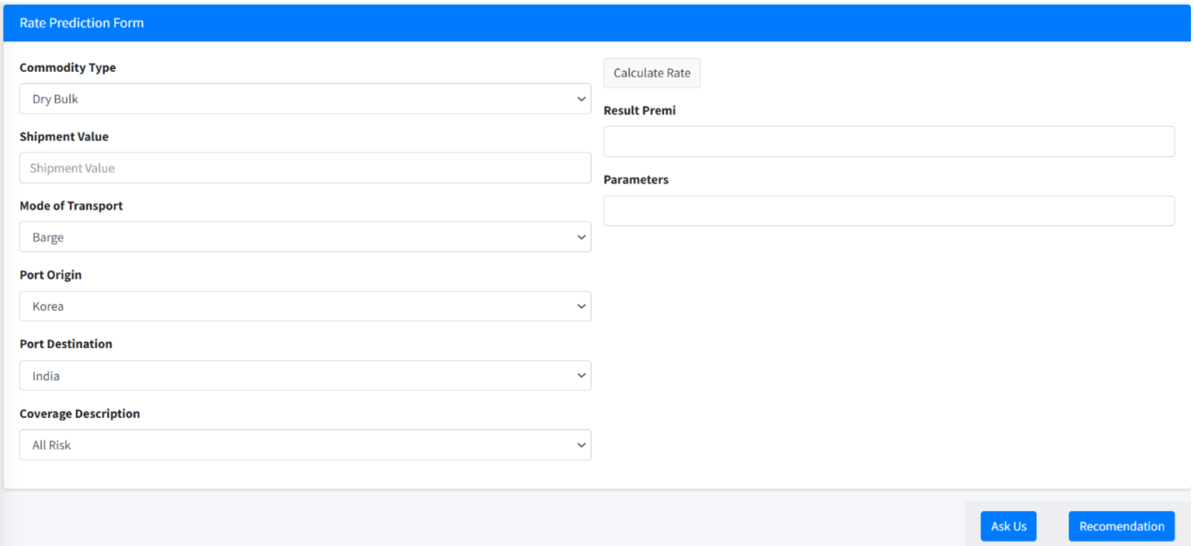

Fig.9: Main Menu Interface Design

If want to consult about filling in the insurance input object, just click the Ask button and the image willappear as below.

Fig.10: Chatbot Menu Design

If want to get other insurance product recommendations, can click the Product Recommendation button, and product recommendations will appear as below.

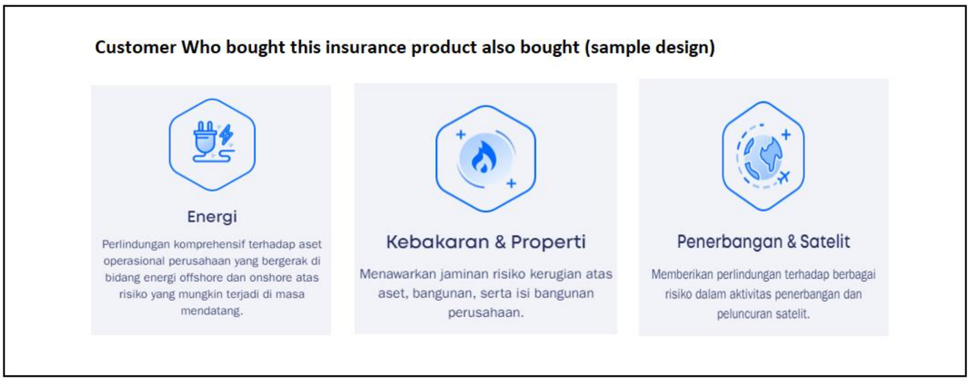

Fig.11: Recommender System Menu Design

4.4. Model testing

In this section discusses the results of the research which is the purpose of *Research Question 3 (RQ3)*, namely testing the model. The tests are as follows.

4.4.1. Premium Rate Prediction Model Testing

For the model used in predicting premium rates, the *Linear Regression* model is the best model because it has *R2 which* is closest to the value of 1 and has the smallest *MSE*, *RMSE*, *MAE* values referring to the discussion of the results of the previous data analysis. For the results of testingthe linear regression model as in table 15.

Table 15: Linear Regression Model Testing Table

Model	MSE	RMSE	MAE	R2
Linear Regression	2.27E-05	0.004761	0.002727	0.993088

4.4.2. Product Recommendation Model Testing

After getting the insurance product association rules as in table 14, then continue to filter for consequent C (*Cargo* product), and obtained as table 16 below

Table 16: Associations with *Consequent Cargo*

No.	Supp	Conf	Lift	Antecedent	Consequent
1	0.002	0.75	6.464	F,L,O	C
2	0.002	0.727	6.268	F,L,O,V	C
3	0.001	0.7	6.033	E,F,O	C
4	0.001	0.778	6.704	F,H,V	C
5	0.001	0.778	6.704	H,L,V	C
6	0.001	0.75	6.464	E,O,V	C
7	0.001	0.75	6.464	E,F,O,V	C

For example, for rule number 1, when F, L, and O are purchased in one transaction, there is a 75% chance that C (Cargo) will be purchased in the next transaction. If re-evaluated from the *lift* value, the rule has a value of 6.464 which represents that the C *itemset* depends on the purchase of F, L and O. If the lift value > 1 then the rule can be said to be strong and interdependent between the premises. If the lift value > 1then the rule can be said to be strong and interdependent between the premise and the conclusion. So for this rule to purchase Product C (Cargo), it can be recommended for insurance products F (*Fire*), L (*Liability)*, and O (*Offshore)*.

4.4.3. Chatbot Model Testing

For testing this chatbot using black box testing, by entering question or chat input. And the output is whether it matches the scenario or not. The scenario is below

Table 17: *Chatbot black box* testing

Question	Results
Question *keywords* found in the *database*	The chatbot brings up answers according to the questions used
Question *keyword* not found in *database*	Chatbot brings up the answer "Chatbot has not beenable to find answers to these questions. For further information, please contact CS

5. Conclusions

Utilizing the Design Science Research approach, the model was developed in five stages. The findings of this research include a conceptual model and prototype for a smart insurance model for the sea cargo business. Based on the test results for the premium rate prediction feature using the linear regression model with an R2 value of 0.993 and an MAE of 0.0002727, As for the product recommendation features, the testing technique uses a lift value with a value above 1, as shown in table 16. As for the chatbot features, the testing uses the black box testing technique as shown in table 17. So it can be concluded that this model can meet the needs of the user.

The contribution of research from a scientific point of view is that it can contribute to the development of science in the field of information systems, especially the ability of artificial intelligence models, which can play a role in insurance performance. Meanwhile, from a practical standpoint, it can contribute to the industry, namely by increasing speed and accuracy in the underwriting decision-making process and increasing customer satisfaction in obtaining insurance risk assessment results and other insurance production recommendations. Meanwhile, from the public side, they can get insurance services quickly, precisely, and according to their needs.

For further research, the features used in the Smart Insurance Model for Sea Cargo Business can be used for other insurance products such as fire, aviation, and others. The difference is in the underwriting factor. For this research, the underwriting factors include origin port, destination port, transportation mode, commodity type, shipping value, insurance coverage clause, and final rate of transportation insurance. Meanwhile, when used in other insurance products such as fire, of course the underwriting factor is different. For the chatbot feature, knowledge can be added with data on frequently asked questions related to these other products. The output feature for the recommendation system will, of course, be different based on the sales history of insurance products. But for the algorithm, there is no change.

References

Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. In Artificial Intelligence Applications and Innovations: 16th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, June 5–7, 2020, Proceedings, Part II 16 (pp. 373-383). Springer International Publishing.

Ahmad Sabri, L. H., Lim, A. H. L., & Goh, H. N. (2022). Click Analysis: How E-commerce Companies Benefit from Exploratory and Association Rule Mining. Journal of System and Management Sciences, 12(5), 36-56.

Ayuso, M., Guillen, M., & Marín, A. M. P. (2016). Using GPS data to analyse the distance travelled to the first accident at fault in pay-as-you-drive insurance. Transportation research part C: emerging technologies, 68, 160-167.

Biddle, R., Liu, S., Tilocca, P., & Xu, G. (2018). Automated underwriting in life insurance: Predictions and optimisation. In Databases Theory and Applications: 29th Australasian Database Conference, ADC 2018, Gold Coast, QLD, Australia, May 24-27, 2018, Proceedings 29 (pp. 135-146). Springer International Publishing.

Cebulsky M, Günther J., Heidkamp P., Brinkmann F., “The Digital Insurance – Facing Customer Expectation in a Rapidly Changing World”, Digital Marketplaces Unleashed pp 359-370, German, 2017

Chan, N. K. W., Lee, A. S. H., & Zainol, Z. (2021, June). Predicting employee health risks using classification ensemble model. In 2021 Fifth International Conference on Information Retrieval and Knowledge Management (CAMP) (pp. 52-58). IEEE.

Chakravarty, T., Ghose, A., Bhaumik, C., & Chowdhury, A. (2013, December). MobiDriveScore—A system for mobile sensor based driving analysis: A risk assessment model for improving one's driving. In 2013 Seventh International Conference on Sensing Technology (ICST) (pp. 338-344). IEEE.

Chew, L. J., Haw, S. C., Subramaniam, S., & Ng, K. W. (2022). A Hybrid Ontology-based Recommender System Utilizing Data Enrichment and SVD Approaches. Journal of System and Management Sciences, 12(5), 139-154.

Davenport, T. H., & Ronanki, R. (2018, January). Artificial Intelligence for the Real World. Harvard Business Review.

Doultani M. "Smart Underwriting – A Personalilses Virtual Agent", ," Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021)”, IEEE Xplore Part Number: CFP21K74-ART; ISBN: 978-0-7381-1327-2

Dubey A, Parida T, Birajdar A, Prajapati AK, Rane S, “Smart Underwriting System: An Intelligent Decision Support System for Insurance Approval & Risk Assessment”, 3rd International Conference for Convergence in Technology, The Gateway Hotel, XION Complex, Wakad Road, Pune, India. Apr 06-08, 2018

Fidelia P., Application of Machine Learning for Estimating Motor Vehicle Insurance Premium, tesis KCA University, 2019

Haw, S. C., Chew, L. J., Ong, K., Ng, K. W., Naveen, P., & Anaam, E. A. (2022). Content-based Recommender System with Descriptive Analytics. Journal of System and Management Sciences, 12(5), 105-120.

Handel, P., Skog, I., Wahlstrom, J., Bonawiede, F., Welch, R., Ohlsson, J., & Ohlsson, M. (2014). Insurance telematics: Opportunities and challenges with the smartphone solution. IEEE Intelligent Transportation Systems Magazine, 6(4), 57-70.

Joram MK, Harrison BK, Joseph N, "A Knowledge-Based System for Life Insurance Underwriting", Information Technology and Computer Science, 2017, 3, 40-49

Kacelan V., Kacelan L., and Buric M. N. A nonparametric data mining approach for risk prediction in car insurance: a case study from the montenegrin market. Economic Research- Ekonomska Istraivanja, pages 545–558, 2017.

Kumar, R. (2019). Research Methodology (Fifth). Los Angelos: SAGE

Lawrence, C., Tuunanen, T., & Myers, M. (2010). Extending Design Science Research Methodology for a Multicultural World. IFIP Advances in Information and Communication Technology (Vol. 318). Springer. https://doi.org/10.1007/978-3-642-12113-5_7

Lee, J. S., & Lee, I. S. (2016). The study of insurance premium rate GIS mapping considering the storm and flood hazard risks. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 8.

Lephoto A, Kogeda O.P. ,Modelling a Rule Based System for Medical Underwriting in an Insurance Industry,Proceedings of the World Congress on Engineering and Computer Science 2014 Vol I WCECS 2014, 22-24 October, 2014, San Francisco, USA

Lertpunyavuttikul, P., Chuenprasertsuk, P., & Glomglome, S. (2017, November). Usage-based Insurance Using IoT Platform. In 2017 21st International Computer Science and Engineering Conference (ICSEC) (pp. 1-5). IEEE.

Lin, W., Alvarez, S. A., & Ruiz, C. (2000). Collaborative recommendation via adaptive association rule mining. Data Mining and Knowledge Discovery, 6(1), 83-105.

Lukas S.Stefani D,Widjaja P. (2019)" Comparing SVM and GLM in calculating insurance premium for flight delay", Pervasive Health: Pervasive Computing Technologies for Healthcare. https://doi.org/10.1145/3369114.3369160

Maier, M., Carlotto, H., Saperstein, S., Sanchez, F., Balogun, S., & Merritt, S. (2020). Improving the accuracy and transparency of underwriting with AI to transform the life insurance industry. AI Magazine, 41(3), 78-93.

Müller, D., & Te, Y. F. (2017, December). Insurance premium optimization using motor insurance policies—A business growth classification approach. In 2017 IEEE International Conference on Big Data (Big Data) (pp. 4154-4158). IEEE.

Mustika, W. F., Murfi, H., & Widyaningsih, Y. (2019, October). Analysis accuracy of xgboost model for multiclass classification-a case study of applicant level risk prediction for life insurance. In 2019 5th International Conference on Science in Information Technology (ICSITech) (pp. 71-77). IEEE.

Pitchayaviwat, T. (2016, October). A study on clustering customer suggestion on online social media about insurance services by using text mining techniques. In 2016 Management and Innovation Technology International Conference (MITicon) (pp. MIT-148). IEEE.

Ramnathan, P. P., Revathy, S., & Mani, V. (2020, May). Smart Property Insurance using IoT. In 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 796-800).

IEEE.

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

Sachan S., Yang J., Dong-LingXu, Benavides D.E.,Li Y, "An explainable AI decision-support-system to automate loan underwriting","Expert Systems with Applications",Volume 144, 15 April 2020.

Santoso C.B., Prabowo H, Warnars H.L.H.S., Fajar A.N., 2021,"Smart Insurance System Model Concept for Marine Cargo Business",2021 International Conference on Data Science and Its Applications (ICoDSA)

Singh, P. K., Singh, R., Muchahary, G., Lahon, M., & Nandi, S. (2019, October). A blockchain-based approach for usage based insurance and incentive in its. In TENCON 2019-2019 IEEE Region 10 Conference (TENCON) (pp. 1202-1207). IEEE.

Surat Edaran Otoritas Jasa Keuangan Nomor 6 /SEOJK.05/2017

Undang-Undang No. 17 Th 2008 tentang Pelayaran

Wengnoon, R., & Limpiyakorn, Y. (2014, May). Extension of Insurance Premium Payment to Mobile Application with QR Code. In 2014 International Conference on Information Science & Applications (ICISA) (pp. 1-4). IEEE.

Xiahou, J., Xiao, M., He, X., & Cui, X. (2021, June). Research on the Applications of Data Mining in the Analysis of Vehicle Insurance Industry. In 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC) (Vol. 4, pp. 1095-1099). IEEE.

Yan W, Bonissone PP. 2006. *Designing a Neural Network Decision System for Automated Insurance Underwriting*. Canada (US): IEEE.