PREDICTION OF BOND PLANNING BASED ON CUSTOMER CHARACTERISTICS IN INDONESIA USING THE RANDOM FOREST ALGORITHM

Hendra Achmadi hendra.achmadi@uph.edu

Abstract

This research aims to identify industries receptive to wealth management products like bond insurance and to pinpoint the key factors influencing bond insurance premium closures. Given Indonesia's low life insurance penetration and density compared to other Asian countries, the study uses a purposive sample of 199 data points from 38 companies across various industries. This research found that the most influential factors in bond purchases are domicile, the second car, industry type, job position , housing type . The methodology that is used in this research is called CRIPS-DM (Cross Industrial Standards Program Data Mining). The first steps what is the purpose of the organization, and the second is what data that needed, and then continue to data preparation, after that modeling, after modeling, it will make an interpretation of the result, and the final steps is deployment, it will plan how it will be implemented in the real world, and the accuracy score from this model is 98 %. From the result of the projection closing bond insurance from each industry, it can be concluded that the most industry that closed the bond insurance is Banking Industry, the second is from insurance and the third is education and the next is education, retail, bond, manufacturing and finance, hospitality, legal, publishing, technology and government and service industries.

Key words: Bond Prediction, CRIPS-DM, Random Forest.

INTRODUCTION

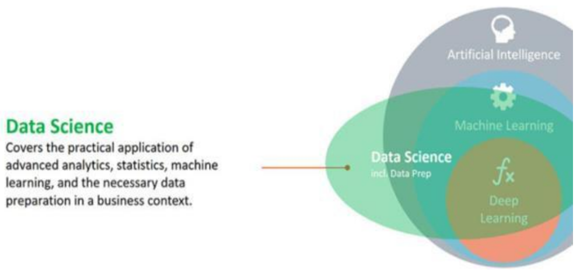
Banking and financial industries prioritize increasing sales effectiveness, often achieved through targeted marketing. Effective targeting requires identifying key industries and the factors influencing customer purchases. However, determining these influential factors can be challenging due to varying sales approaches. Therefore, machine learning algorithms can assist marketing efforts by identifying target industries and predicting which factors are most likely to lead to successful sales closures.

Targeting, driven by historical data, allows insurance companies to focus their operational activities on specific market segments. The insurance industry plays a crucial role in providing risk management and protection instruments. Given the industry's core function of risk management, protecting insurance consumers is paramount, ensuring companies can fulfill their obligations. Effective consumer protection relies on all industry players adhering to the principle of prudence and conducting business responsibly.

While business competition can drive efficiency, improve service quality, and benefit consumers with better and more affordable insurance products, intense competition can also have negative consequences. It can lead to practices that harm consumers and destabilize the industry, such as unsustainable premium pricing, excessive marketing commissions, and a lack of transparency regarding company and product information. This unbondy competition ultimately harms both consumers and the industry as a whole.

LITERATUR VIEW

The evolution of Data Science, starting with Alan Turing's 1950 paper, "Computing Machinery and Intelligence" (Winston, 2017), which explored the concept of thinking computers and introduced the Turing Test – a benchmark requiring computers to convincingly imitate human conversation. The field of Artificial Intelligence (AI) was officially named in 1956 by John McCarthy and Marvin Minsky, marking the beginning of the AI era. A key milestone was reached in 1997 when IBM's Deep Blue computer defeated chess champion Garry



Gambar 2 CRIPS-DM (Academy, 2024)

The CRISP-DM method begins with Business Understanding, defining the objectives of the data mining analysis. Next, Data Understanding focuses on identifying and collecting relevant data to address the business problem. The following stage, Data Preparation, involves cleaning the data (handling outliers), ensuring its reliability and validity (testing for multicollinearity and heteroscedasticity), and removing duplicates. After preparing the data, an appropriate model is selected, followed by evaluation, interpretation of results, and finally, implementation of the derived insights.

Kasparov. In 2006, Geoffrey Hinton coined the term "Deep Learning" to describe an algorithm enabling computers to "see" by identifying objects and text within images and videos (Academy, 2024). It illustrates in figure 3.

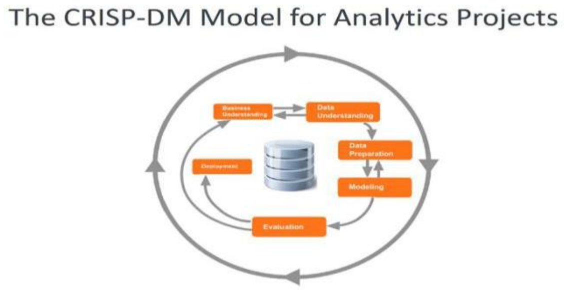


Figure 3. Data Science (Academy, 2024)

The CRISP-DM (Cross Industrial Standards Program for Data Mining) process (Singgalen, 2024). This data mining methodology begins with data preparation and cleaning, involving tasks like data type identification (number, factor, date), removal of special characters, and data transformation to suit the target output. Following this preparation phase, data mining models are applied based on appropriate methods. The final step involves interpreting the knowledge or insights extracted from the processed data. It illustrates in figure 3.

Supervisor Learning

Classification techniques, as described by Duarte et al. (2019), are learning algorithms trained on input data to categorize new observations based on learned patterns. Conversely, supervised learning for regression predicts continuous values. Charbuty and Abdulazeez (2021) highlight the frequent use of decision trees in classification, noting their ability to identify important features for interpretation. Müller and Guido (2017) further explain that supervised learning involves splitting data into training and testing sets to evaluate the accuracy of the trained model.

Random Forest

Random forests, as explained by Ong et al. (2023), are an ensemble learning method that averages the predictions of multiple decision trees to enhance accuracy. Gkikas et al. (2022) elaborate that each tree within the random forest is built independently using a random subset of features at each node. A key advantage of random forests, highlighted by Müller and Guido (2017) and Muhajir and Widiastuti (2022), is their ability to determine feature importance, identifying which features most significantly influence the prediction. This feature importance is derived from the algorithm's consideration of multiple features during tree construction and reflects each feature's contribution to the overall performance of the forest.

METHODOLOGY

3.1 Population

Population from this research from many 38 industries, and with 148 sample data, with purposive sampling, has been collected by google form method. The criteria for purposive sampling is the respondent must have experience to closed insurance police.

```
import pandas as pd
data= pd.read_csv('/SPENDING/SEHAT2.CSV',delimiter=';')
data
```

[1] ✓ 0.2s

	JK	UMUR	BANGSA	DOMISILI	JABATAN	INDUSTRI	INCOME	MAKAN	TRANSPORT	LISTRIK	RUMAH	PE
0	Pria	45 tahun - 55 tahun	Indonesia	Jakarta	mahasiswa	kesehatan	Rp. 29 jt - 53 jt / Bulan	21%-40%	1%-10%	1%-10%	Milik Sendiri	
1	Pria	45 tahun - 55 tahun	Korea	Jakarta	asistant manager	Legal	Rp. 77 jt - 101 jt/Bulan	1%-10%	1%-10%	3700 Watt	Kontrak	
2	Pria	45 tahun - 55 tahun	Korea	Jakarta	Manager	Publishing	Rp 53 jt - 76 jt / Bulan	21%-40%	11%-20%	1300 Watt	Kontrak	
3	Pria	26 tahun - 45 tahun	Indonesia	Jakarta	Art director	Food And Beverage	Rp 5 jt - 29 jt / Bulan	1%-10%	1%-10%	1300 Watt	Milik Keluarga	
4	Pria	45 tahun - 55 tahun	Korea	Jakarta	Manager	Education	Rp 53 jt - 76 jt / Bulan	11%-20%	1%-10%	2200 Watt	Kontrak	

3.2 Business Understanding

Data is collected from 199 respondent from 38 different industries. The question is which of these industries has the most adoption from the insurance products.

Banking	57
Insurance	17
Education	15
Retail	7
Agency	4

Food And Beverage	3
Bond	3
Manufactur	3
Telecommunication	3
Finance	2
Hospitality	2
kesehatan	2
Legal	2
Publishing	2
Service	2
Technology	2
Audit	1
Contruction	1
Dropship kebutuhan rmh tangga	1

Table 1. Industrial Distribution

From Table 1, most of data belongs to Banking (57), Insurance (17) and Education (15) and Retail (7) and the agency (4), and the rest is 2 or 1.

3.2 Data Understanding.

These data was taken from the primary data, whose taken from the questioner, and after that it will put into the data set using python.

3.2 ata Preparation

After the data set is input, data has to decode first, because the data has much contained the category types. The data has to decode into numeric types.

JK	UMUR	BANGS	DOMISILI	JABATAN	INDUSTR	INCOME	MAKAN	TRANSPORT	LISTRIK	RUMAH	PEND	SOCIAL_MA	SJKELUARGA	SMOBIL2	KESEHATAN	STATUS
Pria	45 tahun - 55 tahun	Korea	Jakarta	asistant manage	Legal	Rp. 77 jt - 101 jt/Bulan	1%-10%	1%-10%	3700 Watt	Kontrak	S2	1%-10%	1%-10%	1%-10%	1%-10%	Menikah anak 2 orang
Pria	45 tahun - 55 tahun	Korea	Jakarta	Manager	Publishing	Rp 53 jt - 76 jt / Bulan	21%-40%	11%-20%	1300 Watt	Kontrak	S1	11%-20%	11%-20%	1%-10%	1%-10%	Menikah anak 2 orang
Pria	45 tahun - 55 tahun	Korea	Jakarta	Manager	Education	Rp 53 jt - 76 jt / Bulan	11%-20%	1%-10%	2200 Watt	Kontrak	S2	1%-10%	1%-10%	1%-10%	1%-10%	Menikath anak 3 orang
Pria	17 tahun - 25 tahun	Indonesia	Jakarta	Staff	Education	Rp 5 jt - 29 jt / Bulan	21%-40%	11%-20%	1300 Watt	Milik Keluarga	S1	1%-10%	1%-10%	1%-10%	11%-20%	Single
Wanita	17 tahun - 25 tahun	Indonesia	Jakarta	Staff	Technology	Rp 5 jt - 29 jt / Bulan	1%-10%	11%-20%	3700 Watt	Milik Keluarga	S1	21%-40%	21%-40%	11%-20%	11%-20%	Single
Pria	17 tahun - 25 tahun	Indonesia	Jakarta	Staff	Education	Rp 5 jt - 29 jt / Bulan	21%-40%	1%-10%	2200 Watt	Milik Keluarga	S1	1%-10%	11%-20%	1%-10%	1%-10%	Single
Wanita	17 tahun - 25 tahun	Indonesia	Manado	Staff	Banking	Rp 53 jt - 76 jt / Bulan	21%-40%	11%-20%	3700 Watt	Milik Keluarga	S1	11%-20%	21%-40%	21%-40%	41%-60%	Single
Pria	26 tahun - 45 tahun	Indonesia	Jakarta	Manager	Banking	Rp 5 jt - 29 jt / Bulan	11%-20%	1%-10%	900 Watt	Kontrak	S1	1%-10%	1%-10%	1%-10%	1%-10%	Menikah anak 1 orang
Pria	26 tahun - 45 tahun	Indonesia	Depok	Senior Manager	Banking	Rp 5 jt - 29 jt / Bulan	1%-10%	11%-20%	900 Watt	Kontrak	S2	1%-10%	11%-20%	1%-10%	1%-10%	Menikah anak 1 orang
Wanita	17 tahun - 25 tahun	Indonesia	Depok	Staff	Banking	< Rp 5 jt / Bulan	41%-60%	11%-20%	450 Watt	Kontrak	S1	11%-20%	11%-20%	1%-10%	11%-20%	Single
Wanita	26 tahun - 45 tahun	Indonesia	Jakarta	Staff	Banking	Rp 5 jt - 29 jt / Bulan	11%-20%	1%-10%	1300 Watt	Kontrak	S1	11%-20%	11%-20%	1%-10%	1%-10%	Menikah anak 1 orang
Pria	lebih dari 55 tahun	Indonesia	Jakarta	Senior Manager	Banking	Rp 53 jt - 76 jt / Bulan	1%-10%	1%-10%	900 Watt	Milik Sendiri	S1	1%-10%	1%-10%	1%-10%	1%-10%	Menikah anak 2 orang
Pria	lebih dari 55 tahun	Indonesia	depok	Staff	Banking	Rp 53 jt - 76 jt / Bulan	1%-10%	1%-10%	> 4400 Watt	Milik Sendiri	S3	1%-10%	1%-10%	1%-10%	500 ribu - 1 juta	Menikath anak 3 orang
Pria	45 tahun - 55 tahun	Indonesia	Jakarta	Staff	Banking	Rp. 77 jt - 101 jt/Bulan	11%-20%	1%-10%	> 4400 Watt	Milik Sendiri	S1	11%-20%	11%-20%	11%-20%	2 juta - 3 juta	Menikah anak 2 orang
Pria	17 tahun - 25 tahun	Indonesia	Tangerang	Staff	Banking	Rp 5 jt - 29 jt / Bulan	11%-20%	41%-60%	2200 Watt	Kos Kosan	S1	11%-20%	11%-20%	1%-10%	1 juta - 2 juta	Single
Wanita	26 tahun - 45 tahun	Indonesia	Jakarta	Manager	Banking	Rp 5 jt - 29 jt / Bulan	21%-40%	1%-10%	3700 Watt	Milik Sendiri	S1	1%-10%	1%-10%	1%-10%	500 ribu - 1 juta	Menikah anak 2 orang
Wanita	17 tahun - 25 tahun	Indonesia	Jakarta	Staff	Banking	Rp 5 jt - 29 jt / Bulan	1%-10%	1%-10%	900 Watt	Milik Keluarga	S1	21%-40%	21%-40%	1%-10%	1 juta - 2 juta	Single
Wanita	17 tahun - 25 tahun	Indonesia	Jakarta	Senior Manager	Banking	< Rp 5 jt / Bulan	1%-10%	11%-20%	2200 Watt	Milik Keluarga	S1	1%-10%	21%-40%	1%-10%	500 ribu - 1 juta	Single
Wanita	17 tahun - 25 tahun	Indonesia	Jakarta	Staff	Insurance	Rp 5 jt - 29 jt / Bulan	11%-20%	1%-10%	2200 Watt	Milik Sendiri	S1	1%-10%	1%-10%	21%-40%	500 ribu - 1 juta	Single
Wanita	26 tahun - 45 tahun	Indonesia	Tangerang Se	Senior Manager	Banking	Rp 5 jt - 29 jt / Bulan	21%-40%	21%-40%	2200 Watt	Milik Sendiri	S1	11%-20%	11%-20%	11%-20%	1 juta - 2 juta	Menikah anak 1 orang
Pria	26 tahun - 45 tahun	Indonesia	Tangerang	Manager	Banking	Rp 5 jt - 29 jt / Bulan	11%-20%	11%-20%	1300 Watt	Milik Sendiri	S1	1%-10%	1%-10%	1%-10%	500 ribu - 1 juta	Menikah anak 2 orang
Wanita	17 tahun - 25 tahun	Indonesia	Jakarta	Senior Manager	Banking	Rp 5 jt - 29 jt / Bulan	11%-20%	1%-10%	2200 Watt	Milik Keluarga	S1	11%-20%	11%-20%	1%-10%	500 ribu - 1 juta	Single
Wanita	17 tahun - 25 tahun	Indonesia	Jakarta	Manager	Banking	Rp 5 jt - 29 jt / Bulan	21%-40%	11%-20%	1300 Watt	Kontrak	S1	11%-20%	11%-20%	1%-10%	500 ribu - 1 juta	Single
Wanita	17 tahun - 25 tahun	Indonesia	Jakarta	Manager	Banking	Rp 5 jt - 29 jt / Bulan	21%-40%	1%-10%	2200 Watt	Milik Keluarga	S2	1%-10%	1%-10%	1%-10%	500 ribu - 1 juta	Single
Wanita	26 tahun - 45 tahun	Indonesia	Jakarta	Manager	Banking	Rp. 30 jt - 52 jt / Bulan	21%-40%	1%-10%	4400 Watt	Milik Keluarga	S1	21%-40%	21%-40%	1%-10%	1 juta - 2 juta	Single
Wanita	26 tahun - 45 tahun	Indonesia	Jakarta	Manager	Banking	Rp. 30 jt - 52 jt / Bulan	1%-10%	1%-10%	4400 Watt	Milik Sendiri	S1	1%-10%	1%-10%	11%-20%	3 juta - 5 juta	Menikath anak 3 orang
Pria	26 tahun - 45 tahun	Indonesia	Jakarta	Manager	Insurance	Rp. 30 jt - 52 jt / Bulan	1%-10%	1%-10%	2200 Watt	Milik Sendiri	S1	1%-10%	1%-10%	1%-10%	2 juta - 3 juta	Menikah anak 1 orang
Pria	26 tahun - 45 tahun	Indonesia	Jakarta	Senior Manager	Insurance	Rp. 30 jt - 52 jt / Bulan	21%-40%	11%-20%	2200 Watt	Milik Sendiri	S1	1%-10%	11%-20%	1%-10%	2 juta - 3 juta	Menikah anak 1 orang

Tabke 1. Raw Data

```
from sklearn.preprocessing import LabelEncoder
enc= LabelEncoder()
data['JK']=enc.fit_transform(data['JK'].values)
data['UMUR']=enc.fit_transform(data['UMUR'].values)
data['BANGSA']=enc.fit_transform(data['BANGSA'].values)
data['DOMISILI']=enc.fit_transform(data['DOMISILI'].values)
data['JABATAN']=enc.fit_transform(data['JABATAN'].values)
data['INDUSTRI']=enc.fit_transform(data['INDUSTRI'].values)
data['INCOME']=enc.fit_transform(data['INCOME'].values)
data['MAKAN']=enc.fit_transform(data['MAKAN'].values)
data['TRANSPORT']=enc.fit_transform(data['TRANSPORT'].values)
data['LISTRIK']=enc.fit_transform(data['LISTRIK'].values)
data['PENDIDIKAN']=enc.fit_transform(data['PENDIDIKAN'].values)
data['SOCIAL_MAKANTEMAN']=enc.fit_transform(data['SOCIAL_MAKANTEMAN'].values)
data['SJJKELUARGA']=enc.fit_transform(data['SJJKELUARGA'].values)
data[' SMOBIL2 ']=enc.fit_transform(data[' SMOBIL2 '].values)
data[' STATUS ']=enc.fit_transform(data[' STATUS '].values)
data[' KESEHATAN ']=enc.fit_transform(data[' KESEHATAN '].values)
data['RUMAH']=enc.fit_transform(data['RUMAH'].values)
data[' DEPOSITO ']=enc.fit_transform(data[' DEPOSITO '].values)
data[' REKSADANA ']=enc.fit_transform(data[' REKSA (property) values: ArrayLike
data[' SAHAM ']=enc.fit_transform(data[' SAHAM '].values)
data[' CASH ']=enc.fit_transform(data[' CASH '].values)
data[' OBLIGASI ']=enc.fit_transform(data[' OBLIGASI '].values)
```

```
data
✓ 0.0s
```

	JK	UMUR	BANGSA	DOMISILI	JABATAN	INDUSTRI	INCOME	MAKAN	TRANSPORT	LISTRIK	RUMAH	PENDIDIKAN
0	0	2	0	5	20	39	6	2	0	0	5	4
1	0	2	1	5	19	21	8	0	0	3	1	5
2	0	2	1	5	10	30	4	2	1	1	1	4
3	0	1	0	5	1	9	3	0	0	1	3	4
4	0	2	1	5	10	7	4	1	0	2	1	5
...	...	...	...	...	...	...	...	...	...	...	...	...
143	1	0	0	5	16	31	3	0	3	1	3	4
144	1	0	0	5	16	36	3	1	2	1	3	4
145	1	0	0	5	10	1	3	0	0	1	1	4
146	0	1	0	0	10	31	7	1	0	2	5	3
147	1	1	0	5	15	3	3	1	1	2	5	4

JK	UMUR	BANGSA	DOMISILI	JABATAN	INDUSTRI	INCOME	MAKAN	TRANSPOF	LISTRIK	RUMAH	PENDIDIKA	SOCIAL_M	SJJKELUAR	SMOBIL2	STATUS
0	2	1	5	18	21	6	0	0	2	1	4	0	0	0	1
0	2	1	5	11	30	3	2	1	0	1	3	1	1	0	1
0	2	1	5	11	7	3	1	0	1	1	4	0	0	0	2
0	0	0	5	15	7	2	2	1	0	3	3	0	0	0	3
1	0	0	5	15	35	2	0	1	2	3	3	2	2	1	3
0	0	0	5	15	7	2	2	0	1	3	3	0	1	0	3
1	0	0	8	15	3	3	2	1	2	3	3	1	2	2	3
0	1	0	5	11	3	2	1	0	6	1	3	0	0	0	0
0	1	0	4	14	3	2	0	1	6	1	4	0	1	0	0
1	0	0	4	15	3	0	3	1	4	1	3	1	1	0	3
1	1	0	5	15	3	2	1	0	0	1	3	1	1	0	0
0	3	0	5	14	3	3	0	0	6	5	3	0	0	0	1
0	3	0	27	15	3	3	0	0	7	5	5	0	0	0	2
0	2	0	5	15	3	6	1	0	7	5	3	1	1	1	1
0	0	0	24	15	3	2	1	3	1	2	3	1	1	0	3
1	1	0	5	11	3	2	2	0	2	5	3	0	0	0	1
1	0	0	5	15	3	2	0	0	6	3	3	2	2	0	3
1	0	0	5	14	3	0	0	1	1	3	3	0	2	0	3
1	0	0	5	15	18	2	1	0	1	5	3	0	0	2	3
1	1	0	25	14	3	2	2	2	1	5	3	1	1	1	0
0	1	0	24	11	3	2	1	1	0	5	3	0	0	0	1
1	0	0	5	14	3	2	1	0	1	3	3	1	1	0	3
1	0	0	5	11	3	2	2	1	0	1	3	1	1	0	3
1	0	0	5	11	3	2	2	0	1	3	4	0	0	0	3
1	1	0	5	11	3	5	2	0	3	3	3	2	2	0	3
1	1	0	5	11	3	5	0	0	3	5	3	0	0	1	2
0	1	0	5	11	18	5	0	0	1	5	3	0	0	0	0
0	1	0	5	14	18	5	2	1	1	5	3	0	1	0	0
1	0	0	5	15	12	2	2	3	0	3	3	1	2	2	3

Table 2 . Data Encoding

RESULT

Modeling

After decoding process, it will be ready to make a modeling. The model which chooses based on the data types was the random forest.

```
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
from sklearn.tree import DecisionTreeClassifier
import sklearn.model_selection as ms
X=data[['JK','UMUR','BANGSA','DOMISILI','JABATAN','INDUSTRI','INCOME','MAKAN','TRANSPORT','LISTRIK','RUMAH',
X
y=data[' OBLIGASI ']
y
```

```
import sklearn.ensemble as ens
import sklearn.metrics as met
rf= ens.RandomForestClassifier(n_estimators=100)
X_train, X_test, y_train,y_test=ms.train_test_split(X,y,test_size=0.3)
rf.fit(X_train,y_train)
y_prediksi=rf.predict(X_test)
y_prediksi
```

✓ 0.0s

```
accuracy=met.accuracy_score(y_test,y_prediksi)
print('Accuracy= ',accuracy)
```

✓ 0.0s

```
Accuracy=  0.9833333333333333
```

```
print(met.classification_report(y_test,y_prediksi))
```

✓ 0.0s

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        34
           1       0.83      1.00      0.91         5
           2       1.00      0.75      0.86         4
           7       1.00      1.00      1.00        17

    accuracy                           0.98        60
   macro avg       0.96      0.94      0.94        60
weighted avg       0.99      0.98      0.98        60
```

After the modeling run, it will result the accuracy 98 %, so the model can predict 98 % how much of the bond will be taken from each industry. From table 3. The random forest algorithm has function , is called a feature important. The important feature is a formula that indicates, where factors that will influence the closing with bond product from insurance. The highest factor is DOMICILE (0,033) , this factor is reflected in style of life, the bigger the LISTRIK the more wealthier. The second factor is SMOBIL2, , the allocation for second car is 1-10 % and the second is 11-20 %, So the higher percentage allocation is more wealthier. The third factor is INDUSTRY, the most closing is become from Banking (57), Insurance (17), Education (15) , Retail (7), from table 4. Projection closing bond insurance from each industry. And the Third is DOMICILE , and the most recent from Jakarta (95), Depok (6) , Bogor (6) , Tangerang (5) , Surabaya (4) , so the distribution of insurance is dominated from JABODETABEK. Next from the age, the most closing is become from age 26 to 46 (62) , the second from age 17-25 (48). So in this care, the marketing must approach the candidate customer from age 26 to 46 years old. And the fourth analysis is , the position the most closed the bond insurance is from staff (68), manager (34) and senior manager (19). Finally most of respondents has own house (88), and house of family (20) , so if the respondents has house, it indicates that economic is more secure.

FEATURE	TOTAL
DOMISILI	0.033833536197078476
SMOBIL2	0.023583655398300152
INDUSTRI	0.02126256063027282
JABATAN	0.017109412613677624
RUMAH	0.015907206826568536

Table 3. Feature Important

Industry	Industry Code	Total
Banking	3	57
Insurance	18	17
Education	7	15
Retail	31	7
Agency	1	4
Food And Beverage	9	3
Bond	12	3
Manufactur	22	3
Telecommunication	36	3
Finance	8	2
Hospitality	15	2
kesehatan	39	2
Legal	21	2
Publishing	30	2

Table 4 Projection closing bond insurance from each industry

Jakarta	95
Depok	6
Bogor	6
Tangerang	5
Surabaya	4
Bekasi	4
Semarang	2
Medan	2

Table 5. The most contribution from Domicile

Age	Total
17 tahun - 25 tahun	48
26 tahun - 45 tahun	62
45 tahun - 55 tahun	24
lebih dari 55 tahun	9

Table 6. Distribution of Age

Position	Total
Staff	68
Manager	34
Senior Manager	19
Director	5
assistant manager	2
CEO	2
Art director	1
Dosen	1

Table 7. Distribution of Position

Persentase Alokasi Mobil ke 2	Frequency
1%-10%	119
11%-20%	28
21%-40%	11
None	1
Sudah ada	1
Tdk ada	2
tidak	1

Table 8 Allocaion for Second Car

Housing Type	Frequency
Milik Sendiri	88
Milik Keluarga	70
Kontrak	26
Kos Kosan	12
Apartment sewa	1
Milik Perusahaan	1

Table 9. Allocation for Housing Type

Evaluation

Projections for closing bond reveal the banking industry has the highest closure rate, followed by insurance, education, and retail. Subsequently, closures are seen in bond, manufacturing, finance, hospitality, legal, publishing, technology, government, and service industries. This pattern is attributed to financial literacy levels (Lopus et al., 2019).

CONCLUSION

While traditionally, insurance agents focused on understanding customer needs, leveraging customer behavior offers greater accuracy in predicting bond insurance premiums. Employing machine learning, specifically the random forest algorithm, enhances this predictive capability. Projections show the highest bond insurance closure rates within the banking industry, followed by insurance, education, and then retail. Subsequent closures occur in bond, manufacturing, finance, hospitality, legal, publishing, technology, government, and service industries, achieving a 98% accuracy score.

REFERENCES

Academy, R. (2024). *Data Science*. Rapidminer Academy. https://academy.rapidminer.com/learning-paths/data-science-professional-with-rapidminer

Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20–28. https://doi.org/10.38094/jastt20165

Duarte, V., Zuñiga-Jara, S., & Contreras, S. (2019). *Machine Learning and Marketing: a literature review*. https://ssrn.com/abstract=4006436

Gkikas, D. C., Theodoridis, P. K., & Beligiannis, G. N. (2022). Enhanced Marketing Decision Making for Consumer Behaviour Classification Using Binary Decision Trees and a Genetic Algorithm Wrapper. *Informatics*, *9*(2). https://doi.org/10.3390/informatics9020045

Lopus, J. S., Amidjono, D. S., & Grimes, P. W. (2019). Improving financial literacy of the poor and vulnerable in Indonesia: An empirical analysis. *International Review of Economics Education*, *32*(May), 100168. https://doi.org/10.1016/j.iree.2019.100168

Muhajir, M., & Widiastuti, J. (2022). Random Forest Method Approach to Customer Classification Based on Non-Performing Loan in Micro Business. *Jurnal Online Informatika*, *7*(2), 177–183. https://doi.org/10.15575/join.v7i2.842

Müller, A. C., & Guido, S. (2017). *Introduction to Machine Learning with Python A GUIDE FOR DATA SCIENTISTS Introduction to Machine Learning with Python*.

OJK, O. J. K. (2023). Draft Roadmap Pengembangan Perasuransian Indonesia. *Departemen Pengaturan Dan Pengembangan IKNB, OJK*, 1–63. www.ojk.go.id.

Ong, A. K. S., Cordova, L. N. Z., Longanilla, F. A. B., Caprecho, N. L., Javier, R. A. V., Borres, R. D., & German, J. D. (2023). Purchasing Intentions Analysis of Hybrid Cars Using Random Forest Classifier and Deep Learning. *World Electric Vehicle Journal*, *14*(8). https://doi.org/10.3390/wevj14080227

Singgalen, Y. A. (2024). Implementation of CRISP-DM for Social Network Analysis (SNA) of Tourism and Travel Vlog Content Reviews. *Jurnal Media Informatika Budidarma, 2024*, *8*, 572–583. https://doi.org/10.30865/mib.v8i1.7323

Winston, P. H. (2017). On Computing Machinery and Intelligence. *Boston Studies in the Philosophy and History of Science*, *324*, 265–278. https://doi.org/10.1007/978-3-319-53280-6_11