APPLICATION OF NAÏVE BAYES CLASSIFIER METHOD FOR PREDICTING CLAIMS IN AUTOMOBILE INSURANCE

UNDERGRADUATE THESIS

Submitted as one of the requirements to

obtain

Sarjana Aktuaria

By:

MICHELYNN SOLA GRATIA JIRENE

021202000019

FACULTY OF BUSINESS

ACTUARIAL SCIENCE STUDY PROGRAM

CIKARANG

NOVEMBER 2023

PANEL OF EXAMINERS APPROVAL

The Panel of Examiners declare that the undergraduate thesis entitled **Application of Naïve Bayes Classifier Method for Predicting Claims in Automobile Insurance** that was submitted by Michelynn Sola Gratia Jirene majoring in Actuarial Science from the Faculty of Business was assessed and approved to have passed the Oral Examinations on November 8th, 2023.

Dr. Edwin Setiawan Nugraha, S.Si, M.Sc.

Chair - Panel of Examiner

Maria Yus Trinity Irsan, S.Si.,M.Si.
Examiner I

Promoted by,

Dr. Dadang Amir Hamzah, M.Si.
Advisor

Recommended by,

Maria Yus Trinity Irsan, S.Si.,M.Si.
Head, Acturial Science Study Program

STATEMENT OF ORIGINALITY

In my capacity as an active student of President University and as the author of the thesis/final project/business plan stated below:

Name	: Michelynn Sola Gratia Jirene
Student ID number	: 021202000019
Study program	: Actuarial Science
Faculty	: Business

I hereby declare that my thesis/final project/business plan entitled "Application of Naïve Bayes Classifier Method for Predicting Claims in Automobile Insurance" is, to the best of my knowledge and belief, an original piece of work based on sound academic principles. If there is any plagiarism detected in this thesis/final project/business plan, I am willing to be personally responsible for the consequences of these acts of plagiarism, and will accept the sanctions against these acts in accordance with the rules and policies of President University.

I also declare that this work, either in whole or in part, has not been submitted to another university to obtain a degree.

Cikarang, October 18th, 2023

Michelynn Sola Gratia Jirene

SCIENTIFIC PUBLICATION APPROVAL FOR ACADEMIC INTEREST

As a student of the President University, I, the undersigned:

Name : Michelynn Sola Gratia Jirene

Student ID number : 021202000019

Study program : Actuarial Science

for the purpose of development of science and technology, certify, and approve to give President University a non-exclusive royalty-free right upon my final report with the title:

Application of Naïve Bayes Classifier Method for Predicting Claims in Automobile Insurance

With this non-exclusive royalty-free right, President University is entitled to converse, to convert, to manage in a database, to maintain, and to publish my final report. There are to be done with the obligation from President University to mention my name as the copyright owner of my final report.

This statement I made in truth.

Cikarang, November 8th, 2023

Michelynn Sola Gratia Jirene

ADVISOR APPROVAL FOR JOURNAL/INSTITUTION'S REPOSITORY

As an academic community member of the President's University, I, the undersigned:

Advisor Name	: Dr. Dadang Amir Hamzah, M.Si.
Employee ID number	: 0405108602
Study program	: Actuarial Science
Faculty	: Business

declare that following thesis:

Title of thesis	: Application of Naïve Bayes Classifier Method for Predicting Claims in Automobile Insurance
Thesis author	: Michelynn Sola Gratia Jirene
Student ID number	: 021202000019

will be published in journal / institution's repository / proceeding / unpublish.

Cikarang, October 18th, 2023

Dr. Dadang Amir Hamzah, M.Si.

PLAGIARISM REPORT

Application of Naive Bayes Classifier Method for Predicting Claims in Automobile Insurance

ORIGINALITY REPORT

13%	7%	5%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universidade Nova De Lisboa Student Paper	2%
2	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	1%
3	Submitted to Kaplan University Student Paper	<1%
4	Submitted to University of New South Wales Student Paper	<1%
5	Submitted to University of San Diego Student Paper	<1%
6	Efstathios Kirkos. "Audit-firm group appointment: an artificial intelligence approach", Intelligent Systems in Accounting Finance & Management, 2009 Publication	<1%
7	Submitted to Mercer University Student Paper	<1%

ANTI-PLAGIARISM REPORT

AI Scan

7%

This text is most likely to be written by **a human**

There is a **7%** probability this text was entirely written by AI

APPLICATION OF NAÏVE BAYES CLASSIFIER METHOD FOR PREDICTING CLAIMS IN AUTOMOBILE INSURANCE

By

Michelynn Sola Gratia Jirene

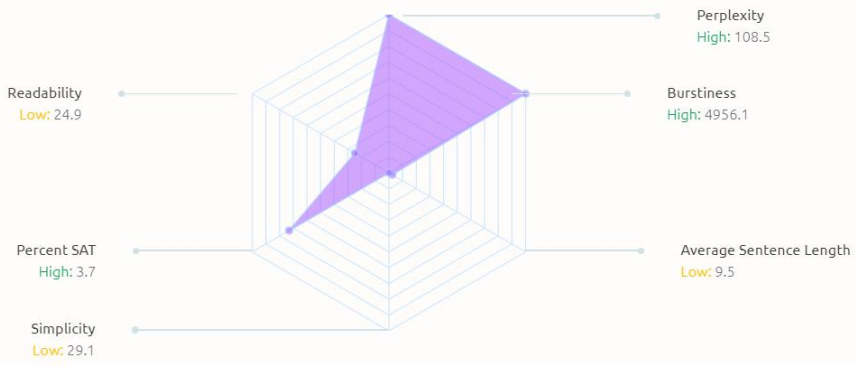
ID no.
021202000019

A Skripsi presented to the

Faculty of Business President Universityin partial fulfillment of the requirements forBachelor Degree in Actuarial Science

0/79 sentences are likely AI generated.

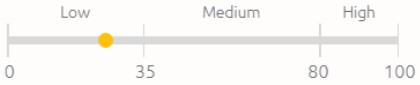
Writing Analysis



These measurements have been normalized on a scale of 1–100 for display on this chart.

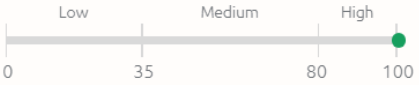
Readability: 24.9

Sentences with short words and low amount of syllables have high readability scores.



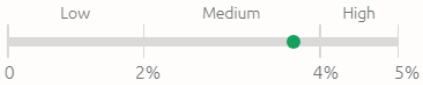
Perplexity: 108.5

How familiar a piece of text is to large language models like ChatGPT.



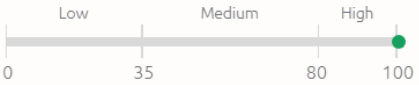
Percent SAT: 3.7%

Measures what percentage of words are SAT words, terms from a standardized college admissions exam known for its labyrinthine vocabulary lists.



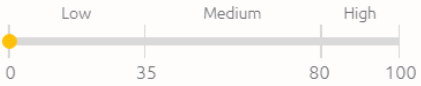
Burstiness: 4956.1

Unique score developed by GPTZero in 2022 that correlates to variance in writing. Humans generally vary their writing patterns over time.



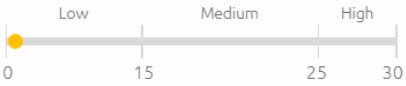
Simplicity: 29.1%

Measures what percentage of words are in the 100 most common words in the English language.



Average Sentence Length: 9.5 words

Unique score that correlates to variance in writing, where humans generally vary writing patterns.



ABSTRACT

Insurance companies often experience difficulties in planning funds. This is caused by the risk of uncertainty in life. Not least, the insurance company went bankrupt because they could not pay their obligations. For this reason, insurance companies need to know the right strategy for setting up reserve funds. One solution that can help insurance companies make decisions and determine strategies is to make claim prediction. In this study, the author will use the Naïve Bayes Classifier method to predict claims in automobile insurance. The Naïve Bayes Classifier itself is a simple probability method where the calculations are based on Bayes' Theorem. The data used is secondary data from Kaggle.com where this data consists of 10,000 samples with 19 features. The prediction results will be divided into two results, namely 0 and 1 where 0 means "no" and 1 means "yes". "yes" or "no" label will inform us whether the customer will claim or not. The data will go through preprocessing in python so that the format is appropriate. The model will be built without feature selection and with feature selection then being compared to determine the best model. Predicted data result will be compared to the actual data and the accuracy of the best model is 82%. Other evaluation method was applied to evaluate how well the model performed by using ROC – AUC score which has a score of 0.87 and 10-fold cross validation which has an average score of 80%. The result of the prediction will help insurance company with underwriting decision and financial approach or planning.

Keywords: *Claim Prediction, Naïve Bayes Classifier, Automobile Insurance*

ABSTRAK

Perusahaan asuransi umumnya mengalami kesulitan dalam perencanaan dana. Hal ini disebabkan oleh resiko ketidakpastian dalam hidup. Tidak sedikit, Perusahaan asuransi mengalami kebangkrutan karena tidak dapat membayar kewajiban mereka. Karena itu, Perusahaan asuransi perlu untuk mengetahui strategi yang tepat dalam mencadangkan dana. Salaah satu solusi yang dapat membantu perusahaan asuransi dalam mengambil keputusan dan menetapkan strategi adalah dengan melakukan prediksi klaim. Dalam penelitian ini, penulis akan menggunakan metode Naïve Bayes Classifier untuk memprediksi klaim pada asuransi automobile. Naïve Bayes Classifier sendiri merupakan metode probabilitas sederhana dimana perhitungannya berdasarkan Teorema Bayes. Data yang digunakan adalah data sekunder dari Kaggle.com dengan 10.000 sampel dan 19 fitur. Hasil prediksi akan dibagi menjadi 2 hasil yaitu 0 dan 1 dimana 0 berarti "tidak" dan 1 berarti "ya". Label "ya" dan "tidak" akan memberikan informasi apakah pelanggan akan melakukan klaim atau tidak. Data akan melalui proses preproses di python agar format sesuai. Model akan dibuat tanpa seleksi fitur dan dengan seleksi fitur lalu dibandingkan untuk menentukan model terbaik. Hasil prediksi data akan dibandingkan dengan data aktual dan akurasi pada model terbaik adalah 82%. Metode evaluasi lain digunakan untuk mengevaluasi seberapa baik performa model dengan menggunakan nilai ROC – AUC dengan nilai 0,87 dan 10-fold cross validation dengan nilai rata-rata 80%. Hasil prediksi akan membantu prerusahaan asuransi dalam keputusan underwriting dan rencana pendekatan finansial.

Kata kunci: Prediksi Klaim, Naïve Bayes Classifier, Asuransi Automobile

ACKNOWLEDGEMENT

First of all, the author would like to express her gratitude to Jesus Christ for His love, mercy and presence for the author is able to finish this research as a final requirement to accomplish bachelor's degree. The author would like to thank President University for the opportunity to study, learn, and obtain bachelor's degree. This thesis entitled "Application of Naïve Bayes Classifier Method for Predicting Claims in Automobile Insurance" cannot be finished without guidance and support from people around me. The author would like to give high appreciation and gratitude to these people:

1. Mrs. Maria Yus Trinity Irsan, S.Si.,M.Si., as head of Actuarial Science study program for her guidance and teaching throughout my study at President University.
2. Mr. Dr. Dadang Amir Hamzah, M.Si., as my thesis advisor who has spared his time and energy to guide me, supporting me, and giving me his advice and knowledge.
3. Lecturers from Actuarial Science study program who have shared their knowledge and teach me as I study at President University.
4. My father who has given me his advice and discussed with me in the process of making this thesis. And my mother who has given me love and support this whole time. My best gratitude for my parents who has given me a lot of prayers and support me mentally to lessen my worries and fear.
5. My best friend Vinesia Tjian who has given motivation from the very beginning until the end so that my days during the making of this thesis become less stressful.
6. Lastly, I would like to thank all the parties who cannot be mentioned one by one for their assistance and support during the completion of this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	: Artificial Intelligence
ML	: Machine Learning
NBC	: Naïve Bayes Classifier
ROC	: Receiver Operator Characteristic Curve
TP	: True Positive
TN	: True Negative
FP	: False Positive
FN	: False Negative
TPR	: True Positive Rate
FPR	: False Positive Rate
OJK	: Otoritas Jasa Keuangan
KDD	: Knowledgeable Discovery in Data
AUC	: Arena Under the Curve
DUIS	: Driving Under Influences
PDF	: Probability Density Function
CDF	: Cumulative Distribution Function