PERFORMANCE OF VARIOUS NAÏVE BAYES USING GRIDSEARCH APPROACH IN PHISHING EMAIL DATASET

JALUR SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi S1 Informatika

disusun oleh

RIZKI RAHMAN

20.11.3559

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2023

PERFORMANCE OF VARIOUS NAÏVE BAYES USING GRIDSEARCH APPROACH IN PHISHING EMAIL DATASET

JALUR SCIENTIST

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi S1 Informatika

disusun oleh

RIZKI RAHMAN

20.11.3559

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2023

HALAMAN PERSETUJUAN

JALUR SCIENTIST

PERFORMANCE OF VARIOUS NAÏVE BAYES USING GRIDSEARCH APPROACH IN PHISHING EMAIL DATASET

yang disusun dan diajukan oleh

Rizki Rahman

20.11.3559

telah disetujui oleh Dosen Pembimbing
pada tanggal 3 Oktober 2023

Dosen Pembimbing,

Ferian Fauzi Abdulloh, **M.Kom**
NIK. 190302276

HALAMAN PENGESAHAN

JALUR SCIENTIST

PERFORMANCE OF VARIOUS NAÏVE BAYES USING GRIDSEARCH APPROACH IN PHISHING EMAIL DATASET

yang disusun dan diajukan oleh

Rizki Rahman

20.11.3559

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Oktober 2023

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Ferian Fauzi Abdulloh, M.Kom **NIK. 190302276**	
Arif Akhsrul Huda, S.Si, M.Eng **NIK. 190302287**	
Erni Seniwati, S.Kom.,M.Cs. **NIK. 190302251**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Oktober 2023

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

Performance of Various Naïve Bayes Using GridSearch Approach In Phishing Email Dataset

Rizki Rahman[1)*], **Ferian Fauzi Abdulloh**[2)]
[1),2)]Universitas Amikom Yogyakarta, Yogyakarta, Indonesia
[1)]riskirahman1116@students.amikom.ac.id, [2)]ferian.fauzi@amikom.ac.id

Submitted : Sep 1, 2023 | **Accepted** : Sep 16, 2023 | **Published** : Oct 1, 2023

Abstract: The background is the increasing cybersecurity threats in the form of phishing attacks that can be detrimental to individuals and organizations. The purpose of this research is to compare the performance of four Naive Bayes variants in classifying phishing emails with a method that involves a data pre-processing stage, phishing emails are collected, cleaned, and converted into appropriate numerical features. Next, the GridSearch approach was used to find the best parameters. This research objective is to understand how each Naive Bayes variant works on phishing email datasets. This phishing detection task is based on the following performance evaluation criteria such as accuracy, precision, recall, and F1-score. In this study, Bernoulli got the best accuracy of 97.34% but when the results obtained a hyperparameter, the results showed an increase with the most optimal results and the best performance is Bernoulli 97.38%. The research results are to provide an in-depth insight into the effectiveness of each variant of Naive Bayes in dealing with phishing email datasets and researchers in selecting the most suitable Naive Bayes variant for phishing detection tasks. In addition, the applied GridSearch method can guide how to find the best parameters for Naïve Bayes models in other contexts. In summary, this study focuses on analyzing the performance of four variants of Naive Bayes Gaussian, Multinomial, Complement, and Bernoulli with the best algorithms Bernoulli 97.38%.

Keywords: Bernoulli, Gaussian, GridSearch, Naïve Bayes, Phishing Email