

ABSTRAK

Nadya Pratiwi Doholio, 2025. *PERBANDINGAN EKSTRAKSI FITUR WORD2VEC DAN COUNTVECTORIZER DENGAN SELEKSI FITUR MUTUAL INFORMATION PADA SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI OPINI PUBLIK DI X(TWITTER).* **SKRIPSI.** Gorontalo. Program Studi Statistika. Jurusan Matematika. Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Negeri Gorontalo.

Pembimbing : **(1) Isran K. Hasan, S.Pd., M.Si.** **(2) Siti Nurmardia Abdussamad, S.Stat., M.Stat.**

Media sosial, khususnya X (Twitter), tidak hanya berfungsi sebagai sarana komunikasi, tetapi juga sebagai platform untuk bertukar pendapat, terutama terkait isu-isu yang bersifat kontroversial seperti perpindahan Ibu Kota Negara (IKN). Untuk memahami opini publik mengenai isu tersebut, dibutuhkan analisis sentimen. Salah satu metode yang sering digunakan untuk klasifikasi sentimen adalah *Support Vector Machine* (SVM), yang dapat meningkatkan kinerjanya jika didukung oleh fitur yang berkualitas. Sehingga diperlukan proses ekstraksi maupun seleksi fitur untuk meningkatkan akurasi klasifikasi SVM. Pemilihan ekstraksi fitur yang tepat sangat penting untuk klasifikasi. Oleh karena itu, penelitian ini bertujuan membandingkan dua metode ekstraksi fitur, yaitu *Word2vec* dan *CountVectorizer*, dengan penambahan seleksi fitur *Mutual Information* untuk meningkatkan performa SVM dalam mengklasifikasikan sentimen publik di media sosial X. Penelitian ini menggunakan data sekunder, yaitu data cuitan X yang diambil menggunakan *package Nitter* pada pemrograman Python dengan kata kunci "perpindahan IKN" dan "perpindahan Ibu Kota Negara" dalam rentang waktu sejak 17 Agustus 2019 - 17 Agustus 2024. Hasil penelitian menunjukkan bahwa SVM dengan ekstraksi fitur *Word2vec* dan seleksi fitur *Mutual Information* menghasilkan kinerja lebih baik dengan akurasi 84%, presisi 90%, *recall* 90%, dan *f1-score* 90%, dibandingkan SVM dengan ekstraksi fitur *CountVectorizer* dan seleksi fitur *Mutual Information* dengan akurasi 80%, presisi 83%, *recall* 92%, dan *f1-score* 87%.

Kata Kunci: Perpindahan IKN, *Word2vec*, *CountVectorizer*, *Mutual Information*, *Support Vector Machine*

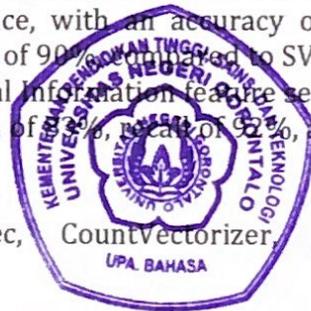
ABSTRACT

Nadya Pratiwi Doholio, 2025. *COMPARISON OF WORD2VEC AND COUNTVECTORIZER FEATURE EXTRACTION WITH MUTUAL INFORMATION FEATURE SELECTION ON SUPPORT VECTOR MACHINE FOR PUBLIC OPINION CLASSIFICATION ON X (TWITTER).* **Undergraduate Thesis.** Gorontalo. Study Program of Statistics. Department of Mathematics. Faculty of Mathematics and Natural Sciences. Universitas Negeri Gorontalo.

The Supervisors: **(1) Isran K. Hasan, S.Pd.,M.Si. (2) Siti Nurmardia Abdussamad, S.Stat., M.Stat.**

Social media, particularly X (Twitter), serves not only as a means of communication but also as a platform for exchanging opinions, especially regarding controversial issues such as the relocation of the IKN (*Ibu Kota Nusantara* or the Nusantara Capital City). To understand public opinion on this issue, sentiment analysis is required. One commonly used method for sentiment classification is the Support Vector Machine (SVM), and its performance can be improved with high-quality features. Therefore, feature extraction and selection processes are needed to enhance SVM classification accuracy. The choice of an appropriate feature extraction method is crucial for classification. Hence, this research aims to compare two feature extraction methods, namely Word2vec and CountVectorizer, with the addition of Mutual Information feature selection to improve SVM performance in classifying public sentiment on X. This research utilizes secondary data, specifically tweets from X, collected using the Nitter package in Python programming with the keywords "*perpindahan IKN*" and "*perpindahan Ibu Kota Negara*" (the Relocation of the Nusantara Capital City) within the period from August 17, 2019, to August 17, 2024. The findings indicate that SVM with Word2vec feature extraction and Mutual Information feature selection achieves better performance, with an accuracy of 84%, precision of 90%, recall of 90%, and F1-score of 90%. SVM with CountVectorizer feature extraction and Mutual Information feature selection, which achieves an accuracy of 80%, precision of 85%, recall of 85%, and F1-score of 87%.

Keywords: *Perpindahan IKN*, Word2vec, CountVectorizer, Mutual Information, Support Vector Machine



LEMBAR PENGESAHAN

Skripsi yang berjudul "PERBANDINGAN EKSTRAKSI FITUR *WORD2VEC*
DAN *COUNTVECTORIZER* DENGAN SELEKSI FITUR *MUTUAL INFORMATION*
PADA *SUPPORT VECTOR MACHINE* UNTUK KLASIFIKASI OPINI PUBLIK DI
X(TWITTER)"

Oleh

NADYA PRATIWI DOHOLIO
NIM. 413421010

Program Studi Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam

Telah dipertahankan di depan dewan penguji

Hari, tanggal : Kamis, 13 Februari 2025

Waktu : 09.00-10.00 WITA

Tempat : Ruang Sidang Statistika Lt. 3, Kampus 4 UNG

A Dewan Pembimbing

1. **Isran K. Hasan, S.Pd., M.Si.** Pembimbing I
NIP. 199012112019031009

2. **Siti N. Abdussamad, S.Stat., M.Stat.** Pembimbing II
NIP. 199503042024212001

B Dewan Penguji

3. **Djihad Wungguli, S.Pd., M.Si.** Penguji I
NIP. 198906122019031018

4. **Nisky Imansyah Yahya, S.Pd., M.Si.** Penguji II
NIP. 199107302020121008

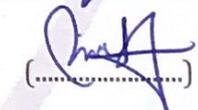
Tanda Tangan





Tanda Tangan





Mengetahui,

Dekan Fakultas Matematika dan IPA



Prof. Dr. Fitryane Lihawa, M.Si

NIP.196912091993032001