

**ANALISIS SENTIMEN KOMENTAR YOUTUBE MENGENAI
IBU KOTA NUSANTARA MENGGUNAKAN PENDEKATAN
SVM DAN RANDOM FOREST**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

FAKHRIYANNUR ADHA

21.11.3932

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**ANALISIS SENTIMEN KOMENTAR YOUTUBE MENGENAI
IBU KOTA NUSANTARA MENGGUNAKAN PENDEKATAN
SVM DAN RANDOM FOREST**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

FAKHRIYANNUR ADHA

21.11.3932

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024**

**ANALISIS SENTIMEN KOMENTAR YOUTUBE MENGENAI
IBU KOTA NUSANTARA MENGGUNAKAN PENDEKATAN
SVM DAN RANDOM FOREST**

yang disusun dan diajukan oleh

Fakhriyannur Adha

21.11.3932

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 19 Desember 2024

Dosen Pembimbing,



Ferian Fauzi Abdellah, M.Kom
NIK. 190302276

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS SENTIMEN KOMENTAR YOUTUBE MENGENAI IBU KOTA
NUSANTARA MENGGUNAKAN PENDEKATAN
SVM DAN RANDOM FOREST

yang disusun dan diajukan oleh

Fakhriyannur Adha

21.11.3932

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Desember 2024

Susunan Dewan Penguji

Nama Penguji

Robert Marco, S.T., M.T., Ph.D
NIK. 190302228

Ria Andriani, S.Kom., M.Kom
NIK. 190302458

Ferian Fauzi Abdullah, S.Kom., M.Kom
NIK. 190302276

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Desember 2024

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Fakhriyannur Adha
NIM : 21.11.3932

Menyatakan bahwa Skripsi dengan judul berikut:

**ANALISIS SENTIMEN KOMENTAR YOUTUBE MENGENAI IBU KOTA
NUSANTARA MENGGUNAKAN PENDEKATAN SVM DAN RANDOM
FOREST**

Dosen Pembimbing : Feriam Fauzi Abdulloh, M.Kom

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan **gagasan, rumusan dan penelitian SAYA sendiri**, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 19 Desember 2024

Yang Menyatakan,



Fakhriyannur Adha

HALAMAN PERSEMBAHAN

Alhamdulillah, segala puji dan syukur hanya kepada Allah SWT yang telah memberikan petunjuk, rahmat, dan kemudahan kepada penulis dalam menyelesaikan skripsi ini. Dengan penuh rasa terima kasih, penulis mempersembahkan skripsi ini kepada semua pihak yang telah memberikan dukungan, motivasi, serta semangat yang luar biasa, sehingga penulisan skripsi ini dapat terselesaikan dengan baik.

Melalui skripsi ini penulis mempersembahkan terima kasih kepada :

1. Dengan penuh rasa syukur kepada Allah SWT atas rahmat dan karuniaNya, penulis diberikan kemampuan dan kemudahan untuk menyelesaikan skripsi ini.
2. Terima kasih yang sebesar-besarnya kepada Ayah, Bapak Sultan Hasanuddin, atas inspirasi, semangat, dan doa yang tiada henti. Terima kasih juga atas kepercayaan yang diberikan untuk menyelesaikan studi hingga meraih gelar sarjana. Ayah adalah sumber kekuatan dan motivasi saya.
3. Terima kasih yang sebesar-besarnya kepada Ibu Fitriani, atas kasih sayang, semangat, doa, dan nasehat yang tak pernah putus. Terima kasih atas kesabaran dan kebesaran hati dalam mendampingi. Ibu adalah pengingat dan penguat yang luar biasa, serta tempatku untuk selalu pulang.
4. Ketiga kakakku, Rany, Nofia, dan Raihan. Terimakasih sudah ikut serta dalam proses penulis menempuh Pendidikan selama ini, terimakasih atas semangat, do'a dan cinta yang selalu diberikan kepada penulis. Terus berkembang dan meraih segala yang terbaik dalam hidup kakak-kakakku.
5. Sahabat kuliah saya yang telah menemani, memberikan dukungan, serta mendengarkan keluh kesah saya, yang selalu ada dalam setiap langkah perjalanan.
6. Terima kasih kepada diri sendiri yang telah tetap kuat dan berjuang menyelesaikan skripsi ini dengan tepat waktu.

KATA PENGANTAR

Alhamdulillah rabbil 'alamin, segala puji dan Syukur penulis panjatkan atas kehadiran Allah subhanahu wa ta 'ala segala rahmat dan karunia-Nya sehingga penulis mampu menyelesaikan skripsi dengan judul "ANALISIS SENTIMEN KOMENTAR YOUTUBE MENGENAI IBU KOTA NUSANTARA MENGGUNAKAN PENDEKATAN SVM DAN RANDOM FOREST" dengan tepat waktu. Skripsi ini ditulis untuk memenuhi syarat kelulusan untuk mencapai Gelar Sarjana pada Fakultas Ilmu Komputer Program Studi Informatika Universitas Amikom Yogyakarta.

Dalam menyelesaikan penulisan skripsi ini, penulis banyak memperoleh bantuan baik, pengajaran dan arahan dari berbagai pihak secara langsung maupun tidak langsung. Dengan ini penulis menyampaikan terimakasih kepada:

1. Bapak prof. Dr. M. Suyanto, M.M. selaku rector Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph.D. selaku Dekan fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Windha Mega Pradnya D, M.Kom. selaku ketua program Studi SI Informatika Universitas Amikom Yogyakarta.
4. Bapak Ferian Fauzi Abdulloh, M.Kom selaku dosen pembimbing saya yang telah memberikan waktu, saran, kritik, motivasi dan bimbingan dalam skripsi ini sehingga dapat menyelesaikan skripsi selesai dengan baik.
5. Bapak Sultan Hasanuddin dan Ibu Fitriani selaku orang tua penulis yang memberikan dukungan, doa dan motivasi kepada penulis.

Penulis menyadari bahwa skripsi ini masih banyak kekurangan karena keterbatasan pengetahuan yang dimiliki oleh penulis. Oleh karena, itu kritik dan saran sangat diharapkan bagi perbaikan penelitian selanjutnya.

Yogyakarta, 9 Desember 2024

Fakhriyannur Adha

DAFTAR ISI

| | |
|---|-------|
| COVER HITAM PUTIH | i |
| COVER WARNA | ii |
| HALAMAN PERSETUJUAN..... | iii |
| HALAMAN PENGESAHAN | iv |
| HALAMAN PERNYATAAN KEASLIAN SKRIPSI | v |
| HALAMAN PERSEMBAHAN | vi |
| KATA PENGANTAR | vii |
| DAFTAR ISI..... | viii |
| DAFTAR TABEL..... | xi |
| DAFTAR GAMBAR..... | xii |
| DAFTAR LAMPIRAN..... | xv |
| DAFTAR LAMBANG DAN SINGKATAN | xvi |
| DAFTAR ISTILAH | xvii |
| INTISARI | xviii |
| <i>ABSTRACT</i> | xix |
| BAB I PENDAHULUAN..... | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 3 |
| 1.3 Batasan Masalah | 3 |
| 1.4 Tujuan Penelitian | 4 |
| 1.5 Manfaat Penelitian | 4 |
| 1.6 Sistematika Penulisan | 5 |
| BAB II TINJAUAN PUSTAKA | 7 |

| | | |
|-----------------------------------|-------------------------------------|----|
| 2.1 | Studi Literatur | 7 |
| 2.1 | Dasar Teori..... | 14 |
| 2.2.1 | Analisis Sentimen | 14 |
| 2.2.1 | Youtube..... | 14 |
| 2.2.2 | Text Mining | 14 |
| 2.2.3 | <i>Support Vector Machine</i> | 15 |
| 2.2.4 | Adaptive Boosting (Adabost) | 16 |
| 2.2.5 | Bagging | 16 |
| 2.2.6 | <i>Random forest</i> | 17 |
| BAB III METODE PENELITIAN | | 19 |
| 3.1 | Dataset..... | 19 |
| 3.2 | Alur Penelitian | 21 |
| 3.2.1 | Scraping Data..... | 22 |
| 3.2.2 | Labeling Data..... | 22 |
| 3.2.3 | Analisis Data Eksplorasi..... | 23 |
| 3.2.4 | Preprocessing Data..... | 23 |
| 3.2.5 | Wordeloud..... | 24 |
| 3.2.6 | Pemodelan Data | 25 |
| 3.3 | Alat dan Bahan..... | 30 |
| 3.3.1 | Alat Penelitian..... | 30 |
| 3.3.2 | Bahan Penelitian | 30 |
| BAB IV HASIL DAN PEMBAHASAN | | 31 |
| 4.1 | Pengambilan Data | 31 |
| 4.2 | Preprocessing Data..... | 33 |
| 4.2.1 | Normalization | 33 |

| | | |
|---------------------|--|----|
| 4.2.2 | Case Folding | 34 |
| 4.2.3 | Tokenizing | 35 |
| 4.2.4 | Stopword Removal..... | 36 |
| 4.2.5 | Stemming | 36 |
| 4.2.6 | Visualisasi Wordcloud | 37 |
| 4.3 | Ekstraksi Fitur Menggunakan TF-IDF..... | 38 |
| 4.3.1 | TF-IDF Pada K-fold..... | 39 |
| 4.3.2 | TF-IDF Pada Data <i>Training</i> 80% dan Data <i>Testing</i> 20% | 39 |
| 4.4 | Evaluasi..... | 40 |
| 4.5.1 | Confusion Matrix pada K-fold..... | 40 |
| 4.5.2 | Confusion Matrix pada Data <i>Training</i> 80% dan Data <i>Testing</i> 20% 55 | |
| 4.5 | Hasil..... | 65 |
| 4.5.1 | Modelling dengan K-fold | 65 |
| 4.5.2 | Modelling dengan Data <i>Training</i> 80% dan Data <i>Testing</i> 20% | 67 |
| BAB V PENUTUP | | 70 |
| 5.1 | Kesimpulan | 70 |
| 5.2 | Saran | 71 |
| REFERENSI | | 72 |
| LAMPIRAN..... | | 75 |

DAFTAR TABEL

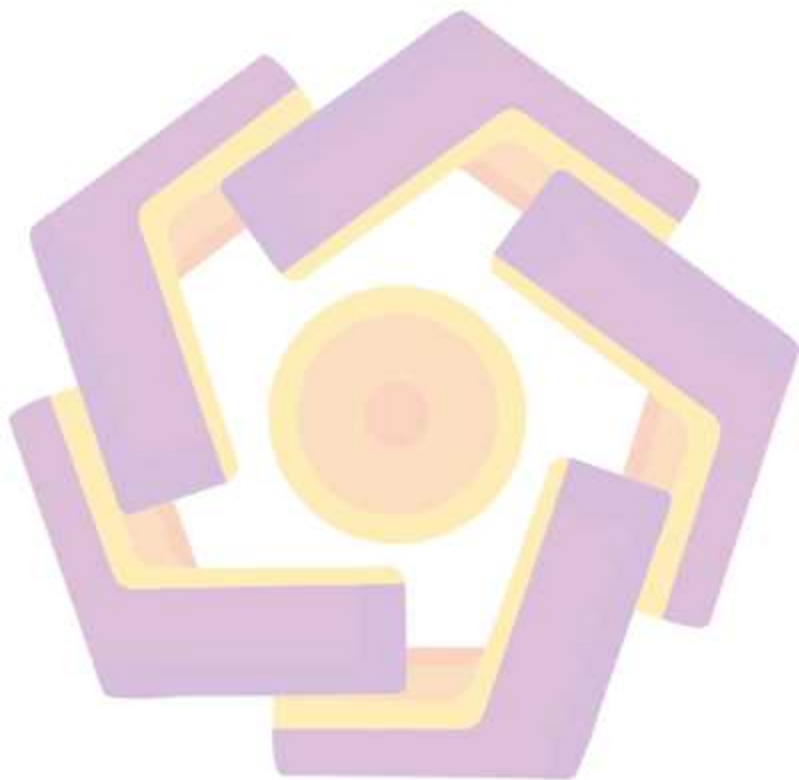
| | |
|--|----|
| Tabel 2. 1 - Keaslian Penelitian | 10 |
| Tabel 3. 1 - Daftar video sumber data | 19 |
| Tabel 3. 2 - Hasil Labeling Data | 22 |
| Tabel 3. 3 - Data Training dan Data Testing | 29 |
| Tabel 3. 4 - Contoh Dataset | 30 |
| Tabel 4. 1 - Data hasil Scrapping | 31 |
| Tabel 4. 2 - Data hasil perapihan data dan labeling | 32 |
| Tabel 4. 3 - Jumlah Dataset Berdasarkan Sentiment | 32 |
| Tabel 4. 4 - Contoh kata normalisasi | 33 |
| Tabel 4. 5 - Hasil dari proses Normalisasi | 34 |
| Tabel 4. 6 - Case Folding | 34 |
| Tabel 4. 7 - Tokenizing | 35 |
| Tabel 4. 8 - Stopword Removal | 36 |
| Tabel 4. 9 - Stemming | 37 |
| Tabel 4. 10 - Tabel Hasil Akurasi Pemodelan Pembagian Data Rasio 80:20 | 68 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 2. 1 - Cara kerja <i>Random forest</i> | 17 |
| Gambar 3. 1 - Grafik Jumlah Dataset | 21 |
| Gambar 3. 2 - Alur Penelitian | 22 |
| Gambar 3. 3- Ilustrasi <i>Support Vector Machine</i> | 25 |
| Gambar 3. 4 - Alur kerja algoritma <i>SVM</i> | 26 |
| Gambar 3. 5 - Alur kerja <i>SVM</i> Rbf dan <i>Random forest</i> | 27 |
| Gambar 4. 1 - Grafik Sentimen | 32 |
| Gambar 4. 2 – Kode dari proses Normalisasi | 33 |
| Gambar 4. 3 – Kode dari proses Case Folding | 34 |
| Gambar 4. 4 – Kode dari proses Tokenizing | 35 |
| Gambar 4. 5 – Kode dari proses Stopword Removal | 36 |
| Gambar 4. 6 – Kode dari proses Stemming | 37 |
| Gambar 4. 7 – Wordcloud Positif | 38 |
| Gambar 4. 8 – Wordcloud Negatif | 38 |
| Gambar 4. 9 - Split data pada K-Fold | 39 |
| Gambar 4. 10 - TF-IDF pada K-Fold | 39 |
| Gambar 4. 11 - TF-IDF pada Data Train dan Data Test | 39 |
| Gambar 4. 12 - Split Data untuk Modelling Data Train dan Testing | 40 |
| Gambar 4. 13 - Kode untuk Confusion Matrix dan Metrik Evaluasi pada <i>SVM</i> RBF | 41 |
| Gambar 4. 14 – Hasil Confusion Matrix dan Metrik Evaluasi 5 Fold pada <i>SVM</i> RBF | 42 |
| Gambar 4. 15 - Kode untuk Confusion Matrix dan Metrik Evaluasi pada <i>SVM</i> Linear | 43 |
| Gambar 4. 16 – Hasil Confusion Matrix dan Metrik Evaluasi 5 Fold pada <i>SVM</i> Linear | 45 |
| Gambar 4. 17 - Kode untuk Confusion Matrix dan Metrik Evaluasi pada <i>SVM</i> Polynomial | 46 |
| Gambar 4. 18 – Hasil Confusion Matrix dan Metrik Evaluasi 5 Fold pada <i>SVM</i> Polynomial | 47 |

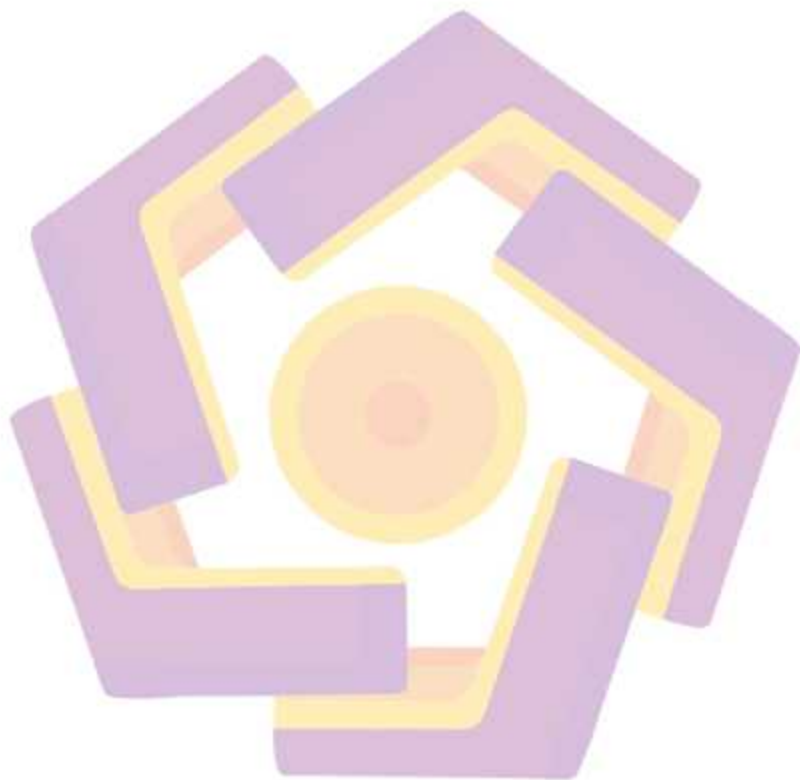
| | |
|---|----|
| Gambar 4. 19 - Kode Confusion Matrix dan Metrik Evaluasi pada Metode Adaboost | 48 |
| Gambar 4. 20 – Hasil Confusion Matrix dan Metrik Evaluasi 5 Fold pada Adaboost | 50 |
| Gambar 4. 21 - Kode Confusion Matrix dan Metrik Evaluasi pada Metode Bagging | 51 |
| Gambar 4. 22 - Hasil Confusion Matrix dan Metrik Evaluasi 5 Fold pada Bagging | 52 |
| Gambar 4. 23 - Kode Confusion Matrix dan Metrik Evaluasi pada Metode <i>Random forest</i> | 53 |
| Gambar 4. 24 - Hasil Confusion Matrix dan Metrik Evaluasi 5 Fold pada <i>Random forest</i> | 55 |
| Gambar 4. 25 – Kode Confusion Matrix dan Metrik Evaluasi pada <i>SVM</i> RBF | 56 |
| Gambar 4. 26 - Hasil Confusion Matrix dan Metrik Evaluasi pada <i>SVM</i> RBF | 57 |
| Gambar 4. 27 - Kode Confusion Matrix dan Metrik Evaluasi pada <i>SVM</i> Linear | 58 |
| Gambar 4. 28 - Hasil Confusion Matrix dan Metrik Evaluasi pada <i>SVM</i> Linear | 58 |
| Gambar 4. 29 - Kode Confusion Matrix dan Metrik Evaluasi pada <i>SVM</i> Polynomial | 59 |
| Gambar 4. 30 - Hasil Confusion Matrix dan Metrik Evaluasi pada <i>SVM</i> Polynomial | 60 |
| Gambar 4. 31 - Kode Confusion Matrix dan Metrik Evaluasi pada Adaboost | 61 |
| Gambar 4. 32 - Hasil Confusion Matrix dan Metrik Evaluasi pada Adaboost | 61 |
| Gambar 4. 33 - Kode Confusion Matrix dan Metrik Evaluasi pada Bagging | 62 |
| Gambar 4. 34 - Hasil Confusion Matrix dan Metrik Evaluasi pada Bagging | 63 |
| Gambar 4. 35 - Kode Confusion Matrix dan Metrik Evaluasi pada <i>Random forest</i> | 64 |
| Gambar 4. 36 - Hasil Confusion Matrix dan Metrik Evaluasi pada <i>Random forest</i> | 65 |
| Gambar 4. 37 - Perbandingan Akurasi Model pada K-Fold Cross Validation | 66 |
| Gambar 4. 38 - Hasil Akurasi Fold dari Setiap Metode dengan <i>SVM</i> RBF,Linear,Polynomial,Adaboost,Bagging, dan <i>Random forest</i> | 66 |

| | |
|--|----|
| Gambar 4. 39 - Grafik Train Akurasi dan Test Akuransi dengan Pemodelan Pembagian Data Rasio 80:20 | 67 |
|--|----|



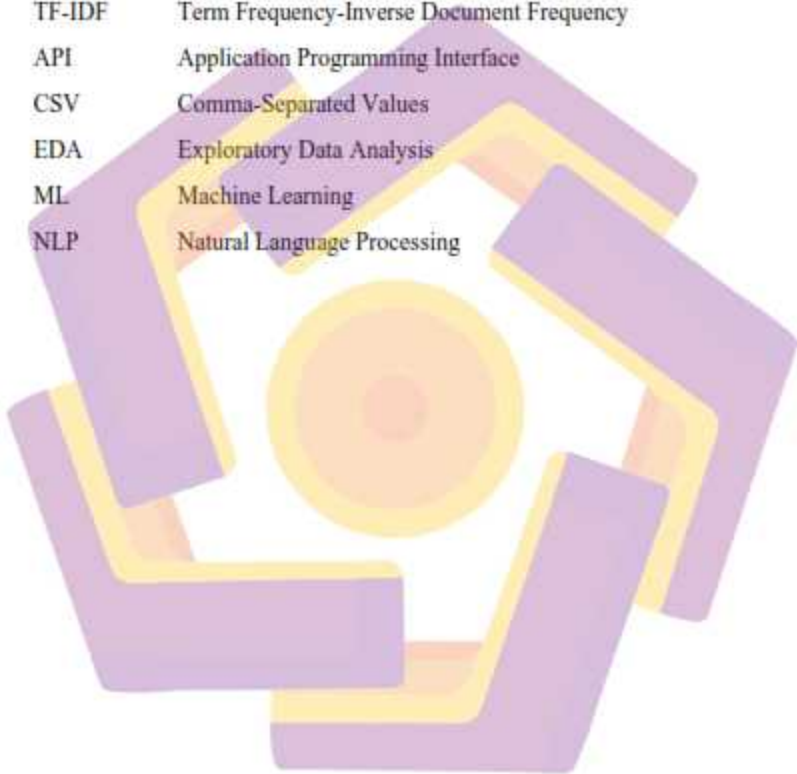
DAFTAR LAMPIRAN

| | |
|-------------------------------------|----|
| Lampiran 1. Profil obyek Penelitian | 10 |
| Lampiran 2. Dokumentasi Penelitian | 11 |



DAFTAR LAMBANG DAN SINGKATAN

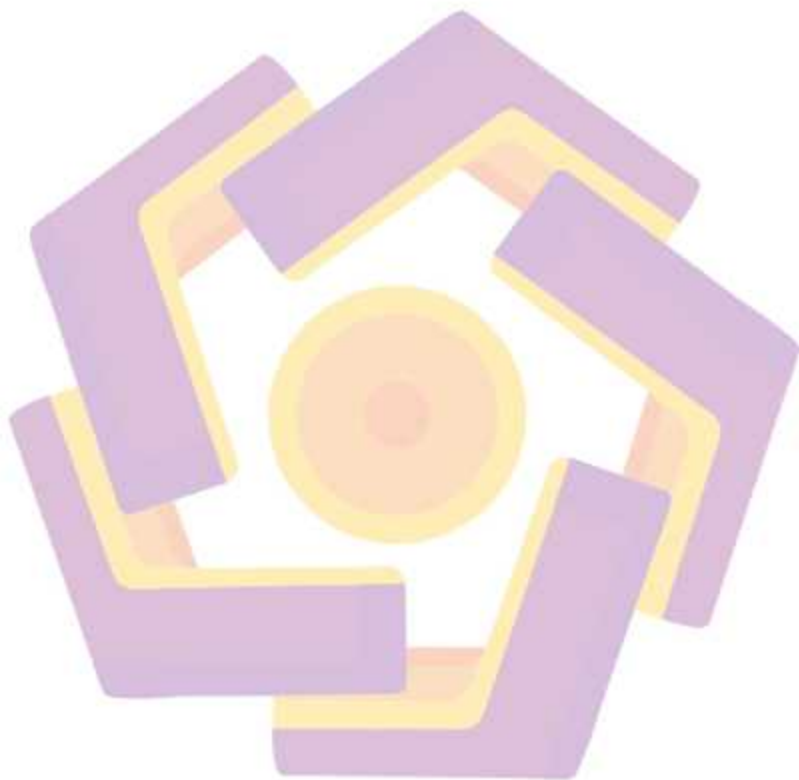
| | |
|--------|---|
| IKN | Ibu Kota Nusantara |
| SVM | Support Vector Machine |
| RBF | Radial Basis Function |
| TF-IDF | Term Frequency-Inverse Document Frequency |
| API | Application Programming Interface |
| CSV | Comma-Separated Values |
| EDA | Exploratory Data Analysis |
| ML | Machine Learning |
| NLP | Natural Language Processing |



DAFTAR ISTILAH

Wordcloud

Representasi visual dari kata-kata dalam teks



INTISARI

Penelitian ini bertujuan untuk menganalisis sentimen komentar pada video YouTube mengenai pemindahan ibu kota negara Indonesia ke Nusantara. Masalah yang diangkat adalah kurangnya pemahaman mengenai opini publik terhadap pembangunan IKN, yang dapat mempengaruhi keputusan kebijakan pemerintah. Dampaknya, informasi yang tidak akurat atau kurang representatif dapat menyesatkan publik dan pengambilan kebijakan.

Untuk menyelesaikan masalah ini, penelitian menggunakan metode analisis sentimen dengan mengaplikasikan teknik machine learning, yaitu kombinasi algoritma Support Vector Machine (SVM) dengan kernel Radial Basis Function (RBF) dan Random Forest. Data yang digunakan adalah komentar-komentar YouTube yang dikategorikan ke dalam sentimen positif dan negatif.

Hasil penelitian menunjukkan bahwa metode pendekatan SVM dan Random Forest memberikan akurasi yang baik dalam mengidentifikasi sentimen publik terhadap pemindahan ibu kota. Penelitian ini bermanfaat bagi pengambilan kebijakan dalam memahami opini publik serta bagi pengembangan teknologi analisis sentimen di masa depan. Penelitian lebih lanjut dapat dilakukan dengan memperluas data atau menerapkan metode lain untuk memperbaiki akurasi analisis sentimen.

Kata Kunci: analisis sentimen, YouTube, SVM, Random Forest, IKN

ABSTRACT

This study aims to analyze the sentiment of comments on YouTube videos regarding the relocation of Indonesia's capital to Nusantara. The problem addressed is the lack of understanding of public opinion about the development of IKN, which can affect government policy decisions. As a result, inaccurate or unrepresentative information may mislead the public and policymakers.

To address this issue, the research utilizes sentiment analysis with a machine learning approach, specifically combining the Support Vector Machine (SVM) algorithm with the Radial Basis Function (RBF) kernel and Random Forest. The data used consists of YouTube comments, which are categorized into positive and negative sentiments.

The results show that the approach method, SVM and Random Forest, provides good accuracy in identifying public sentiment toward the capital relocation. This study is beneficial for policymakers in understanding public opinion and for the future development of sentiment analysis technology. Further research could expand the data or apply other methods to improve sentiment analysis accuracy.

Keywords: *sentiment analysis, YouTube, SVM, Random Forest, IKN*