

**ANALISIS SENTIMEN PUBLIK TERHADAP PROGRAM KIP-KULIAH  
MENGUNAKAN ALGORITMA RANDOM FOREST PADA MEDIA  
SOSIAL X**

SKRIPSI

Program Studi Sistem Informasi (Profesional)

Jenjang Strata-1



Oleh :

Rosdiyanah  
NIM : 09031482326020

**PROGRAM STUDI SISTEM INFORMASI PROFESIONAL  
JURUSAN SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS SRIWIJAYA**

**2025**

**HALAMAN PENGESAHAN**

**PROJEK AKHIR**

**ANALISIS SENTIMEN PUBLIK TERHADAP PROGRAM KIP-KULIAH  
MENGUNAKAN ALGORITMA RANDOM FOREST PADA MEDIA  
SOSIAL X**

Sebagai salah satu syarat untuk penyelesaian studi di

Program Studi S1 Sistem Informasi

Oleh:

**ROSDIYANAH**

**09031482326020**

**Pembimbing 1 : Dinda Lestarini, S.SI., M.T.**  
**NIP. 198912222019032022**

**Mengetahui**

**Ketua Jurusan Sistem Informasi**



**Ahmad Rifai, S.T., M.T.**  
**1979102010121003**

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Rosdiyanah

NIM : 09031482326020

Jurusan : Sistem Informasi

Program Studi : Sistem Informasi Profesional

Judul TA. : Analisis Sentimen Publik Terhadap Program KIP-Kuliah  
Menggunakan Algoritma Random Forest pada Media Sosial X

Hasil Pengecekan *Software Authenticate*/Turnitin: 2%

Menyatakan bahwa laporan skripsi saya merupakan hasil karya saya sendiri dan bukan penjiplakan plagiat. Apabila ditemukan unsur penjiplakan plagiat dalam laporan skripsi saya ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari pihak manapun.



Palembang, 18 Juni 2025



  
**Rosdiyanah**  
**NIM. 09031482326020**

## HALAMAN PERNYATAAN INTEGRITAS

Yang bertanda tangan di bawah ini :

Nama : Rosdiyanah

NIM : 09031482326020

Jurusan : Sistem Informasi

Program Studi : Sistem Informasi Profesional

Judul : Analisis Sentimen Publik Terhadap Program KIP-Kuliah  
Menggunakan Algoritma Random Forest pada Media Sosial X

DOI : <https://doi.org/10.47065/bits.v7i1.7366>

Dengan ini menyatakan bahwa publikasi saya dengan judul :

Analisis Sentimen Publik Terhadap Program KIP-Kuliah Menggunakan Algoritma Random Forest pada Media Sosial X.

Yang diusulkan pada Vol. 7, No 1, Juni (2025): Jurnal *Building of Informatics, Technology and Science* (BITS) Halaman 264–274 bersifat original dan saya sendiri yang bertanggung jawab pada setiap proses submit publikasi tersebut. Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Dengan pernyataan dibuat dengan sesungguhnya dan sebenarnya.

Palembang, 16 Juni 2025

Menyetujui  
Dosen pembimbing,



**Dinda Lestarini, S.SI., M.T.**  
**NIP. 19891222019032022**

Yang menyatakan,



**Rosdiyanah**  
**NIM. 09031482326020**

## HALAMAN PERSETUJUAN TELAH DI UJI

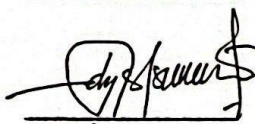
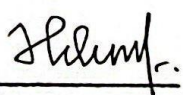
Telah diuji dan lulus pada :

Hari : Kamis  
Tanggal : 26 Juni 2025

Nama : Rosdiyanah  
NIM : 09031482326020  
Judul : Analisis Sentimen Publik Terhadap Program KIP-Kuliah  
Menggunakan Algoritma Random Forest pada Media Sosial X

Komisi Penguji :

1. Pembimbing I : Dinda Lestarini, S.SI., M.T.
2. Ketua : Dinna Yunika Hardiyanti, M.T.
3. Anggota : Hardini Novianti, M.T.



## HALAMAN PERSETUJUAN JURNAL

Telah *Published* jurnal di jurnal *Building of Informatics, Technology and Science* (BITS) Sinta 3 Pada :

Tanggal : 01 Juni 2025

Nama : Rosdiyanah

NIM : 09031482326020

Judul Jurnal : Analisis Sentimen Publik Terhadap Program KIP-Kuliah  
Menggunakan Algoritma Random Forest pada Media Sosial X

Tim Pembimbing :

Pembimbing I : Dinda Lestarini, S.SI., M.T.



## HALAMAN PERSEMBAHAN

“Proses paling panjang yang ditempuh seseorang ialah proses menerima dirinya. Tidak lagi marah, tidak lagi menyesali, dan tidak lagi berandai-andai. *It takes a long time*, bahkan seumur hidup.”

— *The genuine destination is acceptance* —

## KATA PENGANTAR

Dengan segenap kerendahan hati, penulis mengawali untaian kata ini sebagai ungkapan syukur yang tak terhingga ke hadirat Allah SWT. Di setiap detik perjalanan, di sela lelah dan harap, tangan-Nya yang lembut selalu menuntun langkah hingga karya ini dapat terwujud.

Perjalanan menyelesaikan tulisan ini bukan sekadar rentetan huruf dan angka, melainkan kumpulan kisah tentang perjuangan, pembelajaran, dan doa yang tiada henti. Di balik setiap halaman yang tersusun, ada asa yang tumbuh, ragu yang ditepis, serta keyakinan yang terus dijaga.

Dalam kesempatan yang penuh makna ini, izinkan penulis menyampaikan terima kasih yang tulus kepada :

1. Allah SWT yang telah memberikan rahmat, ilmu, kesempatan, kesehatan kemudahan, dan keberkahan yang tak henti mengalir, yang menguatkan penulis hingga sampai pada titik ini.
2. Ayah,Ibu, Kakak dan Adik saya yang menjadi sumber kekuatan dalam setiap detik perjalanan ini. Doa-doa yang senantiasa terpanjat di setiap sujud, cinta yang tak pernah lelah meski jarak dan waktu kadang membentang, serta dukungan yang mengalir tanpa henti, menjadi pijakan kokoh di saat langkah mulai goyah. Dalam setiap keberhasilan yang diraih, ada tangan-tangan tulus kalian yang tak pernah lelah menopang, ada doa yang senyap namun begitu kuat mengiringi langkah, ada pelukan hangat yang diam-diam menguatkan di saat dunia terasa berat. Kalian adalah alasan mengapa setiap keraguan mampu dilawan, mengapa setiap perjuangan layak dijalani, dan mengapa setiap cita-cita pantas untuk terus diperjuangkan sampai menjadi kenyataan.
3. Bapak Prof. DR. Erwin, S.Si., M.Si. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Ahmad Rifai, S.T., M.T. selaku Ketua Program Studi Sistem Informasi Universitas Sriwijaya.
5. Ibu Dwi Rosa Indah, M.T. selaku dosen pembimbing akademik saya yang telah banyak membantu, mendukung dan mengarahkan selama penulis menempuh perkuliahan.

6. Ibu Dinda Lestarini, S.SI., M.T. selaku dosen pembimbing skripsi, yang dengan penuh kesabaran, ketelitian, serta arahan bijaknya telah membimbing dan menuntun penulis hingga mampu menyelesaikan laporan ini dengan sebaik-baiknya.
7. Seluruh Dosen Program Studi Sistem Informasi, yang selama masa perkuliahan telah menanamkan ilmu, membukakan wawasan, serta menyalakan semangat untuk terus belajar dan berkembang.
8. Seluruh karyawan Fakultas Ilmu Komputer, yang dengan ketulusan dan keramahan telah membantu serta mempermudah segala urusan administrasi selama proses penyusunan skripsi ini.
9. Rekan seperjuangan Sistem Informasi Profesional Angkatan 2023 yang selalu menumbuhkan semangat bersama. Terutama kepada Reffina, Vika, dan Dyan yang bukan sekadar rekan di ruang perkuliahan, melainkan sahabat yang telah mewarnai setiap perjalanan dengan kehangatan, tawa, dan ketulusan. Bersama kalian, setiap langkah terasa lebih ringan, setiap tantangan lebih mudah dihadapi, dan setiap kenangan menjadi begitu manis untuk dikenang.
10. Sahabat-sahabat terbaik saya Papak dan Uci yang telah hadir bukan hanya dalam masa-masa mudah, tetapi juga dalam setiap rintik hujan perjuangan. Kehadiran kalian pengingat indah bahwa setiap perjuangan tak pernah dilalui sendirian. Dalam banyaknya persinggungan hidup, hanya segelintir yang mampu hadir lalu menetap dengan tulus, dan kalian merupakan salah satunya yang begitu berharga dalam perjalanan ini.
11. Seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah menjadi bagian dari perjalanan. Setiap dukungan, kebaikan, dan semangat yang kalian berikan, akan selalu penulis kenang dengan penuh rasa syukur.

Semoga karya ini dapat memberi manfaat dan menjadi langkah kecil menuju kontribusi yang lebih besar di masa mendatang.

Palembang, Mei 2025



Rosdiyanah

NIM. 09031482326020

**ANALISIS SENTIMEN PUBLIK TERHADAP PROGRAM KIP-KULIAH  
MENGUNAKAN ALGORITMA RANDOM FOREST PADA MEDIA  
SOSIAL X**

**Oleh**

**Rosdiyanah**

**NIM : 09031482326020**

**ABSTRAK**

Program KIP-K atau Kartu Indonesia Pintar Kuliah adalah bantuan pendanaan oleh pemerintah bagi mahasiswa yang mempunyai kesulitan ekonomi namun ingin merasakan kesempatan mengenyam pendidikan di perguruan tinggi. Program ini tidak lepas dari perbincangan masyarakat, terutama terkait isu penyalahgunaan dana oleh penerima, inkonsistensi pencairan dana, dan pemalsuan berkas pendaftaran. Problematika ini menjadikan pandangan masyarakat bahwa program KIP-K seringkali masih salah sasaran. Penelitian bertujuan untuk mengkaji sentimen atau persepsi masyarakat terhadap program KIPK-K menggunakan algoritma *Random Forest* yang dikombinasikan dengan Word2Vec sebagai teknik pembobotan kata dan *Random Oversampling* (ROS) sebagai teknik *balancing* untuk mengatasi ketidakseimbangan data. *Dataset* yg diperoleh berasal dari media sosial X (*Twitter*) sebanyak 4423 *tweet* dengan kata kunci “kip-k” atau “kipk” dan dengan rentan waktu selama tahun 2024. Kinerja model menunjukkan akurasi yang tinggi sebesar 96,57%, presisi, *recall*, dan *f1-score* di angka yang sama yaitu 97%. Hasil menunjukkan bahwa model mampu menganalisis sentimen dengan akurat dan mempertahankan performa yang seimbang antar kedua kelas sentimen. Berdasarkan temuan penelitian ini, algoritma *Random Forest* yang dikombinasikan dengan Word2Vec dan *Random Oversampling* (ROS) dapat menghasilkan akurasi tinggi dan mampu mengatasi ketidakseimbangan data.

**Kata Kunci:** Analisis Sentimen; kipk; *Random Forest*; Word2Vec; *Random Oversampling*.

**ANALYSIS OF PUBLIC SENTIMENT TOWARD KIP-KULIAH  
PROGRAM USING THE RANDOM FOREST ALGORITHM ON SOCIAL  
MEDIA X**

**By**

**Rosdiyanah**

**NIM : 09031482326020**

**ABSTRACT**

KIP-K program or The Indonesia Smart College Card is a funding assistance from the government for students with economic difficulties who want to experience educational opportunities in higher education. This program can't be separated from public discussion, especially regarding the issue of misuse of funds by recipients, inconsistency in fund disbursement and falsification of registration files. These problems make the public view that the KIP-K program is often still misdirected. The research aims to examine public sentiment or perception towards the KIP-K program using Random Forest algorithm combined with Word2Vec as a word weighting technique and Random Oversampling (ROS) as a balancing technique to overcome data imbalance. The dataset obtained comes via platform X or Twitter) a total of 4423 tweets with the keywords "kip-k" or "kipk" and with a vulnerable time during 2024. The model's performance demonstrated a high accuracy of 96,57%, precision, recall, and f1-score at the same value of 97%. The results indicate that the model is effective in analyzing sentiment accurately and maintaining a balanced performance between the two sentiment classes. Based on research in this study, the Random Forest algorithm combined with Word2Vec and Random Oversampling (ROS) can produce high accuracy and can overcome data imbalance.

**Keywords:** *Sentiment Analysis; kipk; Random Forest; Word2Vec; Random Oversampling.*

## DAFTAR ISI

|                                                        | Halaman      |
|--------------------------------------------------------|--------------|
| <b>COVER .....</b>                                     | <b>i</b>     |
| <b>HALAMAN PENGESAHAN.....</b>                         | <b>ii</b>    |
| <b>HALAMAN PERNYATAAN.....</b>                         | <b>iii</b>   |
| <b>HALAMAN PERNYATAAN INTEGRITAS.....</b>              | <b>iv</b>    |
| <b>HALAMAN PERSETUJUAN TELAH DI UJI.....</b>           | <b>v</b>     |
| <b>HALAMAN PERSETUJUAN JURNAL .....</b>                | <b>vi</b>    |
| <b>HALAMAN PERSEMBAHAN .....</b>                       | <b>vii</b>   |
| <b>KATA PENGANTAR.....</b>                             | <b>viii</b>  |
| <b>ABSTRAK .....</b>                                   | <b>x</b>     |
| <b>ABSTRACT .....</b>                                  | <b>xi</b>    |
| <b>DAFTAR ISI.....</b>                                 | <b>xii</b>   |
| <b>DAFTAR TABEL .....</b>                              | <b>xv</b>    |
| <b>DAFTAR GAMBAR.....</b>                              | <b>xvi</b>   |
| <b>DAFTAR KODE PROGRAM .....</b>                       | <b>xvii</b>  |
| <b>DAFTAR LAMPIRAN .....</b>                           | <b>xviii</b> |
| <b>BAB I PENDAHULUAN.....</b>                          | <b>1</b>     |
| 1.1 Latar Belakang .....                               | 1            |
| 1.2 Rumusan Masalah .....                              | 4            |
| 1.3 Tujuan Penelitian.....                             | 4            |
| 1.4 Manfaat Penelitian.....                            | 4            |
| 1.5 Batasan Masalah.....                               | 5            |
| <b>BAB II ARTIKEL ILMIAH YANG DIPUBLIKASIKAN .....</b> | <b>6</b>     |
| 2.1 Jurnal .....                                       | 6            |
| <b>BAB III TINJAUAN PUSTAKA .....</b>                  | <b>17</b>    |
| 3.1 Analisis Sentimen.....                             | 17           |
| 3.2 Sentimen Publik .....                              | 17           |
| 3.3 Media Sosial X .....                               | 18           |
| 3.4 Program KIP-Kuliah.....                            | 18           |
| 3.5 <i>Data Mining</i> .....                           | 19           |
| 3.6 <i>Text Mining</i> .....                           | 20           |

|                                           |                                               |           |
|-------------------------------------------|-----------------------------------------------|-----------|
| 3.7                                       | <i>Preprocessing</i> .....                    | 21        |
| 3.8                                       | Pembobotan Kata .....                         | 21        |
| 3.9                                       | Klasifikasi <i>Text Mining</i> .....          | 22        |
| 3.10                                      | Algoritma <i>Random Forest</i> .....          | 22        |
| 3.11                                      | Teknik <i>Random Oversampling</i> (ROS) ..... | 24        |
| 3.12                                      | <i>Confusion Matrix</i> .....                 | 24        |
| <b>BAB IV METODOLOGI PENELITIAN .....</b> |                                               | <b>26</b> |
| 4.1                                       | Desain Penelitian .....                       | 26        |
| 4.2                                       | Populasi dan Sampel .....                     | 27        |
| 4.3                                       | Pengumpulan Data .....                        | 27        |
| 4.4                                       | <i>Pre-processing</i> .....                   | 29        |
| 4.4.1                                     | Proses <i>Cleaning</i> .....                  | 29        |
| 4.4.2                                     | Proses <i>Case Folding</i> .....              | 31        |
| 4.4.3                                     | Proses Normalisasi Kata .....                 | 31        |
| 4.4.4                                     | Proses Tokenisasi .....                       | 32        |
| 4.4.5                                     | Proses <i>Stopword Removal</i> .....          | 32        |
| 4.4.6                                     | Proses <i>Stemming</i> .....                  | 33        |
| 4.4.7                                     | Proses Pelabelan .....                        | 33        |
| 4.5                                       | Pembobotan Kata .....                         | 34        |
| 4.6                                       | Teknik <i>Balancing</i> .....                 | 35        |
| 4.7                                       | Pembagian Data .....                          | 36        |
| 4.8                                       | Pemodelan Data .....                          | 36        |
| 4.9                                       | Evaluasi Model .....                          | 36        |
| 4.9.1                                     | Akurasi .....                                 | 37        |
| 4.9.2                                     | <i>Precision</i> .....                        | 37        |
| 4.9.3                                     | <i>Recall</i> .....                           | 37        |
| 4.9.4                                     | <i>F1-Score</i> .....                         | 38        |
| 4.10                                      | Analisis Hasil .....                          | 38        |
| <b>BAB V HASIL DAN PEMBAHASAN .....</b>   |                                               | <b>40</b> |
| 5.1                                       | <i>Dataset</i> .....                          | 40        |
| 5.2                                       | Hasil <i>Preprocessing</i> .....              | 40        |
| 5.3                                       | Hasil Pembobotan Kata .....                   | 41        |

|                                                                             |           |
|-----------------------------------------------------------------------------|-----------|
| 5.4 Hasil Teknik <i>Balancing</i> .....                                     | 42        |
| 5.5 Hasil Pembagian Data .....                                              | 42        |
| 5.6 Hasil Pemodelan Data .....                                              | 43        |
| 5.7 Hasil Evaluasi Model .....                                              | 43        |
| 5.8 Analisis <i>Confusion Matrix</i> dan <i>Classification Report</i> ..... | 44        |
| 5.9 Visualisasi <i>Wordcloud</i> .....                                      | 49        |
| <b>BAB VI PENUTUP .....</b>                                                 | <b>51</b> |
| 6.1 Kesimpulan.....                                                         | 51        |
| 6.2 Saran.....                                                              | 52        |
| <b>DAFTAR PUSTAKA.....</b>                                                  | <b>53</b> |
| <b>LAMPIRAN.....</b>                                                        | <b>55</b> |

## DAFTAR TABEL

|                                                     | Halaman |
|-----------------------------------------------------|---------|
| <b>Tabel 4.1</b> <i>Confusion Matrix</i> .....      | 38      |
| <b>Tabel 5.1</b> <i>Balancing</i> .....             | 42      |
| <b>Tabel 5.2</b> <i>Confusion Matrix</i> .....      | 44      |
| <b>Tabel 5.3</b> <i>Classification Report</i> ..... | 46      |

## DAFTAR GAMBAR

|                                                              | Halaman |
|--------------------------------------------------------------|---------|
| <b>Gambar 4.1</b> Metodologi Penelitian.....                 | 26      |
| <b>Gambar 4.2</b> <i>Dataset Crawling</i> .....              | 29      |
| <b>Gambar 4.3</b> <i>Sampel Data Preproccesing</i> .....     | 29      |
| <b>Gambar 4.4</b> Representasi CBOW .....                    | 35      |
| <b>Gambar 5.1</b> Evaluasi Hasil .....                       | 44      |
| <b>Gambar 5.2</b> <i>Confusion Matrix (Heatmap)</i> .....    | 45      |
| <b>Gambar 5.3</b> <i>Classification Report Metrics</i> ..... | 48      |
| <b>Gambar 5.4</b> Visualisasi Wordcloud .....                | 49      |

## DAFTAR KODE PROGRAM

|                                                       | Halaman |
|-------------------------------------------------------|---------|
| <b>Kode Program 4.1</b> Pengumpulan Data.....         | 28      |
| <b>Kode Program 4.2</b> Proses <i>Cleaning</i> .....  | 30      |
| <b>Kode Program 4.3</b> <i>Case Folding</i> .....     | 31      |
| <b>Kode Program 4.4</b> Normalisasi Kata .....        | 31      |
| <b>Kode Program 4.5</b> Tokenisasi .....              | 32      |
| <b>Kode Program 4.6</b> <i>Stopword Removal</i> ..... | 32      |
| <b>Kode Program 4.7</b> <i>Stemming</i> .....         | 33      |
| <b>Kode Program 5.1</b> Pembobotan Kata.....          | 41      |
| <b>Kode Program 5.2</b> <i>Random Forest</i> .....    | 42      |
| <b>Kode Program 5.3</b> Pembagian Data .....          | 43      |
| <b>Kode Program 5.4</b> Pemodelan Data .....          | 43      |
| <b>Kode Program 5.5</b> Evaluasi Model .....          | 43      |

## DAFTAR LAMPIRAN

|                                                  | Halaman |
|--------------------------------------------------|---------|
| <b>Lampiran 1</b> Kartu Konsultasi .....         | A-1     |
| <b>Lampiran 1</b> Surat Rekomendasi Sidang ..... | B-1     |
| <b>Lampiran 1</b> SK Bebas Pustaka .....         | C-1     |
| <b>Lampiran 1</b> Form Revisi Sidang .....       | D-1     |

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pendidikan tinggi merupakan salah satu fondasi penting dalam pengembangan sumber daya manusia di Indonesia. Dalam upaya meluaskan aksesibilitas pendidikan tinggi, pemerintah Indonesia meluncurkan berbagai program bantuan, salah satunya adalah Program Kartu Indonesia Pintar Kuliah (KIP-K). KIP-K adalah program berkelanjutan dari skema Bidikmisi yang telah ada sejak tahun 2011. Program ini ditujukan untuk membantu mahasiswa dari keluarga kurang mampu agar dapat melanjutkan pendidikan ke jenjang perguruan tinggi (Martins & Toletina, 2024). Menurut data dari Kementerian Pendidikan dan Kebudayaan melalui Portal Informasi Indonesia, anggaran untuk KIP-K terus meningkat dari Rp3,7 triliun pada tahun 2020 menjadi Rp14,69 triliun pada tahun 2025, hal ini menunjukkan komitmen pemerintah dalam mengurangi kesenjangan sosial ekonomi di bidang pendidikan.

Dalam era teknologi modern, media sosial sebagai *platform* utama bagi masyarakat untuk menyuarakan pendapat dan kritik terhadap berbagai kebijakan pemerintah, termasuk program KIP-K. Media sosial X yang sebelumnya biasa dikenal sebagai *Twitter* menjadi salah satu wadah di mana diskusi mengenai program ini berlangsung. Namun, meskipun program ini dirancang untuk membantu mahasiswa yang membutuhkan, banyak komentar negatif yang muncul di media sosial, seringkali terkait dengan penyalahgunaan dana beasiswa. Menurut penelitian sebelumnya, sekitar 79% komentar yang terkait dengan KIP-K bersifat

negatif, menciptakan tantangan bagi pemerintah dalam memastikan bahwa bantuan yang diberikan tepat sasaran dan efektif (Ali & Hendrastuty, 2024).

Analisis sentimen adalah teknik yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini atau perasaan masyarakat terhadap suatu topik salah satunya yaitu tentang Program KIP-K. Dengan menggunakan algoritma pembelajaran mesin, analisis sentimen dapat mengkategorikan komentar publik menjadi sentimen positif, negatif, atau netral. Teknik ini sangat penting dalam memahami persepsi masyarakat, karena dapat memberikan pengetahuan yang berguna bagi pembuat kebijakan untuk mengevaluasi efektivitas program. Dalam konteks KIP-K, analisis sentimen dapat membantu pemerintah memahami bagaimana masyarakat merespons program ini dan apakah program tersebut memenuhi harapan mereka. Menurut (Husada & Paramita, 2021), analisis sentimen dapat digunakan untuk mengevaluasi opini publik dan memberikan masukan yang konstruktif bagi pengambilan keputusan.

Program KIP-K memiliki tujuan yang sangat baik, namun terdapat berbagai tantangan yang dihadapi dalam menjalankannya. Proses seleksi penerima beasiswa sering kali tidak efektif, sehingga mahasiswa yang sangat membutuhkan bantuan justru tidak mendapatkan akses. Selain itu, banyak laporan yang memicu protes publik mengenai penyalahgunaan dana beasiswa oleh penerima yang tidak memenuhi syarat. Penelitian sebelumnya menunjukkan bahwa sentimen negatif mendominasi opini masyarakat terhadap program ini. Dalam penelitian yang menggunakan algoritma *Naive Bayes*, peneliti telah menerapkan teknik SMOTE untuk mengatasi ketidakseimbangan data, yang membantu meningkatkan akurasi model dalam mengklasifikasikan sentimen positif (Pramudita et al., 2024).

Sebaliknya, penelitian yang menggunakan algoritma *Support Vector Machine* (SVM) belum menerapkan teknik SMOTE, sehingga masih menghadapi tantangan dalam mengklasifikasikan sentimen positif secara efektif (Amelia & Sarimole, 2024).

Meskipun penelitian sebelumnya telah menyajikan wawasan mengenai sentimen publik terhadap KIP-K, terdapat kesenjangan dalam hal penggunaan algoritma yang lebih kompleks dan teknik penyeimbangan data. Penelitian yang ada cenderung fokus pada algoritma *Naive Bayes* dan SVM, yang menunjukkan bias terhadap sentimen negatif. Ketidakseimbangan dalam analisis sentimen dapat mengarah pada kebijakan yang tidak efektif. Oleh karena itu, penelitian ini akan menggunakan algoritma *Random Forest* yang dikenal memiliki kemampuan lebih baik dalam menangani data yang tidak seimbang dan memberikan hasil yang lebih akurat (Firnanda, 2025). Selain itu, penerapan teknik *Random Oversampling* (ROS) dipilih, yang lebih terintegrasi dalam algoritma *Random Forest* diharapkan dapat memastikan distribusi data yang lebih seimbang dan meningkatkan performa model.

Dengan demikian, penelitian ini tidak hanya akan melengkapi penelitian sebelumnya, tetapi juga memberikan kontribusi baru yang signifikan dalam memahami persepsi publik terhadap program pendidikan di Indonesia. Hasil dari penelitian ini diharapkan dapat memberikan masukan yang berharga bagi pemerintah dan pemangku kepentingan dalam meningkatkan efektivitas program KIP-K.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana sentimen publik terhadap Program Kartu Indonesia Pintar Kuliah (KIP-K) yang diungkapkan melalui media sosial X?
2. Sejauh mana algoritma Random Forest dan teknik *Random Over Sampler* dapat mengatasi ketidakseimbangan data dan meningkatkan performa model dalam mengklasifikasikan sentimen terkait KIP-K?

## 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, penelitian ini dilakukan dengan beberapa tujuan yaitu sebagai berikut:

- 1 Menganalisis sentimen publik terhadap Program KIP-K yang diungkapkan melalui media sosial X.
- 2 Menggunakan algoritma Random Forest dan teknik *Random Over Sampler* untuk mengatasi ketidakseimbangan data dan meningkatkan performa model klasifikasi sentimen terkait KIP-K.

## 1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan Manfaat yaitu sebagai berikut :

- 1 Secara Teoritis: memberikan kontribusi pada pengembangan ilmu pengetahuan di bidang analisis sentimen dan *machine learning*, serta memperkaya literatur yang ada mengenai analisis sentimen di konteks pendidikan.

- 2 Secara Praktis: memberikan masukan bagi pemerintah dalam meningkatkan mekanisme penyaluran KIP-K agar lebih tepat sasaran dan mengurangi potensi penyalahgunaan, serta memberikan wawasan bagi pemangku kepentingan mengenai persepsi publik terhadap program ini.

### **1.5 Batasan Masalah**

Penelitian ini mempunyai batasan masalah untuk memfokuskan penelitian dan akurasi diantaranya :

- 1 Penelitian ini hanya akan menganalisis sentimen publik terhadap Program KIP-K yang diungkapkan melalui media sosial X dalam periode waktu tertentu selama tahun 2024 yaitu dari 01 Januari 2024 sampai dengan 31 Desember 2024.
- 2 Fokus penelitian ini adalah pada komentar yang berkaitan dengan KIP-K, tanpa mempertimbangkan faktor eksternal yang mungkin mempengaruhi sentimen publik.
- 3 Algoritma yang digunakan dalam penelitian ini terbatas pada *Random Forest* dan tidak membandingkan dengan algoritma lain secara langsung.

## DAFTAR PUSTAKA

- Af'idah, D. I., Dairoh, D., Handayani, S. F., & Pratiwi, R. W. (2021). Pengaruh Parameter Word2Vec terhadap Performa Deep Learning pada Klasifikasi Sentimen. *Jurnal Informatika: Jurnal Pengembangan IT*, 6(3). <https://doi.org/10.30591/jpit.v6i3.3016>
- Ali, H., & Hendrastuty, N. (2024). Comparison of Naïve Bayes Classifier, Support Vector Machine, Random Forest Algorithms for Public Sentiment Analysis of KIP-K Program on Twitter. *Jurnal Teknik Informatika (Jutif)*, 5(6), 1701–1712.
- Amelia, I., & Sarimole, F. M. (2024). Analisis Sentimen Tanggapan Pengguna Media Sosial X Terhadap Program Beasiswa KIP-Kuliah dengan Menggunakan Algoritma Support Vector Machine ( SVM ). *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, 5(3). <https://doi.org/doi: 10.35870/jimik.v5i3.990>
- Amna, W. S., Sudipa, I. G. I. ., Putra, T. A. E. ., Wahidin, A. J. ., Syukrilla, W. A. ., Wardhani, A. K. ., Heryana, N. ., Indriyani, T. ., & Santoso, L. W. (2023). Data Mining. *PT Global Eksekutif Teknologi*.
- Arfian, M. H., Maruhawa, H. S., Putri, A. A., Ratulugina, D., Khoirunnisa, E. R., Hermawan, V., & Syahr, A. J. (2025). Analisis Sentimen Pada Media Sosial Menggunakan Metode Support Vector Machine. *Jurnal Ilmu Teknik Dan Komputer*, 09. <https://doi.org/10.22441/jitkom.v9i1.001>
- Aryanti, R., Misriati, T., & Sagiyanto, A. (2023). Analisis Sentimen Aplikasi Primaku Menggunakan Algoritma Random Forest dan SMOTE untuk Mengatasi Ketidakseimbangan Data. *Journal of Computer System and Informatics (JoSYC)*, 5(1), 218–227. <https://doi.org/10.47065/josyc.v5i1.4562>
- Basar, T. F., Ratnawati, D. E., & Arwani, I. (2022). Analisis Sentimen Pengguna Twitter terhadap Pembayaran Cashless menggunakan Shopeepay dengan Algoritma Random Forest. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(3), 1426–1433.
- Fathoni, F. M., Putra, C. A., & Nurlaili, A. L. (2024). Klasifikasi Penyakit Daun Anggur Menggunakan Metode K-Nearest Neighbor Berdasarkan Gray Level Co-Occurrence Matrix. *Biner : Jurnal Ilmiah Informatika Dan Komputer*, 3(1), 8–15. <https://doi.org/10.32699/biner.v3i1.6332>
- Firnanda, P. A. (2025). Analisis Perbandingan Decision Tree dan Random Forest dalam Klasifikasi Penjualan Produk pada Supermarket. 3(1), 445–461.
- Fitri, E., Yuliani, Y., Rosyida, S., & Gata, W. (2021). Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine. *Jurnal Transformatika*, 18(1), 71–80.
- Husada, H. C., & Paramita, A. S. (2021). Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM). *Teknika*, 10(1), 18–26. <https://doi.org/10.34148/teknika.v10i1.311>
- Irfani, M., & Khomsah, S. (2024). Analisis Sentimen Berbasis Aspek pada EDOM Pembelajaran Menggunakan Metode CNN dan Word2vec. *Justin : Jurnal Sistem Dan Teknologi Informasi*, 12(3). <https://doi.org/10.26418/justin.v12i3.75610>

- Jurafsky, D., & Martin, J. H. (2025). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition. In *Women's Human Rights*. <https://doi.org/10.9783/9780812200027.toc>
- Kinanti, T. A., & Suyono, S. (2023). Fenomena Speak Up pada Media Twitter (Study Deskriptif Korban Penipuan Melalui Gerakan "A Thread"). *Jurnal Bisnis Dan Komunikasi Digital*, 1(1), 12. <https://doi.org/10.47134/jbk.d.v1i1.1912>
- Martins, E. K., & Toletina, N. T. (2024). Evaluasi Pelaksanaan Kebijakan Program KIP-K Di Indonesia. *Professional: Jurnal Komunikasi Dan Administrasi Publik*, 11(1), 331–340. <https://doi.org/10.37676/professional.v11i1.6166>
- Muliana, A. S., Lestari, D., & Raflesia, S. P. (2024). Analysis of Public Sentiment on Election Results using Naive Bayes in Social Media X. *Sistemasi: Jurnal Sistem Informasi*, 13(6), 2467–2478.
- Pramudita, D., Akbar, Y., & Wahyudi, T. (2024). MALCOM: Indonesian Journal of Machine Learning and Computer Science Sentiment Analysis of the Indonesian Smart College Card Program on Social Media X Using the Naive Bayes Algorithm Analisis Sentimen Terhadap Program Kartu Indonesia Pintar Kuliah Pada Med. *Malcom*, 4(October), 1420–1430.
- Putra, T. D., & Oktafiani, D. (2025). Klasifikasi Sentimen Postingan Sosial Media Menggunakan Machine Learning Random Forest dan Naïve Bayes. 5, 2338–2347.
- Putri, N. A. Y., Subagio, R. T., & Asfi, M. (2021). Sistem Pendukung Keputusan Penilaian Kinerja Mahasiswa KIP Kuliah dengan Penerapan Metode TOPSIS dan PROMETHEE. *Jurnal Media Informatika Budidarma*, 5(4), 1394–1404.
- Putri, & Romli, N. A. (2021). Analisis Dampak Adiksi Internet Pada Media Sosial Twitter Di Indonesia Dengan Pendekatan Teori Komunikasi. *Jurnal Komunikasi Universitas Garut: Hasil Pemikiran Dan Penelitian*, 7(1), 582. <https://doi.org/10.52434/jk.v7i1.905>
- Saifullah, S., Fauziyah, Y., & Aribowo, A. S. (2021). Comparison of machine learning for sentiment analysis in detecting anxiety based on social media data. *Jurnal Informatika*, 15(1), 45. <https://doi.org/10.26555/jifo.v15i1.a20111>
- Sihombing, J. (2021). Klasifikasi Data Antropometri Individu Menggunakan Algoritma Naïve Bayes Classifier. *BIOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer*, 2(1), 1–10.