PERBANDINGAN METODE SUPPORT VECTOR MACHINE DAN RANDOM FOREST PADA ANALISIS SENTIMEN PENGGUNA APLIKASI TOPINDOKU

FAKHRI ARIANDI
H1101201049

SKRIPSI

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2026

PERBANDINGAN METODE SUPPORT VECTOR MACHINE DAN RANDOM FOREST PADA ANALISIS SENTIMEN PENGGUNA APLIKASI TOPINDOKU

FAKHRI ARIANDI
H1101201049

Skripsi
Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer pada Program Studi Sistem Informasi

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2026

LEMBAR PENGESAHAN

Judul Tugas Akhir	: Perbandingan Metode *Support Vector Machine* dan *Random Forest* Pada Analisis Sentimen Pengguna Aplikasi Topindoku
Nama Mahasiswa	: Fakhri Ariandi
NIM	: H1101201049
Jurusan/Program Studi	: Sistem Informasi / Sistem Informasi
Tanggal Lulus	: 11 Juni 2026
SK Pembimbing	: No. 3865/UN22.8/TD.06/2024/21 November 2024
SK Penguji	: No. 2997/DST/UN22.8/DT.00.09/2026/11 Juni 2026

Dosen Pembimbing

Pembimbing I	Pembimbing II
Ilhamsyah, S.Si., M.Cs.	Syahru Rahmayudha, S.Kom., M.Kom.
NIP 198405102012121001	NIP 198306172025211055

Dosen Penguji

Ketua Penguji	Anggota Penguji
Dian Prawira, ST., M.Eng.	Renny Puspita Sari, ST., MT.
NIP 198411132015041001	NIP 198704182015042002

Pimpinan Sidang (merangkap anggota penguji)	Sekretaris Sidang (merangkap anggota penguji)
Ilhamsyah, S.Si., M.Cs.	Syahru Rahmayudha, S.Kom., M.Kom.
NIP 198405102012121001	NIP 198306172025211055

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Tanjungpura

Prof. Dr. Gusrizal, S.Si., M.Si.
NIP 197108022000031001

PERNYATAAN INTEGRITAS AKADEMIK

Saya yang bertanda tangan di bawah ini:

Nama	: Fakhri Ariandi
NIM	: H1101201049
Program Studi/ Jurusan	: Sistem Informasi/ Sistem Informasi
Fakultas	: Matematika dan Ilmu Pengetahuan Alam

dengan ini menyatakan bahwa dokumen ilmiah Tugas Akhir yang disajikan ini tidak mengandung unsur pelanggaran integritas akademik sesuai Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 39 Tahun 2021. Apabila di kemudian hari dokumen ilmiah Tugas Akhir ini mengandung unsur pelanggaran integritas akademik sesuai ketentuan perundangan tersebut, maka saya bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Demikian pernyataan ini untuk digunakan sebagaimana mestinya.

Pontianak, 11 Juni 2026



Fakhri Ariandi

H1101201049

PERBANDINGAN METODE SUPPORT VECTOR MACHINE DAN RANDOM FOREST PADA ANALISIS SENTIMEN PENGGUNA APLIKASI TOPINDOKU

ABSTRAK

Ulasan pengguna pada *mobile application marketplace* dapat dimanfaatkan untuk mengetahui sentimen pengguna terhadap suatu aplikasi. Penelitian ini bertujuan untuk membandingkan kinerja metode *Support Vector Machine* dan *Random Forest* dalam analisis sentimen ulasan pengguna aplikasi Topindoku. Data ulasan dikumpulkan dari pengguna aplikasi kemudian diproses melalui tahapan *text preprocessing* yang meliputi *case folding*, *tokenizing*, *stopword removal*, *normalizing*, dan stemming. Penelitian ini juga menerapkan teknik *oversampling* menggunakan SMOTE untuk mengatasi ketidakseimbangan jumlah data pada masing-masing kelas sentimen. Pengujian dilakukan dengan beberapa skenario pembagian data, dan hasil terbaik yang diperoleh kedua metode berada pada skenario *splitting data* 80:20. Metode SVM menghasilkan kinerja terbaik dengan nilai akurasi sebesar 87%, presisi 76%, *recall* 65%, dan *F1-score* 65%. Sementara itu, metode *Random Forest* memperoleh nilai akurasi sebesar 85%, presisi 81%, *recall* 62%, dan *F1-score* 61% pada skenario yang sama. Berdasarkan hasil tersebut, metode *Support Vector Machine* (SVM) menunjukkan kinerja yang lebih unggul dibandingkan *Random Forest* dalam analisis sentimen ulasan pengguna aplikasi Topindoku.

Kata Kunci: Analisis Sentimen, *Support Vector Machine*, *Random Forest*, SMOTE, Topindoku

COMPARISON OF SUPPORT VECTOR MACHINE AND RANDOM FOREST METHODS IN SENTIMENT ANALYSIS OF TOPINDOKU APPLICATION USERS

ABSTRACT

User reviews on mobile application marketplaces can be utilized to identify user sentiment toward an application. This study aims to compare the performance of the Support Vector Machine and Random Forest methods in sentiment analysis of user reviews for the Topindoku application. The review data were collected from application users and processed through several text preprocessing stages, including case folding, tokenizing, stopword removal, normalizing, and stemming. This study also applied the SMOTE oversampling technique to address class imbalance in the sentiment data. Testing was conducted using several data splitting scenarios, and the best results for both methods were obtained using the 80:20 data splitting scenario. The SVM method achieved the best performance with an accuracy of 87%, precision of 76%, recall of 65%, and F1-score of 65%. Meanwhile, the Random Forest method achieved an accuracy of 85%, precision of 81%, recall of 62%, and F1-score of 61% under the same scenario. Based on these results, the Support Vector Machine (SVM) method demonstrated superior performance compared to Random Forest in sentiment analysis of Topindoku application user reviews.

Keywords: *Sentiment Analysis, Support Vector Machine, Random Forest, SMOTE, Topindoku*

PRAKATA

Alhamdulillah rasa syukur penulis panjatkan kehadirat Allah Subhanahu Wa Ta'ala, karena berkat hidayah dan rahmat-Nya, penulis dapat menyelesaikan skripsi yang berjudul "**Perbandingan Metode Support Vector Machine dan Random Forest Pada Analisis Sentimen Pengguna Aplikasi Topindoku**".

Penyusunan penelitian ini merupakan salah satu syarat untuk menyusun skripsi dan mendapatkan gelar sarjana pada program studi S-1 Sistem Informasi FMIPA Universitas Tanjungpura. Dalam menyelesaikan skripsi ini penulis banyak mendapat bantuan, bimbingan dan dorongan dari berbagai pihak. Oleh karena itu pada kesempatan ini penulis ingin mengucapkan banyak terima kasih kepada:

1. Kedua orang tua saya Budiono dan Rumindar Gunati atas perhatian, motivasi, semangat, do'a, biaya, nasehat, dan dukungan yang diberikan.
2. Keluarga saya Nur Satria Wulandari, Muhammad Farhan Budi Pratomo, Fandilatunnisak, dan seluruh keluarga besar dari Ibu dan Ayah Saya yang telah membantu memberikan semangat.
3. Bapak Dr. Gusrizal, S.Si., M.Si. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura Pontianak.
4. Bapak Ilhamsyah, S.Si., M.Cs selaku Pembimbing Pertama atas bimbingan dan arahan yang telah diberikan.
5. Bapak Syahru Rahmayuda, S.Kom., M.Kom selaku Pembimbing Kedua atas ilmu dan saran-saran yang telah diberikan.
6. Bapak Dian Prawira, ST., M.Eng selaku Penguji Pertama dan Ibu Renny Puspitasari, ST., MT. selaku Penguji Kedua atas ilmu yang telah diberikan, saran-saran yang diberikan, dan membantu menggali kemampuan saya.
7. Para sahabat SMP dan SMA saya yaitu Zidan, Raudy, Cesar, Breet, Syauqi, Ami, Farhan, Emm, Azim, Nando, Adit, dan Manyu yang selalu menemani dari awal hingga akhir pembuatan skripsi tanpa perselisihan, pertikaian, kegaduhan, keresahan dan membantu penelitian serta memberikan semangat, dukungan, nasehat, ceramah dalam penulisan skripsi.
8. Para teman se-Sistem Informasi yang telah membantu secara langsung maupun tidak langsung kepada penulis.

9. Pihak yang telah terlibat dalam penelitian maupun penulisan skripsi ini, yang tidak bisa penulis sebutkan satu per satu.

Penulis menyadari bahwa dalam penulisan penelitian ini masih banyak terdapat kekurangan. Oleh karena itu, penulis mengharapkan adanya kritik dan saran yang bersifat membangun demi kesempurnaan skripsi ini. Akhirnya, semoga tulisan ini dapat bermanfaat bagi penulis khususnya dan para pembaca umumnya.

Pontianak, 11 Juni 2026

Penulis,



Fakhri Ariandi

H1101201049

DAFTAR ISI

LEMBAR PENGESAHAN .....Error! Bookmark not defined.
PERNYATAAN INTEGRITAS AKADEMIK.....Error! Bookmark not defined.
ABSTRAK ..... **iv**
ABSTRACT ..... **v**
PRAKATA ..... **vi**
DAFTAR ISI ..... **viii**
DAFTAR TABEL ..... **x**
DAFTAR GAMBAR ..... **xii**
DAFTAR LIST CODE ..... **xiii**
DAFTAR PERSAMAAN ..... **xvi**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Tujuan Penelitian 3
1.4 Batasan Masalah 3
1.5 Manfaat Penelitian 4
1.6 Metodologi Penulisan 4
BAB II STUDI LITERATUR **6**
2.1 Dasar Teori 6
2.1.1. Analisis Sentimen 6
2.1.2 Text Mining 6
2.1.3 Machine Learning 7
2.1.4 Scraping Data 9
2.1.5 Text Preprocessing 9
2.1.6 Support Vector Machine (SVM) 13
2.1.7 Random Forest 15
2.1.8 Confusion Matrix 16
2.1.9 Google Colaboratory 18
2.1.10 Python 18

2.1.11 Aplikasi Topindoku 18
2.2 Kajian Penelitian 19
BAB III METODOLOGI PENELITIAN 23
3.1 Metodologi Penelitian (*IS Research Framework*) 23
3.2 Alat Penelitian 23
3.3 Kerangka Kerja Penelitian 23
BAB IV HASIL DAN PEMBAHASAN 28
4.1 Scraping Data 28
4.2 Labeling 30
4.3 Distribusi dataset 31
4.4 Preprocessing text 32
4.4.1 Cleaning Data & Case Folding 32
4.4.2 Tokenizing 37
4.4.3 Stopword-Removal 39
4.4.4 Filtering token by length 41
4.4.5 Normalization 43
4.4.6 Stemming 46
4.5 Pembobotan kata 48
4.6 Visualisasi data 49
4.6.1 Visualisasi data positif 49
4.6.2 Visualisasi data netral 54
4.6.3 Visualisasi data negatif 59
4.7 Modeling klasifikasi 64
4.7.1 Splitting data 60:40 SVM 64
4.7.2 Splitting data 70:30 SVM 70
4.7.3 Splitting data 80:20 SVM 75
4.7.4 Splitting data 60:40 Random Forest 80
4.7.5 Splitting data 70:30 Random Forest 86
4.7.6 Splitting data 80:20 Random Forest 92
4.7.7 Splitting data 60:40 SVM (SMOTE sebelum Splitting) 97

4.7.8 Splitting data 70:30 SVM (SMOTE sebelum Splitting) 101
4.7.9 Splitting data 80:20 SVM (SMOTE sebelum Splitting) 106
4.7.10 Splitting data 60:40 Random Forest (SMOTE sebelum Splitting) ... 110
4.7.11 Splitting data 70:30 Random Forest (SMOTE sebelum Splitting) ... 115
4.7.12 Splitting data 80:20 Random Forest (SMOTE sebelum Splitting) ... 120
4.8 Analisis Perbandingan Evaluasi SVM dan Random Forest 125
4.8.1 Evaluasi SVM 125
4.8.2 Evaluasi Random Forest 127
4.9 Hasil Perbandingan Evaluasi SVM dan Random Forest 129
BAB V PENUTUP 133
5.1 Kesimpulan 133
5.2 Saran 134
DAFTAR PUSTAKA 135

DAFTAR TABEL

Tabel 2.1 Penelitian terdahulu 19
Tabel 4.1 Hasil scraping data ulasan..... 30
Tabel 4.2 Hasil labeling manual 31
Tabel 4.3 Hasil distribusi data 32
Tabel 4.4 Pemeriksaan data duplikat 32
Tabel 4.5 Penghapusan data duplikat..... 33
Tabel 4.6 Hasil pemeriksaan tipe data 34
Tabel 4.7 Hasil penyesuaian tipe data..... 34
Tabel 4.8 Hasil data cleaning dan case folding..... 36
Tabel 4.9 Hasil Tokenizing..... 38
Tabel 4.10 Hasil Stopwords-removal..... 41
Tabel 4.11 Hasil *Filtering token by length* 42
Tabel 4.12 Normalization 45
Tabel 4.13 Hasil Stemming..... 47
Tabel 4.14 Sepuluh kata positif terbanyak..... 52
Tabel 4.15 Sepuluh kata netral terbanyak..... 57
Tabel 4.16 Sepuluh kata negatif terbanyak..... 62
Tabel 4.17 Hasil Splitting data 60:40 SVM..... 65
Tabel 4.18 Hasil SMOTE 60:40 SVM..... 66
Tabel 4.19 Hasil confusion matrix SVM 60:40 69
Tabel 4.20 Hasil splitting data 70:30 SVM 70
Tabel 4.21 Hasil SMOTE 70:30 SVM..... 72
Tabel 4.22 Hasil confusion matrix 70:30 SVM 75
Tabel 4.23 Hasil Splitting data 80:20 SVM..... 76
Tabel 4.24 Hasil SMOTE 80:20 SVM..... 77
Tabel 4.25 Hasil confusion matrix 80:20 SVM 80
Tabel 4.26 Hasil Splitting data 60:40 random forest 81
Tabel 4.27 Hasil SMOTE 60:40 random forest 82
Tabel 4.28 Hasil confusion matrix 60:40 random forest 86
Tabel 4.29 Hasil Splitting data 70:30 random forest 87

Tabel 4.30 Hasil SMOTE 70:30 random forest 88
Tabel 4.31 Hasil confusion matrix 70:30 random forest 91
Tabel 4.32 Hasil Splitting data 80:20 random forest 92
Tabel 4.33 Hasil SMOTE 80:20 random forest 94
Tabel 4.34 Hasil Confusion Matrix 80:20 random forest 945
Tabel 4.35 SMOTE 60:40 SVM (sebelum splitting data) 98
Tabel 4.36 Hasil splitting data 60:40 SVM (sesudah SMOTE)..... 99
Tabel 4.37 Hasil confusion matrix SVM 60:40 101
Tabel 4.38 SMOTE 70:30 SVM (sebelum splitting data) 102
Tabel 4.39 Hasil splitting data 70:30 SVM (sesudah SMOTE)..... 103
Tabel 4.40 Hasil confusion matrix SVM 70:30 105
Tabel 4.41 SMOTE 80:20 SVM (sebelum splitting data) 106
Tabel 4.42 Hasil splitting data 80:20 SVM (sesudah SMOTE)..... 107
Tabel 4.43 Hasil confusion matrix SVM 80:20 110
Tabel 4.44 SMOTE 60:40 random forest (sebelum splitting data)..... 111
Tabel 4.45 Hasil splitting data 60:40 random forest (sesudah SMOTE)..... 112
Tabel 4.46 Hasil confusion matrix random forest 60:40 114
Tabel 4.47 SMOTE 70:30 random forest (sebelum splitting data)..... 115
Tabel 4.48 Hasil splitting data 70:30 random forest (sesudah SMOTE)..... 116
Tabel 4.49 Hasil confusion matrix random forest 70:30 119
Tabel 4.50 SMOTE 80:20 random forest (sebelum splitting data)..... 120
Tabel 4.51 Hasil splitting data 80:20 random forest (sesudah SMOTE)..... 121
Tabel 4.52 Hasil confusion matrix random forest 80:20 124
Tabel 4.53 Perbandingan skenario evaluasi SVM 125
Tabel 4.54 Perbandingan skenario evaluasi SVM (SMOTE diawal) 127
Tabel 4.55 Perbandingan skenario evaluasi random forest..... 127
Tabel 4.56 Perbandingan skenario evaluasi random forest (SMOTE diawal).... 129
Tabel 4.57 Hasil perbandingan evaluasi SVM dan random forest 129
Tabel 4.58 Hasil perbandingan evaluasi (SMOTE diawal) 130

DAFTAR GAMBAR

Gambar 2.1 Machine Learning (Kristiawan & Widjaja, 2021). 8
Gambar 2.2 Tahapan preprocessing text..... 9
Gambar 2.3 Hyperplane SVM (Kusuma et al., 2022)..... 13
Gambar 2.4 Algoritma Random Forest (Isnayni et al., 2024)..... 15
Gambar 2.5 Confusion Matrix (Zhafran Muflih dkk., 2023)..... 17
Gambar 2.6 Aplikasi Topindoku..... 19
Gambar 3.1 IS Research Framework 24
Gambar 3.2 Diagram tahapan penelitian..... 26
Gambar 4.1 Visualisasi wordcloud kelas positif..... 51
Gambar 4.2 Visualisasi sepuluh kata positif terbanyak 54
Gambar 4.3 Visualisasi wordcloud kelas netral..... 56
Gambar 4.4 Visualisasi sepuluh kata netral terbanyak 59
Gambar 4.5 Visualisasi wordcloud kelas negatif..... 61
Gambar 4.6 Visualisasi sepuluh kata negatif terbanyak 64
Gambar 4.7 Confusion matrix 60:40 SVM..... 69
Gambar 4.8 Confusion matrix 70:30 SVM..... 74
Gambar 4.9 Confusion matrix 80:20 SVM..... 79
Gambar 4.10 Confusion matrix 80:20 SVM..... 79
Gambar 4.11 Confusion matrix 60:40 random forest 85
Gambar 4.12 Confusion matrix 70:30 random forest 91
Gambar 4.13 Confusion matrix 80:20 random forest 96
Gambar 4.14 Confusion matrix 60:40 SVM..... 100
Gambar 4.15 Confusion matrix 70:30 SVM..... 105
Gambar 4.16 Confusion matrix 80:20 SVM..... 109
Gambar 4.17 Confusion matrix random forest 60:40 114
Gambar 4.18 Confusion matrix random forest 70:30 119
Gambar 4.19 Confusion matrix random forest 80:20 124
Gambar 4.20 Perbandingan performa metode SVM dan Random Forest 131

DAFTAR LIST CODE

List Code 4.1 Scraping data 28
List Code 4.2 Scraping data lanjutan 28
List Code 4.3 Scraping data lanjutan 29
List Code 4.4 Distribusi data 31
List Code 4.5 Pemeriksaan data duplikat 32
List Code 4.6 Penghapusan data duplikat 33
List Code 4.7 Pemeriksaan tipe data 34
List Code 4.8 Penyesuaian tipe data 34
List Code 4.9 Cleaning data & case folding 36
List Code 4.10 Tokenizing 38
List Code 4.11 Stopwords-removal 40
List Code 4.12 Filtering token by length 42
List Code 4.13 Normalization 45
List Code 4. 14 Stemming 47
List Code 4.15 Pembobotan TF-IDF 49
List Code 4.16 Visualisasi wordcloud kelas positif 50
List Code 4.17 Analisis sepuluh kata positif terbanyak 52
List Code 4.18 Proses visualisasi sepuluh kata positif terbanyak 53
List Code 4.19 Proses visualisasi wordcloud kelas netral 55
List Code 4.20 Analisis sepuluh kata netral terbanyak 57
List Code 4.21 Proses visualisasi sepuluh kata netral terbanyak 58
List Code 4.22 Proses visualisasi wordcloud kelas negatif 60
List Code 4.23 Analisis sepuluh kata negatif terbanyak 62
List Code 4.24 Proses visualisasi sepuluh kata negatif terbanyak 63
List Code 4.25 Splitting data 60:40 SVM 65
List Code 4.26 SMOTE 60:40 SVM 66
List Code 4.27 Modeling 60:40 SVM 67
List Code 4.28 Confusion Matrix 60:40 SVM 68
List Code 4.29 Splitting data 70:30 SVM 70
List Code 4.30 SMOTE 70:30 SVM 71

List Code 4.31 Modeling 70:30 SVM..... 72
List Code 4.32 Confusion matrix 70:30 SVM..... 73
List Code 4.33 Splitting data 80:20 SVM..... 75
List Code 4.34 SMOTE 80:20 SVM..... 76
List Code 4.35 Modeling 80:20 SVM..... 78
List Code 4.36 Confusion matrix 80:20 SVM..... 79
List Code 4.37 Splitting data 60:40 random forest..... 81
List Code 4.38 SMOTE 60:40 random forest..... 82
List Code 4.39 Modeling 60:40 random forest..... 83
List Code 4.40 Confusion matrix 60:40 random forest..... 84
List Code 4.41 Splitting data 70:30 random forest..... 86
List Code 4.42 SMOTE 70:30 random forest..... 87
List Code 4.43 Modeling 70:30 random forest..... 89
List Code 4.44 Confusion matrix 70:30 random forest..... 90
List Code 4.45 Splitting data 80:20 random forest..... 92
List Code 4.46 SMOTE 80:20 random forest..... 93
List Code 4.47 Modeling 80:20 random forest..... 94
List Code 4.48 Confusion matrix 80:20 random forest..... 96
List Code 4.49 SMOTE 60:40 SVM (sebelum splitting data)..... 97
List Code 4.50 Splitting data 60:40 SVM (sesudah SMOTE)..... 98
List Code 4.51 Modeling SVM 60:40..... 99
List Code 4.52 Evaluasi model SVM 60:40..... 100
List Code 4.53 SMOTE 70:30 SVM (sebelum splitting data)..... 101
List Code 4.54 Splitting data 70:30 SVM (sesudah SMOTE)..... 102
List Code 4.55 Modeling SVM 70:30..... 103
List Code 4.56 Evaluasi model SVM 70:30..... 104
List Code 4.57 SMOTE 80:20 SVM (sebelum splitting data)..... 106
List Code 4.58 Splitting data 80:20 SVM (sesudah SMOTE)..... 107
List Code 4.59 Modeling SVM 80:20..... 108
List Code 4.60 Evaluasi model SVM 80:20..... 108
List Code 4.61 SMOTE 60:40 random forest (sebelum splitting data)..... 110
List Code 4.62 Splitting data 60:40 random forest (sesudah SMOTE)..... 111

List Code 4.63 Modeling random forest 60:40 112
List Code 4.64 Evaluasi model random forest 60:40 113
List Code 4.65 SMOTE 70:30 random forest (sebelum splitting data) 115
List Code 4.66 Splitting data 70:30 random forest (sesudah SMOTE) 116
List Code 4.67 Modeling random forest 70:30 117
List Code 4.68 Evaluasi model random forest 70:30 118
List Code 4.69 SMOTE 80:20 random forest (sebelum splitting data) 120
List Code 4.70 Splitting data 80:20 random forest (sesudah SMOTE) 121
List Code 4.71 Modeling random forest 80:20 122
List Code 4.72 Evaluasi model random forest 80:20 123

DAFTAR PERSAMAAN

Persamaan 2.1 Garis Hyperplane 13
Persamaan 2.2 Hyperplane yang bersinggungan 1 14
Persamaan 2.3 Hyperplane yang bersinggungan 2 14
Persamaan 2.4 Akurasi 17
Persamaan 2.5 Presisi 17
Persamaan 2.6 Recall 17
Persamaan 2.7 F1-Score 17

BAB I
PENDAHULUAN

1.1 Latar Belakang

Era digital saat ini membuat aplikasi daring semakin menjadi bagian penting dalam kehidupan sehari-hari, termasuk dalam kebutuhan transaksi bisnis, layanan publik, dan aktivitas rutin masyarakat. Salah satu aplikasi yang mengalami perkembangan adalah Topindoku, yang digunakan untuk berbagai transaksi digital seperti pengisian pulsa, paket data, dan layanan lainnya. Seiring meningkatnya penggunaan aplikasi tersebut, pengguna juga semakin aktif memberikan ulasan mengenai pengalaman mereka dalam menggunakan aplikasi. Ulasan tersebut dapat berupa komentar positif, negatif, maupun netral yang mencerminkan tingkat kepuasan pengguna terhadap layanan aplikasi.

Ulasan pengguna memiliki peran penting karena dapat menjadi sumber informasi bagi pengembang aplikasi untuk mengetahui kualitas layanan serta kebutuhan pengguna (Tandiapa & Rorimpandey, 2024). Untuk memahami isi ulasan secara lebih efektif, salah satu caranya yaitu dengan melakukan proses pengelompokkan opini pengguna ke dalam kategori positif, negatif, dan netral atau disebut dengan analisis sentimen. Namun, analisis sentimen memiliki beberapa tantangan, seperti ketidakseimbangan jumlah data pada setiap kelas sentimen, terutama ketika jumlah data pada kelas tertentu jauh lebih sedikit dibandingkan kelas lainnya. Kondisi tersebut dapat menyebabkan model *machine learning* menjadi bias terhadap kelas mayoritas. Oleh karena itu, penelitian ini menerapkan teknik *oversampling* menggunakan SMOTE untuk menyeimbangkan distribusi data sehingga proses klasifikasi dapat menghasilkan performa yang lebih optimal.

Dalam menghadapi permasalahan tersebut, *machine learning* menjadi salah satu pendekatan yang efektif dalam analisis sentimen. Penelitian ini menggunakan dua algoritma *machine learning*, yaitu *Support Vector Machine* dan *Random Forest* untuk mengklasifikasikan ulasan pengguna aplikasi Topindoku. *Support Vector Machine* (SVM) memiliki keunggulan dalam memisahkan data yang kompleks dan bekerja secara efektif pada data berdimensi tinggi sehingga sering digunakan pada klasifikasi teks. Sementara itu, *Random Forest* sebagai metode *ensemble* mampu

menghasilkan prediksi yang baik melalui kombinasi banyak pohon keputusan sehingga dapat menangani variasi data ulasan pengguna.

Metode *Support Vector Machine* dan *Random Forest* telah banyak digunakan dalam penelitian analisis sentimen dan menunjukkan hasil yang cukup baik. Penelitian yang dilakukan oleh Samantri (2024), memperoleh nilai akurasi sebesar 77% pada metode SVM dan 76% pada metode *Random Forest*, sehingga menunjukkan bahwa metode SVM memiliki performa yang sedikit lebih unggul dibandingkan *Random Forest*. Berdasarkan penelitian tersebut, penelitian ini tertarik untuk melakukan perbandingan kedua metode menggunakan data ulasan aplikasi Topindoku yang diperoleh melalui proses *scraping data*, *preprocessing*, penerapan SMOTE, *modeling*, dan evaluasi model.

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja metode *Support Vector Machine* dan *Random Forest* dalam menentukan sentimen terhadap ulasan aplikasi Topindoku. Pada penelitian ini dilakukan tahapan preprocessing teks yang meliputi *cleaning data*, *case folding*, *tokenizing*, *stopword removal*, *normalizing*, dan *stemming* untuk mempersiapkan data teks agar dapat diproses dengan baik oleh model *machine learning*. Selanjutnya dilakukan penerapan SMOTE untuk mengatasi ketidakseimbangan data sebelum proses klasifikasi. Setelah seluruh tahapan dilakukan, kedua metode dievaluasi menggunakan beberapa skenario pembagian data untuk mengetahui metode dengan performa terbaik dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Topindoku.

1.2 Rumusan Masalah

Berdasarkan pemaparan latar belakang diatas, rumusan masalah penelitian dapat disimpulkan, yaitu:

1. Bagaimana cara untuk melakukan pengelompokkan opini pengguna aplikasi Topindoku?
2. Bagaimana cara menentukan performa terbaik antara metode *Support Vector Machine* (SVM) dan *Random Forest* dalam melakukan analisis sentimen terhadap ulasan aplikasi Topindoku?
3. Bagaimana kinerja metode *Support Vector Machine* (SVM) dan *Random Forest* dalam klasifikasi ulasan pengguna tentang aplikasi Topindoku?

1.3 Tujuan Penelitian

Adapun tujuan penelitian berdasarkan rumusan masalah diatas yaitu sebagai berikut:

1. Melakukan analisis sentimen pada ulasan pengguna aplikasi Topindoku menggunakan metode *Support Vector Machine* dan *Random Forest*.
2. Membandingkan tingkat akurasi, presisi, *recall*, dan *F1-score* antara metode *Support Vector Machine* (SVM) dan *Random Forest* untuk menentukan metode yang lebih baik dalam analisis sentimen ulasan aplikasi Topindoku.
3. Menganalisis dan membandingkan performa metode SVM dan *Random Forest* dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Topindoku.

1.4 Batasan Masalah

1. Penelitian analisis sentimen ini berfokus pada ulasan pengguna dari aplikasi Topindoku di *Google Play Store*.
2. Ulasan pada penelitian ini menggunakan Bahasa Indonesia.
3. *Support Vector Machine* (SVM) dan *Random Forest* adalah dua perbandingan metode yang digunakan dalam penelitian ini.
4. Menggunakan bahasa pemrograman *Python*.
5. *Google Colaboratory* adalah alat yang digunakan untuk membantu peneliti mengolah data dalam penelitian ini.
6. Data yang digunakan adalah data yang berasal dari ulasan yang relevan di *Google Play Store*.

7. Kelas sentimen yang digunakan adalah positif, netral dan negatif.
8. Penelitian ini menggunakan data label manual sebagai bahan modeling dan evaluasi metode *Support Vector Machine* (SVM) dan *Random Forest*.
9. Data label manual berjumlah 5174 data.

1.5 Manfaat Penelitian

Adapun manfaat dengan adanya penelitian ini antara lain:

1. Bagi Peneliti

   Manfaat bagi peneliti yaitu memperoleh pengetahuan terkait proses menganalisis sentimen aplikasi Topindoku dengan memproyeksikan metode *Support Vector Machine* (SVM) dan *Random Forest* berdasarkan ilmu yang dipelajari pada saat perkuliahan dan pembuatan tugas akhir. Penelitian ini juga berfungsi sebagai langkah untuk memperoleh gelar Sarjana Satu di Program Studi Sistem Informasi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Tanjungpura.

2. Bagi Bidang Ilmu Sistem Informasi

   Manfaat dari penelitian ini untuk bidang ilmu sistem informasi adalah dapat berfungsi sebagai dokumen referensi bagi peneliti berikutnya yang tertarik untuk mengeksplorasi isu serupa atau melakukan studi terbaru mengenai analisis sentimen aplikasi *E-Commerce* dengan metode *Support Vector Machine* (SVM) dan *Random Forest*.

1.6 Metodologi Penulisan

Penulisan skripsi ini mengikuti struktur penulisan sesuai dengan panduan format Sistem Informasi dan terbagi menjadi empat bab yaitu dapat dijabarkan sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini berisikan latar belakang yang meliputi fenomena perkembangan transaksi digital di Indonesia dan gambaran umum permasalahan yang dialami oleh studi kasus dan dua perbandingan metode yang digunakan penulis hingga solusi berupa analisis sentimen dalam mengatasi persoalan pada studi kasus.

BAB II LANDASAN TEORI

Pada bab ini membahas secara singkat landasan teori yang digunakan dalam penelitian dan tinjauan pustaka yang relevan dari penelitian sebelumnya yang digunakan sebagai pembanding dan dasar pengembangan dengan topik penelitian saat ini.

BAB III METODOLOGI PENELITIAN

Pada bab ini berisikan metode yang digunakan dalam proses penelitian untuk menghasilkan *output* yang sesuai dengan tujuan dari penelitan dalam memecahkan masalah dengan cara struktural. Secara umum, penelitian ini menggunakan metode IS *Research* dan perbandingan dua metode SVM dan *Random Forest* yang dijelaskan dalam alur penelitian. Pada bab ini juga menjelaskan alat dan bahan penelitian untuk membangun model analisis sentimen SVM dan *Random Forest* pada studi kasus aplikasi Topindoku.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini membahas tentang hasil dari analisis sentimen aplikasi di *Google Play Store* dengan menerapkan metode *Support Vector Machine* (SVM) dan *Random Forest*. Pembahasan berisi mengenai langkah-langkah penulis dalam melakukan penelitian seperti *scraping data* dari aplikasi Topindoku di *Google Play Store*, *labeling*, *preprocessing data*, *modelling* SVM dan *Random Forest* dengan membagi dataset berdasarkan tiga skenario perbandingan, evaluasi sentimen, serta analisis dan visualisasi atau penggambaran terhadap hasil dari *modeling* yang dibuat.

BAB V PENUTUP

Pada bab ini menghasilkan kesimpulan dari hasil keseluruhan proses penelitian dan saran dari bab-bab yang telah diuraikan sebelumnya.