

**KOMPARASI ALGORITMA NAIVE BAYES DAN SUPPORT
VECTOR MECHINE MENGGUNAKAN LABELING
LEXICON DAN LABELING MANUAL PADA ULASAN
APLIKASI KAI ACCESS BY KAI**

SKRIPSI

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



disusun oleh

RIFKY AKBAR MAULANA

20.11.3627

Kepada

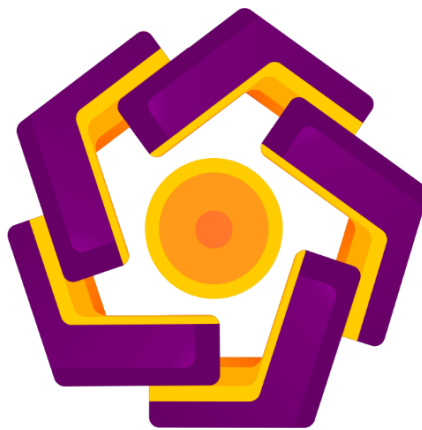
**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**KOMPARASI ALGORITMA NAIVE BAYES DAN SUPPORT
VECTOR MECHINE MENGGUNAKAN LABELING
LEXICON DAN LABELING MANUAL PADA ULASAN
APLIKASI KAI ACCESS BY KAI**

SKRIPSI

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



disusun oleh

RIFKY AKBAR MAULANA

20.11.3627

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**KOMPARASI ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR
MECHINE MENGGUNAKAN LABELING LEXICON DAN LABELING
MANUAL PADA ULASAN APLIKASI KAI ACCESS BY KAI**

yang disusun dan diajukan oleh

Rifky Akbar Maulana

20.11.3627

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 26 Agustus 2024

Dosen Pembimbing,



Ferian Fauzi Abdulloh, M.Kom

NIK. 190302276

HALAMAN PENGESAHAN

SKRIPSI

**KOMPARASI ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR
MECHINE MENGGUNAKAN LABELING LEXICON DAN LABELING
MANUAL PADA ULASAN APLIKASI KAI ACCESS BY KAI**

yang disusun dan diajukan oleh

Rifky Akbar Maulana

20.11.3627

Telah dipertahankan di depan Dewan Penguji
pada tanggal 26 Agustus 2024

Susunan Dewan Penguji

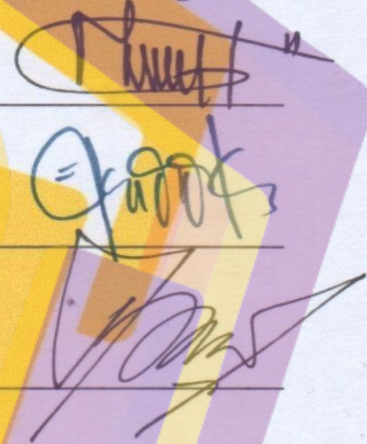
Nama Penguji

Norhikmah, M.Kom
NIK. 190302245

Anna Baita, M.Kom
NIK. 190302290

Ferian Fauzi Abdulloh, M.Kom
NIK. 190302276

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 26 Agustus 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 : Rifky Akbar Maulana
NIM : 20.11.3627

Menyatakan bahwa Skripsi dengan judul berikut:

**KOMPARASI ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MECHINE
MENGUNAKAN LABELING LEXICON DAN LABELING MANUAL PADA
ULASAN APLIKASI KAI ACCESS BY KAI**

Dosen Pembimbing: Ferian 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, 26 Agustus 2024

Yang Menyatakan,



Rifky Akbar Maulana

HALAMAN PERSEMBAHAN

Alhamdulillah segala puji Syukur atas rahmat dan karunia Allah SWT yang telah memberikan petunjuk dan kemudahan bagi penulis dalam menyelesaikan skripsi ini dengan sebaik-baiknya. Dengan ini penulis mempersembahkan skripsi ini kepada semua pihak yang telah membantu dengan memberikan motivasi serta dukungan yang besar sehingga skripsi ini terselesaikan dengan sebaik-baiknya.

Melalui skripsi ini penulis mempersembahkan terima kasih kepada:

1. Dengan penuh rasa Syukur kepada Allah SWT yang telah memberikan rahmat dan karuniaNya sehingga penulis memiliki kemampuan untuk menyelesaikan skripsi ini.
2. Kepada keluarga penulis Bapak Apit Yutantoro, Ibu Sri Lestari, kakak Nurhening Muchonati, kakak Aip Pradipta Farhan, keponakan penulis Giandra Shaka Pradipta yang tak hentinya memberikan doa, cinta kasih dan dukungannya untuk penulis.
3. Sahabat “SUBEDU” yaitu Ririn Sayang, Ageng, Bika, Dias, Faiq, Fariz, Mardiah, Mugen, Rapli, Ryan, Ryo, Sandi, Sindi, Teddy, Tobit, Ulin, Zalza yang telah menemani penulis selama perkuliahan, saling mendukung serta tempat berkeluh kesah.
4. Sahabat penulis Bagus Nugraha yang menemani penulis, memberikan dukungan serta mendengarkan keluh kesah penulis
5. Ririn Harti Ningsih yang memberikan dukungan serta kasih sayang dan menemani penulis dalam menyelesaikan skripsi ini.
6. Teman serta seluruh karyawan UPT ME yang selalu menemani dan mendukung dalam menyelesaikan skripsi ini
7. Terakhir, untuk diri sendiri terima kasih telah kuat berjuang dalam 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 PADA PLAY STORE TERKAIT KELUHAN PENGGUNA APLIKASI ‘KAI ACCESS BY KAI’ MENGGUNAKAN METODE NAÏVE BAYES” 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 S1 Informatika Universitas Amikom Yogyakarta.
4. 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 Apit Yutantoro dan ibu Sri Lestari selaku orang tua penulis yang memberikan dukungan, doa dan motivasi kepada penulis

Yogyakarta, 23 Agustus 2024

Rifky Akbar Maulana

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR ISTILAH	xiv
INTISARI	xv
<i>ABSTRACT</i>	xvi
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.....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Studi Literatur	6
2.2 Dasar Teori	15

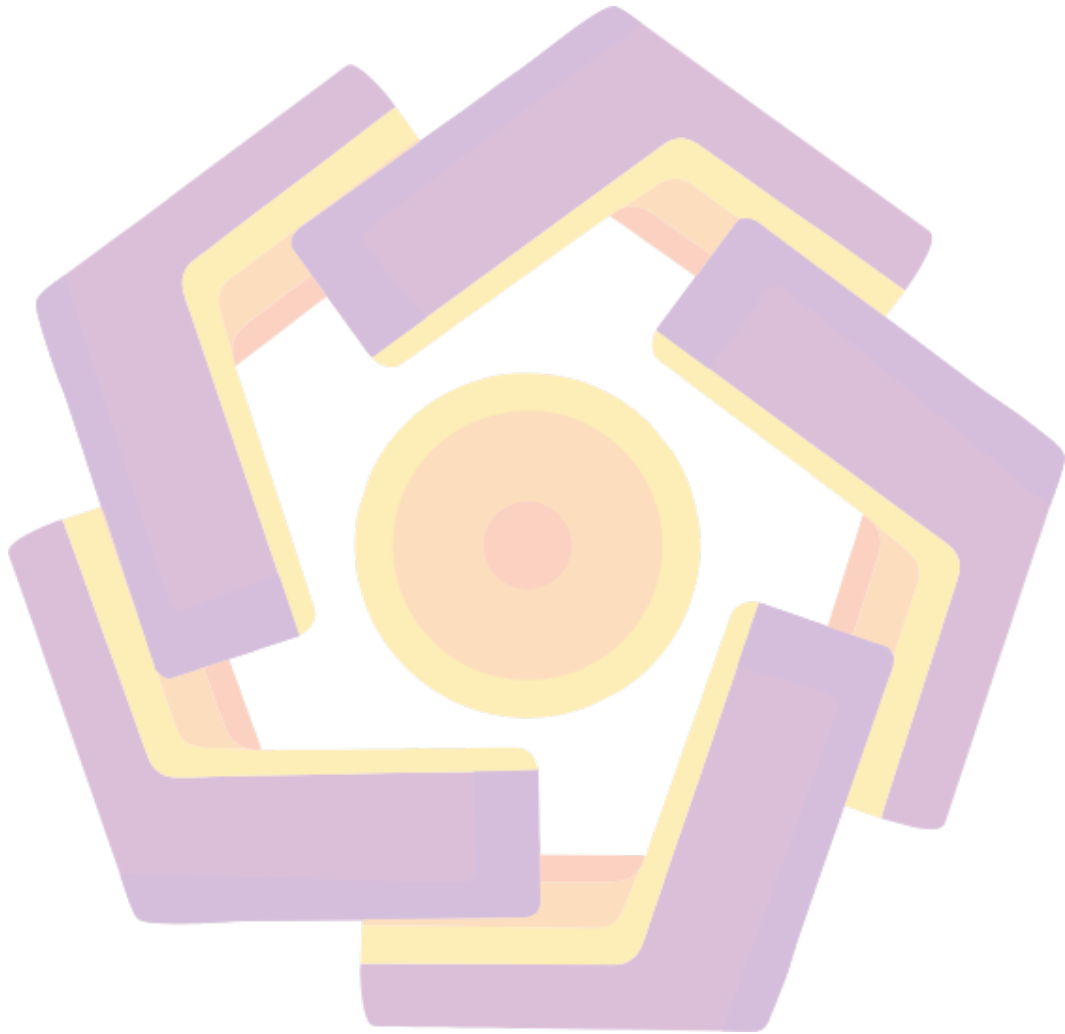
2.2.1 Analisis Sentimen	15
2.2.2 KAI ACCESS	15
2.2.3 Teks Mining	15
2.2.4 Python	16
2.2.5 Lexicon Based.....	16
2.2.6 Naïve Bayes	16
2.2.7 Support Vector Machine (SVM).....	17
2.2.8 Google Colab	18
2.2.9 K-fold Cross Validation	18
2.2.10 Preprocessing Text.....	19
2.2.11 Confusion Matrix	20
BAB III METODE PENELITIAN	22
3.1 Dataset	22
3.2 Alur Penelitian.....	23
3.2.1 Scapping Data.....	24
3.2.2 EDA	24
3.2.3 Pelabelan Data	25
3.2.4 Preprocessing	26
3.2.5 Word Cloud.....	27
3.2.6 Split Data	28
3.2.7 Feature Selection TF-IDF	28
3.2.8 Pemodelan dengan K-Fold.....	29
3.2.9 Komprasi Kinerja Algoritma	30
3.3 Alat dan Bahan	30
3.3.1 Alat penelitian	30
3.3.2 Bahan Penelitian	31
BAB IV HASIL DAN PEMBAHASAN	33

4.1 Scraping Data.....	33
4.2 Cleaning data	33
4.3 Translate Data	33
4.4 Labeling data.....	35
4.4.1 Lexicon	35
4.4.2 labeling manual	36
4.5 Preprocessing data	37
4.5.1 Case Folding	37
4.5.2 Cleansing.....	38
4.5.3 Word Normalization	38
4.5.4 Stop Removal.....	39
4.5.5 Stemming	40
4.5.6 Tokenizing	40
4.6 Wordcloud.....	41
4.6.1 Wordcloud Lexicon	41
4.6.2 Wordcloud Labeling manual.....	42
4.7 K-Fold Cross Validation	43
4.7.1 k-fold Cross Validation menggunakan lexicon.....	43
4.7.2 k-fold Cross Validation mennggunakan labeling manual.....	45
4.8 TF-IDF	46
4.9 Pemodelan data	47
4.9.1 Pelatihan.....	47
4.9.2 Evaluasi.....	48
4.10 Hasil	60
4.10.1 Modelling Labeling dengan Lexicon	61
4.10.2 Modelling Labeling Manual	62
BAB V PENUTUP.....	64

5.1 Kesimpulan.....64

5.2 Saran.....65

REFERENSI66



DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian.....	8
Tabel 2. 2 Ilustrasi Counfusion Matrix	20
Tabel 3. 1 Perbandingan dataset	23
Tabel 4. 1 Hasil Cleaning Data	33
Tabel 4. 2 Hasil Translate Data.....	34
Tabel 4. 3 Hasil Labeling Dataset.....	35
Tabel 4. 4 labeling Manual.....	36
Tabel 4. 5 case folding	37
Tabel 4. 6 Cleansing.....	38
Tabel 4. 7 Word Normalization	38
Tabel 4. 8 Stop Removal.....	39
Tabel 4. 9 Stemming	40
Tabel 4. 10 Tokenizing	40
Tabel 4. 11 hasil 10-Fold Cross Validation dengan lexicon	44
Tabel 4. 12 Hasil 10-Fold Cross Validation dengan labeling manual.....	46
Tabel 4. 13 Hasil dari Algoritma dengan pelabelan Manual dan Lexicon.....	63

DAFTAR GAMBAR

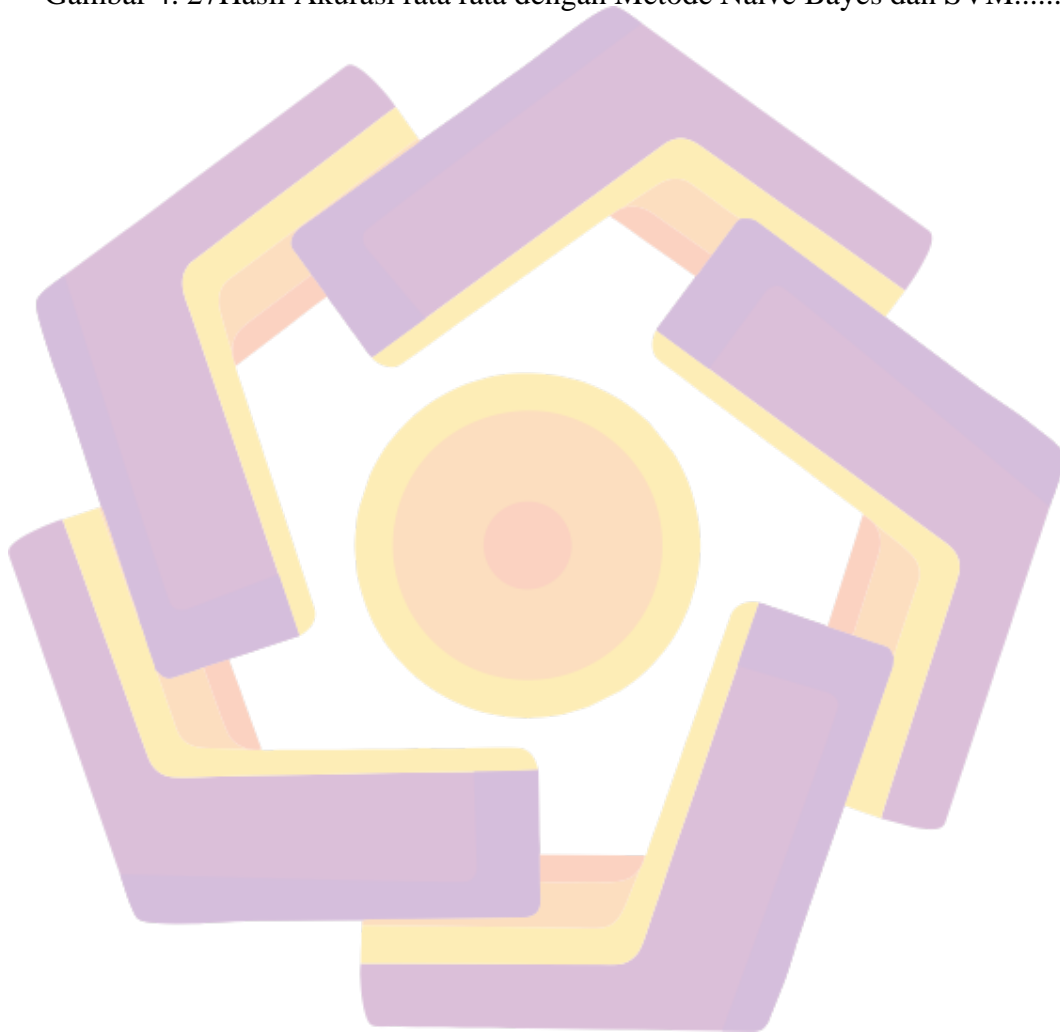
Gambar 2. 1 K-fold cross Validation	19
Gambar 3. 1 Grafik Jumlah Dataset Labelling Manual	22
Gambar 3. 2 Grafik Jumlah Dataset Labelling Lexicon	23
Gambar 3. 3 Alur Penelitian	24
Gambar 3. 4 Preprocessing data.....	26
Gambar 3. 5 Pemodelan data	29
Gambar 4. 1 Query Translate data	34
Gambar 4. 2 Query labeling data lexicon	35
Gambar 4. 3 wordcloud lexicon data positif	41
Gambar 4. 4 wordcloud lexicon data negatif	42
Gambar 4. 5 Wordcloud labeling manual data positif	42
Gambar 4. 6 Wordcloud labeling manual data negatif	43
Gambar 4. 7 Query K-Fold Cross Validation menggunakan lexicon	44
Gambar 4. 8 Perintah K-fold Cross Validation menggunakan labelling manual	45
Gambar 4. 9 Query TF-IDF	46
Gambar 4. 10 Pelatihan Naive Bayes	47
Gambar 4. 11 Pelatihan SVM	48
Gambar 4. 12 Code untuk Confusion Matrix pada Naive Bayes	49
Gambar 4. 13 Confusion Matrix 10-Fold pada Naive Bayes labeling lexicon	51
Gambar 4. 14 Hasil 10-Fold Naive Bayes labeling Lexicon	51
Gambar 4. 15 Code untuk Confusion Matrix SVM Labeling Lexicon.....	52
Gambar 4. 16 Confusion Matrix 10-Fold pada SVM	54
Gambar 4. 17 Confusion Matrix 10-Fold pada SVM	54
Gambar 4. 18 Code untuk Confusion Matrix Naive Bayes Labeling Manual	55
Gambar 4. 19 Confusion Matix 10-Fold pada Naive Bayes labeling manual	57
Gambar 4. 20 Hasil 10-Fold Naive Bayes Labeling manual	57
Gambar 4. 21 Code untuk Confusion Matrix SVM Labeling Manual.....	58
Gambar 4. 22 Confusion Matrix 10-Fold pada SVM labeling Manual	60
Gambar 4. 23 Hasil 10-Fold SVM Labeling Manual.....	60

Gambar 4. 24 Perbandingan dari seriap Fold dengan metode Naive Bayes dan SVM.....61

Gambar 4. 25 Hasil Akurasi rata rata dengan Metode Naive Bayes dan SVM61

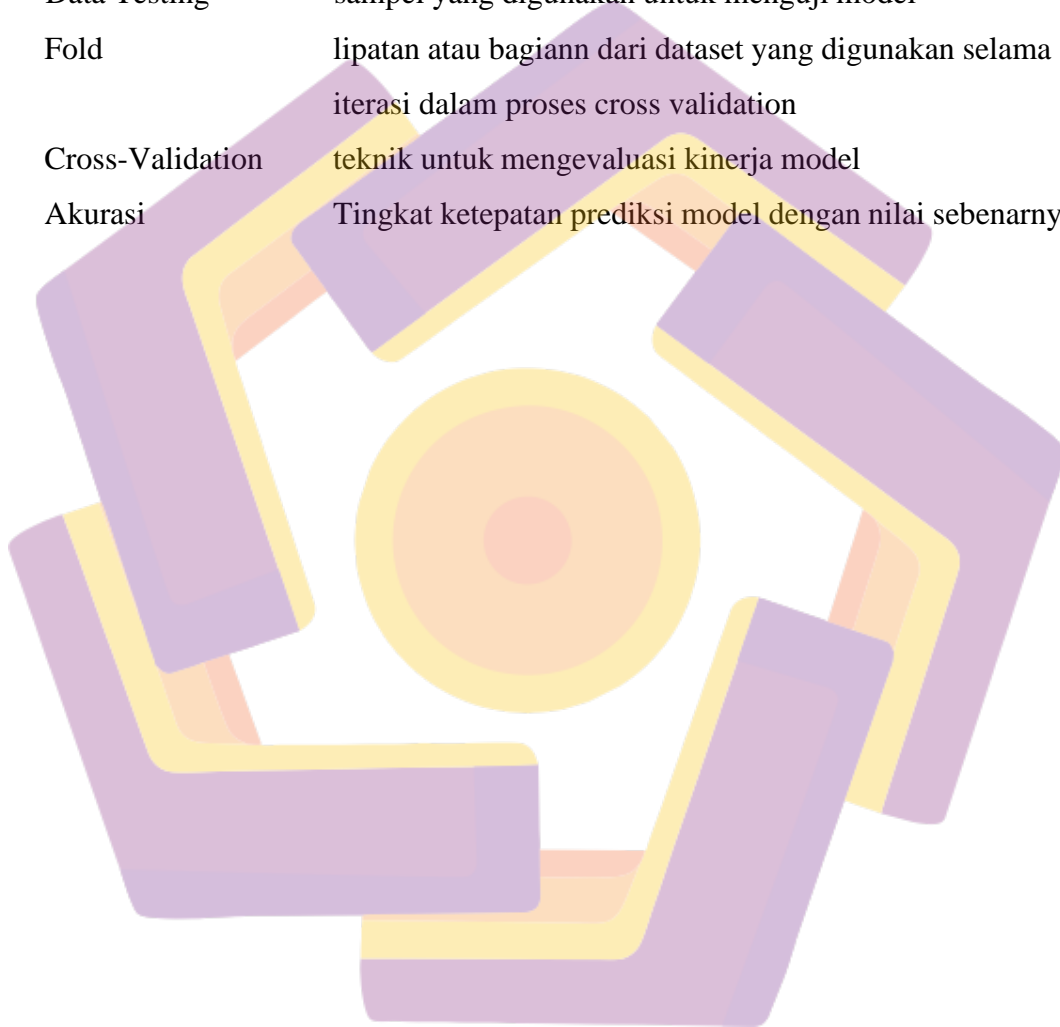
Gambar 4. 26 Perbandingan dari seriap Fold dengan metode Naive Bayes dan SVM.....62

Gambar 4. 27 Hasil Akurasi rata rata dengan Metode Naive Bayes dan SVM.....62



DAFTAR ISTILAH

Algoritma	metode yang direncanakan secara sistematis untuk menyelesaikan masalah
Data Training	sampel yang digunakan untuk melatih model
Data Testing	sampel yang digunakan untuk menguji model
Fold	lipatan atau bagian dari dataset yang digunakan selama iterasi dalam proses cross validation
Cross-Validation	teknik untuk mengevaluasi kinerja model
Akurasi	Tingkat ketepatan prediksi model dengan nilai sebenarnya



INTISARI

Perkembangan teknologi yang pesat telah mengubah cara bisnis dan layanan berkomunikasi dengan pelanggan mereka. Salah satunya adalah aplikasi KAI Access yang diluncurkan oleh PT. KAI untuk memudahkan konsumen dalam memesan tiket. Namun, banyak sekali review negatif terhadap aplikasi tersebut di google playstore. Dibalik keberhasilan aplikasi KAI ACCESS sendiri tentu saja memiliki Tingkat kepuasan berbeda pada pengguna dan pengalaman yang berbeda. Sehingga ulasan yang ada pada Google Playstore memiliki ulasan positif, negative. Ulasan tersebut juga merupakan hal penting untuk perkembangan aplikasi kedepannya. Penelitian ini menggunakan algoritma Naïve Bayes dan SVM untuk analisis sentiment data ulasan aplikasi KAI ACCESS yang didapat dari hasil scraping *Google Play Store*. Dataset yang dikumpulkan diberi label menggunakan pendekatan Lexicon dan manual untuk mempermudah preprocessing dataset dengan *casefolding*, *data cleaning*, *tokenizing*, *stop Removal*, *stemming*. Kemudian data dibagi menggunakan teknik K-Fold Cross Validation untuk membandingkan performa algoritma SVM dan Naïve Bayes. Hasil penelitian menunjukkan bahwa Naïve Bayes mencapai akurasi rata-rata 68,88% dan SVM mencapai 73,95% ketika menggunakan pelabelan Lexicon. Sementara itu, dengan pelabelan manual, Naïve Bayes mencapai akurasi 77,65%, sedangkan SVM mencapai 81,23%. Hal ini menunjukkan bahwa SVM bekerja lebih baik dalam analisis sentimen ulasan aplikasi KAI Access pada Google Play Store, baik dengan menggunakan pelabelan Lexicon maupun manual, karena nilai rata-rata akurasi SVM lebih tinggi dibandingkan dengan Naïve Bayes.

Kata kunci: Analisis Sentiment, KAI ACCESS, SVM, Naïve Bayes.

ABSTRACT

The rapid development of technology has changed the way businesses and services communicate with their customers. One such example is the KAI Access application, which was launched by PT KAI with the objective of facilitating ticket reservations for consumers. However, a considerable number of unfavorable reviews of the application have been published on the Google Play store. It is evident that the success of the KAI ACCESS application is contingent upon the varying levels of satisfaction and experiences of its users. Consequently, the reviews on Google Playstore are a combination of positive and negative evaluations. These reviews are also important for the future development of the application. This study employs Naïve Bayes and SVM algorithms for the sentiment analysis of KAI ACCESS application reviews, obtained through scraping the Google Play Store. The collected dataset is labeled using Lexicon and manual approaches to facilitate dataset preprocessing with case folding, data cleaning, tokenizing, stop removal, stemming. Then the data was divided using K-Fold Cross Validation technique to compare the performance of SVM and Naïve Bayes algorithms. The results demonstrated that the Naïve Bayes algorithm attained an average accuracy of 68.88% and the SVM algorithm achieved 73.95% when Lexicon labeling was employed. In contrast, when manual labeling was employed, Naïve Bayes attained an accuracy of 77.65%, while SVM reached 81.23%. This illustrates that SVM is more effective than Naïve Bayes in sentiment analysis of KAI Access application reviews on the Google Play Store, both with Lexicon and manual labeling, as evidenced by the higher average accuracy value of SVM.

Keyword: Sentiment Analysis, KAI ACCESS, SVM, Naïve Bayes.