

**OPTIMASI METODE MULTINOMIAL NAÏVE BAYES
DENGAN MENGGUNAKAN METODE LEVENSHTTEIN
DISTANCE PADA ULASAN APLIKASI KAI ACCESS**



Oleh :
MUHAMMAD IVAN FARIZ
1910651110

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER**

2023

TUGAS AKHIR

**OPTIMASI METODE MULTINOMIAL NAÏVE BAYES
DENGAN MENGGUNAKAN METODE LEVENSHTEIN
DISTANCE PADA ULASAN APLIKASI KAI ACCESS**

**Disusun untuk Melengkapi Tugas dan Memenuhi Syarat Kelulusan
Program Strata 1 Jurusan Teknik Informatika Fakultas Teknik
Universitas Muhammadiyah Jember**



Oleh :

MUHAMMAD IVAN FARIZ

1910651110

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER**

2023

HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR

**OPTIMASI METODE *MULTINOMIAL NAÏVE BAYES* DENGAN
MENGUNAKAN METODE *LEVENSHTEIN DISTANCE*
PADA ULASAN APLIKASI *KAI ACCESS***

Oleh:

Muhammad Ivan Fariz

1910651110

Telah disetujui bahwa Laporan Tugas Akhir ini untuk diajukan pada Sidang Tugas Akhir sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom.)

di

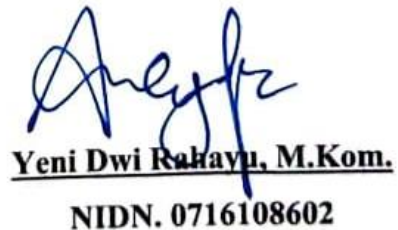
Universitas Muhammadiyah Jember

Menyetujui,

Pembimbing I


Deni Arifianto, M.Kom.
NIDN. 0718068103

Pembimbing II


Yeni Dwi Rahayu, M.Kom.
NIDN. 0716108602

HALAMAN PENGESAHAN
OPTIMASI METODE *MULTINOMIAL NAÏVE BAYES* DENGAN
MENGGUNAKAN METODE *LEVENSHTEIN DISTANCE*
PADA ULASAN APLIKASI *KAI ACCESS*

Oleh :
Muhammad Ivan Fariz
1910651110

Telah mempertanggung jawabkan Proposal Tugas Akhirnya pada sidang Proposal
Tugas Akhir tanggal 5 Agustus 2023 sebagai salah satu syarat kelulusan dan
mendapatkan gelar Sarjana Komputer (S.Kom.)

Di
Universitas Muhammadiyah Jember


Disetujui oleh,

Dosen Penguji :
Penguji I


Ari Eko Wardoyo, S.T., M.Kom.

NIDN. 0014027501

Penguji II


Amalina Maryam Zakiyah S.E., M.Si

NIDN. 0727058705

Mengesahkan,

Dekan Fakultas Teknik



Dr. Ir. Nanang Safful Rizal, S.T., MT., IPM.
NIDN. 0705047806

Dosen Pembimbing
Pembimbing I


Deni Arifiantu, M.Kom.

NIDN. 0718068103

Pembimbing II


Yeni Dwi Rahayu, M.Kom.

NIDN. 0716108602

Mengetahui,

Ketua Program Studi
Teknik Informatika


Ari Eko Wardoyo, S.T., M.Kom.
NIDN. 0014027501

LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini :

Nama : Muhammad Ivan Fariz
Nomor Induk Mahasiswa : 1910651110
Program Studi : Teknik Informatika
Perguruan Tinggi : Universitas Muhammadiyah Jember

Menyatakan dengan sesungguhnya karya ilmiah berupa tugas akhir yang berjudul **“OPTIMASI METODE *MULTINOMIAL NAÏVE BAYES* DENGAN MENGGUNAKAN METODE *LEVENSHTEIN DISTANCE* PADA ULASAN APLIKASI *KAI ACCESS*”** adalah murni hasil tanpa adanya tekanan dan paksaan dari pihak manapun hasil buatan atau karya penulis sendiri, bukan hasil plagiasi ataupun mengambil karya orang lain kecuali dalam bentuk kutipan yang telah dicantumkan sumbernya. Sekian pernyataan ini penulis buat tanpa adanya tekanan ataupun paksaan dari pihak manapun.

Jember, 5 Agustus 2023



Muhammad Ivan Fariz

NIM. 1910651110

ABSTRAK

Fariz, Muhammad Ivan. 2023. Optimasi Metode *Multinomial Naïve Bayes* Dengan Menggunakan Metode *Levenshtein Distance* Pada Ulasan Aplikasi *KAI Access*. Tugas Akhir. Program Sarjana. Program Studi Teknik Informatika. Universitas Muhammadiyah Jember.

Pembimbing: (1) Deni Arifianto, M.Kom. ; (2) Yeni Dwi Rahayu, M.Kom.

KAI Access merupakan aplikasi untuk mempermudah pengguna dalam mengakses layanan dan informasi terkait tiket perjalanan kereta api. *KAI Access* memiliki fitur ulasan beraneka ragam yang merupakan wadah bagi pengguna untuk memberikan *feedback*. Pada penelitian ini akan dilakukan analisis sentimen terhadap data ulasan *KAI Access* sejumlah 8713 data ulasan. Metode yang digunakan yaitu *multinomial naïve bayes* dan metode *levenshtein distance*. Dari kombinasi metode tersebut diharapkan dapat meningkatkan hasil klasifikasi. Seluruh data akan dilakukan pemodelan menggunakan metode *K Fold Cross Validation* dengan nilai $k=2,3,4,5,6,7,8,9$ dan 10. Berdasarkan hasil pengujian menggunakan *multinomial naïve bayes* diperoleh nilai tertinggi yaitu *accuracy* sebesar 92%, tingkat *precision* sebesar 60% dan tingkat *recall* sebesar 45%. Sedangkan kombinasi metode *multinomial naïve bayes-levenshtein distance* mengalami peningkatan pada pengujian K-ke 9 yang awalnya menggunakan *multinomial naïve bayes* saja mendapatkan nilai *accuracy* 83%, *precision* 59% dan *recall* 44%, kemudian meningkat ketika menggunakan kombinasi metode *levenshtein distance* sebesar *accuracy* 84%, *precision* 60% dan *recall* 46%. Proses pengujian yang telah dilakukan dapat disimpulkan bahwa optimasi *multinomial naïve bayes* menggunakan *levenshtein distance* meningkatkan hasil *accuracy*, *precision* dan *recall* sebesar 1-2%.

Kata Kunci: *KAI Access, Analisis Sentimen, Multinomial Naïve Bayes, Levenshtein Distance*

ABSTRACT

Fariz, Muhammad Ivan. 2023. Optimization of the Multinomial Naïve Bayes Method Using the Levenshtein Distance Method on KAI Access Application Reviews. Final Project. Undergraduate Program, Informatics Engineering Study Program. University of Muhammadiyah Jember.

Advisors: (1) Deni Arifianto, M.Kom; (2) Yeni Dwi Rahayu, M.Kom.

KAI Access is an application to facilitate users in accessing services and information related to train travel tickets. KAI Access has a diverse review feature which is a forum for users to provide feedback. In this research, sentiment analysis will be carried out on KAI Access review data totaling 8713 review data. The methods used are multinomial naïve bayes and levenshtein distance method. The combination of these methods is expected to improve classification results. All data will be modeled using the K Fold Cross Validation method with a value of $k = 2, 3, 4, 5, 6, 7, 8, 9$ and 10 . Based on the test results using multinomial naïve bayes, the highest value is obtained, namely accuracy of 92%, precision level of 60% and recall level of 45%. While the combination of the multinomial naïve bayes-levenshtein distance method experienced an increase in the K-9th test which initially used multinomial naïve bayes alone to get an accuracy value of 83%, precision of 59% and recall of 44%, then increased when using a combination of levenshtein distance methods of 84% accuracy, 60% precision and 46% recall. It can be concluded from the testing process that has been carried out that naïve Bayes multinomial optimization using Levenshtein distance increases accuracy, precision and recall by 1-2%.

Keywords: *KAI Access, Sentiment Analysis, Multinomial Naïve Bayes, Levenshtein Distance*

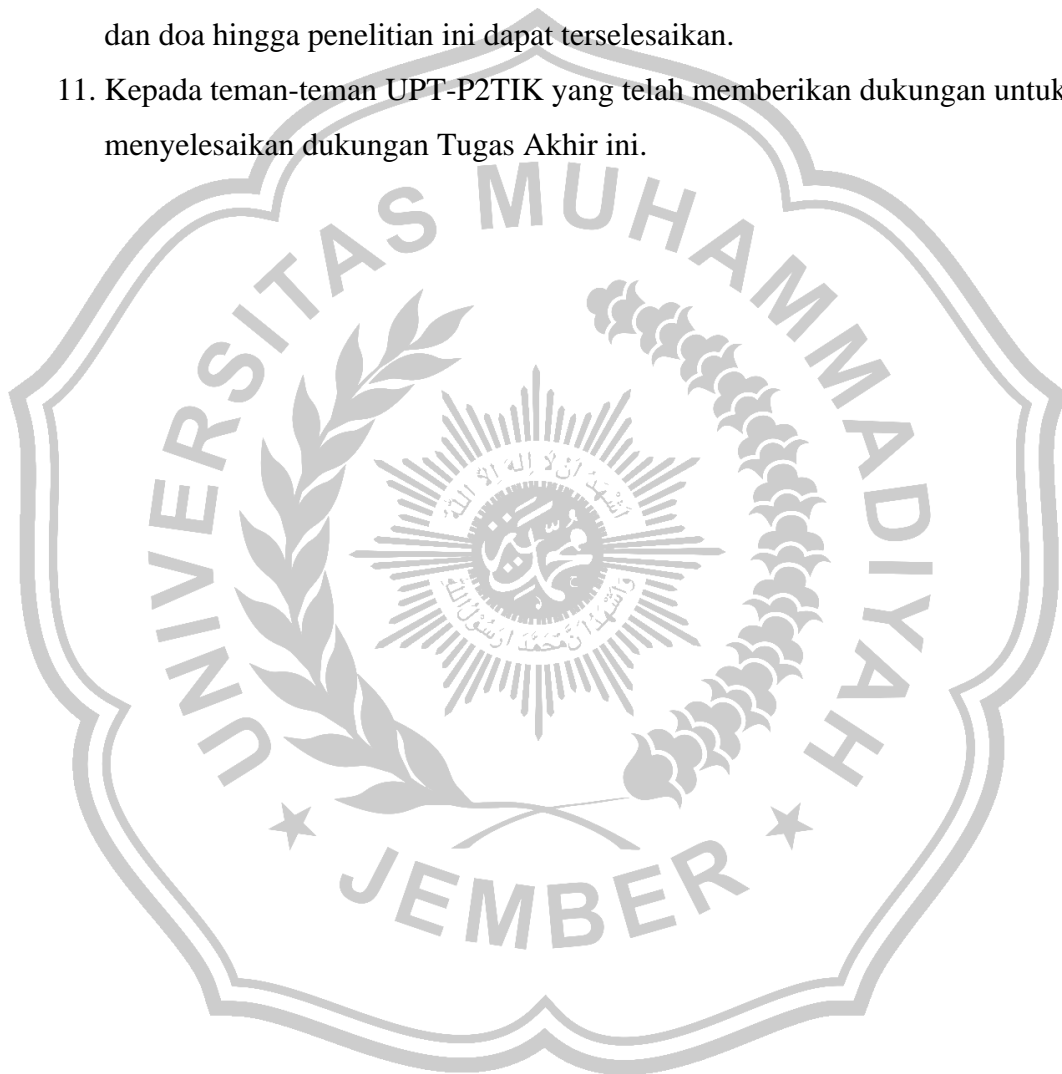
LEMBAR PERSEMBAHAN

Alhamdulillah bersyukur Allah SWT atas segala rahmat, taufiq, dan hidayahnya, atas Ridho dan karunianya penulis dapat menyelesaikan dan diberikan kemudahan dalam menyelesaikan skripsi ini dengan baik.

Atas segala upaya, bimbingan dan arahan dari semua pihak, penulis mengucapkan banyak terimakasih. Sebagai ucapan terima kasih penulis persembahkan kepada:

1. Allah SWT berkat segala ridho, rahmat dan hidayah-Nya penulis dapat menyelesaikan segala urusan dalam menyusun laporan Tugas Akhir dan diberikan kesempatan mendapatkan gelar Sarjana Komputer.
2. Bapak Dr. Ir. Nanang Saiful Rizal, S.T., M.T., IPM. selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jember.
3. Bapak Ari Eko Wardoyo, S.T., M.Kom. selaku Ketua Program Studi Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember.
4. Bapak Deni Arifianto, M.Kom. selaku dosen pembimbing 1 tugas akhir dan Ibu Yeni Dwi Rahayu, M.Kom. selaku dosen pembimbing 2 tugas akhir yang telah meluangkan waktu dan memberikan bimbingan dari awal hingga akhir dengan penuh kesabaran sehingga tugas akhir ini dapat terselesaikan.
5. Bapak Ari Eko Wardoyo, S.T., M.Kom selaku dosen penguji 1 dan Ibu Amalina Maryam Zakiiyyah S.E.,M.Si. selaku dosen penguji 2 yang telah memberikan masukan dan saran dalam penyelesaian tugas akhir ini.
6. Kepada Bapak dan Ibu Dosen Program Studi Teknik Informatika Universitas Muhammadiyah Jember yang telah memberikan ilmunya kepada penulis
7. Kepada kedua orang tua dan adik saya yang selama ini telah membantu saya dalam berbagai hal, baik semangat, perhatian, doa demi kelancaran agar dapat menyelesaikan tugas akhir ini.
8. Kepada Imas Kusnul Alviah S.I.Kom. yang selalu senantiasa ada memberikan semangat, dukungan dan selalu sabar selama menemani penulis dalam menyelesaikan tugas akhir ini.

9. Kepada Zulkifli Aminullah teman seperjuangan yang selalu mendukung, membantu satu sama lain dalam mencapai target kuliah.
10. Kepada Sagtya, Bagas, Ravi, Rivan, Gatot, Yoga dan Zelda teman seperjuangan belajar bersama mulai dari awal kuliah sampai akhir. Serta teman-teman mahasiswa Teknik Informatika Angkatan 2019 dan sahabat-sahabat yang telah menjadi penyemangat, memberikan dukungan, omelan dan doa hingga penelitian ini dapat terselesaikan.
11. Kepada teman-teman UPT-P2TIK yang telah memberikan dukungan untuk menyelesaikan dukungan Tugas Akhir ini.



MOTTO

"Cobaan hidupmu bukanlah untuk menguji kekuatan dirimu. Tapi menakar seberapa besar kesungguhan dalam memohon pertolongan kepada Allah."

(Ibnu Qayyim)

" Saat orang-orang menganggap mu tidak bisa apa-apa, jangan pedulikan. Karena yang bisa mengubah nasibmu adalah kamu, bukan orang lain."

(Rock Lee)



KATA PENGANTAR

Puji syukur saya panjatkan ke hadirat Allah SWT atas rahmat, karunia dan hidayahnya penulis dapat menyelesaikan tugas akhir yang berjudul "**OPTIMASI METODE *MULTINOMIAL NAÏVE BAYES* DENGAN MENGGUNAKAN METODE *LEVENSHTEIN DISTANCE* PADA ULASAN APLIKASI *KAI ACCESS***" disusun sebagai salah satu syarat mengerjakan skripsi pada program S1 di Fakultas Teknik Program Studi Teknik Informatika Universitas Muhammadiyah Jember.

Penulis menyadari dalam penyusunan proposal skripsi ini tidak selesai tanpa dukungan dari Orang tua, Bapak, Ibu pembimbing dan penguji akan sangat sulit untuk menyelesaikan tugas akhir ini. Karena itu pada kesempatan ini saya ingin mengucapkan terima kasih kepada bapak dan ibu pembimbing serta penguji, saya menyadari bahwa proposal tugas akhir ini jauh dari sempurna, semoga proposal tugas akhir ini dapat berguna dan bermanfaat terutama bagi pihak-pihak yang tertarik untuk mengkaji dan mengembangkannya.

Jember, 5 Agustus 2023



Muhammad Ivan Fariz

NIM. 1910651110

DAFTAR ISI

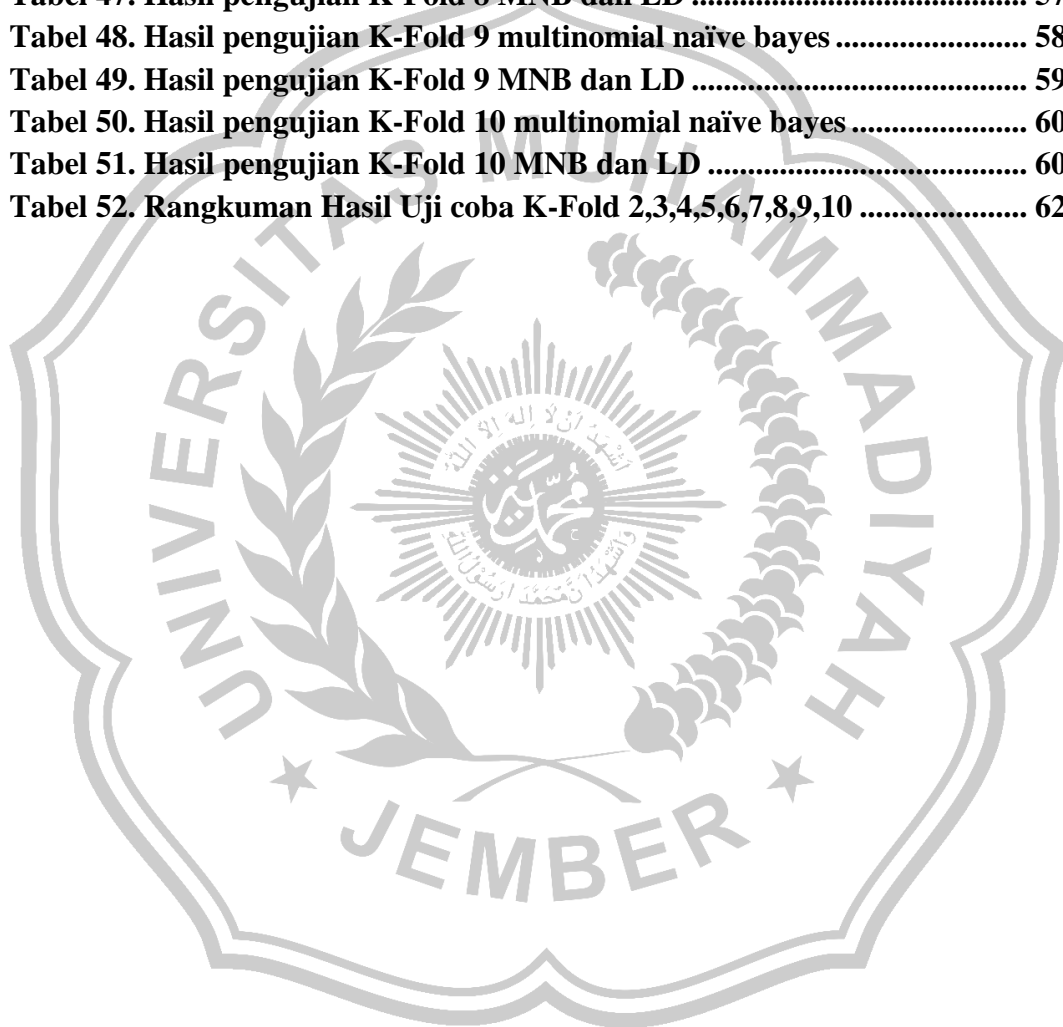
HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR.....	i
HALAMAN PENGESAHAN.....	ii
LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR.....	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
LEMBAR PERSEMBAHAN	vi
MOTTO	viii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Penelitian Terdahulu.....	5
2.2 <i>KAI Access</i>	6
2.3 <i>Google Play</i>	7
2.4 Analisis Sentimen	7
2.5 <i>Text Mining</i>	8
2.6 <i>Preprocessing Text</i>	8
2.7 Metode Klasifikasi.....	9
2.7.1 <i>Naïve Bayes</i>	10
2.7.2 <i>Naïve Bayes Multinomial</i>	10
2.8 <i>Levenshtein Distance</i>	11
2.9 TF-IDF.....	13
2.10 <i>K-Fold Cross validation</i>	14
2.11 <i>Confusion Matrix</i>	15

2.12	<i>Phyton</i>	16
2.13	Data Sampel	17
BAB III METODE PENELITIAN		18
3.1.	Diagram Alur Penelitian	18
3.2.	Studi Pustaka	19
3.3.	Pengumpulan Data	19
3.4.	Pelabelan data	20
3.5.	<i>Text Preprocessing</i>	20
3.6.	Normalisasi Kata dengan Algoritma <i>Levenshtein Distance</i>	24
3.7.	Implementasi <i>Naïve Bayes Multinomial</i>	27
3.7.1.	Pembobotan TF-IDF	28
3.7.2.	<i>Multinomial Naïve Bayes</i>	29
3.8.	Pengujian Validasi dan Evaluasi	32
BAB IV HASIL DAN PEMBAHASAN		36
4.1.	Pengumpulan Data	36
4.2.	Validasi Sentimen	36
4.3.	<i>Text Preprocessing</i>	38
4.4.	Implementasi Metode <i>Levenshtein Distance</i>	43
4.5.	Pembobotan TF-IDF dan Implementasi Metode <i>Multinomial Naïve Bayes</i>	47
4.6.	Pengujian Data	49
4.7.	Hasil Evaluasi dan Validasi	62
4.8.	Hasil Analisis	64
BAB V KESIMPULAN DAN SARAN		68
5.1.	Kesimpulan	68
5.2.	Saran	68
DAFTAR PUSTAKA		69
LAMPIRAN-LAMPIRAN		73
BIODATA DIRI		83

DAFTAR TABEL

Tabel 1. Contoh Model 3-Fold Cross Validation	15
Tabel 2. Confusion Matrix	15
Tabel 3. Contoh data komentar scrapping aplikasi KAI Access	19
Tabel 4. Contoh pelabelan data secara manual	20
Tabel 5. Tahapan cleaning	20
Tabel 6. Tahapan case folding	21
Tabel 7. Tahapan tokenizing	22
Tabel 8. Tahapan stopword removal.....	23
Tabel 9. Tahapan stemming	23
Tabel 10 Hasil Scrapping	25
Tabel 11. Perhitungan normalisasi kata levenshtein distance	26
Tabel 12. Perhitungan normalisasi kata levenshtein distance	26
Tabel 13. Perhitungan normalisasi kata levenshtein distance	27
Tabel 14. Contoh normalisasi levensthein distance	27
Tabel 15. Contoh dokumen yang akan dihitung nilai TF-IDF	28
Tabel 16. Data beserta label	29
Tabel 17. Nilai probabilitas term dokumen uji	30
Tabel 18. Hasil klasifikasi.....	32
Tabel 19. Hasil kriteria untuk Accuracy.....	32
Tabel 20. Hasil kriteria untuk Precision & Accuracy Positif.....	33
Tabel 21. Hasil kriteria untuk Precision & Accuracy Negatif.....	33
Tabel 22. Hasil kriteria untuk Precision & Accuracy Netral.....	33
Tabel 23. Evaluasi dan Validasi K-Fold.....	34
Tabel 24. Data hasil scrapping.....	36
Tabel 25. Validator penentuan sentimen ulasan aplikasi.....	37
Tabel 26. Hasil validasi sentimen.....	37
Tabel 27. Jumlah kategori data	38
Tabel 28. Hasil case folding dan cleaning	38
Tabel 29. Hasil Tokenezing	39
Tabel 30. Hasil Filtering	41
Tabel 31. Hasil Stemming.....	42
Tabel 32. Hasil Pengujian Metode Levenshtein Distance.....	44
Tabel 33. Contoh beberapa pengujian keakuratan hasil LD.....	45
Tabel 34. Hasil pengujian K-Fold 2 multinomial naïve bayes	49
Tabel 35. Hasil Pengujian K-Fold 2 MNB dan LD	50
Tabel 36. Hasil pengujian K-Fold 3 multinomial naïve bayes	51
Tabel 37. Hasil pengujian K-Fold 3 MNB dan LD	51
Tabel 38. Hasil pengujian K-Fold 4 multinomial naïve bayes	52

Tabel 39. Hasil pengujian K-Fold 4 MNB dan LD	52
Tabel 40. Hasil pengujian K-Fold 5 multinomial naïve bayes	53
Tabel 41. Hasil pengujian K-Fold 5 MNB dan LD	53
Tabel 42. Hasil pengujian K-Fold 6 multinomial naïve bayes	54
Tabel 43. Hasil pengujian K-Fold 6 MNB dan LD	55
Tabel 44. Hasil pengujian K-Fold 7 multinomial naïve bayes	56
Tabel 45. Hasil pengujian K-Fold 7 MNB dan LD	56
Tabel 46. Hasil pengujian K-Fold 8 Multinomial naïve bayes.....	57
Tabel 47. Hasil pengujian K-Fold 8 MNB dan LD	57
Tabel 48. Hasil pengujian K-Fold 9 multinomial naïve bayes	58
Tabel 49. Hasil pengujian K-Fold 9 MNB dan LD	59
Tabel 50. Hasil pengujian K-Fold 10 multinomial naïve bayes	60
Tabel 51. Hasil pengujian K-Fold 10 MNB dan LD	60
Tabel 52. Rangkuman Hasil Uji coba K-Fold 2,3,4,5,6,7,8,9,10	62



DAFTAR GAMBAR

Gambar 1. Tahapan Penelitian	18
Gambar 2. Flowchart Levenshtein Distance	25
Gambar 3. Perhitungan pembobotan TF-IDF	28
Gambar 4. Beberapa kata penting TF IDF	48
Gambar 5. Rata rata Perbandingan Accuracy	65
Gambar 6. Rata rata Perbandingan Precision.....	65
Gambar 7. Rata-rata perbandingan Recall.....	66

