

**PERBANDINGAN METODE SVM DAN MULTINOMIAL
NAÏVE BAYES UNTUK KLASIFIKASI KOMENTAR
MASYARAKAT DKI JAKARTA**

SKRIPSI

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



disusun oleh

RANIYA RAKARAHAYU PUTRI

20.11.3803

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**PERBANDINGAN METODE SVM DAN MULTINOMIAL
NAÏVE BAYES UNTUK KLASIFIKASI KOMENTAR
MASYARAKAT DKI JAKARTA**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

RANIYA RAKARAHAYU PUTRI

20.11.3803

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN METODE SVM DAN MULTINOMIAL
NAÏVE BAYES UNTUK KLASIFIKASI KOMENTAR
MASYARAKAT DKI JAKARTA**

yang disusun dan diajukan oleh

Raniya Rakarabaya Putri

20.11.3803

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 19 Februari 2024

Dosen Pembimbing,



Nuri Cahyono, M.Kom
NIK. 190302278

HALAMAN PENGESAHAN

SKRIPSI

PERBANDINGAN METODE SVM DAN MULTINOMIAL NAÏVE BAYES UNTUK KLASIFIKASI KOMENTAR MASYARAKAT DIKI JAKARTA

yang disusun dan diajukan oleh

Raniya Bakarrahayu Putri
20.11.3803

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Februari 2024

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Krisnawati, S.Si., M.T.
NIK. 190302030



Yuli Astuti, M.Kom
NIK. 190302146



Nuri Cahyono, M.Kom
NIK. 19030278



Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Februari 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 : Raniya Rakarabaya Putri
NIM : 20.11.3803

Menyatakan bahwa Skripsi dengan judul berikut:

PERBANDINGAN METODE SVM DAN MULTINOMIAL NAÏVE BAYES UNTUK KLASIFIKASI KOMENTAR MASYARAKAT DKI JAKARTA

Dosen Pembimbing : Nuri Cahyono 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 Februari 2024

Yang Menyatakan,



UNIVERSITAS AMIKOM
YOGYAKARTA
MITRA
TEMPEL
CCALX04400458

Raniya Rakarabaya Putri

HALAMAN PERSEMBAHAN

Bismillahirrahmanirrahim, dengan nama Allah Yang Maha Pengasih lagi Maha Penyayang, penulis letakkan persembahan ini. Teriring doa dan syukur yang tak terhingga atas segala limpahan rahmat dan petunjuk-Nya yang menghampiri dalam setiap langkah hingga penulis mampu menyelesaikan tugas akhir ini tepat pada waktunya. Tugas akhir ini penulis persembahkan untuk:

1. Pilar ku yang kokoh, Ayahanda Safruddin, Penulis tahu ada beban yang beliau pikul sendiri, tanpa penulis sadari. Beliau memang tidak sempat menyelesaikan bangku perkuliahannya karena adanya suatu halangan. Namun, dengan kesungguhan dan doa beliau, penulis mampu menyelesaikan studi hingga jenjang sarjana. Setiap langkah yang penulis ambil, setiap pencapaian yang penulis raih, akan menjadi bukti atas keteguhan dan cinta kasih yang tak terhingga dari Ayah. Terima kasih atas segala pengorbanan dan doa yang selalu mengiringi langkah penulis. Dengan penuh rasa hormat dan cinta, penulis mengucapkan terima kasih.
2. Surgaku, Ibunda Sukrah. Penulis menyadari betapa besar pengorbananmu untuk keluarga, mengejar impian dan membimbing kami tanpa mengenal lelah. Setiap senyummu, setiap doa yang engkau panjatkan, menjadi pelipur lara dan kekuatan yang tak tergantikan. Terima kasih, mama, atas segala cinta, dukungan, dan doa yang tak pernah pudar. Dengan rasa cinta dan hormat yang dalam, penulis berjanji untuk selalu menjaga nama baikmu dan meneruskan warisan cintamu kepada generasi selanjutnya.
3. Teman-teman yang sudah tulus membantu. Tiada kata yang mampu menggambarkan betapa berharganya bantuan dan dukungan yang kalian berikan dalam perjalanan menyusun tugas akhir ini. Dengan penuh rasa haru, penulis ingin menyampaikan ucapan terima kasih yang tak terhingga atas segala bantuan dan dukungan dari kalian semua.

KATA PENGANTAR

Dengan segala kerendahan hati, penulis ingin menyampaikan ucapan terima kasih yang tak terhingga kepada Allah SWT atas segala rahmat, karunia, dan petunjuk-Nya yang senantiasa mengiringi perjalanan penelitian ini. Shalawat serta salam tak terhingga senantiasa tercurah kepada baginda Rasulullah SAW, sebagai suri teladan bagi seluruh umat manusia.

Tak lupa, penulis juga ingin menyampaikan penghargaan yang tulus kepada Bapak/Ibu Dosen Pembimbing, Tim Dosen Penguji, dan seluruh staf pengajar Universitas Amikom Yogyakarta yang telah memberikan arahan, bimbingan, serta dukungan dalam penyusunan tugas akhir ini. Terima kasih juga kepada semua pihak yang terlibat dalam memberikan sumbangan dana dan fasilitas yang sangat berarti dalam menyelesaikan penelitian ini. Ucapan terima kasih yang sebesar-besarnya juga disampaikan kepada kedua orang tua yang senantiasa memberikan doa, semangat, dan dukungan moral dalam setiap langkah perjalanan ini. Semoga penelitian ini dapat memberikan kontribusi yang bermanfaat bagi perkembangan ilmu pengetahuan, terutama dalam pemahaman sentimen masyarakat terhadap pelayanan publik di DKI Jakarta. Terima kasih.

Yogyakarta, 16 Februari 2024

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI	xvi
<i>ABSTRACT</i>	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6

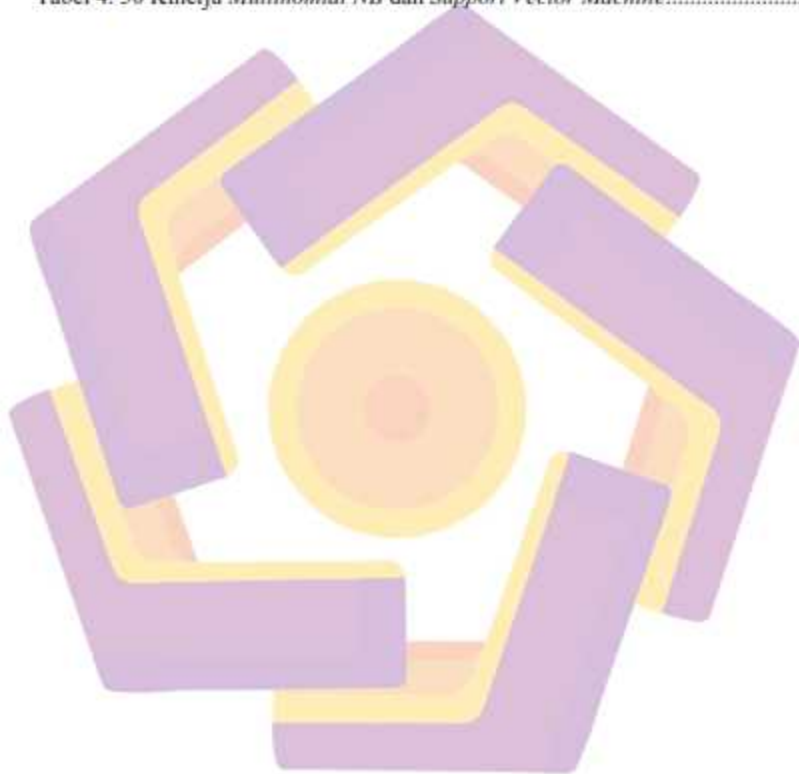
2.2	Dasar Teori.....	15
2.2.1	<i>Text Mining</i>	15
2.2.2	Analisis Sentimen.....	15
2.2.3	<i>Text Preprocessing</i>	16
2.2.4	Pembobotan Kata TF-IDF.....	16
2.2.5	<i>Super Vector Machine</i>	17
2.2.6	<i>Naïve Bayes</i>	18
2.2.7	<i>Multinomial Naïve Bayes</i>	18
2.2.8	<i>Grid Search Cross Validation</i>	19
2.2.9	Ketetapan Klasifikasi.....	19
2.2.10	K-Cross Vold Validation.....	20
2.2.11	Instagram.....	20
2.2.12	@dkijakarta.....	21
BAB III METODE PENELITIAN.....		22
3.1	Objek Penelitian.....	22
3.2	Alur Penelitian.....	22
3.3	Data Penelitian.....	24
3.3.1	Data Primer.....	24
3.3.2	Data Sekunder.....	24
3.4	Pengumpulan Data.....	25
3.5	Metode Analisis.....	25
BAB IV HASIL DAN PEMBAHASAN.....		31
4.1	Pengumpulan Data.....	31
4.2	<i>Pre-processing Data</i>	32
4.2.1	Case Folding.....	33

4.2.2	Cleaning	35
4.2.3	Normalisasi Kata	36
4.2.4	Stemming	37
4.2.5	Tokenizing	38
4.2.6	Stopword	39
4.2.7	Labeling	40
4.2.8	Pembobotan Kata TF-IDF	44
4.3	Model	48
4.3.1	<i>Super Vector Machine</i>	48
4.3.2	<i>Multinomial Naïve Bayes</i>	49
4.4	Evaluasi	50
4.4.1	<i>Evaluasi Metode Support Vector Machine</i>	50
4.4.2	<i>Evaluasi Metode Multinomial Naïve Bayes</i>	56
4.5	Hasil	60
BAB V PENUTUP		64
5.1	Kesimpulan	64
5.2	Saran	65
REFERENSI		66
LAMPIRAN		70

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	9
Tabel 2. 2 <i>Confusion Matrix</i>	20
Tabel 4. 1 Data Awal Penelitian	33
Tabel 4. 2 Hasil Proses <i>case folding</i>	34
Tabel 4. 3 Hasil Proses Cleaning	35
Tabel 4. 4 Contoh Kamus Kata	36
Tabel 4. 5 Hasil Normalisasi Kata	36
Tabel 4. 6 Hasil Proses Stemming	37
Tabel 4. 7 Hasil Proses Tokenizing	38
Tabel 4. 8 Hasil Proses Stopword	39
Tabel 4. 9 Hasil Analisis Orientasi Sentimen	41
Tabel 4. 10 Sampel data komentar yang akan diekstrasi	45
Tabel 4. 11 Hasil perhitungan <i>Term-Frequency</i> pada index ke 5	46
Tabel 4. 12 Hasil perhitungan <i>Inverse Document Frequency</i> pada index ke-5	47
Tabel 4. 13 Top term IF-IDF	48
Tabel 4. 14 Perbandingan Data Train dan Data Test Metode SVM	48
Tabel 4. 15 Grid Search Parameter Metode SVM	49
Tabel 4. 16 Perbandingan Data Train dan Data Test Metode Multinomial NB	50
Tabel 4. 17 Grid Search Parameter Metode Multinomial NB	50
Tabel 4. 18 Akurasi, Presisi, Recall SVM	51
Tabel 4. 19 Hasil <i>Confussion Matrix SVM</i>	52
Tabel 4. 20 Hasil <i>Grid Search CV</i> Tiap Kernel	53
Tabel 4. 21 Akurasi, Presisi, Recall Kernel RBF	54
Tabel 4. 22 Hasil <i>Confussion Matrix hyperparameter SVM</i>	54
Tabel 4. 23 Hasil dari 10-folds cross validation SVM	55
Tabel 4. 24 Akurasi, Presisi, Recall <i>Multinomial Naive Bayes</i>	56

Tabel 4. 25 Hasil <i>Confussion Matrix</i> MNB.....	56
Tabel 4. 26 Hasil <i>Grid Search CV</i> Tiap <i>alpha</i>	57
Tabel 4. 27 Akurasi, Presisi, Recall <i>alpha 2.0</i>	58
Tabel 4. 28 Hasil <i>Confussion Matrix hyperparameter</i> MNB	58
Tabel 4. 29 Hasil dari 10-folds cross validation MNB	59
Tabel 4. 30 Kinerja <i>Multinomial NB</i> dan <i>Support Vector Machine</i>	61



DAFTAR GAMBAR

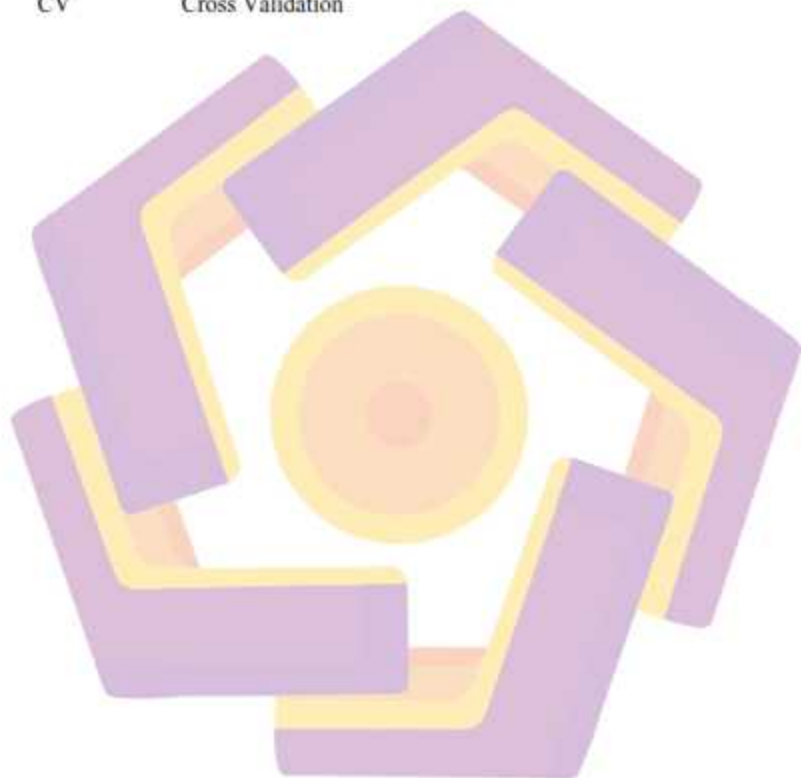
Gambar 2. 1 <i>Hyperplane Supervector Machine</i>	17
Gambar 3. 1 Alur Penelitian	23
Gambar 3. 2 Diagram Alur Labeling	28
Gambar 4. 1 Skrip <i>Instagram Profile Post Extractor</i>	31
Gambar 4. 2 Skrip <i>Instagram Post Commenters Export</i>	32
Gambar 4. 3 Frekuensi Ulasan Berdasarkan Sentimen.....	42
Gambar 4. 4 <i>Word Cloud</i> Ulasan Positif	43
Gambar 4. 5 <i>Word Cloud</i> Ulasan Negatif.....	44
Gambar 4. 6 Proses perhitungan <i>Term-Frequency</i>	46
Gambar 4. 7 Proses perhitungan <i>Inverse Document Frequency</i>	47
Gambar 4. 8 Visualisasi Perbandingan Akurasi SVM dan MNB.....	62

DAFTAR LAMPIRAN

Lampiran 1 Case Folding.....	70
Lampiran 2 Code Hapus Emoji.....	70
Lampiran 3 Output Hapus Emoji.....	71
Lampiran 4 Code Pembersihan Text.....	71
Lampiran 5 Output Pembersihan Text.....	72
Lampiran 6 Normalisasi Kata.....	72
Lampiran 7 Output Normalisasi Kata.....	73
Lampiran 8 Code Stemming.....	73
Lampiran 9 Output Stemming.....	73
Lampiran 10 Tokenizing.....	74
Lampiran 11 Stopword.....	74
Lampiran 12 Code labeling.....	75
Lampiran 13 Output Labeling.....	75
Lampiran 14 Visualisasi Polarity.....	75
Lampiran 15 Code Wordcloud Positive.....	76
Lampiran 16 Code Wordcloud Negative.....	76
Lampiran 17 TF-IDF.....	77
Lampiran 18 Code SVM tanpa Hyperparameter.....	77
Lampiran 19 Code SVM dengan Hyperparameter.....	77
Lampiran 20 Code K-fold Cross Validation SVM.....	78
Lampiran 21 Code MNB tanpa Hyperparameter.....	79
Lampiran 22 Code Hyperparameter MNB.....	79
Lampiran 23 Code MNB dengan Hyperparameter.....	80
Lampiran 24 Code K-fold Cross Validation MNB.....	81

DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machines
MNB	Multinomial Naïve Bayes
RBF	Radial Basis Function
CV	Cross Validation



DAFTAR ISTILAH



Vektor	Besaran yang mempunyai arah
Pre-Processing	Persiapan Data
Tokenizing	Pembagian Teks
Stemming	Pengubahan Kata
Stopword	Kata Dihapus
Lexicon Labeling	Pelabelan Kamus
Supervised Learning	Pembelajaran Terpantau
Grid Search	Pencarian Grid
Akurasi	Tingkat Ketepatan
Presisi	Ketepatan
Dataset	Kumpulan Data
TF-IDF	Frekuensi kemunculan dokumen kata

INTISARI

Penelitian ini membahas perbandingan kinerja dua metode klasifikasi, yaitu Support Vector Machine (SVM) dan Multinomial Naive Bayes (MNB), dalam menganalisis sentimen terhadap komentar masyarakat pada akun media sosial DKI Jakarta. Proses pemodelan melibatkan pembagian data latih dan uji dengan variasi rasio, sementara metode SVM dievaluasi menggunakan kernel linear dan radial basis function (RBF) dengan grid search cross-validation. Proses labeling menggunakan pendekatan lexicon based dengan Indonesia Sentimen Lexicon (InSet Lexicon) untuk menghitung bobot setiap kata pada ulasan, yang kemudian diterapkan pada dataset. Hasil penelitian menunjukkan bahwa SVM memberikan akurasi yang sedikit lebih tinggi daripada MNB, dengan mencapai 82%. Parameter optimal untuk SVM adalah $C=100$ dan $\gamma=0.1$ pada kernel RBF, sementara untuk MNB, parameter $\alpha=2.0$ dan $\text{fit_prior}=\text{True}$ memberikan kinerja optimal dengan akurasi 80%. Evaluasi dilakukan menggunakan confusion matrix dan k-fold cross validation dengan 10-folds cross validation. Hasil dari 10-folds cross validation SVM menunjukkan akurasi, presisi, dan recall tertinggi pada n ke-5 sebesar 86%, 86%, dan 86%. Sedangkan hasil dari 10-folds cross validation MNB menunjukkan akurasi, presisi, dan recall tertinggi pada n ke-3 sebesar 82%, 83%, dan 82%.

Kata kunci: *sentimen analisis, multinomial naïve bayes, super vector machine, grid search*

ABSTRACT

This research discusses the performance comparison of two classification methods, namely Support Vector Machine (SVM) and Multinomial Naive Bayes (MNB), in analyzing sentiment towards public comments on DKI Jakarta social media accounts. The modeling process involves dividing the training and test data by varying the ratio, while the SVM method is evaluated using a linear kernel and radial basis function (RBF) with grid search cross-validation. The labeling process used a lexicon-based approach with the Indonesia Sentiment Lexicon (InSet Lexicon) to calculate the weight of each word in the reviews, which was then applied to the dataset. The results show that SVM provides slightly higher accuracy than MNB, reaching 82%. The optimal parameters for SVM are $C=100$ and $\gamma=0.1$ on the RBF kernel, while for MNB, the parameters $\alpha=2.0$ and $\text{fit_prior}=\text{True}$ provide optimal performance with 80% accuracy. Evaluation is done using confusion matrix and k-fold cross validation with 10-folds cross validation. The results of 10-folds cross validation SVM showed the highest accuracy, precision, and recall at the 5th n of 86%, 86%, and 86%. While the results of 10-folds cross validation MNB show the highest accuracy, precision, and recall at the 3rd n of 82%, 83%, and 82%.

Keyword: *sentiment analysis, multinomial naïve bayes, super vector machine, grid search*