



PERBANDINGAN METODE NAÏVE BAYES CLASSIFIER DAN K-NEAREST NEIGHBOR (K-NN) TERHADAP ANALISIS SENTIMEN PADA KEBIJAKAN PEMEBERLAKUAN PEMBATAAN KEGIATAN MASYARAKAT (PPKM)

Skripsi

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Statistika

Oleh

**M. ABDURRAFI HAMDI
B2A219045**

**PROGRAM STUDI S1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS MUHAMMADIYAH SEMARANG
2022**

PERSETUJUAN PEMBIMBING

Skripsi dengan Judul “Perbandingan Metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor (K-NN)* Terhadap Analisis Sentimen pada Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM)” yang disusun oleh :

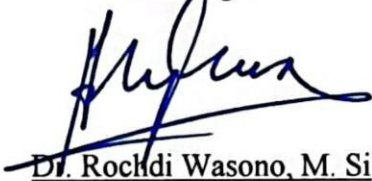
Nama : M. Abdurrafi Hamdi

NIM : B2A219045

Program Studi : S1 Statistika

telah disetujui oleh dosen pembimbing pada tanggal : 2 Maret 2022

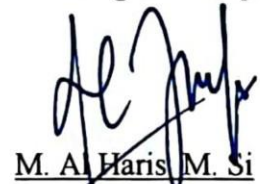
Pembimbing Utama



Dr. Rochdi Wasono, M. Si

NIK. 28.6.1026.119

Pembimbing Pendamping



M. Al Haris, M. Si

NIK. 28.6.1026.442

Mengetahui,

Ketua Program Studi



Indah Manfaati Nur, M. Si
NIK. 28.6.1026.221

PENGESAHAN KELULUSAN

Skripsi dengan Judul “Perbandingan Metode *Naïve Bayes Classifier* dan *K Nearest Neighbor (K-NN)* Terhadap Analisis Sentimen pada Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM)” yang disusun oleh:

Nama : M. Abdurrafi Hamdi

NIM : B2A219045

Program Studi : S1 Statistika

telah diujikan dalam Sidang Panitia Ujian Skripsi Program Sarjana, Universitas Muhammadiyah Semarang pada tanggal : 19 Juli 2022

Panitia Ujian

Ketua Tim Penguji



Indah Manfaati Nur, M. Si
NIK. 28.6.1026.221

Anggota Tim Penguji I



Tiara Wahyuni Utami, M. Si
NIK. 28.6.1026.341

Anggota Tim Penguji II



Dr. Rochol Wasono, M. Si NIK.
28.6.1026.119

Anggota Tim Penguji III



M. Al Haris, M. Si
NIK. 28.6.1026.442

Mengetahui

Ketua Program Studi




Indah Manfaati Nur, M. Si
NIK. 28.6.1026.221

PERNYATAAN KEASLIAN

Dengan ini saya menyatakan bahwa :

1. Karya tulis saya, skripsi ini, adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (sarjana, magister, dan/atau doktor), baik di Universitas Muhammadiyah Semarang maupun di perguruan tinggi lain.
2. Karya tulis ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan Tim Pembimbing dan masukan Tim Penguji.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya ini, serta sanksi lainnya sesuai dengan norma yang berlaku di perguruan tinggi ini.

Semarang, 19 Juli 2022

Yang membuat pernyataan,



(M. Abdurrafi Hamdi)

NIM. B2A219045

KATA PENGANTAR

Assalamu'alaikum Wr.Wb

Segala puji dan syukur penulis ucapkan kehadiran Allah SWT dan mengharapkan ridho yang telah melimpahkan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Perbandingan Metode *Naïve Bayes Classifier* dan *K Nearest Neighbor (K-NN)* Terhadap Analisis Sentimen Pada Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM)” Skripsi ini disusun sebagai salah satu persyaratan meraih gelar Sarjana Pendidikan pada Program Studi Pendidikan Matematika Universitas Muhammadiyah Semarang. Shalawat dan salam disampaikan kepada junjungan alam Nabi Muhammad SAW, mudah-mudahan kita semua mendapatkan safaatNya di yaumul akhir nanti, Amin.

Penulis menyadari sepenuhnya, bahwa dalam penyelesaian skripsi ini tidak terlepas dari bantuan dan bimbingan berbagai pihak. Untuk itu pada kesempatan ini penulis ingin menyampaikan terima kasih dan penghargaan yang setinggi-tingginya kepada:

1. Ibu Dr. Eny Winaryati, M.Pd, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang.
2. Ibu Indah Manfaati Nur, S.Si, M.Si, selaku Ketua Program Studi Statistika Universitas Muhammadiyah Semarang.
3. Bapak Dr. Rochdi Wasono, M.Si, selaku Dosen Pembimbing Pertama yang telah meluangkan waktu, tenaga dan pikirannya untuk memberikan bimbingan dan pengarahan dalam penyusunan skripsi ini.
4. Bapak M. Al Haris, S.Si, M.Si, selaku Dosen Pembimbing Pendamping yang telah meluangkan waktu, tenaga dan pikirannya untuk memberikan bimbingan dan pengarahan dalam penyusunan skripsi ini.
5. Segenap Dosen dan Civitas Akademik Universitas Muhammadiyah Semarang yang telah membantu penulis selama menempuh pendidikan

di kampus ini.

6. Kedua orang tua tercinta Ayah Isrul Hamdi, dan Umi Saidah Sembiring yang selalu memberikan kasih sayang, do'a, dan support baik secara moril maupun materil.
7. Kepada semua pihak yang sudah membantu dalam penyelesaian Skripsi ini. Terkhusus untuk Fadhli Syamily Mahmud Melba, Tanti Yulianita dan Rizqa El Husna yang turut banyak membantu dalam proses penyelesaian skripsi ini.

Penulis menyadari akan segala keterbatasan dan kekurangan dari isi maupun tulisan skripsi ini. Oleh karena itu, kritik dan saran yang bersifat membangun dari semua pihak akan diterima dengan senang hati. Semoga hasil penelitian ini dapat memberikan manfaat dan kontribusi bagi pengembangan pembelajaran statistika.

Wassalamu'alaikum Wr.Wb

Semarang, Juli 2022

M. Abdurrafi Hamdi

ABSTRAK

Hamdi, M. Abdurrafi, 2022, Perbandingan Metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor (K-NN)* Terhadap Analisis Sentimen Pada Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM), Skripsi, Program Studi Statistika, Universitas Muhammadiyah Semarang. Pembimbing: I. Dr. Rochdi Wasono, M.Si, II. M. Al Haris, M.Si.

Untuk menekan angka Covid 19 di Indonesia pemerintah membuat kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) yang diterapkan oleh Pemerintah Indonesia sejak awal tahun 2021, pemberlakuan kebijakan ini menimbulkan banyak sentimen publik yang berkembang terutama pada media sosial *Twitter*. Penelitian ini bertujuan untuk mengetahui bagaimana respon masyarakat terhadap kebijakan PPKM di media sosial *Twitter*, sebagai acuan pemerintah dalam menerapkan kebijakan yang serupa kedepan. Solusi yang bisa digunakan untuk penelitian ini adalah dengan menggunakan analisis sentimen, data penelitian yang diperoleh diklasifikasikan dengan menggunakan metode *Naïve Bayes Classifier (NBC)* yang merupakan metode algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat, dan *K-Nearest Neighbor (K-NN)* yaitu metode dengan pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama. Hasil dari penelitian menunjukkan, pada pengujian dengan menggunakan metode adalah *Naïve Bayes Classifier (NBC)*, diperoleh nilai akurasi sebesar 50%, sedangkan pada *K-Nearest Neighbour (K-NN)*, nilai akurasi yang diperoleh adalah sebesar 60%.

Kata Kunci: Analisa Sentimen, *Naïve Bayes Classifier (NBC)*, *K-Nearest Neighbour (K-NN)*, PPKM, *Twitter*.

ABSTRACT

Hamdi, M. Abdurrafi, 2022, *Comparison of Naïve Bayes Classifier and K-Nearest Neighbor (K-NN) Methods on Sentiment Analysis on Policy for Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM)*, Thesis, Statistics Study Program, University of Muhammadiyah Semarang. Supervisor: I. Dr. Rochdi Wasono, M.Si, II. M. Al Haris, M.Sc.

To suppress Covid 19 somewhat in Indonesia, the government made a policy of Enforcement of Community Activity Restrictions (PPKM) which has been implemented by the Government of Indonesia since early 2021, the implementation of this policy has caused a lot of public sentiment to develop, especially on Twitter social media. This study aims to find out how the public responds to the PPKM policy on Twitter social media, as the government's desire to implement a similar policy in the future. The solution that can be used for this research is to use sentiment analysis, the research data obtained are classified using the Naïve Bayes Classifier (NBC) method which is an algorithm method used to find the highest probability value to classify the test data in the most appropriate category, and K - Nearest Neighbor (K-NN) is an approach method to find cases by approaching the distance between new cases and old cases. The results of the study show that in testing using the Naïve Bayes Classifier (NBC) method, it obtained an accuracy value of 50%, while in K-Nearest Neighbor (K-NN), the accuracy value obtained was 60%.

Keywords: Sentiment Analysis, Naïve Bayes Classifier (NBC), K-Nearest Neighbor (K-NN), PPKM, Twitter.

DAFTAR ISI

	Halaman
LEMBAR JUDUL	i
PERSETUJUAN PEMBIMBING.....	ii
ABSTRAK	iii
ABSTRACT.....	iv
DAFTAR ISI.....	v
DAFTAR TABEL.....	vii
DAFTAR GAMBAR	viii
BAB I PENDAHULUAN	
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	4
1.5 Batasan Penelitian	5
BAB II TINJAUAN PUSTAKA	
2.1 PPKM	6
2.2 <i>Twitter</i>	7
2.3 <i>Data Mining</i>	8
2.4 Analisis Sentimen.....	9
2.5 Preprocessing.....	10
2.6 <i>K-Nearest Neighbor</i> (K-NN).....	10
2.7 <i>Naïve Bayes Classifier</i>	12
2.8 Pembobotan TF-IDF.....	13
2.9 Evaluasi Performa Klasifikasi	14
BAB III METODOLOGI PENELITIAN	

3.1	Sumber Data	17
3.2	Variabel Penelitian dan Struktur Data.....	17
3.3	Langkah Penelitian.....	17
3.4	Diagram Alir Penelitian.....	19
3.5	Jadwal Penelitian.....	19
BAB IV ANALISIS DAN PEMBAHASAN		
4.1	Analisis Deskriptif.....	20
4.2	Tahap <i>Preprocessing</i> Data	22
4.3	Tahap Pelabelan Data.....	26
4.4	TF-IDF.....	27
4.5	Metode Klasifikasi	29
4.6	Evaluasi Performa	37
BAB V KESIMPULAN DAN SARAN		
5.1	Kesimpulan.....	39
5.2	Saran.....	39
DAFTAR PUSTAKA		41



DAFTAR GAMBAR

Gambar	Halaman
2.1 Lingkup Kajian Data Mining	8
3.1 Diagram Alir Penelitian	19
4.1 Wordcloud Sentimen Positif	20
4.2 Tabel Frekuensi Sentimen Positif	20
4.3 Wordcloud Sentimen Negatif	21
4.4 Tabel Frekuensi Sentimen Negatif	21



DAFTAR TABEL

Tabel	Halaman
3.1 Struktur Data	18
4.1 Struktur Data Sebelum Praproses Text	22
4.2 Struktur Data Pada Tahap <i>Case Folding</i>	23
4.3 Struktur Data Pada Tahap <i>Cleaning</i>	24
4.4 Struktur Data Pada Tahap <i>Stopword Removal</i>	25
4.5 Struktur Data Pada Tahap <i>Tokenizing</i>	25
4.6 Struktur Data Pada Tahap <i>Stemming</i>	26
4.7 Pelabelan Kelas Sentimen.....	26
4.8 Hasil Pelabelan Kelas Sentimen.....	27
4.9 Simulasi Perhitungan TF-IDF.....	28
4.10 Struktur Data Simulasi <i>Naïve Bayes Classifier</i>	29
4.11 Simulasi Perhitungan Probabilitas <i>Naïve Bayes Classifier</i>	31
4.12 Struktur Data Simulasi <i>K-Nearst Neighbour</i>	34
4.13 Simulasi Pembobotan <i>K-Nearst Neighbour</i>	35
4.14 Simulasi Perhitungan Panjang Vektor	36
4.15 Simulasi Perhitungan Kemiripan Dokumen.....	37
4.16 <i>Confusion Matrix</i> Algoritma <i>Naïve Bayes Classifier</i>	37
4.17 <i>Confusion Matrix</i> Algoritma <i>K-Nearst Neighbour</i>	38
4.18 Evaluasi Performa	38