

**PERBANDINGAN KINERJA ALGORITMA MULTINOMIAL
NAÏVE BAYES (MNB) DAN BERNOULLI NAÏVE BAYES (BNB)
PADA PENGKLASIFIKASIAN KOMENTAR CYBERBULLYING**

SKRIPSI

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



disusun oleh
FRITZ ZONE
19.11.2928

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023**

**PERBANDINGAN KINERJA ALGORITMA MULTINOMIAL
NAÏVE BAYES (MNB) DAN BERNOULLI NAÏVE BAYES (BNB)
PADA PENGKLASIFIKASIAN KOMENTAR CYBERBULLYING**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
FRITZ ZONE
19.11.2928

Kepada
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN KINERJA ALGORITMA MULTINOMIAL NAÏVE
BAYES (MNB) DAN BERNOULLI NAÏVE BAYES (BNB) PADA
PENGKLASIFIKASIAN KOMENTAR CYBERBULLYING**

yang disusun dan diajukan oleh

Fritz Zone

19.11.2928

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 9 September 2022

Dosen Pembimbing,



Windha Mega Pradnya Duhita, M.Kom

NIK. 190302185

HALAMAN PENGESAHAN

SKRIPSI

PERBANDINGAN KINERJA ALGORITMA MULTINOMIAL NAÏVE
BAYES (MNB) DAN BERNOULLI NAÏVE BAYES (BNB) PADA
PENGKLASIFIKASIAN KOMENTAR CYBERBULLYING

yang disusun dan diajukan oleh

Fritz Zone

19.11.2928

Telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Mei 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Moch Farid Fauzi, M.Kom
NIK. 190302284



Ika Nur Fitri, M.Kom
NIK. 190302268



Windha Mega Pradnya Dhuhita, M.Kom
NIK. 190302185



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 Mei 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Fritz Zone
NIM : 19.11.2928

Menyatakan bahwa Skripsi dengan judul berikut:

PERBANDINGAN KINERJA ALGORITMA MULTINOMIAL NAÏVE BAYES (MNB) DAN BERNOULLI NAÏVE BAYES (BNB) PADA PENGKLASIFIKASIAN KOMENTAR CYBERBULLYING

Dosen Pembimbing : Windha Mega Pradnya Dhuhita, 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, 23 Mei 2023

Yang Menyatakan,



Fritz Zone

HALAMAN PERSEMBAHAN

Dengan tulus dan penuh rasa syukur, penulis ingin menyampaikan persembahan ini kepada mereka yang penulis cintai, hormati, dan berjasa dalam perjalanan penulisan skripsi ini. Setiap langkah yang penulis tempuh tidak akan lengkap tanpa bantuan, dukungan, dan cinta dari mereka.

Pertama-tama, penulis ingin mengucapkan terima kasih kepada Tuhan Yang Maha Esa, atas rahmat-Nya yang tiada henti memberikan petunjuk, kekuatan, dan keberkahan dalam perjalanan kami menyelesaikan skripsi ini.

Penulis juga ingin mengucapkan terima kasih sebesar-besarnya kepada kedua orang tua yang telah menjadi tiang yang kokoh dan sumber inspirasi sepanjang hidup penulis. Terima kasih kepada kedua orang tua yang selalu memberikan dukungan penuh, doa, dan cinta tanpa syarat.

Tak lupa, penulis ingin mengungkapkan rasa terima kasih yang dalam kepada dosen pembimbing atas kesabaran, bimbingan, dan arahan yang berharga dalam mengarahkan langkah-langkah penulis. Terima kasih juga kepada seluruh dosen dan staf di Universitas Amikom Yogyakarta yang telah memberikan ilmu pengetahuan dan wawasan yang tak ternilai selama masa perkuliahan.

Penulis juga ingin mengucapkan terima kasih kepada keluarga, kerabat, dan teman-teman penulis, terlebih kepada orang terkasih yang telah selalu memberikan semangat, dukungan dan doa sehingga penulis dapat menyelesaikan skripsi ini.

Penulis menyadari bahwa segala yang telah kami capai tidak mungkin terlaksana tanpa campur tangan dari-Nya, kehadiran semua pihak yang telah penulis sebutkan dan yang tidak dapat disebutkan satu-satu. Semoga persembahan ini dapat mencerminkan rasa terima kasih penulis yang mendalam dan menjadi ungkapan kebahagiaan serta penghargaan penulis kepada semua yang telah memberikan inspirasi dan dorongan bagi kesuksesan skripsi ini.

Yogyakarta, 23 Mei 2023

Penulis

KATA PENGANTAR

Puji dan syukur kepada Tuhan Yang Maha Esa atas berkat dan rahmat-Nya, sehingga penulis menyelesaikan skripsi yang berjudul “Perbandingan Kinerja Algoritma Multinomial Naïve Bayes (MNB) dan Bernoulli Naïve Bayes (BNB) pada Pengklasifikasian Komentar Cyberbullying” sebagai syarat untuk memperoleh gelar sarjana komputer dari program studi Informatika Universitas Amikom Yogyakarta. Melalui penelitian ini, penulis berharap dapat memberikan kontribusi yang signifikan terhadap pengembangan bidang penelitian ini.

Penulis juga mengucapkan terima kasih yang tak terhingga kepada kedua orang tua yang telah memberikan dukungan moral, doa, dan kasih sayang yang tak tergantikan. Terima kasih atas pengertian, kesabaran, dan semangat yang diberikan dalam setiap langkah penulis menuju penyelesaian skripsi ini.

Penulis juga ingin mengungkapkan rasa terima kasih yang mendalam kepada dosen pembimbing atas bimbingan, arahan, dan pengawasannya yang cermat sepanjang proses penulisan skripsi ini. Beliau telah meluangkan waktu, pengetahuan, dan pengalaman untuk membimbing penulis dengan sungguh-sungguh, memperbaiki karya ini, dan memberikan masukan berharga yang telah memperkaya isi skripsi ini.

Penulis juga ingin menyampaikan rasa terima kasih kepada dosen penguji yang telah memberikan saran, kritik, dan masukan yang konstruktif. Kontribusi tersebut telah membantu penulis untuk meningkatkan kualitas penelitian ini.

Akhir kata, penulis berharap bahwa skripsi ini dapat memberikan manfaat dan memberi sumbangsih dalam pengembangan ilmu pengetahuan serta pemahaman di bidang penelitian serupa. Semoga hasil penelitian ini dapat menjadi pijakan bagi penelitian selanjutnya dan memberikan kontribusi positif bagi masyarakat.

Yogyakarta, 23 Mei 2023

Penulis

DAFTAR ISI

| | |
|--|-----------|
| HALAMAN JUDUL | i |
| DAFTAR ISI..... | vii |
| DAFTAR TABEL..... | ix |
| DAFTAR GAMBAR..... | x |
| DAFTAR LAMBANG DAN SINGKATAN | xi |
| DAFTAR ISTILAH..... | xii |
| INTISARI | xiii |
| ABSTRACT..... | xiv |
| BAB I PENDAHULUAN | 1 |
| 1.1 Latar Belakang..... | 1 |
| 1.2 Rumusan Masalah..... | 2 |
| 1.3 Batasan Masalah | 2 |
| 1.4 Tujuan Penelitian | 2 |
| 1.5 Manfaat Penelitian | 3 |
| 1.6 Sistematika Penulisan | 3 |
| BAB II TINJAUAN PUSTAKA | 5 |
| 2.1 Studi Literatur..... | 5 |
| 2.2 Dasar Teori..... | 12 |
| BAB III METODE PENELITIAN | 21 |
| 3.1 Gambaran Umum Penelitian..... | 21 |
| 3.2 Alur Penelitian..... | 22 |
| 3.3 Alat dan Bahan..... | 30 |
| BAB IV HASIL DAN PEMBAHASAN | 32 |
| 4.1. Implementasi <i>Preprocessing</i> | 32 |
| 4.2. Implementasi <i>Feature Extraction</i> | 35 |
| 4.3. Implementasi <i>K-Fold Cross Validation</i> | 36 |
| 4.4. Implementasi <i>Algoritma Naïve Bayes</i> | 36 |
| 4.5. Implementasi <i>Confusion Matrix</i> | 36 |
| 4.6. Implementasi <i>Pengujian</i> | 36 |
| 4.7. Implementasi <i>Sistem</i> | 52 |
| BAB V PENUTUP | 55 |
| 5.1 Kesimpulan..... | 55 |
| 5.2 Saran | 55 |
| REFERENSI | 56 |

DAFTAR TABEL

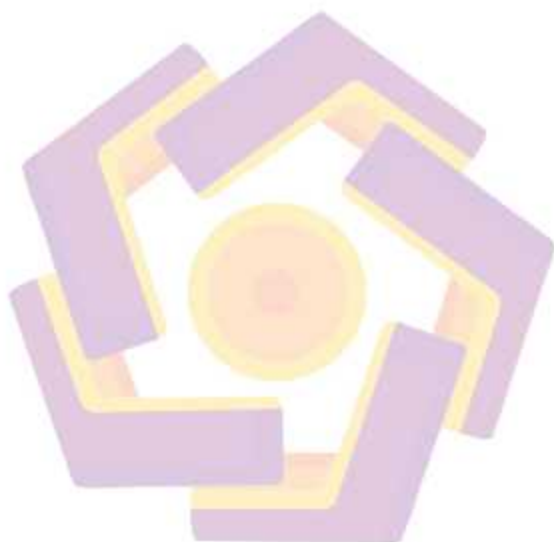
| | |
|---|----|
| Tabel 2. 1. Keaslian Penelitian | 8 |
| Tabel 2. 2. Contoh teks | 18 |
| Tabel 2. 3. Bentuk list dari teks | 18 |
| Tabel 2. 4. Frekuensi kata berdasarkan teks pada dokumen | 18 |
| Tabel 2. 5. <i>Cross Validation</i> | 19 |
| Tabel 2. 6. <i>Confusion Matrix</i> | 19 |
| Tabel 3. 1. <i>Dataset Cyberbullying</i> | 24 |
| Tabel 3. 2. Rincian <i>Dataset</i> | 30 |
| Tabel 3. 3. Spesifikasi <i>Hardware</i> | 30 |
| Tabel 3. 4. <i>Software</i> yang digunakan | 31 |
| Tabel 4. 1. Proses <i>Cleaning</i> | 32 |
| Tabel 4. 2. Proses <i>Case Folding</i> | 33 |
| Tabel 4. 3. Proses <i>Tokenizing</i> | 33 |
| Tabel 4. 4. Proses <i>Normalization</i> | 34 |
| Tabel 4. 5. Proses <i>Stopword Removal</i> | 34 |
| Tabel 4. 6. Proses <i>Stemming</i> | 35 |
| Tabel 4. 7. Pengujian Skenario 1 <i>Multinomial Naïve Bayes fold 10</i> | 39 |
| Tabel 4. 8. Pengujian Skenario 1 <i>Bernoulli Naïve Bayes fold 10</i> | 42 |
| Tabel 4. 9. Pengujian Skenario 2 <i>Multinomial Naïve Bayes fold 13</i> | 44 |
| Tabel 4. 10. Pengujian Skenario 2 <i>Bernoulli Naïve Bayes fold 10</i> | 47 |
| Tabel 4. 11. Rangkuman Nilai Akurasi Berdasarkan Hasil Pengujian | 48 |
| Tabel 4. 12. Rangkuman Nilai Precision Berdasarkan Hasil Pengujian | 48 |
| Tabel 4. 13. Rangkuman Nilai Recall Berdasarkan Hasil Pengujian | 49 |
| Tabel 4. 14. Rangkuman Nilai F1-Score Berdasarkan Hasil Pengujian | 49 |
| Tabel 4. 15. Rangkuman Waktu Pemrosesan Berdasarkan Hasil Pengujian | 49 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 2. 1. Tahapan <i>Text Mining</i> | 13 |
| Gambar 3. 1. Alur Penelitian | 23 |
| Gambar 3. 2. Tahapan Preprocessing..... | 25 |
| Gambar 3. 3. Alur Proses <i>Feature Extraction</i> | 28 |
| Gambar 3. 4. <i>Confusion Matrix Model</i> | 29 |
| Gambar 4. 1. Skenario 1 Multinomial Naïve Bayes <i>fold 5</i> | 37 |
| Gambar 4. 2. Skenario 1 Multinomial Naïve Bayes <i>fold 10</i> | 38 |
| Gambar 4. 3. Skenario 1 Multinomial Naïve Bayes <i>fold 13</i> | 38 |
| Gambar 4. 4. Skenario 1 Bernoulli Naïve Bayes <i>fold 5</i> | 40 |
| Gambar 4. 5. Skenario 1 Bernoulli Naïve Bayes <i>fold 10</i> | 41 |
| Gambar 4. 6. Skenario 1 Bernoulli Naïve Bayes <i>fold 13</i> | 41 |
| Gambar 4. 7. Skenario 2 Multinomial Naïve Bayes <i>fold 5</i> | 43 |
| Gambar 4. 8. Skenario 2 Multinomial Naïve Bayes <i>fold 10</i> | 43 |
| Gambar 4. 9. Skenario 2 Multinomial Naïve Bayes <i>fold 13</i> | 44 |
| Gambar 4. 10. Skenario 2 Bernoulli Naïve Bayes <i>fold 5</i> | 46 |
| Gambar 4. 11. Skenario 2 Bernoulli Naïve Bayes <i>fold 10</i> | 46 |
| Gambar 4. 12. Skenario 2 Bernoulli Naïve Bayes <i>fold 13</i> | 47 |
| Gambar 4. 13. Hasil Multinomial Naïve Bayes Skenario 1..... | 50 |
| Gambar 4. 14. Hasil Multinomial Naïve Bayes Skenario 2..... | 50 |
| Gambar 4. 15. Hasil Bernoulli Naïve Bayes Skenario 1..... | 51 |
| Gambar 4. 16. Hasil Bernoulli Naïve Bayes Skenario 2..... | 51 |
| Gambar 4. 17. Halaman Pengujian (1)..... | 52 |
| Gambar 4. 18. Halaman Pengujian (2)..... | 53 |
| Gambar 4. 19. Halaman Pelatihan (1)..... | 53 |
| Gambar 4. 20. Halaman Pelatihan (2)..... | 54 |

DAFTAR LAMBANG DAN SINGKATAN

| | |
|--------|---|
| KBBI | Kamus Besar Bahasa Indonesia |
| MNB | Multinomial Naïve Bayes |
| BNB | Bernoulli Naïve Bayes |
| BoW | Bag of Words |
| TF-IDF | Term Frequency – Inverse Document Frequency |



DAFTAR ISTILAH

| | |
|---------------|--|
| Netizen | Sebutan lain dari warga internet untuk orang yang aktif |
| Cyberbullying | Tindakan intimidasi yang terjadi di media sosial |
| Dataset | Sekumpulan data mentah berupa tabel yang dapat diolah lebih lanjut |
| Bullying | Bentuk intimidasi kepada orang lain |
| Algoritma | Rangkaian instruksi – instruksi yang rumit untuk memecahkan masalah |
| Data Latih | Data yang digunakan untuk melatih mesin agar dapat memahami pola |
| Data Uji | Data yang digunakan untuk menguji hasil dari pelatihan |
| Prefix | Huruf atau beberapa huruf yang ada di awal kata dan merubah arti kata asli |
| Suffix | Huruf atau beberapa huruf yang ditambahkan dibelakang kata |

INTISARI

Pada era sekarang, ditengah pesatnya perkembangan teknologi, ada banyak orang yang menyalahgunakan teknologi untuk melakukan tindakan kurang menyenangkan terhadap orang lain, salah satunya *bullying* yang dilakukan menggunakan media sosial yang disebut sebagai *cyberbullying*. Maka peneliti melakukan pengklasifikasian data komentar pada media sosial apakah termasuk *bullying* atau tidak. Tujuan dari penelitian ini untuk melakukan klasifikasi data komentar pada sosial media termasuk *cyberbullying* atau tidak dengan membandingkan terlebih dahulu kinerja antara algoritma Naïve Bayes model Multinomial dan Bernoulli dalam mengklasifikasikan data komentar tersebut. Peneliti membandingkan algoritma Naïve Bayes Classifier model Multinomial dan Bernoulli untuk mendapatkan model terbaik. Peneliti juga membandingkan penggunaan metode *feature extraction Bag of Words* dan TF-IDF untuk meningkatkan akurasi dari algoritma yang digunakan. Hasil dari penelitian yang dilakukan adalah algoritma Naive Bayes model Multinomial memperoleh akurasi lebih tinggi dan memperoleh rata - rata waktu pemrosesan yang lebih cepat dibandingkan model Bernoulli pada skenario 1 dan 2. Penggunaan metode *feature extraction Bag of Words* juga dapat meningkatkan akurasi lebih signifikan dibanding TF-IDF.

Kata kunci: Cyberbullying, Klasifikasi, Perbandingan, Multinomial Naïve Bayes, Bernoulli Naive Bayes

ABSTRACT

In the current era, amidst the rapid development of technology, there are many people who misuse technology to engage in unpleasant actions towards others, including bullying conducted through social media, known as cyberbullying. Therefore, the researcher conducted the classification of comment data on social media to determine whether they constitute bullying or not. The aim of this research is to classify whether comment data on social media, including cyberbullying or not, by comparing the performance of the Naïve Bayes Multinomial and Bernoulli models in classifying the comment data. The researcher compared the Naïve Bayes Classifier models of Multinomial and Bernoulli to obtain the best model. Additionally, the researcher compared the use of feature extraction methods such as Bag of Words and TF-IDF to enhance the accuracy of the algorithms used. The results of the research indicate that the Naïve Bayes Multinomial model achieves higher accuracy and faster processing time compared to the Bernoulli model in scenarios 1 and 2. Furthermore, the utilization of the Bag of Words feature extraction method significantly improves accuracy compared to TF-IDF.

Keyword: *Cyberbullying, Classification, Comparison, Multinomial Naïve Bayes, Bernoulli Naïve Bayes*