SERTIFIKAT

Kementerian Riset dan Teknologi/
Badan Riset dan Inovasi Nasional

RISTEK-BRIN

Petikan dari Keputusan Menteri Riset dan Teknologi/
Kepala Badan Riset dan Inovasi Nasional
Nomor 148/M/KPT/2020
Peringkat Akreditasi Jurnal Ilmiah Periode II Tahun 2020
Nama Jurnal Ilmiah

InComTech: Jurnal Telekomunikasi dan Komputer

E-ISSN: 25796089

Penerbit: Universitas Mercu Buana

Ditetapkan sebagai Jurnal Ilmiah

TERAKREDITASI PERINGKAT 3

Akreditasi Berlaku selama 5 (lima) Tahun, yaitu
Volume 4 Nomor 2 Tahun 2020 sampai Volume 9 Nomor 1 Tahun 2025

Jakarta, 03 Agustus 2020
Menteri Riset dan Teknologi/
Kepala Badan Riset dan Inovasi Nasional
Republik Indonesia,

Bambang P. S. Brodjonegoro

InComTech
Jurnal Telekomunikasi dan Komputer

Vol. 12 No. 1, April 2022
p-ISSN: 2085-4811, e-ISSN: 2579-6089

InComTech : Jurnal Telekomunikasi dan Komputer	Volume 12	Nomor 1	Halaman 01-83	Jakarta April 2022	p-ISSN: 2085-4811 e-ISSN: 2579-6089

pISSN: 2085-4811 - eISSN: 2579-6089

InComTech
Jurnal Telekomunikasi dan Komputer

Vol. 12, No. 1, April 2022

Editorial Address :
Magister Teknik Elektro, Universitas Mercu Buana
Jl. Raya Meruya Selatan, Kembangan, Jakarta 11650
Tlp/Fax: 021-31935454 / 021-31934474
http://www.publikasi.mercubuana.ac.id/index.php/Incomtech

InComTech: Jurnal Telekomunikasi dan Komputer is a peer-reviewed journal academics, practitioners and other activist in the field on information, telecommunication and computers (ICT) to publish their works. Areas covered by this journal include technology, business, and regulation in the field of ICT, such as IP Technology, Wireless Technology, Internet of Things, Microwaves, Digital Broadcasting, Optical Fiber, ICT Business Strategy, Human Resources ICT, Business Planning, NGN Regulation, Security in ICT, cyberlaw.

pISSN: 2085-4811 - eISSN: 2579-6089

InComTech
Jurnal Telekomunikasi dan Komputer

Vol. 12, No. 1, April 2022

InComTech: Jurnal Telekomunikasi dan Komputer is a peer-reviewed journal academics, practitioners and other activist in the field on information, telecommunication and computers (ICT) to publish their works. Areas covered by this journal include technology, business, and regulation in the field of ICT, such as IP Technology, Wireless Technology, Internet of Things, Microwaves, Digital Broadcasting, Optical Fiber, ICT Business Strategy, Human Resources ICT, Business Planning, NGN Regulation, Security in ICT, cyberlaw.

pISSN: 2085-4811 - eISSN: 2579-6089

InComTech
Jurnal Telekomunikasi dan Komputer

Vol. 12, No. 1, April 2022

TABLE OF CONTENTS

InComTech: Jurnal Telekomunikasi dan Komputer
vol.12, no.1, April 2022, 60-70
http://publikasi.mercubuana.ac.id/index.php/Incomtech
P-ISSN: 2085-4811 E-ISSN: 2579-6089

Analisis Performa Algoritma Random Forest dan Naive Bayes Multinomial pada *Dataset* Ulasan Obat dan Ulasan Film

Rolly Maulana Awangga*, Nuha Hanifatul Khonsa'

Program Studi D-IV Teknik Informatika, Politeknik Pos Indonesia, Bandung, Indonesia

* Email Penulis Koresponden: awangga@poltekpos.ac.id

Abstrak:
Kemudahan akses informasi memberikan peluang pertukaran informasi antar individu maupun kelompok. Kemudahan akses tersebut memberikan dampak dengan munculnya banyak opini terhadap suatu produk atau topik terhangat. Data opini ulasan dapat diolah menjadi data informasi baru yang memiliki nilai lebih bagi perusahaan maupun pemanfaat data. Pengolahan data ulasan dapat dilakukan dengan menggunakan *machine learning* dengan algoritma klasifikasi untuk mendapatkan analisis sentimen terhadap produk tertentu. *Dataset* yang digunakan pada penelitian ini adalah *dataset* ulasan obat dan ulasan film untuk melakukan analisis sentimen dengan mengulas performansi algoritma Random Forest dengan menggunakan beberapa pohon keputusan yang sama yang disatukan dan Naïve Bayes Multinomial menggunakan perhitungan probabilitas pada tingkat akurasi dan waktu latih data. Dalam *preprocessing* untuk pengolahan data dan penyesuaian tipe data pada metode yang akan digunakan dengan menggunakan *CountVectorizer* untuk mengubah token kata menjadi vektor dan mengubah data fitur menjadi tipe *array*. Pembagian data latih dan uji dengan rasio 75:25. Dengan hasil akurasi data terbaik 0,57% dengan menggunakan algoritma Naïve Bayes Multinomial pada data ulasan film. dan latih waktu terlama pada algoritma Random Forest sehingga disarankan untuk dapat menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) sebagai *term* pembobotan kata untuk mendapatkan hasil akurasi yang lebih baik pada penelitian selanjutnya.

***Kata Kunci**:*
Random Forest;
Naïve Bayes;
Klasifikasi;
CountVectorizer;
Preprocessing;

Riwayat Artikel:
Diserahkan 9 Februari 2022
Direvisi 19 Maret 2022
Diterima 23 Maret 2022
Dipublikasi 29 April 2022

DOI:
10.22441/incomtech.v12i1.14770

1. PENDAHULUAN

Perkembangan internet semakin pesat dimana kemudahan akses informasi memberikan dampak yang besar terhadap keberlangsungan bermasyarakat. Kemudahan akses informasi yang ada memberikan peluang kemunculan dan perkembangan analisis data. Akses informasi dan berbagi informasi menjadi kegiatan yang umum dalam memberikan timbal balik yang tinggi. Setiap orang memiliki kesempatan dalam akses informasi dan membagikan informasi. Setiap informasi memiliki nilai yang dapat ditambang menjadi data analisa informasi baru yang lebih berguna. Timbal balik informasi individu menjadi data opini yang dapat diolah dengan penambangan manfaat data. Data opini sebagai besar didapat dari media sosial seperti Twitter, Instagram, Facebook, dan media sosial lainnya[1]. Selain media sosial data opini dapat didapatkan dari situs penyedia ulasan dimana opini ulasan produk dikumpulkan.

Dengan adanya data opini ulasan suatu produk atau tema tertentu maka data opini tersebut dapat diolah menjadi data informasi produk yang berguna bagi perusahaan terkait. Adanya data ulasan yang besar maka diperlukan teknologi dalam pengolahannya. Dalam pengolahan data dapat digunakan *machine learning* sebagai teknologi dalam pengolahan data terkomputerisai.

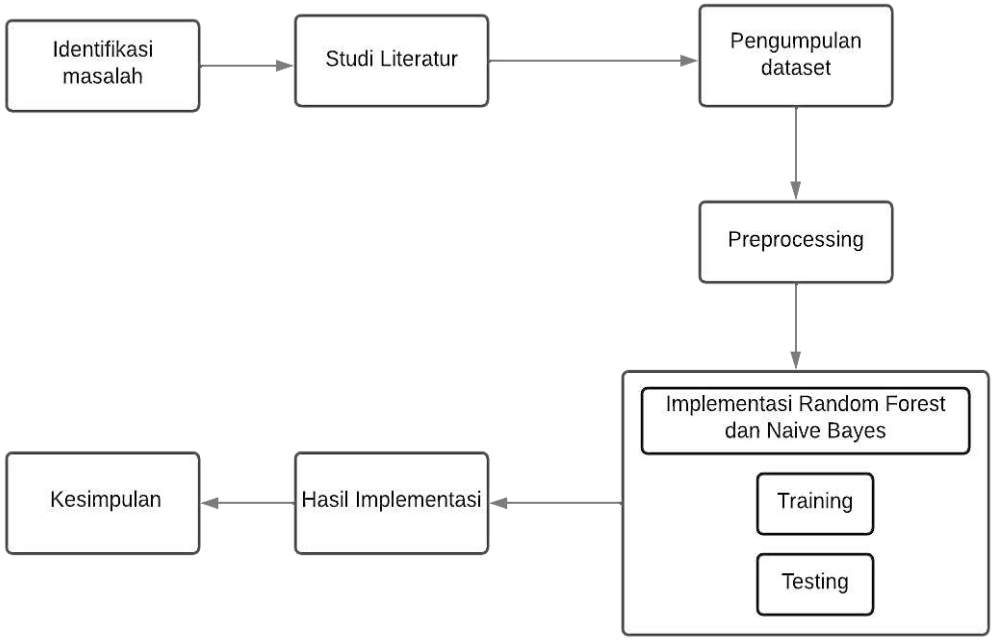
Penggunaan *machine learning* menjadi salah satu *tools* dalam melakukan analisis sentimen. Salah satu bidang penelitian pada *Natural Language Processing* (NLP)[2]. Dalam penerapannya analisis sentimen dapat diterapkan pada data ulasan, komentar dan opini. Penelitian analisis sentimen sebelumnya mengenai analisis sentimen terhadap aplikasi Ruang Guru menggunakan algoritma Naïve Bayes, Random Forest dan Support Vector Machine dengan menggunakan *dataset* ulasan aplikasi Ruang Guru pada Google Play Store dengan rentang waktu 30 hari pada bulan Maret tahun 2020 dihasilkan bahwa klasifikasi menggunakan Random Forest mendapat akurasi tertinggi yaitu 97,16% disusul oleh Support Vector Machine dengan akurasi 96,01% dan Naïve Bayes dengan akurasi 94,16% [3]. Selanjutnya mengenai penelitian sebelumnya tentang perbandingan akurasi analisis sentimen *tweet* terhadap pemerintah Provinsi DKI Jakarta pada masa pandemi dengan *dataset* yang didapat dari *crawling* data pada Twitter @dkijakarta dengan rentang waktu 9 April sampai 15 April 2020 data 14208 baris data. Dengan dilakukan pelabelan secara mandiri. Dilakukan *preprocessing* dan pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) dilanjutkan proses klasifikasi dengan didapatkannya hasil akurasi pada masing-masing model klasifikasi yaitu, algoritma Random Forest dengan akurasi sebesar 75,81%, algoritma Naïve Bayes dengan hasil akurasi 75,22%, dan algoritma Support Vector Machine 77,58% [4]. Selanjutnya penelitian oleh Kamoltep Moolthaisong pada *conference paper* mengenai analisis emosi dan Klasifikasi ulasan film menggunakan penambangan data (*data mining*) diperoleh hasil evaluasi dengan membandingkan persentase akurasi yang didapat dari penggunaan Naïve Bayes, Random Forest, dan Decision Tree. Nilai persentase akurasi tertinggi berasal dari Naïve Bayes sebesar 80.25% [5]. Pada penelitian selanjutnya oleh Asmita Singh dkk pada jurnal yang membahas mengenai dampak tipe *dataset* yang berbeda pada algoritma pengklasifikasian memaparkan bahwa perubahan jumlah atribut dalam *dataset* tidak berpengaruh pada Random Forest, sedangkan Naïve Bayes memiliki akurasi yang berfluktuasi naik dan turun[6].

Berdasarkan tinjauan penelitian di atas, penelitian ini bertujuan untuk ulasan performansi dua algoritma yaitu Random Forest *classifier* sebagai model algoritma

yang terdiri dari beberapa Random Forest dan Naïve Bayes Multinomial sebagai model algoritma yang menerapkan perhitungan probabilitas dalam pengklasifikasian nilai sentimen dari hasil akurasi tiap model dan waktu latih data dalam implementasinya pada dua *dataset* ulasan yang berbeda dengan perbedaan jumlah kelas sentimen pada ulasan obat dan *dataset* ulasan film. Perbedaan penelitian ini dengan penelitian sebelumnya terdapat pada penggunaan *dataset* dimana pada penelitian sebelumnya *dataset* yang digunakan hanya satu *dataset* sedangkan pada penelitian ini digunakan dua *dataset* dengan jumlah kelas sentimen yang berbeda dalam upaya mengetahui performa terbaik pada kedua algoritma yang digunakan dengan melihat hasil akurasi dan waktu latih data.

2. METODE

Metodologi penelitian sebagai alur dari kegiatan penelian sebagai prosedur yang harus dipenuhi oleh peneliti sebagai analisa teoritis dari sebuah cara [7]. Metode penelitian yang digunakan dalam penelitian ini akan diwakilkan oleh *flowchart* pada Gambar 1 berikut.



Gambar 1. Diagram Alir Penelitian

2.1 Identifikasi Masalah

Setelah mengidentifikasi masalah, permasalahan yang ditemukan mengenai ulasan performa terkait dua algoritma yaitu Random Forest dan Naïve Bayes pada 2 *dataset* yang berbeda jumlah kelasnya pada pengklasifikasian sentimen. Ulasan performa tersebut bertujuan untuk mengetahui algoritma mana yang memiliki performa terbaik dengan melihat tingkat akurasi dan lama waktu latih data.

2.2 Studi Literatur

Tahap selanjutnya dalam penelitian ini adalah dengan memperkaya referensi penelitian dengan mengambil dari buku referensi, konferensi paper dan jurnal [8] yang berhubungan dengan pembahasan tema yaitu pengklasifikasian analisis sentimen dengan algoritma Random Forest dan Naïve Bayes Multinomial.

2.3 Pengumpulan *Dataset*

Tahap penelitian selanjutnya dengan mengumpulkan *dataset* ulasan sentimen sebagai data sekunder. *Dataset* yang digunakan pada penelitian ini adalah *dataset* ulasan obat terdapat pada *uci machine learning* dengan alamat *website* https://archive.ics.uci.edu/ml/*datasets*/Drug+Ulasan+*Dataset*+%28Drugs.com%29 dengan jumlah data latih 161297 dengan 6 kolom atribut dan data uji 53766 dengan 7 kolom serta *dataset* ulasan film IMDB pada Kaggle https://www.kaggle.com/lakshmi25npathi/sentiment-analysis-of-imdb-film-ulasans/data dengan jumlah data 50000 baris 2 kolom atribut.

2.4 *Preprocessing*

Tahap *preprocessing* sebagai tahap pengolahan data dengan menghapus data yang tidak perlu untuk meningkatkan kualitas data [8]. Adapun tahap *preprocessing* pada penelitian ini sebagai berikut:

a. Pembersihan Data

Pada tahap pembersihan data sebagai proses pengolahan data sesuai kebutuhan, konsisten dengan membersihkan data melalui beberapa tahapan yaitu tahap dalam membersihkan *noise* data [8] meliputi data kosong dan kata tidak penting dengan menggunakan *stopword removel* [9]. Menghapus *tag* html, mengubah huruf menjadi huruf kecil *case folding*, *stemming* menjadikan kata dasar. Setelah melakukan pembersihan data maka data didokumentasikan dalam kolom data bersih untuk di ditransformasi.

b. Transformasi Data

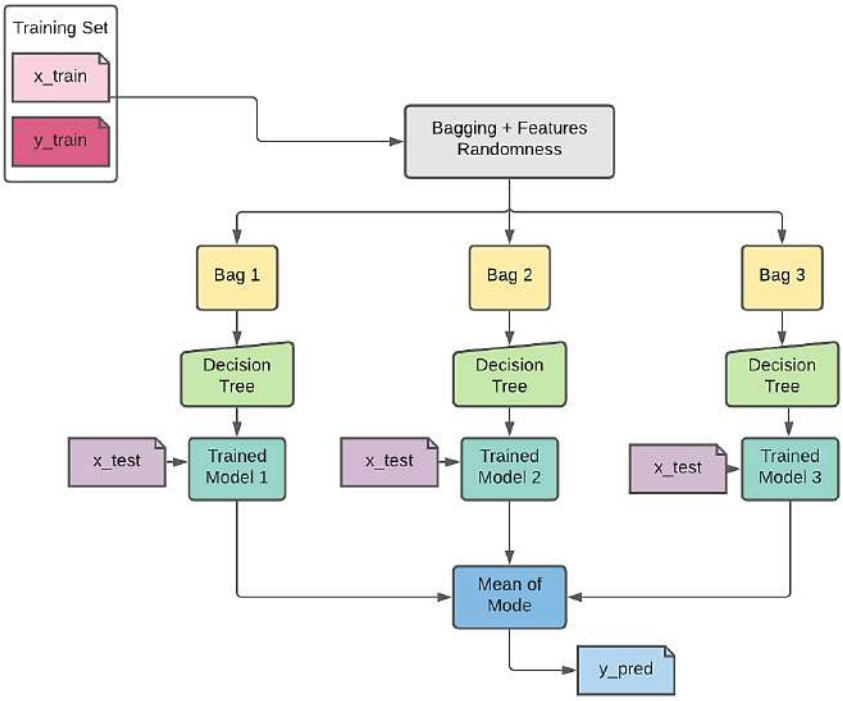
Tahap selanjutnya dalam *preprocessing* dengan melakukan tranformasi data dimana data yang sudah bersih diubah formatnya sesuai kebutuhan penambangan data (*data mining*) [10]. Data diubah dari data *string* menjadi vektor lalu data vektor tersebut di ubah menjadi data *array* agar dapat dilakukan implementasi klasifikasi dengan algoritma klasifikasi.

2.5 Implementasi Random Forest dan Naïve Bayes Multinomial

Implementasi algoritma dilakukan dengan algoritma Random Forest terlebih dahulu kemudian dilanjutkan dengan implementasi Naïve Bayes pada *dataset* pertama ulasan obat dan dilanjutkan implementasi algoritma Random Forest dan Naïve Bayes Multinomial pada *dataset* ulasan film.

1. Random Forest sebagai implementasi dari *homogeneous ensemble learning* atau pembelajaran *ensamble* dengan model yang sama dengan menerapkan Decision Tree [9]. Selain *bagging* Random Forest juga menerapkan *features randomness* dimana tiap *bagging* yang dihasilkan akan mengadopsi *features* dari sampel latih

secara acak. Kita dapat artikan tiap model Decision Tree akan dilatih dengan *bag* yang berisikan *dataset* yang beragam. Berikut merupakan diagram alur dari algoritma Random Forest pada Gambar 2 sebagai berikut.



Gambar 2. Diagram Proses Random Forest

Dari diagram pada Gambar 2 dijabarkan bahwa Random Forest dapat dibangun menggunakan *bagging* dengan pemilihan atribut acak [9]. Pada setiap *bag* di latih menggunakan model yang sama yaitu Decision Tree sehingga menghasilkan model terlatih dengan hasil yang berbeda beda dikarenakan *dataset* yang berbeda walau menggunakan model yang sama. Selanjutnya prediksi tiap model akan di satukan menggunakan *majority voting* untuk mendapatkan nilai prediksi (y_pred) [10].

2. Naïve Bayes Multinomial sebagai salah satu algoritma klasifikasi yang didasarkan pada Teorema Bayes [11]. Teori ini digunakan untuk memprediksi peluang dimasa depan yang didasarkan pengalaman atau kondisi sebelumnya. Metode ini menggunakan pengklasifikasian metode probabilitas[10] dan statistik. Metode ini menggunakan probabilistik objek dimana objek memilki memiliki ciri-ciri tertentu yang termasuk dalam kelas atau pengelompokan tertentu. Dengan memperkirakan himpunan peluang dan menghitung kemunculan dalam himpunan data [12]. serta pada klasifikasi jenis teks *Naïve Bayes* memiliki performa yang cukup baik [13]. hal itu lah alasan mengapa algoritma Naive Bayes disebut naif. Dipilihnya

metode *Naïve Bayes* karena memiliki keunggulan seperti kesederhanaan perhitungan, presisi tinggi dan kecepatan pada pemrosesan [14].

Pada dasarnya ada beberapa model Naïve Bayes untuk klasifikasi salah satunya Naïve Bayes Multinomial. Sebagai salah satu metode *supervised learning* [15] Naïve Bayes Multinomial memerlukan pelabelan data sebelum data di-latih. Nilai probabilitas dari kemunculan nilai target label Y dapat dihitung dengan cara [16]:

$$P(x|y) = \frac{P(x|y)\ .\ P(y)}{P(x)} \tag{2.1}$$

- $P(x|y)$: mengkalkulasikan probabilitas dari kemunculan sekumpulan nilai fitur X bila diketahui nilai target label y. Lalu nilai tersebut di kali dengan $P(y)$
- $P(y)$: probabilitas nilai target label y. Selanjutnya nilai tersebut di bagi dengan $P(x)$
- $P(x)$: nilai dari probabilitas dari sekumpulan nilai fitur x.

2.6 Akurasi dan Waktu Latih

Analisis hasil penelitian dilakukan dengan menganalisa hasil akurasi dari setiap performansi algoritma Random Forest dan Naïve Bayes Multinomial pada *dataset* ulasan obat dan ulasan film serta menganalisa lama waktu latih setiap *dataset* dengan menggunakan algoritma tersebut.

2.7 Kesimpulan

Penarikan simpulan sebagai tahap akhir dari penelitian ini dengan menarik kesimpulan hasil penelitian dan pengujian pada penelitian ini.

3. HASIL DAN PEMBAHASAN

3.1 Data Deskripsi dan *Preprocessing*

Pada penelitian ini *dataset* pertama dari UCI *machine learning drug review dataset (drugs.com) dataset* kedua dari Kaggle IMDB *Dataset of 50K Movie Reviews.*

UCI *machine learning drug review dataset (drugs.com)* berisi dua *dataset* yaitu *drug_train_raw* sebagai *dataset* latih dan *drug_test_raw* sebagai *dataset* uji dimana pada penelitian ini kedua *dataset* pada *drug_review* digabungkan untuk mendapat hasil analisis yang lebih baik dengan jumlah data 215063 baris dengan 7 kolom. Data yang digunakan dalam analisis sentimen yaitu data ulasan dan *rating* dengan data *rating* dibagi menjadi tiga klasifikasi positif, negatif dan netral seperti pada Tabel 1 berikut ini.

Tabel 1. Nilai Sentimen

Sentimen	***Rating***	**Nilai Sentimen**
Positif	*Rating* >= 7	2
Netral	6 <= *Rating* <7	1
Negatif	*Rating* <=5	0

Pada data dilakukan cek data kosong dimana terdapat data kosong pada kolom *condition* yang mana data kosong harus dihapus guna mendapat hasil analisis yang baik. Selanjutnya pada kolom data ulasan dilakukan *preprocessing* dengan menggunakan *stopword corpus nltk* dengan list kata yang umum untuk menjaga sentimen ulasan. Pada tahap ini dilakukan penghapusan data dimana kondisi dengan kurang dari 5 obat akan dihapus. Sehingga 156432 jumlah data. Data tersebut diubah menjadi huruf kecil, menghapus selain huruf, *tag* html dan menetapkan stemming penggunaan kata dasar menggunakan *nltk library* yaitu *nltk.stem.snowball.SnowballStemmer.*

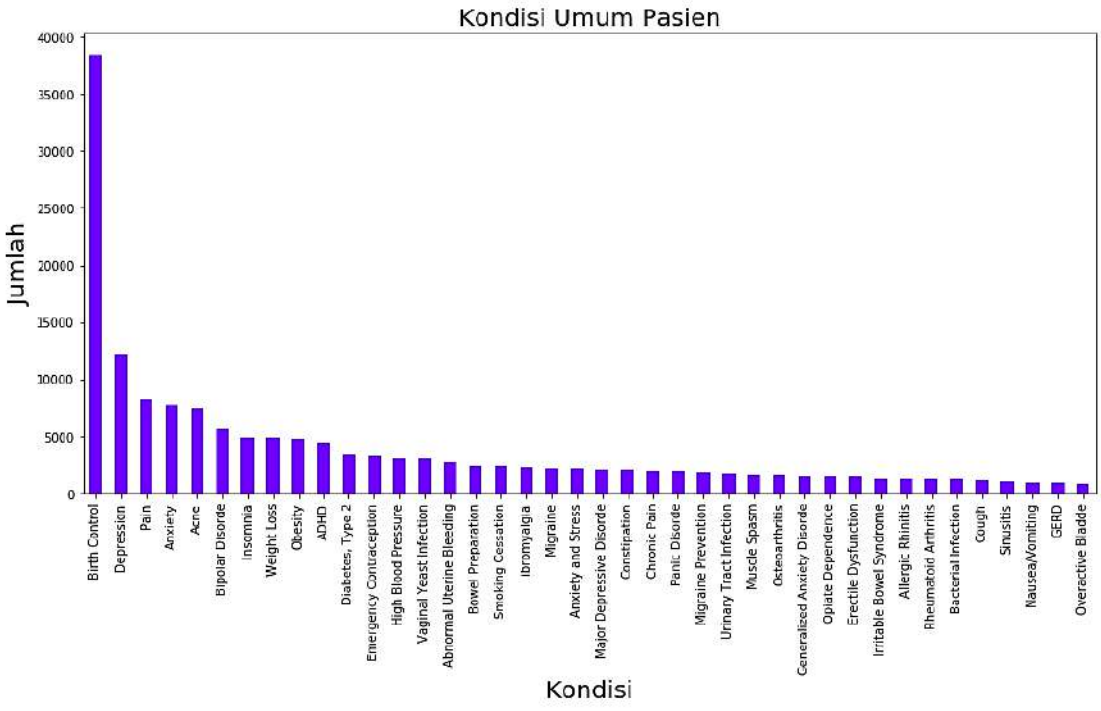
Selanjutnya pada *preprocessing* data diubah menjadi matrik dengan melakukan vektorisasi dengan menggunakan *CountVectorizer* dan mengubah hasil vektorisasi menjadi data *array* sebagai metode untuk mengubah data teks menjadi nilai numerik [11]. Data latih dan uji dibagi dengan rasio 75:25.

Selanjutnya pada *dataset* kedua Kaggle IMDB *Dataset of 50K Movie Reviews* berisi 50.000 *dataset* ulasan dengan dua kolom data yaitu ulasan dan sentimen. Sentimen dibagi menjadi dua kelas yaitu sentimen positif dan negatif. Pada kolom sentimen dilakukan label *encoder* guna mengubah label positif dan negatif diwakilkan oleh angka yaitu 1 untuk positif dan 0 untuk negatif. Jumlah data sentimen terdiri dari 25000 data positif dan 25000 data negatif.

Preprocessing data dilakukan pada kolom ulasan dengan tahapan seperti pada *dataset* sebelumnya. Dengan hasil pembersihan data pada kolom *review_clean*. Data latih dan uji dibagi dengan rasio perbandingan 75:25 dengan jumlah data latih 37500 dengan 3 kolom dan data uji 12500 dengan 3 kolom.

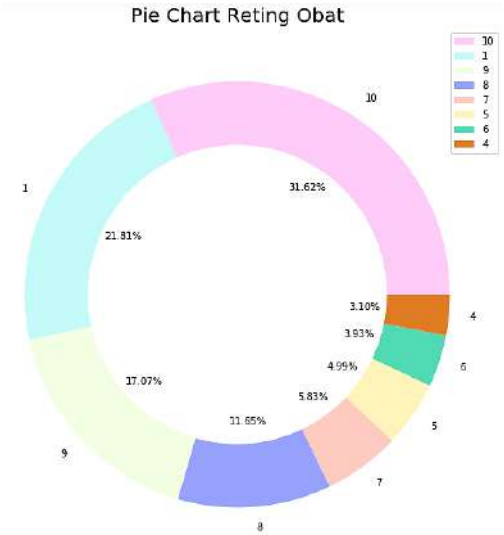
3.2 Hubungan Antar Data

Memahami hubungan antar data sebagai sarana untuk menentukan teknik statistik untuk analisis data yang sesuai. Pada *dataset* ulasan obat hubungan antar data ditemukan bahwa kondisi paling umum pada *dataset* yaitu *birth control* dengan jumlah kondisi sekitar 38000 kondisi seperti pada Gambar 3 di bawah ini.

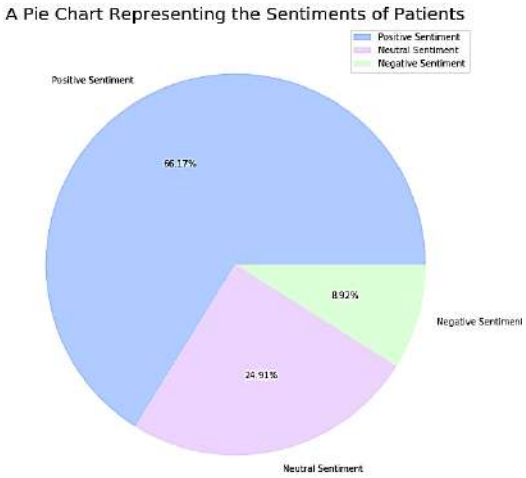


Gambar 3. Kondisi Umum Pasien

Visualisasi data *rating* dengan menggunakan *pie chart* guna mengetahui jumlah pada tiap *rating*-nya pada Gambar 4. Dilanjutkan dengan distribusi hasil pembagian kelas sentimen sesuai dengan *rating* yang divisualisasikan dengan *pie chart* seperti pada Gambar 5 sebagai berikut:



Gambar 4. *Pie Chart Rating* Obat



Gambar 5. *Pie Chart* Sentimen

Memahami hubungan antar data pada *dataset* ulasan film yang mana terdapat dua kolom yaitu kolom ulasan dan sentimen. Maka kita dapat memvisualisasikan data sentimen dengan menggunakan *pie chart* pada Gambar 6 sebagai berikut.

Gambar 6. Data Sentiment Film

3.3 Hasil Implementasi Algoritma Random Forest dan Naïve Bayes

Setelah semua data dibersihkan dan memahami hubungan antar data maka kita melakukan *CountVectorizer* sebagai konversi teks menjadi vektor dimana setiap kata akan memiliki nilai sehingga data dapat diolah pada model implementasi algoritma. Sebelum melakukan implementasi algoritma maka setiap *dataset* dibagi menjadi dua bagian yaitu data latih sebagai data latih model algoritma dan data uji sebagai data untuk menguji model algoritma. Dalam implementasinya data latih dan data test dibagi dengan rasio 75:25. Selanjutnya *dataset* ulasan obat akan diimplementasikan pada model algoritma Random Forest dilanjutkan dengan implementasi model algoritma Naive Bayes Multinomial. Setelah didapatkan hasil akurasi dan waktu latih data menggunakan *dataset* ulasan obat maka selanjutnya implementasi kedua model algoritma menggunakan *dataset* ulasan film dengan pembagian *dataset* berupa data latih dan data uji dengan rasio 75:25 sehingga didapatkan hasil implementasi berupa akurasi dan waktu latih data pada Tabel 2 berikut ini.

Tabel 2. Perbandingan Hasil Akurasi

		Random Forest	***Naïve Bayes Multinomial***
Ulasan **obat**	Latih time	1480.49045586586	308.73597264289856
	Accuracy	0.43737854147489	0.5769663495959906
Ulasan film	Latih time	120.35102891921997	7.296995162963867
	Accuracy	0.56272	0.56216

Dari hasil akurasi pada Tabel 2 di atas dapat kita lihat bahwa pada ulasan obat akurasi terbaik pada algoritma Naïve Bayes Multinomial dengan hasil akurasi 0.57% dengan waktu data 308 detik sedangkan Random Forest memiliki akurasi lebih kecil 0.43% dengan waktu latih lebih lama yaitu 1480 detik. Selanjutnya untuk *dataset* ulasan film akurasi data bernilai sama yaitu 0,56% tetapi waktu latih berbeda, dimana Naïve Bayes melakukan waktu latih lebih cepat yaitu 7 detik sedangkan Random Forest memakan waktu latih selama 120 detik. Untuk penelitian selanjutnya kita dapat menambahkan metode algoritma Neural Network

sebagai metode tambahan untuk dibandingkan hasil akurasi dan waktu latihnya sehingga menambah ulasan performansi untuk mengetahui algoritma dengan akurasi terbaik dan waktu latih data terbaik dalam klasifikasi analisis sentimen.

4. KESIMPULAN

Berdasarkan hasil penelitian ini menghasilkan kesimpulan bahwa implementasi algoritma Random Forest memerlukan waktu latih data yang lebih lama dibanding dengan algoritma Naïve Bayes pada *dataset* ulasan obat selama 1172 detik lebih lama. Namun pada saat jumlah kelas hanya dua yaitu kelas negatif dan postif, Random Forest dan Naïve Bayes memiliki hasil akurasi yang sama pada *dataset* ulasan film dengan hasil akurasi 0,56%. Tetapi untuk waktu latih memiliki perbedaan. Naïve Bayes Multinomial dapat melakukan pelatihan lebih cepat dengan waktu latih hanya 7 detik sedangkan Random Forest selama 120 detik. Sehingga disarankan untuk dapat menggunkan TF-IDF sebagai *term* pembobotan kata untuk mendapatkan hasil akurasi yang lebih baik pada penelitian selanjutnya.

REFERENSI

[1] F. P. Rachman, "Perbandingan Model Deep Learning untuk Klasifikasi Sentiment Analysis dengan Teknik Natural Languange Processing," *J. Teknol. dan Manaj. Inform.*, vol. 7, pp. 103–112, 2021.

[2] C. Colón-Ruiz and I. Segura-Bedmar, "Comparing deep learning architectures for sentiment analysis on drug ulasans," *J. Biomed. Inform.*, vol. 110, no. February, p. 103539, 2020, doi: 10.1016/j.jbi.2020.103539.

[3] E. Fitri, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *J. Transform.*, vol. 18, no. 1, p. 71, 2020, doi: 10.26623/transformatika.v18i1.2317.

[4] R. D. Himawan and E. Eliyani, "Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi," *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 1, p. 58, 2021, doi: 10.26418/jp.v7i1.41728.

[5] K. Moolthaisong and W. Songpan, "Emotion Analysis and Classification of Film Ulasans Using Data Mining," *2020 Int. Conf. Data Sci. Artif. Intell. Bus. Anal. DATABIA 2020 - Proc.*, pp. 89–92, 2020, doi: 10.1109/DATABIA50434.2020.9190363.

[6] A. Singh, M. N., and R. Lakshmiganthan, "Impact of Different Data Types on Classifier Performance of Random Forest, Naïve Bayes, and K-Nearest Neighbors Algorithms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 12, pp. 1–11, 2017, doi: 10.14569/ijacsa.2017.081201.

[7] A. F. Kurniawan, S. F. Pane, and R. M. Awangga, "Prediksi Jumlah Penjualan Rumah di Bojongsoang ditengah Pandemi Covid-19 dengan Metode ARIMA," *J. Media Inform. Budidarma*, vol. 5, no. 4, pp. 1479–1487, 2021, doi: 10.30865/mib.v5i4.3121.

[8] C. Prianto, N. H. Harani, and I. Firmansyah, "Analisis Sentimen Terhadap Kandidat Presiden Republik Indonesia Pada Pemilu 2019 di Media Sosial Twitter," *J. Media Inform. Budidarma*, vol. 3, no. 4, p. 405, 2019, doi: 10.30865/mib.v3i4.1549.

[9] M. Farid and D. Fitrianah, "Rekomendasi Pemilihan Restoran Berdasarkan Rating Online Menggunakan Algoritma C4.5," *J. Telekomun. dan Komput.*, vol. 11, no. 1, p. 9, 2021, doi: 10.22441/incomtech.v11i1.9791.

[10] D. E. Dila, Y. arum Sari, and M. T. Furqon, "Pembentukan Daftar Stopword Menggunakan Zipf Law dan Pembobotan Augmented TF-Probability IDF pada Klasifikasi Dokumen Ulasan Produk," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, pp. 406–412, 2020.

[11] D. Irawan, E. B. Perkasa, Y. Yurindra, D. Wahyuningsih, and E. Helmud, "Perbandingan Klassifikasi SMS Berbasis Support Vector Machine, Naive Bayes Classifier, Random Forest dan Bagging Classifier," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 10, no. 3, pp. 432–437, 2021, doi: 10.32736/sisfokom.v10i3.1302.

[12] A. I. Kusumarini, P. A. Hogantara, M. Fadhlurohman, and N. Chamidah, "Perbandingan Algoritma Random Forest, Naïve Bayes, Dan Decision Tree Dengan Oversampling Untuk Klasifikasi Bakteri E. Coli," no. April, pp. 792–799, 2021.

[13] R. Leonardo, J. Pratama, and C. Chrisnatalis, "Perbandingan Metode Random Forest Dan Naïve Bayes Dalam Prediksi Keberhasilan Klien Telemarketing," *J. Teknol. Dan Ilmu Komput. Prima*, vol. 3, no. 2, pp. 1–5, 2020.

[14] E. Indrayuni, "Klasifikasi Text Mining Ulasan Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes," *J. Khatulistiwa Inform.*, vol. 7, no. 1, pp. 29–36, 2019, doi: 10.31294/jki.v7i1.1.

[15] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.

[16] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 640, 2021, doi: 10.30865/mib.v5i2.2937.