ANALISIS SENTIMEN PEJUALAN JAFRA DALAM PANDEMI COVID-19 DENGAN ALGORITMA KLASIFIKASI

Petrus Damianus Batlayeri[1], Windu Gatta[2]

[1]Program Studi Magister Ilmu Komputer, [2,]Program Studi Magister Ilmu Komputer, Universitas Nusa Mandri Jakarta

Jl. l. Kramat Raya No.18, RW.7, Kwitang, Kec. Senen Kota Jakarta Pusat
[1] 14002337@nusamandiri.ac.id, [2]windu@nusamandiri.ac.id

Abstract

At *this time in the era of big data, the use of social media often make posts on his social media accounts in the form of opinions on events and goods around him, one of which is making a post that gives an opinion on an item so that we know the impact or public view of a product in this case that is jafra. The stages in this study began with data collection conducted by collecting tweet data on social media twitter as many as 1,000 tweets related to jafra products, then pre-processing to find the words that often appear in tweets. This study aims to determine public sentiment on the impact of sales of Jafra products during the covid-19 pandemic, thus helping sales efforts to conduct research on public opinion. Classification algorithms such as Naive Bayes (NB), K-Nearest Neighbor (k-NN) and Decision Tree are proposed by many researchers for use in text sentiment analysis. The three algorithms and methods will be tested with two inputs using positive and negative Tokenize and Transform Case comments, accuracy obtained by Naïve Bayes accuracy algorithm: 74.92%, k-NN accuracy: 76.22%, Decision Tree accuracy: 77.85% The results showed that Decisoin Tree algorithm gets the best and accurate results*

Keywords : *Jafra, Naïve Bayes, K-Nearest Neighbour, Decision Tree*

Abstrak

Pada masa ini di zaman big data , penggunaan media sosial sering kali membuat postingan di akun media social miliknya berupa opini-opini terhadap kejadian dan barang disekitarnya, salah satunya yaitu membuat suatu postingan yang memberikan opini pada suatu barang sehingga kita jadi tahu dampak atau pandangan publik dari suatu produk pada kasus ini yaitu jafra. Adapun beberapa tahapan-tahapan dalam penelitian ini diawali dari pengumpulan data yang dilakukan dengan mengcollect data tweet pada media social twitter sebanyak 1.000 tweet yang berkaitan dengan produk jafra, selanjutnya dilakukan pra-prosesing untuk mencari kata-kata yang sering muncul didalam *tweet*. Penelitian ini bertujuan untuk menentukan sentimen publik terhadap dampak dari penjualan produk Jafra dimasa pandemic virus covid-19 ini, sehingga membantu usaha penjualan untuk melakukan riset atas opini publik. Klasifikasi algoritma seperti *Naive Bayes* (NB), *K-Nearest Neighbor* (k-NN) dan *Decision Tree* yang diusulkan oleh banyak peneliti untuk digunakan dalam analisis sentimen teks. Ketiga algoritma dan metodenya, akan diuji dengan dua masukan dengan menggunakan komentar *Tokenize* and *Transform Case* yang positif dan negatif , akurasi yang didapat algoritma *Naïve Bayes accuracy*: 74.92%, k-NN *accuracy*: 76.22%, *Decision Tree accuracy*: 77.85% Hasil penelitian menunjukkan bahwa algoritma *Decisoin Tree* mendapatkan hasil terbaik dan akurat

Kata kunci : *Jafra, Naïve Bayes, K-Nearest Neighbour, Decision Tree*

1. PENDAHULUAN

Pada masa pandemi covid ini merupakan masa dimana terjadi ketidakstabilan dari berbagai masa aspek kehidupan, mulai dari aspek kesehatan, social, dan juga aspek ekonomi. Dampak dari adanya pandemi tersebut mempengaruhi segala aspek dan salah satunya adalah penjualan produk kosmetik yang ada di Indonesia, dari dampak tersebut banyak produsen menanggapi permasalahan penjualan tersebut disosial media dengan berbagai sentiment baik maupun positif. Pada Analisis sentimen yang disertakan dalam *data mining* untuk membantu menentukan komentar mana yang positif atau negatif di seluruh sistem. Analisis sentimen adalah teknologi berbantuan komputer yang dapat mendukung dan menganalisis opini tekstual, metode untuk memahami, mengekstraksi, dan menghasilkan informasi emosional, seperti halnya penambangan teks[1]

Teknologi media sosial dapat mengambil banyak bentuk, termasuk majalah, forum internet, weblog, podcast, foto, dan foto [5]. Lainnya menggunakan media sosial sebagai referensi data yang berfungsi sebagai sumber penting opini dan reaksi masyarakat, ukuran kesadaran, dan tolok ukur layanan agensi dan perusahaan. Saat ini, proses pengambilan data di media sosial Twitter dapat ditangani oleh beberapa algoritma yang sulit untuk media sosial, karena penulis percaya bahwa opini adalah platform yang berharga. Hambatan untuk mengumpulkan data dari media sosial adalah bahasa formal atau informal, sehingga Anda perlu menafsirkan dan menganalisis pendapat dan pendapat yang Anda terima agar dapat digunakan sebagai pengambil keputusan. Penulis yang merupakan metode pengolahan data dari opini Twitter saat ini mencoba untuk membandingkan dan menggunakan tiga metode yang lebih akurat dan dapat diproses. Metode tersebut antara lain KNearest Neighbor, Decision Tree, dan Naive Bayes menggunakan aplikasi RapidMiner. Oleh karena itu, pada penelitian ini penulis menyarankan untuk menggunakan pohon keputusan, naive Bayes, dan K-nearest neighbor sebagai pembanding proses klasifikasi yang menyediakan preprocessing untuk pengolahan data tweet. Hal ini dimaksudkan untuk meningkatkan akurasi pengelompokan tweet. .. Berdasarkan jumlah tweet.

2. TINJAUAN PUSTAKA DAN TEORI

2.1. Penelitian terkait

Pada penelitian Upma Kumari dkk, pada tahun 2017 tentang Analis Sistem Tinjauan Produk Ponsel Pintar Menggunakan Teknik Svm *Clasification* dengan hasil akurasi yang didapat dari metode svm adalah 90.99% [2].

Dari penelitian Pradep Kumar Tiwari dkk, pada tahun 2018 tentang Identifikasi Bot: Membantu Analis untuk Data yang Tepat di Twitter dengan menggunakan metode *naïve bayes* dan Bot Clasifier dengan hasil precision pada Bot *clasifier* 80.2% dan precision naïve bayes 79[3].

Berdasarkan penelitian dari Ratih Yulia Ayuningtyas dkk pada tahun 2019 tentang Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata TMII Menggunakan Naïve Bayes dan yang menggunakan 2 metode yaitu naïve bayes dengan naïve bayes dan PSO, dengan hasil dari penelitian tersebut hasil akurasi dari naïve bayes yaitu 70% sedangkan hasil dari akurasi *naïve bayes* dan PSO sebesar 94.02%[4].

Adapun penelitian lain dari Oman Somantri dkk pada tahun 2019 tentang Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis *Text Mining*, dengan hasil yang telah didapatkan metode dengan menggunakan *Naïve Bayes* tingkat akurasi yang lebih baik dibandingkan dengan decision tree yaitu mempunyai tingkat akurasi sebesar 73,33%[5].

Pada penelitian Moh. Nazrul Adziz tentang *Sentiment Analysis and Topic Modelling for Identification of Government Service Satisfaction,* dengan hasil yang telah didapatkan bahwa akurasi dengan metode SVM 75%.[6]

2.2. Algorima K-Nearest Neighbors

K-Nearest Neighbor (K-NN) Untuk proses penyelesaian masalah klasifikasi K-nearest neighbor (KNN) adalah yang paling sederhana dari klasifikasi lainnya. Teknik yang digunakan dalam JST ini adalah dengan mengklasifikasikan data menggunakan objek-objek dengan nilai-nilai terdekat yang berdekatan. Hasil yang diperoleh dari proses JST lebih tinggi atau terbaik jika setiap pembobotan kesamaan cosinus digunakan dalam perhitungan setiap suku. Klasifikasi teks dengan metode KNN memberikan nilai yang lebih baik ketika ekspresi kesamaan cosinus digunakan untuk menimbang setiap kata dalam dokumen teks yang sedang diproses. Sebelum menghitung nilai cosinus similarity perlu dilakukan langkah-langkah proses

pemerataan kata (tf, df, idf, tfidf) yang terdapat pada Persamaan 1 sd 3. Setelah pembobotan kata selesai, langkah selanjutnya adalah menggunakan rumus cosinus similarity untuk menghitung kemiripan antar dokumen. Persamaan persamaan cosinus ditunjukkan oleh persamaan berikut:[7].

$$CosSim(q, d_j) = \frac{d_j \cdot q}{|d_j| \cdot |q|} = \frac{\sum_{i=1}^{t}(w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^{t} w_{ij}^2} \cdot \sqrt{\sum_{i=1}^{t} w_{ij}^2}} \quad (1)$$

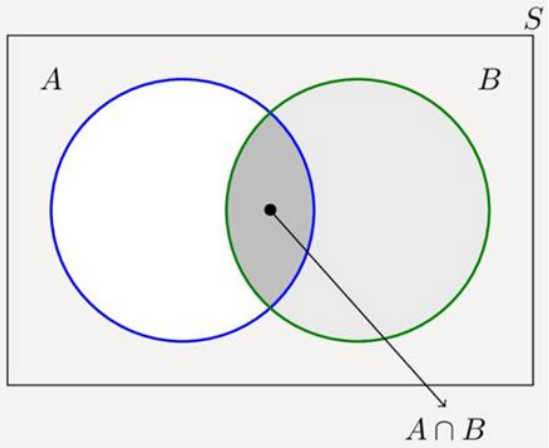
Keterangan : CosSim(q, d): Nilai kemiripan antara dokumen uji, (q) dengan dokumen latih ke j (d j) t: Jumlah term (kata) d : dokumen q : kata kunci (query) wij : Bobot term (kata) ke i pada dok. latih j wiq : Bobot term (kata) ke i pada dok.uji q

2.3. *Algoritma Naïve Bayes*

Pengklasifikasi naif Bayes adalah metode klasifikasi berdasarkan teorema Bayes. Metode klasifikasi probabilistik dan statistik yang diusulkan oleh ilmuwan Inggris Thomas Bayes, yang memprediksi peluang masa depan berdasarkan pengalaman masa lalu, dikenal sebagai teorema Bayes. Fitur utama dari pengklasifikasi Bayes naif ini adalah membuat asumsi yang sangat kuat (sederhana) tentang independensi semua kondisi dan peristiwa. Menurut [15], NBC merupakan salah satu algoritma dari teknik data mining yang menggunakan teorema Bayes untuk klasifikasinya. Teorema Bayes adalah pendekatan statistik dasar untuk pengenalan pola. Naive Bayes didasarkan pada asumsi yang disederhanakan bahwa nilai atribut secara kondisional independen satu sama lain sehubungan dengan nilai output. Dengan kata lain, mengingat output, probabilitas pengamatan kolektif adalah produk dari probabilitas individu. Kemudian gunakan teorema ini untuk mengklasifikasikan ke dalam data mining. Teorema Bayes memiliki bentuk umum berikut:[8].

$$P(H|X)=(P(X|H)P(H))/(p(x)) \quad (2)$$

Keterangan :
X = Data dengan class yang belum diketahui
H = Hipotesis data X merupakan suatu class spesifik
P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)
P(H) = Probabilitas hipotesis H (prior prob.)
P(X|H) = Probabilitas X berdasarkan kondisi tersebut
P(X) = Probabilitas dari X
P(X|H)P(H)
P(X)



Gambar 1. Theorema Naïve Bayes

2.4. *Algoritma Decision Tree*

Decision Tree adalah Pohon keputusan struktur diagram alur seperti pohon, di mana setiap simpul internal mewakili pengujian atribut, setiap cabang mewakili hasil pengujian, dan simpul daun mewakili kelas atau distribusi kelas. Aliran pohon keputusan berlanjut dari simpul akar ke simpul daun yang berisi ramalan[9].

Nilai gain adalah information gain yang digunakan untuk mencari variabel/atribut pada dataset (S) yang digunakan sebagai root/node dan node cabang, dan merupakan atribut dengan nilai gain tertinggi. Anda dapat menggunakan konsep entropi, koefisien Gini, dan kesalahan klasifikasi untuk mencari nilai perolehan (information gain). Nilai profit maksimum yang diperoleh dari atribut-atribut dataset (data pelatihan) pertama-tama digunakan untuk mencari atribut yang layak mendapatkan akar (pohon keputusan) dari pohon keputusan. Kemudian proses pencarian atribut yang akan menjadi cabang diulangi sampai menemukan daun yang menjadi label kelasnya. Entropi yang informasinya diperoleh dengan nilai entropi adalah ekspresi untuk menghitung keseragaman atribut (A) dari data sampel (S).

$$Entrophy(S) = \sum_{i=1}^{n} p_i log_2 p_i \quad (3)$$

S = Himpunan kasus dalam dataset
A = Fitur (atribut)
n = jumlah partisi atribut S
Pi = proposi dari Si terhadap S

Maka :

$$Entrophy(S) = \sum_{i=1}^{n} p(i|s) log_2 p(i|s) \quad (4)$$

Gain(S.A) adalah informasi dari attribute A pada koleksi contoh

$$Gain(S,A) = Entrop\ y(S) \sum_{i=1}^{n} \frac{|S_i|}{|S|} \quad (5)$$
$$Entrop\ y(S_i)$$

Information Gain dengan Nilai Gini Index

$$Ginit(t) = 1\ \sum_{i=1}^{n} [p(i|s)]^2 \quad (6)$$

$$Gain(S,A) = Entropy(S) \sum_{i=1}^{n} \frac{|S_i|}{|S|} Gini(S_i) \quad (7)$$

Information Gain dengan classification error, nilai C. Error diperoleh dari nilai value attribut yang terkecil dari class label

$$C.Error(S) = 1 \quad max_i[p(i|s)] \quad (8)$$

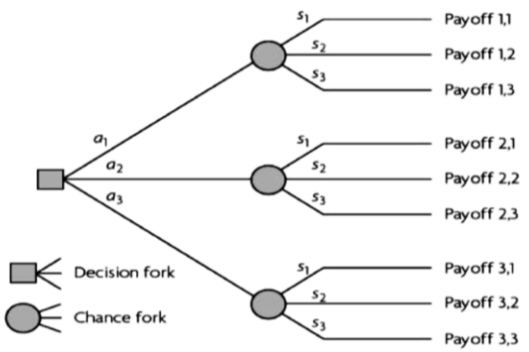
$$Gain(S,A) = Error(S) \sum_{i=1}^{n} \frac{|S_i|}{|S|} Error(S_i) \quad (9)$$

Layaknya sebuah pohon, Seperti halnya pohon, pohon keputusan memiliki akar (root/node), batang/cabang (branch node), dan daun (leaf). Strategi yang dapat Anda gunakan untuk membangun pohon keputusan menggunakan pohon keputusan adalah:

1) Pohon dimulai sebagai simpul tunggal (root) yang mewakili semua data.
2) Setelah root node dibuat, data pada root node diukur menggunakan information gain untuk memilih atribut yang akan digunakan sebagai atribut pembagi.
3) Cabang dibuat sebagai pembagi dari atribut yang dipilih dan data didistribusikan ke setiap cabang.
4) Algoritma ini terus menggunakan proses yang sama (secara rekursif) untuk membangun pohon keputusan. Jika atribut dipilih sebagai simpul split/cabang, atribut ini tidak akan lagi dimasukkan dalam perhitungan nilai perolehan informasi.
5) Proses pembagian rekursif berakhir ketika salah satu kondisi berikut terpenuhi:
   - Semua data dari subcabang berada dalam kelas yang sama.
   - Semua atribut sudah digunakan, tetapi datanya masih berada di kelas lain. Dalam hal ini, data yang mewakili sebagian besar kelas akan digunakan sebagai label kelas untuk simpul daun.
   - Tidak ada data untuk cabang baru. Dalam hal ini, simpul daun dari cabang sebelumnya dipilih dan data yang mewakili sebagian besar kelas digunakan sebagai label kelas

.

Pohon keputusan adalah metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang mewakili aturan. Aturannya mudah dimengerti dalam bahasa alami. Itu juga dapat direpresentasikan dalam database seperti Structure Query Language (SQL) untuk mengambil catatan untuk data tertentu. Pohon keputusan adalah struktur yang dapat digunakan untuk membagi kumpulan data besar menjadi kumpulan data yang lebih kecil dengan menerapkan seperangkat aturan keputusan. Di pohon keputusan, setiap simpul daun menandai label kelas. Sebuah node yang bukan node terakhir terdiri dari root dan node internal yang terdiri dari kondisi pengujian atribut untuk beberapa record dengan properti yang berbeda. Pohon keputusan adalah struktur diagram alur seperti pohon, di mana setiap simpul internal mewakili pengujian atribut, setiap cabang mewakili hasil pengujian, dan simpul daun mewakili kelas atau distribusi kelas. Alur pohon keputusan berlanjut dari simpul akar ke simpul daun yang berisi ramalan.



Gambar 1. Bentuk Dari Decision Tree

2.5 *Term Frequency (TF)*

Term Frequency merupakan frekuensi kemunculan kata dalam dokumen teks. Term frekuensi (tft, d) didefinisikan sebagai banyaknya kemunculan term t dalam dokumen d

2.6 *Document Frequency (DF)*

Document Frequency merupakan kata yang digunakan dalam banyak dokumen. Kata ini tidak seinformatif kata tersebut.

2.7 *Invers Documents Frequency (IDF)*

Invers Document Frequency merupakan frekuensi kemunculan istilah di seluruh dokumen teks. Term yang jarang muncul di seluruh dokumen teks memiliki nilai frekuensi dokumen terbalik yang lebih besar dari nilai term yang sering muncul.

2.8 *Term Frequency-Invers Documents Frequency (TF-IDF)*

Term Frequency merupakan nilai tfidf suatu kata yang merupakan gabungan dari nilai tf dan idf dalam perhitungan pembobotan. Metode TFIDF merupakan metode pembobotan berupa metode terintegrasi term frequency (TF) dan inverse document frequency (IDF). Metode TFIDF dapat dirumuskan sebagai berikut dengan persamaan (1) dan (2). sebagai berikut :

$$w(t,d)=tf(t,d)*idf \quad (1)$$

$$idf = log\ (N\ /df\ ) \quad (10)$$

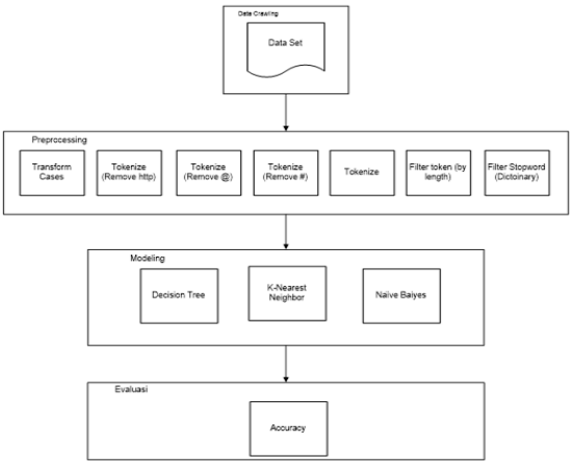
$$idf=log\ (N\ /d) \quad (11)$$

Dimana, tf (t , d) adalah kemunculan kata t pada dokumen d, N adalah jumlah dokumen pada kumpulan dokumen. Dan df = jumlah dokumen yang mengandung term t [10].

3. METODOLOGI PENELITIAN

3.1. Tahapan penelitian

Beberapa proses dilakukan di dalam penelitian ini diantaranya pengambilan data yang berasal dari twitter atau *Crawling Text* dengan *query* "Penjualan Jafra dan Jafra menggunakan alat bantu *rapidminer*, lalu dilanjutkan dengan melakukan *cleansing dataset* yang diantara seperti menghapus data yang sama, menambahkan atribut positif dan negatif, setelah itu mulai pemrosesan data yang berupa komentar dengan mengilangkan tanda baca, atau opini yang dianggap tidak berkaitan dengan objek pelenitian, mengganti huruf besar dengan huruf kecil agar tidak ada eror dalam pemrosesan data, menghilangkan kata https:, dan memfilter komentar dengan bahasa indonesia, setelah itu masuk ke pemodelan algoritma yang ada, penulis menggunakan 3 algoritma pembanding seperti K-Nearest Neighbor, Naive Bayes, dan Decision Tree, lalu setelah itu diproses dan menghasilkan akurasi, tahapan tersebut dapat ditunjukan pada Gambar 2.



Gambar 2. Alur Penelitian

3.2. Pre-Processing

Tujuan pada tahap ini adalah pengolah kata. Ini termasuk sensitivitas huruf, sanitasi, terjemahan, rooting, penghapusan stopword, dan tokenization. Tahap pertama adalah pra-pemrosesan untuk memproses data yang diterima dan klasifikasi untuk menghilangkan data yang bising[11].

3.3. *Case Folding*

Level ini peka terhadap huruf besar atau kecil dan mengubah huruf besar atau huruf besar menjadi huruf kecil. [12].

3.4. *Cleaning*

Tahap keempat adalah pembersihan. Pada fase ini, Anda juga dapat menangani penghapusan karakter tertentu, menghapus URL

atau tautan, dan bahkan menghapus nama pengguna dan tagar

3.5. *Stopword Removal*

Pada tahap ini, proses penyaringan kata-kata yang sering dan jarang terjadi kadang-kadang disebut sebagai stopwords. Metode ini paling baik karena menghilangkan kata-kata yang jarang muncul untuk menjaga efisiensi klasifikasi [13].

3.6. *Tokenisasi*

Tahap keenam adalah tokenization. Proses ini memisahkan atau memotong semua kata dalam sebuah kalimat. Tujuan utama dari pemecahan kalimat adalah untuk membentuk sebuah kalimat [12]. Tokenisasi digunakan untuk mengambil semua istilah unik yang terkandung dalam data sampel dan pada saat yang sama mencoba untuk menyusunnya dengan data input bahasa non-standar (misalnya, oleh KBBI dan EYD[11].

4. HASIL DAN PEMBAHASAN

Pada fase ini, penulis akan menggali informasi dari media sosial, yaitu data tweet yang diperoleh dari Twitter, dengan menggali dan memposting tweet ke komunitas sehubungan dengan analisis sentimen penjualan kosmetik dalam kasus Jafra. Anda perlu mengerti. , Klasifikasi Tweet dilakukan untuk mengidentifikasi tweet positif dan negatif. Negatif. Berdasarkan hal tersebut, algoritma Knearest Neighbor, Decision Tree, dan Naive Bayes digunakan sebagai pendekatan model klasifikasi tweet

4.1. Pengumpulan Data

Proses pengumpulan data ini dimulai dengan web crawling. Dalam proses ini, RapidMiner 9.6 dapat digunakan untuk mengolah data dari Twitter yang berisi banyak komentar opini publik tentang ekonomi pasar hingga 318 tweet

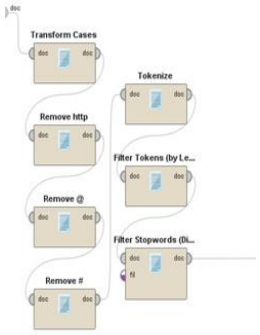
Gambar 3. Proses Pengumpulan Data

4.2 *Data Cleansing*

Proses pembersihan data adalah proses menghapus data dari catatan duplikat. Dari 1000 tweet yang diperoleh dari Twitter, menghapus elemen duplikat, menghapus "RT" di teks, dan diproses, datanya adalah 127 tweet.

4.3 *Pre-Processing*

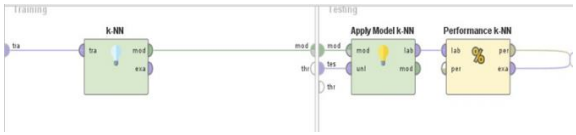
Proses prapemrosesan adalah proses yang menggunakan parameter pembuatan vektor TFIDF dari alat RapidMiner. Proses ini adalah tahap konversi case untuk merubah dari huruf besar ke huruf kecil, proses untuk menghilangkan http untuk menghilangkan data terkait http, proses untuk menghilangkan @ untuk menghilangkan simbol @ pada kalimat, proses untuk menghilangkan karakter teks ## Ini juga melalui proses penghapusan. Masalahnya adalah tokenize ini untuk menghapus angka dan tanda baca dalam teks, menyaring token (berdasarkan panjang) untuk membatasi karakter dalam teks dari minimal 4 karakter hingga maksimal 25 karakter, dan filter Stowords



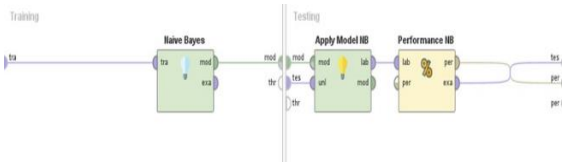
Gambar 4. Proses Pre-Processing Data

4.4 *Cross Validation*

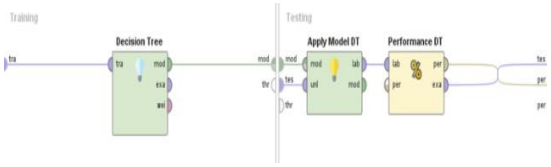
Langkah selanjutnya adalah melakukan validasi Corss menggunakan tiga algoritma yaitu KNN, Decision Tree, dan Naive Bayer. Setiap algoritma melalui proses training dan testing, dan setiap algoritma diberikan cross-validation fold sebesar K = 10. Langkah selanjutnya adalah mendapatkan nilai akurasi, akurasi, dan brand awareness melalui proses modeling. Semua hasil ini berasal dari algoritma Knn, pohon keputusan, dan pemrosesan naive Bayer.



Gambar 5. Proses Training Dan Testing Data Algoritma K-Nearest Neighbors



Gambar 6. Proses Training Dan Testing Pada Algoritma Naïve Bayes



Gambar 7. Proses Training Dan Testing Pada Algoritma Decision Tree

Tabel 1. NILAI ACCURACY, RECALL DAN PRECISION PADA ALGORITMA K-NN

	Akurasi	Presisi	Recall
K-NN	76.22 %	41.67 %	22.39 %

Tabel 2. NILAI ACCURACY, RECALL, DAN PRECISION PADA ALGORITMA NAÏVE BAYES

	Akurasi	Presisi	Recall
NB	74.95%	45.00%	67.16%

Tabel 3. NILAI ACCURACY, RECALL DAN PRECISION PADA ALGORITMA DECISION TREE

	Akurasi	Presisi	Recall
DT	77.85%	46.15%	8.96%

4.5. Perbandingan Hasil Pengujian

Dari hasil evaluasi 3 algoritma diatas, maka didapatkan selisij akurasi, dapat dilihat pada tabel berikut ini :

Tabel 4. PERBANDINGAN HASIL EVALUASI ALGORITMA K-NN, NAÏVE BAYES, DAN DECISION TREE

	K-NN	NB	DT
Akurasi	76.22%	74.92%	77.85%
Presisi	41.67%	45.00%	46.15%
Recall	22.39%	67.16%	8.96%

Metode decision tree lebih unggul dikarenakan memiliki integrasi yang mudah ke dalam sistem basis data sehingga memiliki akurasi yang baik dan dapat menemukan kombinasi data yang tidak terduga. Area keputusan yang sebelumnya kompleks dan sangat global dapat dibuat lebih sederhana dan lebih spesifik, disamping itu dapat menghilangkan perhitungan yang tidak diperlukan. Karena dengan metode ini, sampel hanya diuji berdasarkan kriteria atau kelas tertentu sehingga pemilihan fitur yang fleksibel dari node internal yang berbeda dan fitur yang dipilih membedakan kriteria dari kriteria lain di node yang sama

5. Kesimpulan Dan Saran

Dari penelitian diatas dapat disimpulkan bahwa. Ketiga algoritma dan metodenya, akan diuji dengan dua masukan dengan menggunakan komentar Tokenize and Transform Case yang positif dan negatif , akurasi yang didapat dari algoritma Naïve Bayes yaitu 74.92%, akurasi untuuk metode K-NN yaitu 76.22%, dan akurasi yang didapat dari algoritma Decision Tree yaitu 77.85%. Hasil penelitian menunjukkan bahwa algoritma Decision Tree mendapatkan hasil terbaik dan akurat. Untuk menakikan keakuratan dari penelitian bisa coba dengan menggunakan smote agar bisa memperbesar akurasi dalam penelitian ini.

6. UCAPAN TERIMA KASIH

Ucapan terimakasih kami sampaikan kepada Universitas Nusa Mandiri yang telah memberikan support selama ini, dan juga Pak Windu Gatta yang sudah membimbing untuk menerapkan konsep ke paper ini

Daftar Pustaka :

[1] N. Boudad, R. Faizi, R. Oulad Haj Thami, and R. Chiheb, "Sentiment analysis in Arabic: A review of the literature," *Ain Shams Eng. J.*, vol. 9, no. 4, pp. 2479–2490, 2018, doi: 10.1016/j.asej.2017.04.007.

[2] X. Jdxu, J. Frp, and G. J. Frp, "6hqwlphqw $qdo\vlv ri 6pduw 3krqh 3urgxfw 5hylhz xvlqj 690 &odvvlilfdwlrq 7hfkqltxh," pp. 1469–1474, 2017.

[3] T. Velayutham and P. K. Tiwari, "Bot identification: Helping analysts for right data in twitter," *Proc. - 2017 3rd Int. Conf. Adv. Comput. Commun. Autom. (Fall), ICACCA 2017*, vol. 2018-Janua, pp. 1–5, 2018, doi: 10.1109/ICACCAF.2017.8344722.

[4] R. Y. Hayuningtyas and R. Sari, "Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata Tmii Menggunakan Naïve Bayes Dan Pso," *J. Techno Nusa Mandiri*, vol. 16, no. 1, pp. 37–42, 2019, doi: 10.33480/techno.v16i1.115.

[5] O. Somantri and D. Dairoh, "Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 2, p. 191, 2019, doi: 10.26418/jp.v5i2.32661.

[6] L. U. L. Lupdqwr and D. Dul, "Sentiment Analysis and Topic Modelling for Identification of Government Service Satisfaction," pp. 125–130.

[7] W. E. Nurjanah, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, pp. 1750–1757, 2017.

[8] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.

[9] M. S. Mustafa, M. R. Ramadhan, and A. P. Thenata, "Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Creat. Inf. Technol. J.*, vol. 4, no. 2, p. 151, 2018, doi: 10.24076/citec.2017v4i2.106.

[10] J. Riany, M. Fajar, and M. P. Lukman, "Penerapan Deep Sentiment Analysis pada Angket Penilaian Terbuka Menggunakan K-Nearest Neighbor," *Sisfo*, vol. 06, no. 01, pp. 147–156, 2016, doi: 10.24089/j.sisfo.2016.09.011.

[11] V. I. Santoso, G. Virginia, and Y. Lukito, "Penerapan Sentiment Analysis Pada Hasil Evaluasi Dosen Dengan Metode Support Vector Machine," *J. Transform.*, vol. 14, no. 2, p. 72, 2017, doi: 10.26623/transformatika.v14i2.439.

[12] D. S. Indraloka and B. Santosa, "Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia," *J. Sains dan Seni ITS*, vol. 6, no. 2, pp. 6–11, 2017, doi: 10.12962/j23373520.v6i2.24419.

[13] H. Saif, Y. He, M. Fernandez, and H. Alani, "Contextual semantics for sentiment analysis of Twitter," *Inf. Process. Manag.*, vol. 52, no. 1, pp. 5–19, 2016, doi: 10.1016/j.ipm.2015.01.005.