

ANALISIS SENTIMEN DAN KATEGORISASI REVIEW PELANGGAN PADA CAFE KOPI PASTE DENGAN METODE NAIVE BAYES DAN K-NEAREST NEIGHBOR

SENTIMENT ANALYSIS AND CATEGORIZATION OF CUSTOMER REVIEWS ON KOPI PASTE CAFE USING NAIVE BAYES AND K- NEAREST NEIGHBOR METHODS

Vincentius Riandaru Prasetyo^{1)*}, Ihza Akbar Ryanda²⁾, Delta Ardy Prima³⁾

^{1), 2), 3)} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Surabaya

Jl. Raya Kalirungkut, Kali Rungkut, Kec. Rungkut, Kota Surabaya, Jawa Timur

Email : vincent@staff.ubaya.ac.id¹⁾, s160417112@student.ubaya.ac.id²⁾, delta@staff.ubaya.ac.id³⁾

Abstrak

Minum kopi telah menjadi budaya di kalangan masyarakat Indonesia, budaya minum kopi ini sudah ada sejak zaman penjajahan Belanda. Seiring berkembangnya bangsa Indonesia, penikmat kopi di Indonesia pun semakin meningkat. Untuk memenuhi kebutuhan masyarakat Indonesia, maka lahirlah banyak tempat minum kopi (coffee shop) atau biasa disebut cafe. Karena banyaknya cafe yang ada, maka setiap cafe harus berlomba-lomba meningkatkan kualitasnya agar bisnisnya tetap terus berjalan. Kopi Paste merupakan cafe yang berdiri sejak 2017 di Sidoarjo. Kopi Paste harus tetap dapat bersaing dengan cafe lainnya yang ada di Sidoarjo dengan terus meningkatkan kualitas dari produk, tempat dan pelayanan. Hal inilah yang menjadi latar belakang pembuatan sistem customer relationship management pada cafe Kopi Paste. Tujuan dari pembuatan sistem ini untuk meningkatkan loyalitas pelanggan sehingga dapat meningkatkan penjualan dan dapat terus bersaing dengan cafe lainnya yang ada di Sidoarjo. Metode yang digunakan dalam pembuatan sistem ini menggunakan metode K-nearest neighbor (KNN) dalam melakukan kategorisasi data review yang diberikan oleh pelanggan. Lalu sistem juga menggunakan metode Naive Bayes Classifier untuk melakukan analisis sentimen dari data yang sudah dikategorikan sebelumnya. Data review yang digunakan berasal dari sosial media dan juga dari google review Kopi Paste yang telah diberikan oleh pelanggan sebelumnya. Metode Naive Bayes Classifier berhasil menganalisis sentimen dengan akurasi sebesar 93,33% dan metode K-Nearest Neighbor berhasil melakukan kategorisasi data dengan akurasi sebesar 77%.

Kata kunci: analisis sentimen, kategorisasi, naive bayes, k-nearest neighbor

Abstract

Drinking coffee has been a culture in Indonesia, this culture has existed since the netherland invasion. Coffee enthusiasts in Indonesia grew as the time passed. To fill the needs of coffee in Indonesia, so many coffee shops were born. Because of this, every coffee shop must compete with the other to give the best that they got in order for the business to go. Kopi Paste has been a coffee shop since 2017 and is still going in Sidoarjo. Kopi Paste must still keep competing with the other with improving the quality of the product, place and services. That was the background of making this customer relationship management system. The purpose of this is to improve the customers loyalty in order to improve sales of Kopi Paste and keep competing with the other coffeeshops in Sidoarjo. This system uses K-nearest neighbor(KNN) to categorize the reviews from customers. Naive bayes use for sentiment analysis the data that has been categorized before. Data was from social media and google review of Kopi Paste that customers before. Naive Bayes Classifier did the sentiment analysis with 93,33% accuracy and K-Nearest Neighbor did the categorized data with 77% accuracy.

Keywords: sentiment analysis, categorization, naive bayes, k-nearest neighbor

1. PENDAHULUAN

Budaya minum kopi atau sering disebut ngopi di masyarakat Indonesia sudah tidak asing lagi didengar di masyarakat Indonesia. Berdasarkan buku Ragam Minuman Indonesia kopi pertama kali masuk dan dikenalkan di Indonesia pada tahun 1696. Dan kopi mulai ditanam dan dikembangkan pertama kali di Indonesia terletak di daerah Jakarta Timur yang sekarang sering disebut dengan Pondok Kopi. Sekitar tahun 1699 kopi mulai menyebar ke penjuru nusantara

seperti di Bali, Sumatra, Papua. Sejak saat itulah masyarakat Indonesia mulai mengenal budaya ngopi [1].

Era milenial saat ini budaya ngopi menjadi tren baru yang meningkat di kalangan anak muda. Berdasarkan data Pusat Data dan Sistem Informasi Pertanian Kementerian Pertanian, pertumbuhan minum kopi nasional meningkat 8.22% per tahunnya pada tahun 2016 hingga 2019 [2]. Budaya ngopi seolah menjadi gaya hidup baru bagi mereka yang sebelumnya tidak tertarik pada dunia kopi. Khususnya bagi kalangan mahasiswa kopi adalah teman yang wajib sembari mengerjakan tugas. Demi memenuhi permintaan masyarakat terhadap budaya ngopi, *Coffee Shop* atau sering disebut dengan kafe dituntut untuk memberikan kualitas produk kopi dan tempat yang nyaman untuk mereka yang ngopi sembari mengerjakan tugas ataupun hanya sekedar ngobrol dengan kerabatnya. Setiap kafe akan berlomba-lomba untuk memberikan kualitas terbaiknya agar menarik perhatian pelanggannya yang mencari tempat yang nyaman dan produk yang berkualitas.

Kecenderungan yang dimiliki pelanggan saat ini yang ingin tidak hanya transaksional, beli barang, dapat barang, selesai, tetapi ingin lebih, beli barang, mendapat *service*, dapat barang, mendapat *service*, beli lagi. Hubungan dua arahlah yang diinginkan pelanggan dengan perusahaan yang mampu memberikan pelayanan tambahan tanpa terasa mahal. Loyalitas pelanggan dapat mempengaruhi penjualan dan kelanjutan bisnis itu sendiri. Apabila pelanggan puas dan senang atas suatu produk atau layanan, maka memungkinkan pelanggan tersebut dapat kembali dan kelanjutan bisnis itu di masa depan akan berlanjut. Demi meningkatkan loyalitas pelanggan, pihak pemilik kafe butuh mendengarkan saran dan pendapat dari pelanggannya. Saat ini media yang mungkin dapat menampung pendapat pelanggan adalah media sosial seperti *Twitter*, *Instagram* atau *Facebook*. Akan tetapi akan sangat susah bagi pengelola kafe untuk mendapatkan dan menganalisa satu per satu *review* pelanggan tersebut. Oleh karena itu dibutuhkan sistem yang dapat menganalisis sentimen *review* dari pelanggan dan mampu menunjang keputusan dalam hal operasional, layanan, maupun marketing terhadap *review* yang diberikan oleh pelanggan. Metode yang digunakan dalam melakukan analisis sentiment dan kategorisasi *review* yaitu *Naïve Bayes Classifier* dan *K-Nearest Neighbor* (KNN).

Metode *Naïve Bayes* merupakan metode yang sangat sederhana dan efisien. Selain itu, metode *Naïve Bayes* adalah teknik *machine learning* yang populer untuk klasifikasi teks, dan memiliki performa yang baik pada banyak domain [3]. Sama halnya dengan metode *Naïve Bayes*, metode KNN adalah juga merupakan salah satu metode sering digunakan untuk klasifikasi teks dan data. Pada metode ini dilakukan klasifikasi terhadap objek berdasarkan data yang jaraknya paling dekat dengan objek tersebut [4]. Pada penelitian ini, metode *Naïve Bayes Classifier* akan digunakan untuk melakukan analisis sentiment *review*, sedangkan untuk kategorisasi *review* akan menggunakan metode KNN. Metode *Naïve Bayes* dipilih untuk melakukan analisis sentiment karena berdasarkan penelitian yang dilakukan oleh Gunawan, et al [5], metode *Naïve Bayes* memiliki akurasi yang lebih baik dibandingkan metode lainnya. Untuk kategorisasi *review* pelanggan, penelitian ini menerapkan metode KNN karena metode ini memiliki akurasi yang baik untuk melakukan kategorisasi teks [6].

Penelitian terkait analisis sentimen terhadap *review* pelanggan di suatu kafe atau restoran pernah dilakukan juga sebelumnya. Laksono, et al [7] bertujuan untuk menganalisis ulasan pelanggan restoran dari *TripAdvisor* online di sepuluh restoran terbaik Surabaya. Tujuan lain adalah untuk menemukan metode terbaik untuk menganalisis data ulasan pelanggan restoran dengan membandingkan metode *Naïve Bayes* dan analisis sentimen *TextBlob* karena keduanya memiliki perbedaan metode yang mendasar dalam hal perhitungan. Hasil penelitian tersebut menunjukkan bahwa metode *Naïve Bayes* memiliki nilai akurasi 72,06% dan sedikit lebih baik (2,94%) dibandingkan analisis sentimen menggunakan *TextBlob*.

Penelitian lainnya yang dilakukan oleh Sharif, et al [8] mengembangkan sistem otomatis untuk analisis sentimen teks bahasa Bengali untuk ulasan restoran untuk mengklasifikasikan opini menjadi dua kelas: sentimen positif atau negatif. Tiga algoritme *machine learning* seperti *decision*

tree, *random forest*, dan *Multinomial Naïve Bayes* diimplementasikan untuk mengklasifikasikan sentimen dari teks ulasan bahasa Bengali. Penelitian tersebut menguji sistem yang dikembangkan dengan 1000 teks ulasan restoran yang ditulis dalam bahasa Bengali. Hasil percobaan menunjukkan bahwa sistem yang dikembangkan dapat mengklasifikasikan ulasan restoran dengan akurasi 80,48% menggunakan metode *Multinomial Naïve Bayes*. Percobaan lainnya juga dilakukan dengan membandingkan metode *Multinomial Naïve Bayes* dengan metode lainnya. Hasil percobaan menghasilkan kesimpulan bahwa metode *Multinomial Naïve Bayes* memiliki akurasi lebih tinggi dibandingkan dengan metode *Random Forest* dan *Decision Tree* [8].

Selain melakukan analisis sentiment, Zahoor, et al [9] pada penelitiannya juga melakukan kategorisasi teks terhadap ulasan pelanggan terhadap suatu restoran. Penelitian tersebut berusaha menganalisis ulasan pelanggan tentang berbagai restoran di seluruh Karachi – salah satu kota terbesar di Pakistan. Pada penelitian ini, ulasan pelanggan dikumpulkan dari komunitas Facebook yang sangat populer - panduan SWOT untuk restoran Karachi. Penelitian ini mengklasifikasikan setiap komentar sebagai komentar positif atau negatif. Selain itu, dengan menggunakan teknik kategorisasi teks, komentar secara otomatis diklasifikasikan menurut *feedback* tentang rasa makanan, suasana, layanan, dan harga. Beberapa algoritma digunakan untuk klasifikasi, termasuk *Naïve Bayes Classifier*, *Logistic Regression*, *Support Vector Machine (SVM)*, dan *Random Forest*. Perbandingan kinerja dari algoritma-algoritma tersebut didapatkan hasil terbaik, yaitu akurasi 95%, dicapai dengan menggunakan algoritma *random forest*.

2. DASAR TEORI

2.1 Analisis Sentimen

Analisis sentimen adalah tugas pengelompokan atau klasifikasi teks yang tidak terstruktur yang sering berisi pendapat orang tentang topik tertentu dengan tujuan untuk lebih memahami pendapat orang tentang topik tersebut. Misalnya, sebuah organisasi melakukan analisis sentimen pada halaman media sosial mereka (mis., Twitter, Facebook) untuk memahami bagaimana masyarakat umum memandang organisasi dan produk/layanannya. Berbeda dengan sistem penilaian di platform belanja online umum seperti Amazon, Shopee dan Tokopedia, orang mengungkapkan pendapat mereka di media sosial sebagai teks terbuka kualitatif. Ada kalanya organisasi perlu memahami polarisasi pendapat (misalnya, apakah seseorang mengungkapkan pengalaman negatif atau positif). Dalam memahami polarisasi ribuan, jutaan atau lebih pendapat memakan waktu dan dengan demikian, sejumlah besar sumber daya diperlukan jika tugas ini akan dilakukan secara manual. Oleh karena itu, analisis sentimen telah menjadi fokus banyak penelitian dalam beberapa tahun terakhir untuk secara otomatis mengelompokkan atau mengklasifikasikan banyak pendapat tentang topik tertentu [10].

Dalam hal jenis tugas, analisis sentimen biasanya digunakan untuk mengklasifikasikan dan/atau mengelompokkan kumpulan teks yang tidak terstruktur. Tugas klasifikasi berfokus pada pengelompokan sekumpulan teks (misalnya opini) ke dalam himpunan bagian dari kelompok yang telah ditentukan sebelumnya (sering diberi label sebagai kelas). Sebaliknya, tugas pengelompokan berfokus pada pengelompokan sekumpulan teks berdasarkan kesamaan atau pola di dalam teks tanpa kelompok yang telah ditentukan sebelumnya [10].

2.2 Preprocessing

Preprocessing bertujuan untuk mendapatkan data yang bersih guna meningkatkan akurasi proses pendeteksian [11]. Langkah pertama dalam *preprocessing* adalah menghapus simbol-simbol yang sering muncul pada suatu komentar, seperti #, @, *emoticon*, dan tautan situs web. Hal ini dikarenakan simbol memiliki sedikit pengaruh pada proses klasifikasi. Tahap selanjutnya adalah *case folding*. Tahap ini mengubah semua huruf dalam satu kata menjadi huruf kecil. Hal ini dilakukan untuk menghindari penggunaan kata yang memiliki arti yang sama, seperti “*EnAk*” dan “*enak*”.

Stemming adalah langkah selanjutnya dalam *preprocessing*. Ini merupakan proses penentuan akar kata yang akan diimplementasikan menggunakan algoritma Nazief Andriani yang akan digunakan pada penelitian ini dengan *library Sastrawi 1.0.1*. Algoritma Nazief Andriani dikembangkan berdasarkan aturan morfologi Indonesia yang luas, yang dikumpulkan ke dalam

satu kelompok dan dienkapsulasi menjadi afiks yang diperbolehkan dan afiks yang tidak diperbolehkan. Algoritma ini menggunakan kamus kata dasar dan mendukung pencatatan, yaitu penyusunan kembali kata-kata yang memiliki proses stemming yang berlebihan [11]. Wahyudi, et.al [12] dalam penelitiannya menyimpulkan bahwa algoritma Nazief Andriani memberikan hasil yang lebih baik daripada algoritma Porter untuk mendukung sistem pencarian informasi. Langkah selanjutnya setelah *stemming* adalah *stop word removal*. Langkah ini digunakan untuk menghilangkan kata-kata yang tidak berpengaruh pada proses kategorisasi, seperti: *yang, dan, atau, ke, dari*, dan sebagainya. Tokenisasi adalah langkah terakhir dalam proses *preprocessing*. Tokenisasi adalah proses membagi kalimat menjadi kata-kata. Hasil tokenisasi akan disimpan sebagai data bersih.

2.3 Ekstraksi Fitur

Ekstraksi fitur bertujuan untuk menghilangkan duplikasi kata yang terdapat pada hasil *preprocessing* sehingga akan diperoleh sekumpulan kata yang unik. Ekstraksi fitur akan dihitung menggunakan metode TF-IDF (*Term Frequency and Inverse Document Frequency*). Metode TF-IDF dihitung berdasarkan persamaan (1) [13].

$$tf.idf_{t,d} = tf_{t,d} \times \log \frac{N}{df_t} \tag{1}$$

Nilai $tf_{t,d}$ adalah jumlah kemunculan dari term t dalam dokumen d . Nilai N adalah jumlah data yang digunakan, sedangkan df_t adalah jumlah dokumen yang mengandung term t .

2.4 Naïve Bayes

Naïve Bayes Classifier adalah metode klasifikasi berdasarkan teorema Bayes. *Naïve Bayes Classifier* dikenal lebih baik dari beberapa metode klasifikasi lainnya karena ciri utama *Naïve Bayes* adalah asumsi independensi yang sangat kuat dari setiap kondisi atau peristiwa. Selain itu, modelnya yang sederhana dan mudah dibuat. Alasan terakhir, model dapat diimplementasikan untuk set data yang besar [14]. Perhitungan *Naïve Bayes* dilakukan berdasarkan persamaan (2).

$$P(c | x) = \frac{P(x|c) \cdot P(c)}{P(x)} \tag{2}$$

Di mana $P(c/x)$ adalah probabilitas posterior kelas (target) yang diberikan prediktor (atribut), sedangkan $P(c)$ adalah probabilitas prior suatu kelas. Selanjutnya, $P(x/c)$ adalah *likelihood* yang merupakan probabilitas prediktor yang diberikan suatu kelas dan $P(x)$ adalah probabilitas prior suatu prediktor.

2.5 K-Nearest Neighbor

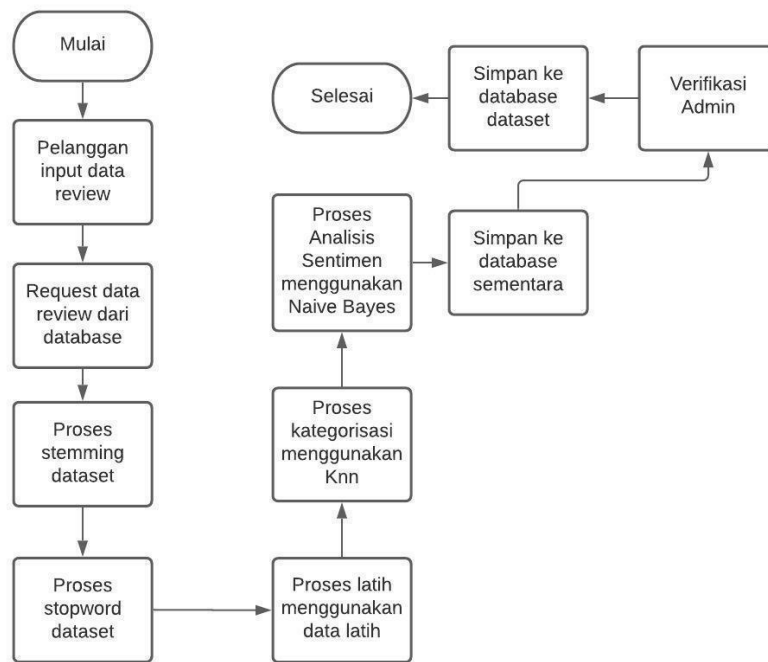
Metode *K-Nearest Neighbor* (KNN) adalah metode klasifikasi yang menghitung jarak terpendek antara data *training* dan data uji. Keakuratan metode KNN ditentukan oleh fitur-fitur yang dimilikinya. Keakuratan metode KNN menurun seiring dengan berkurangnya jumlah fitur yang relevan. Metode KNN cocok untuk data *training* yang besar dan memiliki *noise* yang banyak. Jarak antara sampel masukan dan data *training* dapat dihitung dengan menggunakan berbagai metode, termasuk *Euclidean, Manhattan, dan Cosine Similarity* [11].

Pada penelitian ini, jarak antara data *training* dan data uji dihitung dengan menggunakan metode *cosine similarity* karena memiliki tingkat akurasi yang baik. Metode *cosine similarity* dihitung menggunakan persamaan (3), di mana X dan Y merupakan nilai TF-IDF dari setiap term yang ada pada teks X dan Y . Pada metode *cosine similarity*, apabila nilai yang dihasilkan semakin mendekati 1, maka similaritasnya semakin tinggi. Sebaliknya, apabila nilai yang dihasilkan semakin mendekati 0, maka semakin tidak mirip [15].

$$Sim_{x,y} = \frac{\sum_{i=1}^n x_i \times y_i}{\sqrt{\sum_{i=1}^n (x_i)^2 \times \sum_{i=1}^n (y_i)^2}} \tag{3}$$

3. METODOLOGI PENELITIAN

Sistem yang dikembangkan pada penelitian ini akan mengklasifikasikan *review* dari pelanggan kafe kopi paste ke dalam kelas positif atau negatif. Hal ini bermanfaat bagi pemilik kafe kopi paste untuk mengorganisir *review* pelanggan agar lebih mudah dalam menanggapi *review* dari pelanggan. Tidak hanya analisis sentimen, sistem juga mengkategorikan *review* secara otomatis. Kategori akan terbagi menjadi tiga yaitu: produk, tempat dan pelayanan. *Dataset* yang digunakan adalah data yang telah disiapkan sebelumnya yang didapatkan dari *google review*. Teknik pengumpulan *dataset* dilakukan dengan teknik *crawling*. *Crawling* merupakan suatu teknik akuisisi data dari internet dengan cara mengekstrak struktur HTML dari suatu situs web. Teknik ini digunakan untuk memperoleh informasi yang ditampilkan pada halaman website dengan cepat [16]. Tahapan dari analisis sentimen dan kategorisasi *review* dapat dilihat pada gambar 1.



Gambar 1. Tahapan Proses Analisis Sentimen Dan Kategorisasi Review

Sebelum dilakukan analisis sentimen, sistem akan melakukan kategorisasi data yang akan dimulai dengan *preprocessing*, dengan melakukan *stemming* dan *stopword removal*. Setelah itu akan dilakukan ekstraksi fitur dengan metode *tf-idf*. Selanjutnya, sistem akan melakukan proses kategorisasi data menggunakan metode *K-Nearest Neighbour*. Setelah kategorisasi dilakukan, sistem akan melakukan proses analisis sentimen menggunakan metode *Naïve Bayes* untuk data *review* tersebut. Hasil dari analisis sentimen dan kategorisasi data akan disimpan dalam *database* sementara karena dibutuhkan verifikasi secara langsung dari admin sistem untuk memastikan kategori dan *review* sudah sesuai. Admin dapat merubah kategori dan sentimen yang dilakukan oleh sistem secara manual untuk memastikan bahwa kategori dan sentimen tersebut sudah benar. Apabila admin telah melakukan verifikasi, maka data *review* ditambahkan pada *database dataset* yang selanjutnya akan digunakan untuk *data training* ke depannya.

4. PENGUJIAN DAN PEMBAHASAN

Pada penelitian ini, pengujian dilakukan dengan dua cara, yaitu menguji akurasi dari model klasifikasi metode KNN dan *Naïve Bayes*, serta validasi kepada pemilik kafe Kopi Paste. Pengujian akurasi model klasifikasi dilakukan dengan menggunakan metode *k-fold cross validation*. Validasi kepada pemilik Kopi Paste dilakukan dengan wawancara setelah pemilik Kopi Paste selesai mencoba sistem yang telah dibuat.

Pada pengujian akurasi model klasifikasi dilakukan dengan metode *k-fold cross validation* dan menggunakan 100 *dataset* dengan parameter $k = 5$. Parameter k disini adalah jumlah perulangan yang akan dilakukan untuk mendapatkan nilai akurasi dari setiap kelompok *data testing*. Hal ini berarti terdapat 5 kali iterasi dengan 10 data yang akan menjadi *datatest* tiap iterasinya. Hasil akurasi dari tiap perulangan ini akan dirata-rata untuk menentukan hasil akurasi akhir sistem ini. Tabel 1 menunjukkan hasil akurasi dari tiap *fold* dan rata-rata akurasi untuk pengujian model klasifikasi metode *naive bayes* dan *k-nearest neighbor*.

Tabel 1. Pengujian Akurasi Model Klasifikasi Dengan Seluruh Dataset

Nilai fold	Akurasi	
	Naïve Bayes	KNN
1	100%	85%
2	83,33%	80%
3	100%	60%
4	83,33%	80%
5	100%	80%
Rata-rata	93,33%	77%

Berdasarkan hasil pengujian yang ditunjukkan pada tabel 1, memperlihatkan bahwa metode *Naive bayes* memiliki performa yang sangat baik dengan rata-rata akurasi sebesar 93,33% dengan menggunakan keseluruhan *dataset* pada sistem. Metode KNN memiliki rata-rata akurasi yaitu 77%. Nilai rata-rata ini terbilang cukup baik mengingat metode ini melakukan kategorisasi data yang memiliki 3 macam *class* yang berbeda. Pengujian akurasi model klasifikasi lainnya juga dilakukan dengan mengurangi data *training* yang digunakan pada pembentukan model klasifikasi. Hal ini dilakukan untuk melihat pengaruh jumlah data yang digunakan saat proses pembentukan model klasifikasi. Parameter k yang digunakan pada metode *k-fold cross validation* adalah 4, dengan data *training* setiap *fold*-nya adalah 50 dan data ujinya sebanyak 50 data. Tabel 2 menunjukkan hasil akurasi dari tiap *fold* dan rata-rata akurasi untuk pengujian kedua model klasifikasi metode *naive bayes* dan *k-nearest neighbor*.

Tabel 2. Pengujian Akurasi Model Klasifikasi Dengan Sebagian Dataset

Nilai fold	Akurasi	
	Naïve Bayes	KNN
1	80%	68%
2	88%	80%
3	76%	68%
4	84%	84%
Rata-rata	82%	75%

Berdasarkan hasil pengujian akurasi yang ditunjukkan pada tabel 2, dapat dilihat bahwa akurasi model klasifikasi dari metode *naive bayes* dan *k-nearest neighbor* mengalami penurunan dari sebelumnya. Hal ini membuktikan bahwa banyaknya data yang digunakan saat proses *training*, berpengaruh terhadap akurasi model yang dihasilkan. Pengujian selanjutnya dilakukan melalui validasi dengan pemilik Kopi Paste. Validasi tersebut dilakukan untuk mengetahui tingkat kesulitan dalam menggunakan aplikasi, desain sistem, dan fungsi sistem.

Berdasarkan validasi yang dilakukan, pemilik Kopi Paste menyatakan bahwa sistem yang dibuat mudah digunakan dan cukup familiar dengan tombol-tombol dan peletakkan objek lainnya. Selain itu, sistem juga dapat membantu dalam memudahkan untuk melakukan *filter review* negatif

atau positif, beserta kategorinya. Hal ini dapat membantu pihak pengelola Kopi Paste untuk fokus pada *review* negatif untuk segera ditindak lanjuti, sehingga lebih mudah untuk menganalisis data untuk kemajuan Kopi Paste kedepannya. Pemilik Kopi Paste juga menyatakan bahwa desain pada sistem terlihat menarik, rapi, dan elegan. Secara keseluruhan, pemilik Kopi Paste puas dengan sistem yang telah dibuat, terlebih lagi Kopi Paste memiliki *website* sebagai media promosi dan dapat meningkatkan kredibilitas Kopi Paste itu sendiri.

5. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan sebelumnya, dapat ditarik kesimpulan bahwa metode klasifikasi yang digunakan sudah cukup baik dalam melakukan tugasnya yaitu *naive bayes* untuk analisis sentimen dengan rata-rata akurasi sebesar 93,33% dan *k-nearest neighbor* untuk kategorisasi data dengan rata-rata akurasi sebesar 77%. Selain itu, secara keseluruhan pemilik Kopi Paste mengaku puas dengan sistem yang telah dikembangkan.

Saran untuk pengembangan ke depan, dapat menambahkan dataset yang berasal dari media sosial seperti Twitter, Facebook, dan Instagram untuk memperkaya isi dari dataset, sehingga dapat meningkatkan akurasi dari model klasifikasi. Selain itu, sistem dapat dikembangkan agar tampilan saat dibuka menggunakan perangkat mobile atau smartphone dapat terlihat lebih menarik.

DAFTAR PUSTAKA

- [1] I. Fathoni, " Analisis Daya Tarik Budidaya dan Pengolahan Kopi untuk Pengembangan Agrowisata di Desa Babadan Banjarnegara," Jurnal Kawistara, vol. 10, no. 3, pp.310-327, 2020.
- [2] D. S. T. Manalu, V. B. Sebayang, Suharno, W. H. Situmeang and F. Adrian, "Business Development Strategy Of Pt Bogor Kopi Indonesia In Bogor Regency, West Java," Journal of Integrated Agribusiness, vol. 5, no. 1, pp. 28-36, 2023.
- [3] B. Ning, W. Junwei and H. Feng, "Spam Message Classification Based on the Naïve Bayes Classification Algorithm," IAENG International Journal of Computer Science, vol. 46, no. 1, 2019.
- [4] M. Syarifuddin, "Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naive Bayes Dan KNN," Inti Nusa Mandiri, vol. 15, no. 1, pp. 23-28, 2020.
- [5] D. Gunawan, D. Riana, D. Ardiansyah, F. Akbar and S. Alfarizi, "Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023," Jurnal Teknik Komputer AMIK BSI, vol. VI, no. 1, pp. 121-129, 2020.
- [6] D. Utami, Kusri and Sudarmawan, "Analisis Metode Wordtovec Pada Performa Klasifikasi Kategori Berita Menggunakan Metode SVM dan KNN," Jurnal Explore STMIK Mataram, vol. 11, no. 2, pp. 82-85, 2021.
- [7] R. A. Laksono, K. R. Sungkono, R. Sarno and C. S. Wahyuni, "Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes," in 12th International Conference on Information & Communication Technology and System (ICTS), Surabaya, 2019.
- [8] O. Sharif, M. M. Hoque and E. Hossain, "Sentiment Analysis of Bengali Texts on Online Restaurant Reviews Using Multinomial Naïve Bayes," in 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, 2019.
- [9] K. Zahoor, N. Z. Bawany and S. Hamid, "Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning," in 21st International Arab Conference on Information Technology (ACIT), Giza, 2020.
- [10] Jimmy and V. R. Prasetyo, "Sentiment Analysis on Feedback of Higher Education Teaching Conduct An Empirical Evaluation of Methods," in International Conference on Informatics, Technology, and Engineering 2021 (InCITE 2021), Surabaya, 2021.
- [11] V. R. Prasetyo and A. H. Samudra, "Hate Speech Content Detection System on Twitter using KNearest Neighbor Method," in International Conference on Informatics,

- Technology, and Engineering 2021 (InCITE 2021), Surabaya, 2021.
- [12] D. Wahyudi, T. Susyanto and D. Nugroho, "Implementasi dan Analisis Algoritma Stemming Nazief & Adriani dan Porter Pada Dokumen Berbahasa Indoensia," *Jurnal Ilmiah SINUS*, vol. 15, no. 2, pp. 49-56, 2017.
- [13] V. R. Prasetyo, "Searching Cheapest Product On Three Different ECommerce Using K-Means Algorithm," in *2018 International Seminar on Intelligent Technology and Its Application*, Bali, 2018.
- [14] N. Salmi and Z. Rustam, "Naïve Bayes Classifier Models for Predicting the Colon Cancer," in *9th Annual Basic Science International Conference 2019 (BaSIC 2019)*, Malang, 2019.
- [15] V. R. Prasetyo, M. Widiastri and M. M. Angkiriwang, "Sistem Berbasis Web Untuk Koreksi Soal Esai Dengan Association Rules," *Jurnal TEKNIKA*, vol. 11, no. 1, pp. 62-68, 2022.
- [16] V. R. Prasetyo, M. Mercifia, A. Averina, L. Sunyoto and Budiarjo, "Prediksi Rating Film Pada Website Imdb Menggunakan Metode Neural Network," *Jurnal Ilmiah NERO*, vol. 7, no. 1, pp. 1-8, 2022.