

Analisis Sentimen Masyarakat dalam Pembangunan City Walk Kota Tegal di Media Sosial Facebook Menggunakan Algoritma Naïve Bayes dan SVM

Ali Sofyan^{*1}, Ria Indah Fitria², Fatiah Isralestina³

^{1,2,3}Program Studi Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Pancasakti Tegal

Email: ^{*1}alisofyan@upstegal.ac.id, ²riaindah@upstegal.ac.id, ³fatiah.isralestina@gmail.com

(Naskah masuk: 30 Agustus 2023, diterima untuk diterbitkan: 30 Desember 2023)

Abstrak: Kota Tegal sebagai kota perdagangan dan industri tentunya berdampak pada peningkatan ekonomi masyarakat dan pemerintah daerahnya, sehingga berdampak juga pada peningkatan pembangunan. Seperti contohnya pembangunan City Walk di Kota Tegal banyak pendapat masyarakat yang disampaikan di media sosial. Kebebasan menyampaikan pendapat tentunya akan memunculkan beragam komentar. Penelitian ini dilakukan untuk memberikan informasi tentang akurasi model klasifikasi dengan menggunakan algoritma klasifikasi dengan menggunakan data persepsi masyarakat terhadap pembangunan City Walk Kota Tegal. Penelitian ini juga dapat digunakan sebagai masukan untuk pengambilan keputusan dalam pengembangan proyek pembangunan di Kota Tegal. Penelitian ini adalah penelitian data teks atau bisa disebut penelitian teks mining dimana teks mining adalah proses ekstraksi informasi dari kumpulan data teks. Pada penelitian ini data teks yang digunakan berasal dari pendapat masyarakat yang disampaikan melalui media sosial facebook. Penelitian ini bertujuan untuk mencari akurasi terbaik dari model algoritma klasifikasi dengan mengklasifikasikan pendapat yang disampaikan masyarakat menjadi pendapat positif, negatif. Metode yang digunakan Naïve Bayes dan SVM untuk mengklasifikasikan komentar pembangunan City Walk di kota Tegal. Setiap komentar akan dilabeli sebagai komentar positif dan negatif. Dari hasil uji dengan model Naïve Bayes didapatkan akurasi tertinggi 70.19% dengan sufled sampling. Sedangkan di metode SVM mendapatkan akurasi tertinggi 80.77% di stratified sampling.

Kata Kunci – Sentimen Analisis; Teks Mining; City Walk Kota Tegal

Community Sentiment Analysis In Development City Walk Tegal City on Facebook Social Media Using Naïve Bayes and SVM Algorithm

Abstract: Tegal City as a trade and industrial city certainly has an impact on improving the economy of the community and local government, so that it also has an impact on increasing development. For example, the construction of the City Walk in Tegal City has expressed many public opinions on social media. Freedom to express opinions will of course give rise to various comments. This research was conducted to provide information about the accuracy of the classification model using a classification algorithm using data on public perception of the City Walk development in Tegal City. This research can also be used as input for decision making in developing development projects in Tegal City. This research is text data research or can be called text mining research where text mining is the process of extracting information from a collection of text data. In this research, the text data used comes from public opinion conveyed via the social media Facebook. This research aims to find the best accuracy of the classification algorithm model by classifying opinions expressed by the public into positive and negative opinions. The method used by Naïve Bayes and SVM to classify comments on the City Walk development in the city of Tegal. Each comment will be labeled as positive and negative. From the test results with the Naïve Bayes model, the highest accuracy was 70.19% with suffled sampling. Meanwhile, the SVM method got the highest accuracy of 80.77% in stratified sampling.

Keywords – Sentiment Analysis; Text Mining; City Walk Tegal City

1. PENDAHULUAN

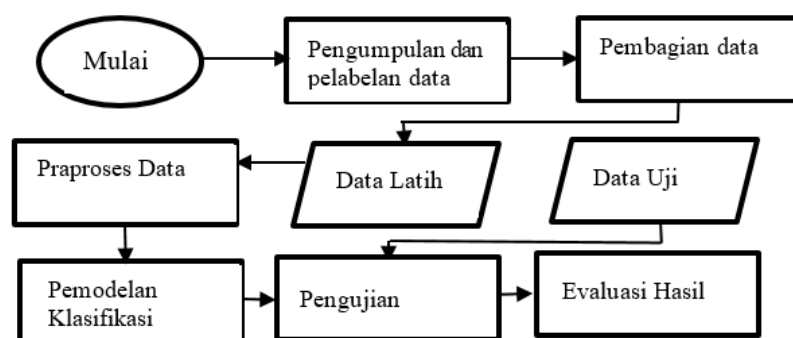
Kota Tegal adalah kota yang memiliki luas yang sempit dibandingkan dengan kota-kota lain di Indonesia. Perdagangan dan jasa merupakan sektor utama perekonomian kota Tegal. Kota ini menjadi tempat pengolahan akhir dan pemasaran berbagai produk dari kawasan Jawa Tengah bagian barat. Usaha kecil dan menengah yang cukup pesat kemajuannya adalah industri logam rumahan di kawasan jalan Cempaka, dan kerajinan batik Tegal di kelurahan Kalinyamat. Selain itu Kota Tegal juga memiliki banyak pusat perbelanjaan seperti mall yang berdiri di beberapa titik pusat ekonomi Kota Tegal. Kota Tegal juga memiliki banyak obyek wisata, seperti wisata pantai, kuliner, dan beberapa spot menarik di pusat Kota Tegal. Seperti yang terbaru yaitu pembangunan kawasan jalan Ahmad Yani Kota Tegal atau pembangunan Malioboro Kota Tegal yang sekarang menjadi nama City Walk Kota Tegal. Pembangunan City Walk yang merubah bentuk jalan menjadi lebih sempit dan menambah lebar trotoar pengguna jalan. Selain itu merubah arus kendaraan di Jalan Ahmad Yani menjadi satu arah[1].

Perubahan yang dilakukan pada kondisi jalan Ahmad Yani memunculkan berbagai pendapat, komentar para pengguna dan para pelaku usaha yang ada di Jalan Ahmad Yani tersebut. Selain penyampaian protes saat pembangunan juga banyak komentar yang disampaikan di media sosial. Salah satunya media sosial facebook, dimana masyarakat yang melihat pembangunan menyampaikan pendapat yang beragam. Beberapa pendapat yang disampaikan ada yang menyatakan mendukung ada juga yang menyatakan tidak mendukung. Ragamnya pendapat yang disampaikan oleh masyarakat bias menjadi masukan kepada instansi pemerintah dalam pembangunan Kota Tegal ke depan. Agar dapat mengetahui pendapat yang pro atau mendukung dan kontra atau tidak mendukung maka diperlukan sentiment analysis terhadap komentar yang ada di media sosial[2].

Analisis sentimen berfungsi untuk menemukan pola pada kalimat lalu mengelompokkan ke dalam sentimen positif atau negatif. Hasil dari sentimen positif merupakan reaksi pro sedangkan hasil dari sentimen negatif merupakan reaksi kontra. Beberapa penelitian sebelumnya pernah melakukan penelitian mengenai analisis sentimen. Oman Somantri & Dairoh, "Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining[7]. Pada penelitian membahas tentang sentimen analisis pengunjung wisata di Kota Tegal. Data yang diambil bersumber pada testimoni yang disampaikan di google map. Pengujian dilakukan dengan Naïve Bayes dan Decision Tree didapatkan akurasi tertinggi dengan nilai tertinggi 77.50% pada Naïve Bayes sedangkan pada Decision Tree didapatkan akurasi 60.83%. selain itu pada penelitian "Support Vector Machine Berbasis Feature Selection Untuk Sentiment Analysis Kepuasan Pelanggan Terhadap Pelayanan Warung dan Restoran Kuliner Kota Tegal"[6][10]. Penelitian ini mengukur kepuasan pelanggan terhadap layanannya. Data diambil di situs tripadvisor. Hasil akurasi dengan menggunakan SVM 72.45%.

2. METODE PENELITIAN

2.1. Tahapan Penelitian



Gambar 1. Tahapan Penelitian

Data diambil dari group facebook masyarakat, dengan menggunakan proses pencarian di Facebook. Pencarian menggunakan kata kunci “City Walk Kota Tegal” dan “Malioboro Kota Tegal” Proses pelabelan data dilakukan dengan cara mengambil 100 yang kemudian dilakukan praprosesing data. Kemudian peneliti memberikan label pada setiap komentar. Label yang diberikan ada 2 kelas, yaitu kelas positif, kelas negatif. Kelas positif merupakan kelas yang berisikan data yang mengandung kata bermakna positif, pernyataan setuju, dan dukungan. Kelas negatif merupakan kelas dengan data yang mengandung kata bermakna negatif, ejekan, dan kontra.

2.2. *Praproses Data*

Data yang terkumpul belum bisa diolah secara langsung. Biasanya data perlu melalui proses penyempurnaan data sebelum digunakan. Hal ini bertujuan untuk meningkatkan kualitas dari perhitungan akurasi dan juga menghindari banyak data yang tidak berguna[8].

1. Pembersihan Data

Proses ini adalah menghilangkan data data yang tidak penting dan tidak berpengaruh, iasanya berupa karakter-karakter tanda baca yang memang bukan hal yang diperlukan untuk perhitungan akurasi.

2. Normalisasi Data

Merubah kata kata yang tidak dengan digantungan dengan kata yang baku. Sehingga mendapatkan makna asli dari kata tersebut.

3. Stemming Data

Proses menghilangkan imbuhan baik yang berada di awal kata ataupun yang berada di akhir

4. Penghapusan Stopword

Tahap stopwords adalah tahapan untuk menghilangkan teks yang tidak penting atau kurang bermakna contoh kata sambung. Stopword merupakan kata-kata dengan frekuensi kemunculan yang tinggi Tala (2003). Stoplist yang digunakan berasal dari analisis kata dasar yang dilakukan oleh Tala (2003).

2.3. *Evaluasi*

Evaluasi dengan akurasi (accuracy)[9] adalah salah satu metrik yang umum digunakan untuk mengukur sejauh mana model klasifikasi dapat mengklasifikasikan data dengan benar. Adapun rumus yang digunakan adalah:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \times 100 \quad (1)$$

Selain itu juga menggunakan validasi k-Fold Cross Validation, dimana kita bisa mengukur tingkat akurasi berdasarkan perbedaan pembagian data sebanyak subset yang kita tentukan nilai k-Fold nya

3. HASIL DAN PEMBAHASAN

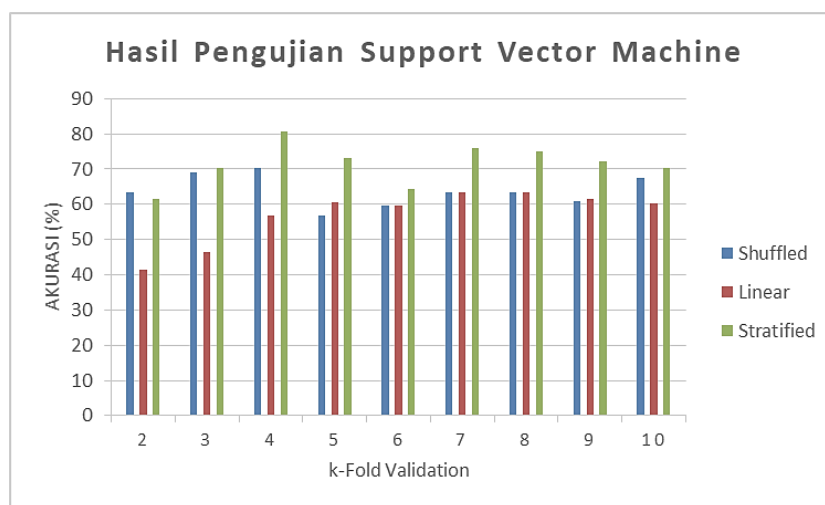
Pengujian data ini dilakukan dengan menggunakan perangkat lunak RapidMiner yang sudah banyak digunakan peneliti sebelumnya. Adapun spesifikasi komputer yang digunakan peneliti yaitu dengan menggunakan processor i7 dan RAM 16GB dan operating sistwm Wndows 10 Professional.

3.1. *Hasil Pengujian dengan Algoritma Support Vector Machine (SVM)*

Pada pengujian dengan SVM digunakan k-Fold dengan range 2 sampai dengan 10 dengan SMOTE Upsampling bisa didapatkan hasil sebagai berikut :

Tabel 1. Pengujian dengan metode Support Vector Machine

TABEL PENGUJIAN SUPPORT VECTOR MACHINE			
k-Fold	Metode Sampling		
	Shuffled	Linear	Stratified
2	63.46%	41.35%	61.54%
3	69.19%	46.36%	70.20%
4	70.19%	56.73%	80.77%
5	56.81%	60.57%	73.24%
6	59.64%	59.64%	64.32%
7	63.33%	63.47%	75.99%
8	63.46%	63.46%	75.00%
9	61.03%	61.62%	72.31%
10	67.36%	60.36%	70.45%



Gambar 2. Grafik Pengujian dengan Support Vector Machine

Dari tabel pengujian diatas maka dapat dilihat bahwa, nilai akurasi tertinggi pada tiap k-Fold pengujian mendapatkan hasil yang berbeda beda pada tiap sampling nya. Bisa dilihat ditabel dan gambar diatas bahwasanya nilai akurasi tertinggi untuk Shuffled Sampling didapatkan pada k-Fold=3 dengan akurasi 69.19%. Sedangkan pada Linear Sampling didapatkan akurasi tertinggi di k-Fold=7 dengan akurasi 63,47%. Di Stratifiel Sampling dengan akurasi tertinggi dengan nilai akurasi 80,77% berada di k-Fold=4.

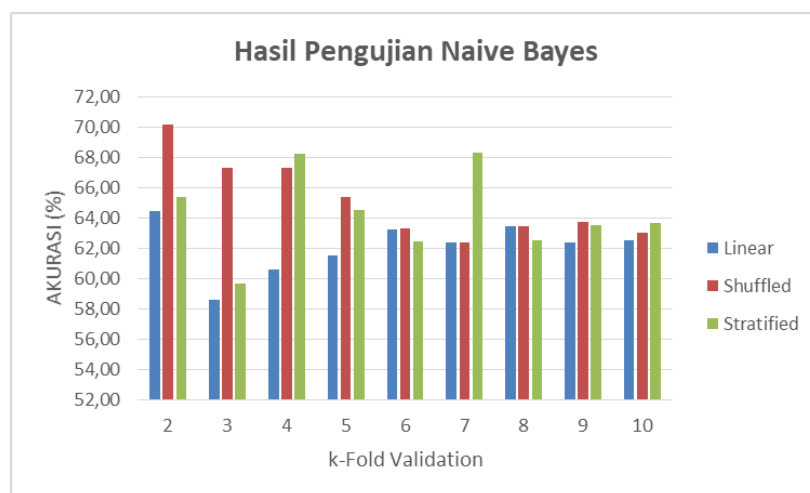
3.2. Hasil Pengujian dengan Algoritma Naïve Bayes

Dengan melakukan pengujian yang sama dengan metode yang berbeda yaitu dengan Naïve Bayes hasil yang didapatkan bisa dilihat pada tabel berikut:

Untuk pengujian dengan metode Naïve Bayes didapatkan data data seperti tabel dan gambar 2 diatas. Dimana untuk pengujian dengan Linier Sampling nilai tertinggi akurasi dapatkan pada k-Fold=2 dengan nilai akurasi 64.42% sedangkan untuk pengujian dengan Sufled Sampling didapatkan akurasi tertinggi 70.19% pada k-Fold=2 juga seperti Linier Sampling. Untuk Stratifiel Sampling didapatkan nilai akurasi tertinggi di k-Fold=7 dengan nilai akurasi 68.30%. sehingga akurasi terbaik untuk Naïve Bayes tertinggi didapatkan pada k-Fold=2 dengan nilai akurasi 70.19%.

Tabel 2. Pengujian dengan metode Naïve Bayes

TABEL PENGUJIAN NAÏVE BAYES			
k-Fold	Metode Sampling		
	Linear	Shuffled	Stratified
2	64.42	70.19	65.38
3	58.57	67.28	59.64
4	60.58	67.31	68.27
5	61.52	65.38	64.52
6	63.24	63.29	62.42
7	62.38	62.38	68.30
8	63.46	63.46	62.50
9	62.37	63.72	63.55
10	62.55	63.00	63.64



Gambar 3. Grafik Pengujian dengan Naïve Bayes

3.3. Hasil Evaluasi 2 Model Algoritma

Setelah dilakukan percobaan dengan menggunakan metode Support Vector Machine dan Naïve Bayes maka dapat disimpulkan bahwa nilai tertinggi pada pengujian mendapatkan kenaikan dan penurunan pada uji k-Fold. Hal ini merata pada Uji Sampling kedua metode tersebut. sehingga bila dilihat pada tabel uji kedua algoritma tersebut maka dapat disimpulkan bahwa untuk Support Vector Machine (SVM) nilai tertinggi akurasi diraih 80.77% dengan k-Fold=4 di Stratified Sampling dan untuk Naïve Bayes didapatkan nilai akurasi tertingginya 70.19% dengan k-Fold=2 di Shuffled Sampling. Atau bisa dilihat pada tabel berikut :

Tabel 3. Perbandingan Hasil

Model	Akurasi
Support Vector Machine	80,77%
Naïve Bayes	70,19%

4. KESIMPULAN

Hasil penelitian menunjukkan bahwa baik Support Vector Machine (SVM) maupun Naïve Bayes memiliki potensi untuk digunakan dalam analisis yang dilakukan. Kedua metode mampu memberikan nilai akurasi yang dapat diandalkan dalam konteks pengujian yang dilakukan.

Dari hasil pengujian, terlihat bahwa performa kedua algoritma dapat berfluktuasi akibat variasi dalam metode k-Fold dan jenis sampling yang digunakan. Hal ini mengindikasikan bahwa

pemilihan parameter seperti nilai k-Fold dan metode sampling memiliki pengaruh pada performa model.

Support Vector Machine (SVM) berhasil mencapai nilai akurasi tertinggi sebesar 80.77% dengan k-Fold=4 dalam metode Stratified Sampling. Sementara itu, Naïve Bayes meraih akurasi tertinggi sebesar 70.19% dengan k-Fold=2 dalam metode Shuffled Sampling.

DAFTAR PUSTAKA

- [1] Fannisa, S., Fauzi, M., & Adinugroho, S. (2018). Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*.
- [2] Haryanto, D., Muflikhah, L., & Fauzi, M. (2018). Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*.
- [3] Hidayatullah, A. F. (2014). Analisis Sentimen Dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter. *Seminar Nasional Informatika 2014 (semnasIF 2014)*. Yogyakarta.
- [4] Wilianto, L., Rakhmat Umbara, F., & Hendro Pudjiantoro, T. (2017). Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode Naïve Bayes Classifier Studi Kasus Jawa Barat. *Prosiding SNATIF Ke -4*.
- [5] Hamzah A, "Sentiment Analysis Untuk Memanfaatkan Saran Kuesioner Dalam Evaluasi Pembelajaran Dengan Menggunakan Naive Bayes Classifier (Nbc)" in *proc SNAST 2014*, 15 November 2014.
- [6] Oman Sumantri, Dyah Apriliany "Support Vector Machine Berbasis Feature Selection Untuk Sentiment Analysis Kepuasan Pelanggan Terhadap Pelayanan Warung dan Restoran Kuliner Kota Tegal" *Jurnal Teknologi Informasi dan Ilmu Komputer*, Vol 5 No. 5, 2018
- [7] Oman Somantri & Dairoh, "Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining", *JEPIN Vol 5, No 2* (2019).
- [8] Sihombing, L., Hannie, H. dan Dermawan, B. (2021) "Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma Naïve Bayes Classifier," *Edumatic: Jurnal Pendidikan Informatika*, 5, hal. 233–242. doi: [10.29408/edumatic.v5i2.4089](https://doi.org/10.29408/edumatic.v5i2.4089)
- [9] M. T. Akter, M. Begum, and R. Mustafa, "Bengali Sentiment Analysis of E-commerce Product Reviews using K-Nearest Neighbors," in *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, 2021, pp. 40–44, doi: [10.1109/ICICT4SD50815.2021.9396910](https://doi.org/10.1109/ICICT4SD50815.2021.9396910)
- [10] S. F. N. H. R. JAYADI, "Sentiment Analysis Of Indonesian E-Commerce Product Reviews Using Support Vector Machine Based Term Frequency Inverse Document," *vol. 99, no. 17*, pp. 4316–4325, 2022