



Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis *Text Mining*

Oman Somantri^{#1}, Dairoh^{*2}

[#]Program Studi Teknik Informatika, Politeknik Negeri Cilacap

Jln. Dokter Soetomo No.01 Karangcengis Sidakaya, Cilacap, Indonesia

¹oman_mantri@yahoo.com

^{*}Program Studi Teknik Informatika, Politeknik Harapan Bersama Tegal

Jln. Mataram No.09 Pesurungan Lor Margadana Tegal, Indonesia

²zaroh31@yahoo.co.id

Abstrak—Mendapatkan sebuah informasi untuk meningkatkan pelayanan dan strategi dalam pengelolaan tempat kunjungan wisata masih sedikit dan sulit dikarenakan informasi yang didapatkan masih terbatas. Media sosial memiliki peranan dalam memberikan sebuah data terhadap penilaian kunjungan wisata, sama halnya dengan tempat-tempat wisata yang berada di wilayah Tegal dan sekitarnya. Pada penelitian ini model sentiment analysis diusulkan sebagai solusi untuk mengatasi permasalahan. Tujuan dari penelitian ini adalah mencari model sistem untuk memberikan sebuah informasi pendukung keputusan bagi para wisatawan dan pengelola tempat wisata untuk dijadikan sumber informasi terhadap tempat wisata yang ada. Metode penelitian yang digunakan adalah dengan melakukan eksperimen untuk mendapatkan model yang diharapkan. Pada penelitian ini metode *Naïve Bayes* dan *Decision Tree* diterapkan untuk klasifikasi teks sehingga mendapatkan model terbaik yang dihasilkan untuk diimplementasikan pada sistem sehingga model yang didapatkan dapat dijadikan sebuah sistem pendukung keputusan untuk pengembangan sistem cerdas pada pihak terkait dalam upaya peningkatan nilai jual potensi daerah khususnya bidang pariwisata. Hasil penelitian menunjukkan bahwa model yang didapatkan setelah dilakukan eksperimen didapatkan tingkat akurasi *naïve bayes* menghasilkan 77,50% lebih baik dibandingkan dengan menggunakan *Decision Tree* yang menghasilkan tingkat akurasi 60,83%.

Kata kunci— sentiment analysis, naïve bayes, decision tree, tempat wisata

I. PENDAHULUAN

Sumber devisa setiap negara yang saat ini menjadi sedang digalakan dimanapun adalah tempat pariwisata. Saat ini Indonesia mendapatkan sumber devisa yang tinggi salah satunya karena adanya pariwisata. Tempat wisata yang dijadikan tujuan destinasi para wisatawan sangat berpengaruh terhadap banyak tidaknya wisatawan yang datang ke tempat tersebut. Setiap tempat mempunyai daya tarik sendiri, sehingga sudah sewajarnya apabila pengelolaan terhadap daerah wisata tersebut harus

menjadi fokus utama. Sebagai salah satu kota di Jawa Tengah, Tegal merupakan salah satu kota yang mempunyai daerah tempat tujuan wisata yang menjanjikan dan saat ini sudah banyak orang yang mengenal daerah tempat-tempat wisata di kota ini seperti misalnya wisata pemandian air panas Guci, pantai Alam Indah, dan yang lainnya. Sebuah pengelolaan yang tepat pada bidang pariwisata akan meningkatkan jumlah kedatangan wisatawan, sehingga akan meningkatkan sumber pendapatan daerah tersebut.

Media sosial sebagai bagian dari perkembangan teknologi dewasa ini mempunyai peranan yang sangat signifikan terhadap berbagai hal. Setiap hal apapun baik itu informasi, curhatan pribadi, penilaian terhadap sesuai dan banyak lainnya masuk dalam media sosial [1]. Saat ini setiap orang sudah mulai menjadikan media sosial sebagai parameter dalam melakukan sebuah penilaian, hal ini menjadi sangat penting karena dengan informasi yang didapatkan akan dijadikan orang sebagai referensi untuk melakukan hal tersebut seperti diantaranya *review* film [2] dan *review* tempat wisata [3]. Sama halnya dengan *review* penilaian tempat tujuan wisata dimana para wisatawan yang susah pernah mengunjungi tempat tersebut biasanya melakukan *review* dengan membuat sebuah status atau pernyataan terhadap tempat yang mereka kunjungi, penilaian tersebut bisa saja positif, negatif ataupun bahkan netral atau biasa saja. Penilaian wisatawan terhadap tempat kunjungan wisata ini menjadi salah satu hal yang penting karena dapat dijadikan sebagai bagian dari pendukung keputusan baik itu wisatawan maupun pengelola tempat tersebut, wilayah Kota Tegal dan sekitarnya salah satunya.

Sentiment analysis merupakan sebuah teknologi komputerisasi yang dapat membantu dan menganalisis sebuah kalimat opini tekstual yang cara bekerjanya adalah dengan memahami dan mengekstraknya seperti halnya *text mining* sehingga menghasilkan sebuah informasi sentiment [4]. *Text mining* merupakan sebuah teknik atau cara untuk mengekstraksi informasi yang berasal dari data

teks yang tidak terstruktur dan berguna [5]. *Sentiment analysis* dilakukan untuk melihat kecenderungan seseorang berpandangan apakah beropini positif atau negatif, bahkan beropini netral yang dijadikan sebagai sebuah pendukung keputusan [6]. Penelitian terhadap sentimen tersebut seperti yang telah dilakukan untuk penilaian tempat penjualan makanan kuliner [7] yang kemudian hasil penelitian tersebut dikembangkan [8].

Sebelumnya penelitian dilakukan oleh Bigne, E, dkk [9] melakukan penelitian terkait sebuah pendukung keputusan yang dengan menggunakan *text mining* yang berasal dari media sosial twitter dengan menerapkan sebuah *artificial intelligent* algoritma *Neural Network* terhadap tempat wisata di Kota negara Spanyol. Pada penelitian ini yang menjadi target utama adalah hotel di tempat wisata tersebut.

Penelitian lain dilakukan oleh Park, S.B, dkk memanfaatkan data yang berasal dari twitter untuk diterapkan pada sebuah travel wisata kapal pesiar [10]. Data yang ada analisis dan diolah dengan menggunakan pola *bigdata analytic* dengan tujuan sebagai pendukung keputusan dalam marketing wisatawan. Pada penelitian ini hasil analisis dari setiap tweets dimasukan kedalam tiga tipe grup pengguna, yaitu *commercial*, *news/blog*, dan *private*. Berdasarkan hasil penelitian yang didapatkan digunakan sebagai strategi dalam marketing. Penelitian selanjutnya dilakukan oleh Statiswaty, dkk, melakukan penelitian analisis sentimen terhadap wisata bahari yang terdapat di Sulawesi Tenggara. Pada penelitian ini dimanfaatkan data yang besala dari media sosial twitter dengan menerapkan metode yang berbasis *Lexicon*. Berdasarkan hasil penelitian didapatkan sistem yang ada mampu memberikan informasi tentang opini positif, negatif dan netral masyarakat terhadap wisata bahari di Sulawesi tenggara dengan tingkat akurasi sebesar 77,06% [11].

Pada tahun 2018, Fanisa, S., dkk dengan menggunakan metode *Naïve Bayes* dan seleksi fitur *Query Expansion* Rangkings melakukan penelitian untuk *sentiment analysis* terhadap pariwisata di Kota Malang. Dari hasil pengujian penelitian ini uji akurasi dengan menggunakan variasi rasio seleksi fitur menghasilkan akurasi terbaik 86,6% [12]. Penelitian lain adalah terkait dengan penilaian opini terhadap sebuah produk dengan menggunakan *Naïve bayes* [13].

Sedikit berbeda pada penelitian yang dilakukan pada artikel penelitian ini, model *sentiment analysis* dipergunakan untuk menganalisis review penilaian wisatawan terhadap tempat tujuan wisata khususnya di kota tegal dan sekitarnya dengan cara mengklasifikasikan antara yang beropini positif dan negatif. Pada penelitian ini diterapkan algoritma *Naïve Bayes* dan *Decision Tree* untuk mendapatkan model terbaik dengan tingkat akurasi tertinggi untuk mengklasifikasikan sentimen opini dari data yang telah didapatkan.

Pada penelitian ini bertujuan untuk mencari model terbaik sentiment anlysisi terbaik berbasis *text mining* dengan tingkat akurasi tinggi dengan menggunakan *Naïve*

bayes dan *Decision Tree* untuk klasifikasi sentimen penilaian wisatawan terhadap tujuan wisata di kota Tegal sebagai pendukung keputusan dalam pengelolaan tempat wisata sehingga lebih berpotensi menjadi tujuan wisata yang menjanjikan.

II. NAÏVE BAYES & DECISION TREE

A. Naïve Bayes

Salah satu metode yang sering dipergunakan dalam ilmu statistika dalam menghitung sebuah peluang dari suatu hipotesis adalah *naïve bayes*. Algoritma ini bekerja dengan cara mengklasifikasikan kelas berdasarkan pada probabilitas sederhana dimana dalam hal ini diasumsikan setiap atribut pada data yang ada sifatnya saling terpisah. Sebagai salah satu metode yang cukup sederhana, metode ini mencoba mengklasifikasikan data berdasarkan probabilitas P atribut x dari setiap kelas y suatu data [14]. Adapun persamaan rumus metode tersebut adalah seperti pada persamaan (1) dan (2) berikut:

$$P = (y_k | x_1, x_2, \dots, x_n) \quad (1)$$

$$P(y_k | x_a) = \frac{P(y_k)P(x_a|y_k)}{P(x_a)} \quad (2)$$

Dimana:

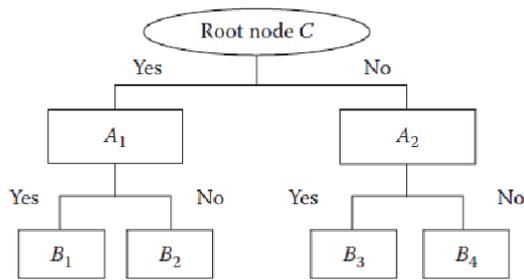
- $Y_k P(x_a/y_k)$: kategori kelas
- $P(y_k)$: probabilitas kelas
- $P(x_a)$: probabilitas kemunculan dokumen

Berdasarkan hasil yang didapatkan, kemudian dilakukan proses pemilihan kelas yang optimal sehingga dipilih suatu nilai peluang terbesar dari setiap probabilitas kelas yang ada. Dalam pembobotan suatu atribut kelas maka dapat meningkatkan pengaruh satu prediksi, dan dengan memperhitungkan bobot atribut terhadap kelas maka akan menjadi sebuah titik point dasar ketepatan dalam klasifikasi dan bukan hanya probabilitas melainkan juga pada bobot setiap antar kelas, melalui persamaan berikut:

$$y(x_i) = \arg \max P(y) \prod_{i=1}^n P(x_i | y) \quad (3)$$

B. Decision Tree

Metode *Decision tree* adalah sebuah metode klasifikasi representasi dari sebuah *tree* atau pohon keputusan, dimana sebuah atribut direpresentasikan sebagai *node*, dan sebagai nilainya adalah cabang pohon tersebut, serta kelas direpresentasikan sebagai kelas. Dalam hal ini *node* yang paling atas dari sebuah *decision tree* dinamakan sebagai *root*, seperti digambarkan pada Gambar 1.



Gambar. 1 Gambaran struktur sebuah decision tree

Salah satu algoritma *decision tree* adalah C4.5. Sebuah *decision tree* sangat tergantung kepada sebuah aturan *if-then*, tetapi tidak membutuhkan parameter dan matrik. Algoritma C4.5 merupakan algoritma berbasis *supervised learning* dan merupakan pengembangan dari algoritma ID3 (*Iterative Dichotomiser*).

III. METODE PENELITIAN

A. Data Penelitian

Data penelitian yang digunakan adalah data teks terkait dengan komentar-komentar opini yang diperoleh dari situs www.google.com/maps terhadap opini review tempat tujuan wisata di kota Tegal dan sekitarnya tahun 2017 sampai dengan 2018. Data yang diperoleh adalah berjumlah sebanyak 120 data file teks dengan isi jumlah kata yang berbeda tiap file-nya. Review data opini yang digunakan adalah salah satu tempat wisata di Tegal yaitu pemandian air panas Guci, seperti pada Gambar 2.

- Untuk harga kamar 1 juta lebih dimusim liburan, pelayanan nya jelek, gua bawa barang banyak gotong sendiri ke lantai 2, pesen indomie goreng gak di aduk bumbunya, plis dah ini cuaca dingin ya mie nya kaku, jadi susah ngaduknya, jadi asin trus hambar rasa mie nya. Besok nya gua lebih kecewa lagi dengan lagunya, ini liburan keluar live music nya lagu galau atau dangdut dengan setelan bass yang kegedean, sulit buat tenang untuk istirahat setelah capek berenang.
- Tempatnya bersih,nyaman,mantap deh..
- Klo pulkam pasti selalu kesini pemandangan yg keren dan tiket masuk terjangkau dan jalan.y gak terlalu jauh dari rumah mertua
- Fasilitas bersih, full cctv, no preman

Gambar. 2 Contoh data opini teks yang digunakan

B. Praprocessing data

Tahapan *praprocessing* data dilakukan untuk mendapatkan data yang sesuai dengan yang diinginkan sebelum dilakukannya proses analisis data kedalam model yang diinginkan. Pada tahapan ini dilakukan beberapa hal

untuk mendapatkan data yang sesuai diantaranya adalah proses tokenisasi, stopword, dan pembobotan data.

1) *Tokenisasi*: Pada tahapan ini adalah dilakukan pemisahan kata dari sebuah kalimat menjadi token-token yang dipisahkan oleh spasi ataupun tanda baca.

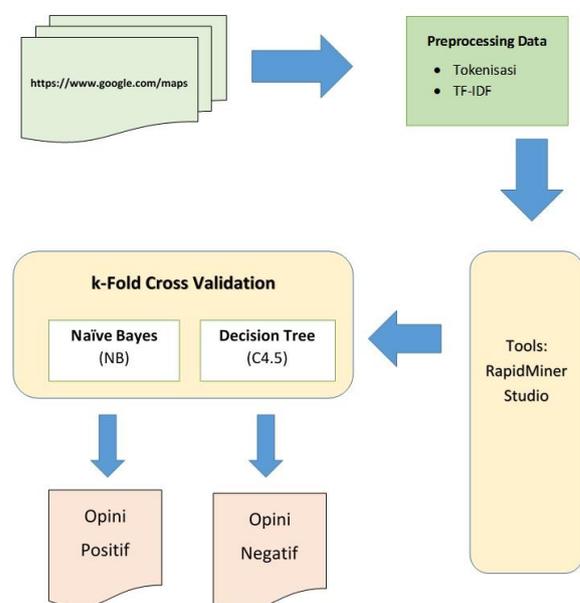
2) *Filter Tokens (by length)*: Proses ini adalah pembatasan panjang karakter minimum dan maksimum yang digunakan pada tahapan *praprocessing* data. Pada penelitian ini ditentukan karakter minimum= 3 dan karakter maksimum=25.

3) *Filter Stopword*: Proses menghilangkan kata yang tidak sesuai dengan yang diinginkan, pada tahapan ini dilakukan untuk mengoptimalkan data teks terhadap kata-kata yang tidak sesuai. Pada proses ini setiap kata yang akan diproses adalah hanya kata berbahasa Indonesia.

4) *Pembobotan TF-IDF*: Proses ini dilakukan untuk mendapatkan nilai bobot dari data inputan teks yang akan dimasukkan kedalam model yang akan digunakan. Nilai bobot pada *Term Frequency-Inversed Document Frequency* (TF-IDF) tahapan ini didapatkan berdasarkan frekuensi kata yang muncul.

C. Validasi dan Evaluasi Sistem

Untuk melihat hasil yang diinginkan pada model yang diharapkan, validasi data menggunakan *k-Fold Cross Validation* [15]. Pada proses ini memperlihatkan sebuah hasil tingkat akurasi dari model yang diusulkan. Evaluasi sistem dilakukan dengan membandingkan komparasi tingkat akurasi klasifikasi opini positif dan negatif dengan menggunakan dua model yang berbeda yaitu *Naive bayes* dan *Decision Tree* sehingga memperlihatkan model yang diharapkan dengan tingkat akurasi yang terbaik. Kerangka metode yang diusulkan seperti diperlihatkan pada Gambar 3.



Gambar. 3 Kerangka metode yang diusulkan

IV. HASIL DAN PEMBAHASAN

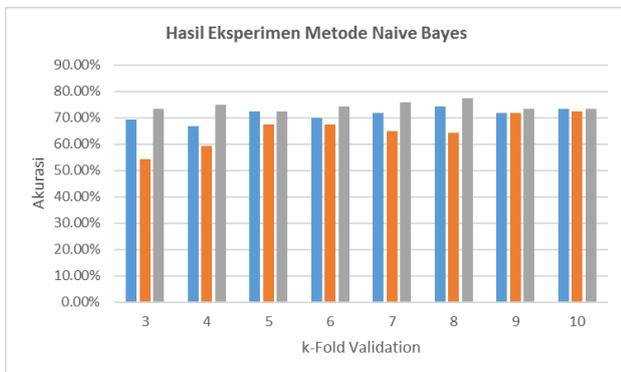
Pada penelitian ini untuk analisis data menggunakan software RapidMiner untuk mendapatkan model yang sesuai dan diinginkan. Spesifikasi komputer menggunakan sistem operasi Window 7 SP1 dan memori Core i5 4Gb.

A. Hasil Eksperimen Model Naïve Bayes

Pada hasil eksperimen dalam mencari model terbaik dengan menggunakan Naïve Bayes, telah didapatkan hasil seperti pada Tabel I dan Gambar 4.

TABEL I
HASIL EKSPERIMEN METODE NAÏVE BAYES

K-Fold	Metode Sampling		
	Shuffled	Linear	Stratified
3	69.17%	54.17%	73.33%
4	66.67%	59.17%	75.00%
5	72.50%	67.50%	72.50%
6	70.00%	67.50%	74.17%
7	71.62%	64.99%	75.96%
8	74.17%	64.17%	77.50%
9	71.67%	71.61%	73.26%
10	73.33%	72.50%	73.33%



Gambar. 4 Grafik hasil eksperimen metode naïve bayes

Berdasarkan hasil yang didapatkan pada Tabel I dan Gambar 4 terlihat bahwa tingkat akurasi terbaik dari model Naïve Bayes adalah sebesar 77,50% dengan menggunakan k-Fold=8 secara stratified sampling. Untuk tingkat akurasi tertinggi secara liner sampling adalah sebesar 72,50% dengan menggunakan k-fold=10, dan secara suffled sampling menggunakan k-fold=8 tingkat akurasi tertinggi yang dihasilkan adalah 74.17%.

B. Hasil Decision Tree

Untuk mendapatkan model terbaik, pada metode Decision tree ditentukan terdapat parameter yang telah ditetapkan yaitu criterion, diantaranya adalah gain ratio, information gain, dan gini index. Terjadi perbedaan tingkat akurasi dari model yang dihasilkan sesuai dengan parameter dari masing-masing criterion tersebut. Hasil eksperimen dari tingkat kaurasi model yang dihasilkan adalah seperti pada Tabel II, Tabel III, dan Tabel IV.

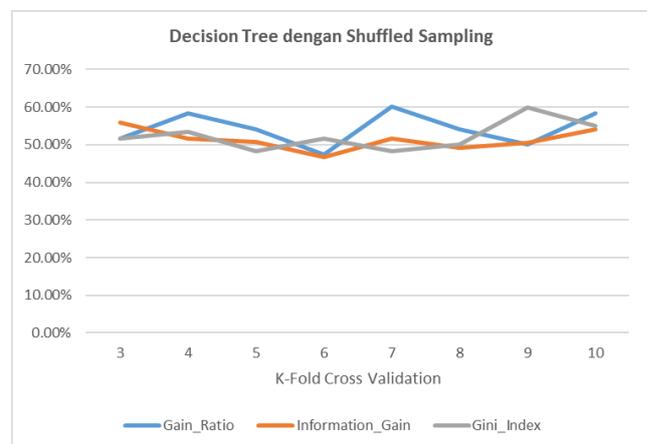
Tingkat akurasi tertinggi yang diperoleh pada Tabel II dengan menggunakan Decision tree menggunakan parameter criterion Gain_ratio secara shuffled sampling adalah sebesar 60,68% dengan fold=7. Untuk hasil yang didapatkan dengan menggunakan criterion Information gain adalah tertinggi sebesar 55,83% dan Gini Index tertinggi sebesar 59,83%.

TABEL II
HASIL DECISION TREE DENGAN SHUFFLED SAMPLING

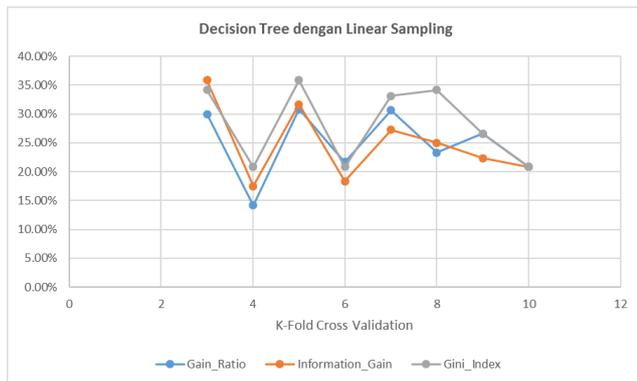
Fold	Gain_Ratio	Information_Gain	Gini_Index
3	51.67%	55.83%	51.67%
4	58.33%	51.67%	53.33%
5	54.17%	50.83%	48.33%
6	47.50%	46.67%	51.67%
7	60.08%	51.63%	48.32%
8	54.17%	49.17%	50.00%
9	50.00%	50.49%	59.89%
10	58.33%	54.17%	55.00%

TABEL III
HASIL DECISION TREE DENGAN LINEAR SAMPLING

Fold	Gain_Ratio	Information_Gain	Gini_Index
3	30.00%	35.83%	34.17%
4	14.17%	17.50%	20.83%
5	30.83%	31.67%	35.83%
6	21.67%	18.33%	20.83%
7	30.63%	27.26%	33.15%
8	23.33%	25.00%	34.17%
9	26.56%	22.34%	26.56%
10	20.83%	20.83%	20.83%



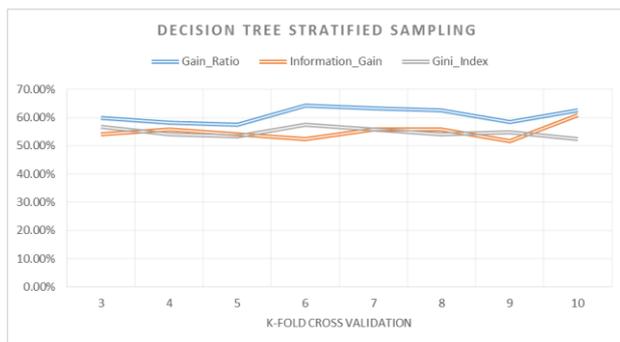
Gambar. 5 Grafik hasil eksperimen Decision Tree Suffled sampling



Gambar. 6 Grafik hasil eksperimen Decision Tree Linear Sampling

TABEL IV
HASIL DECISION TREE DENGAN STRATIFIED SAMPLING

Fold	Gain_Ratio	Information_Gain	Gini_Index
3	60.00%	54.17%	56.67%
4	58.33%	55.83%	54.17%
5	57.50%	54.17%	53.33%
6	64.17%	52.50%	57.50%
7	63.35%	55.84%	55.88%
8	62.50%	55.83%	54.17%
9	58.42%	51.59%	54.88%
10	62.50%	60.83%	52.50%



Gambar. 7 Grafik hasil eksperimen Decision Tree Stratified Sampling

Hasil tingkat akurasi yang diperoleh dengan menggunakan decision tree secara *linear sampling* secara keseluruhan masih dikatakan sangat kecil hal ini bisa terlihat pada Tabel III untuk tingkat akurasi tertinggi diperoleh sebesar 35,83% dengan menggunakan *criterion Gini Index* dan *Information gain*, hal yang tidak berbeda dihasilkan dengan menggunakan *criterion Gain ratio* hanya menghasilkan 30,83%.

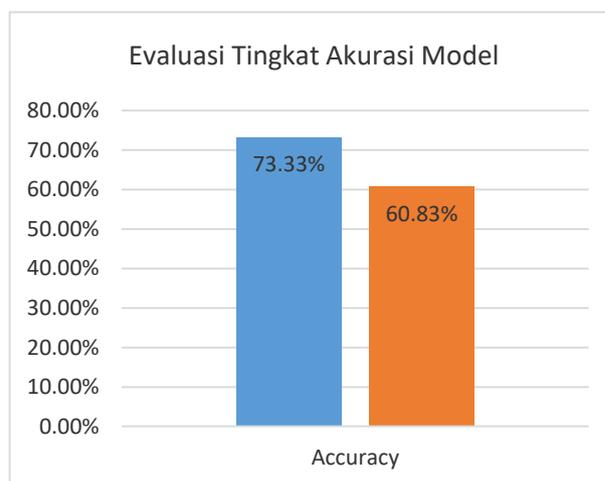
Secara *stratified sampling* pada IV dengan metode *Decision tree* diperoleh tingkat akurasi tertinggi sebesar 60,83% dengan menggunakan k-Fold=10 dan *criterion Information Gain*. Untuk hasil terendah menghasilkan model dengan tingkat akurasi sebesar 57,50% dengan menggunakan *criterion Gini_indek* dan Fold=6.

C. Evaluasi Model

Dari hasil eksperimen yang telah didapatkan terhadap dua model yang menggunakan algoritma yang berbeda yaitu *Naïve Bayes* (NB) dan *Decision Tree* (DT), maka tingkat akurasi yang tertinggi diperlihatkan pada Tabel V.

TABEL V
EVALUASI MODEL NB DAN DT

No.	Model	Accuracy
1	Naïve Bayes (NB)	73.33%
2	Decision Tree	60.83%



Gambar. 8 Grafik evaluasi model yang diusulkan

Berdasarkan pada Tabel V dan Gambar 4, terlihat bahwa berdasarkan hasil evaluasi tingkat akurasi tertinggi dari model yang telah didapatkan adalah sebesar 73,33% dengan menggunakan metode *Naïve Bayes*(NB), sedangkan *Decision Tree* hanya menghasilkan tingkat akurasi sebesar 60,83 lebih kecil dibandingkan dengan metode NB.

Hasil peningkatan akurasi tertinggi *Naïve Bayes* yang didapatkan diperoleh berdasarkan pengaturan nilai-nilai parameter yang dilakukan sehingga tingkat akurasi yang didapatkan lebih baik dibandingkan dengan *Decision Tree*.

V. KESIMPULAN

Sebuah pendukung keputusan dalam menentukan tingkat sentiment opini pengunjung wisata memberikan sebuah informasi yang dapat dijadikan sebagai pendukung keputusan dalam proses pengelolaan tempat wisata, dengan diterapkannya metode *text mining* dalam pembuatan model *sentiment analysis* penilaian tujuan tempat wisata di Tegal dan sekitarnya memberikan sebuah solusi yang tepat dalam membuat sebuah keputusan bagi para pengambil kebijakan. Metode *Naïve Bayes* dan *Decision Tree* telah diterapkan dan mampu dijadikan sebagai model dalam *sentiment analysis* penilaian tempat wisata, berdasarkan hasil yang telah didapatkan metode dengan menggunakan *Naïve Bayes* mempunyai

tingkat akurasi yang lebih baik dibandingkan dengan *decision tree* yaitu mempunyai tingkat akurasi sebesar 73,33%.

Tingkat akurasi yang didapatkan masih diperlukan adanya optimasi sehingga untuk penelitian selanjutnya diperlukan adanya sebuah optimasi baik itu dengan menggunakan metode *feature weight* atau *feature selection* sehingga diharapkan dapat menemukan model terbaik yang dapat menaikkan tingkat akurasi dari hasil yang ada sekarang.

UCAPAN TERIMA KASIH /

Penelitian ini merupakan bagian dari kolaborasi penelitian antara Politeknik Harapan Bersama Tegal dan Politeknik Negeri Cilacap. Terimakasih disampaikan kepada seluruh pihak yang telah membantu dan memfasilitasi kegiatan penelitian ini.

REFERENSI

- [1] W. Fan and M. D. Gordon, "Unveiling the Power of Social Media Analytics," *Commun. ACM*, vol. 12, no. JUNE 2014, pp. 1–26, 2013.
- [2] A. S. H. Basari, B. Hussin, I. G. P. Ananta, and J. Zeniarja, "Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization," *Procedia Eng.*, vol. 53, pp. 453–462, 2013.
- [3] B. Zeng and R. Gerritsen, "What do we know about social media in tourism? A review," *Tour. Manag. Perspect.*, vol. 10, pp. 27–36, Apr. 2014.
- [4] N. Boudad, R. Faizi, R. Oulad Haj Thami, and R. Chiheb, "Sentiment analysis in Arabic: A review of the literature," *Ain Shams Eng. J.*, 2017.
- [5] B. S. Kumar and V. Ravi, *A survey of the applications of text mining in financial domain*, vol. 114. Elsevier B.V., 2016.
- [6] B. Heredia, T. M. Khoshgoftaar, J. Prusa, and M. Crawford, "Cross-Domain Sentiment Analysis: An Empirical Investigation," in *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*, 2016, pp. 160–165.
- [7] O. Somantri and D. Apriliani, "Feature Weights Menggunakan Particle Swarm Optimization Untuk Sentiment Analysis Penilaian Kepuasan Pelanggan Makanan Kuliner," in *Seminar Nasional Sains & Teknologi Informasi (SENSASI)*, 2018, pp. 697–704.
- [8] O. Somantri and D. Apriliani, "Support Vector Machine Berbasis Feature Selection Untuk Sentiment Analysis Kepuasan Pelanggan Terhadap Pelayanan Warung dan Restoran Kuliner Kota Tegal," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, pp. 537–548, Oct. 2018.
- [9] E. Bigné, E. Oltra, and L. Andreu, "Harnessing stakeholder input on Twitter: A case study of short breaks in Spanish tourist cities," *Tour. Manag.*, vol. 71, no. July 2018, pp. 490–503, Apr. 2019.
- [10] S. B. Park, C. M. Ok, and B. K. Chae, "Using Twitter Data for Cruise Tourism Marketing and Research," *J. Travel Tour. Mark.*, vol. 33, no. 6, pp. 885–898, Jul. 2016.
- [11] S. Statiswaty, R. Rusnia, N. R.- SemanTIK, and U. 2018, "Analisis Sentimen Wisata Bahari Di Sulawesi Tenggara Memanfaatkan Media Sosial Twitter Dengan Menggunakan Metode Lexicon," *Ojs.Uho.Ac.Id*, vol. 3, no. 2, pp. 161–168, 2017.
- [12] S. Fanissa, M. A. Fauzi, and S. Adinugroho, "Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2766–2770, 2018.
- [13] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, Dec. 2018.
- [14] Y. Ganisaputra and R. Tan, "Naïve Bayes Classifier," pp. 173–188.
- [15] M. van der Gaag *et al.*, "The five-factor model of the Positive and Negative Syndrome Scale II: A ten-fold cross-validation of a revised model," *Schizophr. Res.*, vol. 85, no. 1–3, pp. 280–287, 2006.