Jurnal INFORMA Politeknik Indonusa Surakarta p-ISSN : 2442-7942, e-ISSN 2716-5051
Vol. 6 Nomor 1 Juni 2020

Analisa Sentimen Twitter Menggunakan Algoritma Klasifikasi Pada Promosi Wisata Museum Sangiran Kabupaten Sragen

Yoga Dwi Pambudi[1)], Wing Wahyu Winarno[2)], Andi Sunyoto[3)]

Magister Teknik Informatika, Universitas AMIKOM Yogyakarta
Jl. Ring Road Utara, Condong Catur, Sleman, Yogyakarta
yoga.pambudi@students.amikom.ac.id[1)], wing@amikom.ac.id[2)], andi@amikom.ac.id[3)]

Abstrak

Museum Sangiran merupakan tempat wisata andalan di Kabupaten Sragen yang sudah terkenal dalam kanca nasional hingga internasional. Dalam rangka membangun minat dan mempengaruhi niat wisatawan dalam meningkatkan kunjungan wisata di Museum Sangiran maka, diperlukan berbagai macam strategi untuk menyampaikan pesan promosi wisata pada masyarakat. Perkembangan teknologi informasi yang semakin meningkat drastis di era Revolusi Industri 4.0 saat ini sangat berpengaruh dalam berbagai hal termasuk dalam bidang pariwisata dengan memanfaatkan sosial media twitter sebagai jalur promosi. Dalam sosial media twitter terdapat banyak opini yang berisikan pendapat, pikiran maupun kritikan dari para pengguna. Pada penelitian ini berfokus pada opini yang muncul mengenai keadaan dan pelayanan di Kawasan wisata Museum Sangiran Kabupaten Sragen dengan menggunakan analisis sentimen data Twitter dengan harapan memperoleh presepsi masyarakat terhadap kualitas pelayanan yang dapat digunakan sebagai bahan tambahan evaluasi pengelolaan kawasan wisata. Penggunaan metode *Naive Bayes Classifier (NBC)* digunakan untuk klasifikasi serta menerapkan seleksi fitur menggunakan *Particle Swarm Optimization (PSO)* untuk mengurangi atribut yang tidak relevan. Hasil pengujian pada penelitian ini menunjukan bahwa algoritma *Naive Bayes Classifier (NBC)* dengan seleksi fitur *Particle Swarm Optimization (PSO)* dengan nilai akurasi sebesar 87,91%.

Kata kunci: Twitter, Analisis Sentimen, NBC, PSO, Klasifikasi

1. PENDAHULUAN

Menurut (Mayfield, 2008) media sosial adalah mengenai bagaimana menjadi manusia seutuhnya. Manusia dapat menemukan informasi dan inspirasi lebih cepat dari sebelumnya. Ide-ide baru, layanan jasa, model bisnis dan teknologi muncul dan berkembang begitu cepat di media sosial. Peranan media sosial pada kehidupan masyarakat khususnya generasi muda sudah sangat melekat pada kegiatan sehari-hari. Pandangan atau penilaian masyarakat juga sangat mudah digiring melalui sebaran informasi dari sosial media padahal belum diketahui kebenaran informasi tersebut baik itu berupa informasi yang positif, negatif atau bermakna provokatif.

Periklanan melalui twitter menawarkan kesempatan bisnis yang memungkinkan terhubungnya pelanggan dan perusahaan membangun hubungan. Jika pelanggan menggunakan twitter maka perusahaan harus berada disana juga. (Gunelius, 2011).

Sentimen analisis merupakan sebuah studi komputasi yang berhubungan dengan pendapat dan berorientasi dengan pengolahan bahasa alami (Kumar & Sebastian, 2012). Sedangkan menurut (Dave, dkk, 2003) memaparkan bahwa Sentimen adalah informasi tekstual yang berada di dalam *web* dan berisi tentang fakta dan opini. Sentimen merupakan pernyataan atau ungkapan subjektif masyarakat yang mencerminkan persepsi seseorang terhadap suatu peristiwa atau kejadian.

(Liu, 2011) Dalam proses analisis sentimen pada twitter menggunakan cara pengambilan data dalam jumlah yang besar atau sering disebut dengan *Crawling data*. *Crawling data* adalah proses pengambilan sejumlah besar halaman *web* dengan cepat ke dalam suatu tempat penyimpanan lokal dan mengindeksnya berdasar sejumlah kata kunci. Hal ini sangat membantu dalam penelitian untuk mendapatkan data besar secara daring.

Analisis sentimen pada twitter diambil menggunakan twitter *crawler* otomatis pada opini-opini yang diambil langsung dari halaman twitter. Hasil dalam analisis sentimen akan mendapatkan presepsi masyarakat terhadap kualitas pelayanan di Museum Sangiran Kabupaten Sragen. Selanjutnya dari perolehan data tersebut akan dilakukan klasifikasi. Menurut (Adam, dkk., 2002) menerangkan bahwa klasifikasi memiliki dua proses yaitu membangun model klasifikasi dari sekumpulan kelas data yang sebelumnya sudah didefinisikan *(training data set)* dan model tersebut digunakan untuk klasifikasi tes data serta mengukur akurasi dari model.

Dalam rangka membangun minat wisatawan untuk mengunjungi Museum Sangiran Kabupaten Sragen maka, pada penelitian ini memiliki tujuan untuk mengukur pandangan serta interaksi terhadap produk wisata yang ditawarkan kepada masyarakat luas melalui sosial media twitter dengan analisis data menggunakan sentimen data twitter dan mengklasifikasinya menggunakan algoritma *Naive Bayes Classifier (NBC)* dengan seleksi fitur *Particle Swarm Optimization (PSO)*.

2. METODE PENELITIAN

Dalam menyelesaikan penelitian ini dilakukan secara sistematis dengan tahapan-tahapan metodologi sebagai berikut:

a. Studi Pustaka

Penelitian dilakukan dengan melakukan studi kepustakaan, dengan mengumpulkan beberapa bahan referensi yang terkait dengan penelitian, baik melalui buku, artikel paper, jurnal, makalah dan mengunjungi beberapa situs yang terdapat di initernet terkait dengan analisis sentimen data twitter menggunakan algoritma klasifikasi *Naive Bayes Classifier (NBC)*.

b. Pengumpulan Data

Pada tahap ini dilakukan pengambilan data mengenai segala bentuk informasi yang berkaitan dengan informasi Museum Sangiran secara khusus melalui sosial media twitter dan media daring yang lain sebagai data tambahan atau data sekunder dan pengambilan data melalui studi lapangan ke Museum Sangiran. Proses pengambilan data yang digunakan dalam penelitian ini dilakukan dengan tiga metode yaitu metode wawancara, metode observasi dan metode dokumentasi.

c. Klasifikasi Data

Pada penelitian ini, data yang telah diperoleh kemudian diklasifikasikan dengan berdasarkan pada metode algoritma klasifikasi yang sudah ditentukan untuk dapat dilakukan analisis lebih lanjut dan menghasilkan nilai akhir yang dapat menentukan nilai akurasi.

d. Kesimpulan

Tahapan akhir yaitu penyampaian kesimpulan atas hasil dari penelitian yang telah dilakukan dan pemberian saran dan ide untuk penelitian yang akan dilakukan berikutnya.

3. TINJAUAN PUSTAKA

a. *Naive Bayes Classifier (NBC)*

Konsep dasar yang digunakan oleh *Naive Bayes* adalah teorema Bayes yang dimana dalam klasifikasi menerapkan perhitungan nilai probabilitas p(C = ci | D = dj), yaitu probabilitas kategori ci jika diketahui dokumen dj. Klasifikasi yang dilakukan untuk menentukan kategori c ϵ C dari suatu dokumen d ϵ D dimana C = {c1,c2,c3, ..., ci} dan juga D = { d1,d2,d3, ..., di}. Penentuan dari kategori sebuah dokumen dilakukan dengan mencari nilai maksimum dari :

$$p(C = ci \mid D = dj)$$

pada P= {p (C = ci | D = dj | c ϵ C dan d ϵ D}.

Nilai probabilitas p (C = ci | D = dj) dapat dihitung dengan persamaan (Mitchell., 2010).

$$p(C = c_i \mid D = d_j) = (p(C = c_i) / \cap D = d_j$$
$$= \frac{p(D = d_j \mid C = c_i) \times p(C = c_i)}{p(D = d_j)}$$

Dengan p merupakan nilai probabilitas dari kemunculan dokumen dj jika diketahui dokumen tersebut berkategori ci, p (C = ci) adalah nilai probabilitas kemunculan kategori ci, dan p (D = dj) adalah nilai kemunculan dokumen dj. Naïve bayes menganggap sebuah dokumen sebagai kumpulan dari kata-kata yang menyusun dokumen tersebut, dan tidak memperhatikan urutan kemunculan kata pada dokumen. Sehingga perhitungan probabiltas

p(D = dj | C = ci) dapat dituliskan sebagai berikut (Mitchell., 2010) :

$$p\,(C = ci \mid D = dj) = \frac{\prod_k p(W_k \mid C=c_i)\; x\; p(C=c_i)}{p(W_1, W_2, W_3, \ldots, W_k \quad \ldots, W_n)}$$

Perhitungan adalah hasil perkalian dari probabilitas kemunculan semua kata pada dokumen dj.

Secara teoritis, *Naive Bayes Classifier (NBC)* seringkali bekerja jauh lebih baik dan memiliki tingkat kesalahan yang minimum dibandingkan dengan metode klasifikasi yang lain meski dengan menggunakan rancangan-rancangan yang "naive" dan dengan asumsi yang disederhanakan (Rennie, dkk., 2003).

4. HASIL DAN PEMBAHASAN

Pada penelitian ini dilakukan *text processing* pada *dataset*. Tahapan awal ini bertujuan untuk mengubah data teks yang tidak terstruktur dan sembarang menjadi data yang terstruktur. Proses *processing* yang dilakukan dalam tahapan ini adalah sebagai berikut:

a. *Case Folding*

Case folding adalah mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter (Ronen Feldman, 2007).

b. *Tokenizing*

Tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Dalam *tokenizing* juga dapat membuang beberapa karakter yang dianggap sebagai tanda baca.

c. *Stopword Removal atau Filtering*

Tahap pemotongan *string input* berdasarkan tiap kata yang menyusunnya. Dalam *tokenizing* juga dapat membuang beberapa karakter yang dianggap sebagai tanda baca.

Pemilihan algoritma *Naive Bayes Classifier* yang digabungkan dengan metode pemilihan fitur diharapkan dapat mempengaruhi hasil klasifikasi yang signifikan. Proses selanjutnya adalah proses pemilihan fitur menggunakan *Particle Swarm Optimization* (PSO) dengan parameter *Term Frequency* (TF).

Term frequency merupakan salah satu metode untuk menghitung bobot tiap *term* dalam teks. Dalam metode ini, tiap *term* diasumsikan memiliki nilai kepentingan yang sebanding dengan jumlah kemunculan *term* tersebut pada teks (Mark Hall & Lloyd Smith, 1999). Bobot sebuah term t pada sebuah teks dirumuskan dalam persamaan berikut:

$$W(d,t) = TF(d,t)$$

Dimana *Term Frequency* dapat memperbaiki nilai *recall* pada pengambilan informasi namun tidak selalu dapat memperbaiki nilai dari *precision.* Karena disebabkan oleh *Term* yang *frequent* sering muncul di banyak teks. Oleh karena itu term dengan nilai frekuensi yang tinggi disarakan untuk dibuang dari *set term.*

Tahapan selanjutnya adalah tahap kasifikasi menggunakan Algoritma *Naive Bayes Classifier (NBC)* untuk mengklasifikasi data uji pada kategori yang paling tepat (Feldman & Sanger, 2007).

Dalam algoritma *Naïve Bayes Classifier* setiap dokumen direpresentasikan dengan pasangan atribut "x1, x2, x3,...xn" dimana x1 adalah kata pertama, x2 adalah kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori tweet. Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (VMAP), dimana persamaannya adalah sebagai berikut:

$$V_{MAP} = \underset{VjeV}{argmax} \frac{P(X_1, X_2, X_3, \ldots \ldots X_n \mid V_j) P(V_j)}{P(X_1, X_2, X_3, \ldots\ldots X_n)}$$

Untuk P (x1, x2, x3,...xn) nilainya konstan untuk semua kategori (Vj) sehingga persamaan dapat ditulis sebagai berikut:

$$\bar{V}_{MAP} = \underset{VjeV}{\bar{a}rgmax} \prod_{i-1}^{n} P(x_i \mid V_j) P(V_j)$$

Keterangan:

Vj = Kategori tweet j =1, 2, 3,…n.

Dimana dalam penelitian ini

j1 = kategori tweet sentimen negatif,

j2 = kategori tweet sentimen positif, dan

j3 = kategori tweet sentimen netral

P(xi|Vj) = Probabilitas xi pada kategori Vj
P(Vj) = Probabilitas dari Vj

Untuk P(Vj) dan P(xi|Vj) dihitung pada saat pelatihan dimana persamaannya adalah sebagai berikut:

$$P(V_j) = \frac{|docs\ j|}{|conto\ |}$$
$$P(X_i|V_j) = \frac{nk+1}{n+|kosakata|}$$

Keterangan :
|docs j| = jumlah dokumen setiap kategori j
|contoh| = jumlah dokumen dari semua Kategori
nk = jumlah frekuensi kemunculan setiap kata
n = jumlah frekuensi kemunculan kata dari setiap kategori
|kosakata| = jumlah semua kata dari semua kategori

Pada penelitian ini menggunakan seleksi fitur PSO menggunakan parameter *Term Frequency* (TF) dengan *Naïve Bayes Classifier* karena hasil seleksi fitur didapatkan terms lebih banyak yaitu sebanyak 173 kata yang sudah terseleksi. Hasil pengujian seleksi fitur *Particle Swarm Optimization* (PSO) terbukti dapat meningkatkan akurasi algoritma *Naïve Bayes Classifier*.

Pengujian menggunakan seleksi fitur *Particle Swarm Optimization* (PSO) menggunakan parameter *Term Frequency* (TF) dengan *Naïve Bayes Classifier* sebesar 87,91%.

5. PENUTUP

a. Kesimpulan

Pada penelitian ini menghasilkan data-data yang diperoleh dari seleksi fitur *Particle Swarm Optimization* (PSO) menghasilkan dataset yang terseleksi sebanyak 173 kata yang sudah terseleksi sehingga dapat menghasilkan proses klasifikasi yang lebih efektif dan akurat.

Hasil dari pengujian menggunakan parameter *Term Frecuency* (TF) pada klasifikasi algoritma *Naive Bayes Classifier* (NBC) sebesar 87,91%.

b. Saran

Saran untuk pengembangan pada penelitian selanjutnya dapat menggunakan algoritma klasifikasi yang lain sehingga dapat menghasilkan nilai akurasi klasifikasi yang lebih tinggi.

6. REFERENSI

Antony, Mayfield. (2008). *What is Social Media?*. London: iCrossing.

Gunelius, Susan. (2011). *30 Minute Social Media Marketing.* United States: McGraw Hill

Kumar, A., dan Sebastian, T.M. (2012). Sentimen Analysis on Twitter, IJCSI *International Journal of Computer Science Issues,* Vol. 9, No 3, July 2012, ISSN (Online): 1694-0814.

Liu, B. (2011). Web Crawling. In Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (pp. 311-362). Chicago: Springer.

Feldman, R., & Sanger. J. (2007). *Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data.* New York: Cambridge University Press.