Preferensi Konsumen Terhadap Produk By.U Dan MPWR Dengan Analisis Sentimen Berbasis Multinomial Naïve Bayes

by Triawan Cahyanto

Submission date: 13-Feb-2022 04:33AM (UTC+0800)
Submission ID: 1760843116
File name: 6434-17858-1-PB.pdf (445.19K)
Word count: 4295
Character count: 25980

Jurnal Smart Teknologi
Vol. 3, No. 1, November 2021, Halaman 32 – 40
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

Preferensi Konsumen Terhadap Produk By.U Dan MPWR Dengan Analisis Sentimen Berbasis Multinomial Naïve Bayes

Consumer Preferences On By.U And MPWR Products Using Sentimen Analysis Based On Multinomial Naïve Bayes

Alta Randika Setiawan P[1)], Bagus Setya Rintyama[2)], Triawan Adi Cahyanto[3)]
[1]Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
email: altarandika05@gmail.com
[2]Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
email: bagus.setya@unmuhjember.ac.id
[3]Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
email: triawanac@unmuhjember.ac.id

Abstrak

Kebutuhan internet dibutuhkan untuk dapat mengakses informasi di era saat ini. Provider terus meningkatkan produk yang akan mereka jual, salah satunya yaitu membuat provider serba digital. by.U adalah salah satu layanan provider yang baru sebagai penyedia layanan serba digital pertama di Indonesia yang diluncurkan Telkomsel, dan juga terdapat MPWR yang diluncurkan oleh Indosat. Untuk mengetahui preferensi konsumen antara provider by.U dan MPWR dibutukan dari opini pelanggan. Opini pelanggan bisa didapatkan dari sumber twitter. Selanjutnya tweet akan digolongkan ke dalam 4 aspek fitur harga, jaringan, produk dan layanan menggunakan algoritma Cosine Similarity. Dan selanjutnya data akan di analisis sentimen menggunakan metode Multinomial Naïve Bayes. Hasil akurasi terbaik pada dataset by.U 82 %, 80 % untuk presisi dan recall sebesar 75%. Sedangkan untuk dataset MPWR memiliki nilai akurasi terbaik sebesar 87.50 %, presisi 90.91 % dan untuk recall sebesar 90.91 %.

***Keywords**: Twitter, Analisis Sentimen, Cosine Similarity, Multinomial Bayes.*

Abstract

The need for the internet is needed to be able to access information in the current era. Providers continue to improve the products they will sell, one of which is to make providers completely digital. by.U is one of the new service providers as the first all-digital service provider in Indonesia launched by Telkomsel, and there is also an MPWR launched by Indosat. To find out the consumer preferences between by.U and MPWR providers, it is necessary from the customer's opinion. Customer opinion can be obtained from twitter. And then, tweets will be classified into 4 aspects of price, network, product and service features using the Cosine Similarity algorithm. And then the data will be analyzed for sentiment using the Multinomial Naïve Bayes method. The best accuracy results on the dataset by.U are 82%, 60% for precision and a recall of 75%. Meanwhile, the MPWR dataset has the best accuracy value of 87.50%, precision of 90.91% and for recall of 90.91%.

***Keywords**: Twitter, Sentimen Analysis, Cosine Similarity, Multinomial Bayes*

1. PENDAHULUAN

Pertumbuhan pada media sosial secara online seperti pada twitter dapat memunculkan informasi tekstual yang masih harus digali nilai informasi pada twitter. Informasi tekstual dikategorikan antara fakta dan opini. Fakta adalah ekspresi objektif, sedangkan opini adalah ekspresi subjektif [1]. Opini tersebut dapat digunakan dan dimanfaatkan sebagai bahan data pengembangan produk mereka, agar dapat saling bersaing dan bertahan terhadap pesaing lain. Salah satunya

Jurnal Smart Teknologi
Vol. 3, No. 1, November 2021, Halaman 32 – 40
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

perusahaan di bidang telekomunikasi yaitu provider dalam hal internet [2].

Internet dibutuhkan untuk dapat mengakses informasi di era saat ini menggunakan provider. By.U salah satu layanan provider yang baru sebagai penyedia layanan serba digital pertama di Indonesia yang diluncurkan Telkomsel, dan juga terdapat MPWR yang diluncurkan Indosat. Untuk mengetahui preferensi konsumen antara provider by.U dan MPWR dibutukan dari opini pelanggan.

Pelanggan sangat dibutuhkan oleh perusahaan, maka dari itu perusahaan provider harus mencari cara untuk meningkatkan pelayanan dengan mencari informasi tentang kepuasan konsumen. Twitter merupakan suatu media yang dapat menjadi penyebar informasi secara cepat. Informasi yang didapatkan berupa komentar, opini, kritik maupun saran yang bersifat positif dan negatif [2]. Aspek aspek pendorong kepuasan diantaranya adalah kualitas produk, harga, layanan dan kemudahan penggunaan atau jaringan [3].

Data berupa tweet nantinya akan di analisis sentimen menggunakan metode Multinomial Naïve Bayes dengan kategori positif dan negatif dan algoritma untuk untuk mengelompokkan tweet menggunakan algoritma Cosine Similarity dengan menghitung kemiripannya berdasarkan Kata Kunci atau kata kunci.

Penelitian terkait sudah pernah dilakukan oleh Haqqizar N & Larasyati T.K, (2019) mengenai "analisis sentiment terhadap layanan telekomunikasi telkomsel di twitter dengan metode Naïve Bayes", penelitian tersebut menggunakan 151 data tweet dan berisi tentang mengklasifikasi tweet opini pada layanan telkomsel tanpa ada klasifikasi fitur dan memperoleh nilai akurasi tertinggi 70.21%, precision 70.11%, dan recall 70.33%. Sedangkan pada penelitian Rizki Tri Wahyuni, Dhidik Prastiyanto, dan Eko Supraptono (2017) yang berjudul "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi" mengklasifikasikan dokumen secara otomatis ke dalam folder berbeda pada database agar lebih mengelola dokumen yang ada dan tingkat ketepatan senilai 98% dari data berjumlah 50 dokumen skripsi dan terdiri 9 kategori skripsi. Dari sumber penelitian yang didapat,

Berdasarkan latar belakang yang dijelaskan tentang preferensi konsumen terhadap produk by.U dan MPWR , maka penulis akan membuat penelitian dengan judul "Preferensi Konsumen terhadap Produk by.U dan MPWR dengan Analisis Sentimen Berbasis *Multinomial Naïve Bayes*"

2. TINJAUAN PUSTAKA

2.1 Analisis Sentimen

Analisis sentimen adalah suatu proses mengolah dan menafsirkan sikap atau emosi serta data terhadap suatu masalah atau peristiwa untuk mendapatkan informasi. Analasis sentimen harusnya dilakukan untuk mengetahui dan melihat opini atau pendapat terhadap suatu masalah atau peristiwa oleh personal seseorang, apakah berpendapat opini cenderung positif atau opini negatif [4]. Tugas pada analisis sentimen adalah menggolongkan nilai polaritas dari teks yang ada pada dokumen, kalimat atau fitur aspek dapat bersifat positif, negatif atau pun netral[5].

2.2 Text Mining

Text mining merupakan bidang ilmu baru untuk mendapatkan informasi yang bermanfaat dari dokmen/teks asli yang belum dimanipulasi. Dapat dikatakan bahwa text mining merupakan proses menganalisa teks agar mendapatkan suatu informasi berguna untuk tujuan hal tertentu. Text mining itu sendiri adalah varian kreasi baru hasil perkembangan data mining. Dalam proses text mining terdapat istilah *preprocessing data*, yaitu suatu proses pendahuluan yang diterapkan terhadap data pada teks. Proses text mining pada umumnya melakukan tahapan *tokenizing*, *filtering*, dan *stemming* [4].

Text mining fokus terhadap data berbentuk teks atau dokumen,dengan memanfaatkan informasi yang ada pada teks, sumber datanya berasal dari teks atau dokumen tertentu, dengan bertujuan untuk mencari hasil kata-kata yang mengkriteriakan golongan isi teks atau dokumen, dan sehingga dapat

Jurnal Smart Teknologi
Vol. 3, No. 1, November 2021, Halaman 32 – 40
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

dilakukan analisa hubungan antar teks atau dokumen tersebut [4].

2.3 Text Preprocessing

Text preprocessing merupakan suatu proses mengatur, mengolah dan menyeleksi informasi dengan cara terstuktur agar informasi tersebut dapat diklasifikasikan sesuai kebutuhan pemakai. Tahapan pada proses *text preprocessing* diawali dari *case folding, cleansing, tokenizing, normalisasi* bahasa, *stopword removal* dan diakhiri tahapan *stemming* [6].

2.4 Provider by.U

Provider by.U adalah layanan telekomunikasi yang relatif baru dan menarik yang diluncurkan oleh Telkomsel dan mengklaim sebagai penyedia digital pertama di Indonesia. Semua layanan dilakukan secara digital dengan aplikasi by.U yang terdapat pada smartphone, sehingga memudahkan pengguna dalam menggunakan sesuai kebutuhan. Provider by.U menawarkan layanan pemilihan nomor, penetapan kuota internet, pengiriman sim card serta layanan live chat semua serba digital atau online [7]. Dan By.U juga merupakan layanan selular prabayar digital pertama di Indonesia yang menyediakan pengalaman digital end-to-end untuk seluruh kebutuhan telekomunikasi. by.U dikembangkan khusus untuk segmen Gen Z di Indonesia, yang saat ini diproyeksikan berjumlah sekitar 44 juta orang. Generasi ini selalu melakukan semua aktivitas secara online[8].

2.5 Provider MPWR

Provider MPWR adalah layanan telekomunikasi yang relatif baru dan menarik yang diluncurkan oleh Indosat disertai dengan pelayanan serba digital. Layanan yang ditawarkan pemilihan nomor bebas, pemilihan kuota internet, live chat serta terdapat bonus kuota untuk pembelian pertama sebesar 10gb (*mpwr.id*). Dan juga MPWR merupakan digital lifestyle telco pertama di Indonesia yang memberikan empowerment kepada Gen Z dan milenial pada generasi muda Indonesia melalui paket data yang simple dan menawarkan layanan lifestyle di dalam aplikasinya[9].

2.6 Twitter

Twitter adalah layanan suatu microblogging yang telah rilis pada tanggal 13 Juli 2006 secara resmi [10]. Twitter dapat dijadikan sumber yang hampir tak terbatas pada penggunaaan text classification. Pesan yang terdapat di dalam *tweet* memiliki banyak nilai attribute dan bersifat unik, sehingga beberapa hal pembeda dari media sosial lain [10].

2.7 Vector Space Model

Vector Space Model atau VSM adalah model Information Retrieval yang memperlihatkan Kata Kunci dan dokumen sebagai suatu vektor pada ruangan multi dimensi. Kesamaan antara*Kata Kunci* dan dokumen dapat dihitung dengan menggunakan vektor Kata Kunci dan vektor dokumen. Perhitungan kemiripan atau similaritas antara vektor Kata Kunci dan vektor dokumen dapat dilihat menggunakan sudut yang paling kecil. Similaritas antara vektor Kata Kunci dan vektor dokumen akan diukur atau di hitung dengan pendekatan rumus cosine similarity [11]

Menurut [11] pengukuran Cosine Similarity :

$$Sim\ (di,dj) = \frac{Di\ Dj}{||D1||\ ||D2||} = \frac{Wiq.Wij}{\sqrt{Wiq^2} * \sqrt{Wij^2}}$$

Keterangan :

Sim (di,dj)	=	similaritas antara kata kunci dan dokumen
‖D1‖	=	panjang vektor dokumen 1
‖D2‖	=	panjang vektor dokumen 2
Wij	=	bobot term yang ada dida lam dokumen
Wiq	=	bobot kata kunci yang ada didalam dokumen

2.8 TFIDF

TFI-DF atau *Term Frequency Inverse-Document Frequency* merupakan pembobotan yang dapat dilakukan setelah ekstrasi dokumen. Proses *TF-IDF* merupakan perhitungan bobot yang dilakukan dengan cara integrasi antara nilai *Term Frequency(TF)* dan

Jurnal Smart Teknologi
Vol. 3, No. 1, November 2021, Halaman 32 – 40
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

nilai term *Inverse Document Frequency(IDF)*. Hal ini dilakukan berguna untuk mencari jumlah setiap kata yang telah diketahui(*TF*) setelah dikalikan dengan berapa banyak dokumen dimana suatu kata itu muncul (*IDF*) [12].

Menurut [12], rumus dalam perhitungan dalam metode *TF-IDF* :

$$W_{dt} = tf_d \times idf_t = tf_d \times \log(N/df_t)$$

Keterangan :

W_{dt} = Nilai bobot *term* ke-t pada Dok. d
tf_d = Jumlah munculnya *term* t pada Dok. d
N = Jumlah dokumen secara keseluruhan
df_t = Jumlah dokumen yang mengandung *term*

2.9 *Cosine Similarity*

Cosine Similarity adalah metode yang pakai untuk menghitung similaritas antara dua objek. Penghitungan metode ini di dasari pada perhitungan *vector space similarity measure*. Metode ini menghitung similaritas antara dua objek (contoh: *D1* dan *D2*) yang dirubah menjadi dua *vector* dengan menggunakan kata kunci (*keywords*) pada dokumen [13].

Menurut [13] rumus Cosine Similarity sebagai berikut :

$$\text{Cosine a} = \frac{A.B}{|A||B|} = \frac{\sum_{i=n}^{n} A_i \, X \, B_i}{\sqrt{\sum_{i=1}^{n}(Ai)^2 \, X \, \sum_{i=1}^{n}(Bi)^2}}$$

Keterangan

A = Vektor A, yang akan dibandingkan kemiripannya
B = Vektor B, yang akan dibandingkan kemiripannya
A • B = *dot product* antara vektor A dan vektor B
|A| = panjang vektor A
|B| = panjang vektor B
A||B| = *cross product* antara |A| dan |B|

2.10 *Multinomial Naïve Bayes (MNB)*

Multinomial Naive Bayes merupakan satu kondisi probabilitas yang dilakukan tanpa memperhitungkan urutan pada kata dan informasi yang telah ada pada dokumen atau kalimat pada umumnya. Dalam algoritma tersebut juga menghitung jumlah kata yang muncul pada dokumen [14].

Menurut [14], model *Multinomial Naïve Bayes* menggunakan rumus sebagai berikut

$$\mathbf{P(c|term\ dok\ d)} = \mathbf{P(c)}\times\mathbf{P}(t1|c)\times\mathbf{P}(t2|c)\times\mathbf{P}(tn|c)$$

P(c|term dok d) = Probabilitas suatu dokumen dalam kelas c
P(c) = Probabilitas *prior* dari kelas c
P(tn|c) = Probabilitas kata ke-n pada kelas c
tn = kata ke n pada dokumen

Menurut [14], untuk menentukan probabilitas *prior* dari kelas c dihitung dengan menggunakan rumus:

$P(c) = \frac{Nc}{N}$
Nc = Banyak kelas c pada semua dokumen
N = Banyak seluruh dokumen

2.11 Confusion Matrix

Confusion Matrix merupakan tempat yang berisi suatu informasi klasifikasi aktual dan hasilnya telah diprediksi yang dilakukan oleh system klasifikasi. Kinerja pada sistem tersebut pada umumnya dievaluasi secara rinci dengan menggunakan data yang terdapat pada matriks. Tabel berikut menunjukkan confusion matrix untuk 2 kelas

2.12 Cross Validation

Cross validation adalah metode untuk mengevaluasi dan membandingkan pembelajaran dari algoritma (learning algorithms) dengan membagi data menjadi dua bagian, satu bagian digunakan untuk *training*.

2.13 Uji Perfomansi

Pengujian performa model dilakukan dengan pengujian terhadap nilai yang dihasilkan dari kelas hasil prediksi dan nilai asli dari suatu kelas. uji performansi dilakukan dengan mencocokkan antara kedua nilai tersebut. Terdapat beberapa rumus umum yang digunakan untuk menghitung performa model yang dibuat yaitu nilai pada akurasi, *precision*, *recall* dan F-Measure [15].

Jurnal Smart Teknologi
Vol. 3, No. 1, November 2021, Halaman 32 – 40
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

a) Akurasi

Akurasi adalah metode untuk pengujian yang berdasar pada tingkat kedekatan pada nilai aktual dan nilai prediksi. Jika sudah mengetahui hasil jumlah klasifikasi maka akan mendapatkan hasil prediksi akurasi.

Rumus nilai akurasi dapat dihitung sebagai berikut [15].

$$Akurasi = \frac{TP + TN}{TP + FP + FN + TN}$$

b) Precision

Presisi adalah metode pengujian dengan menghitung perbandingan hasil jumlah pada informasi yang relevan berasal dari system dengan nilai jumlah seluruh informasi yang terambil oleh sistem baik relevan maupun tidak. Rumus *precision* dapat dihitung sebagai berikut[15]:

$$Precision = \frac{TP}{TP\ +\ FP}$$

c) Recall

Recall merupakan sebuah system dalam memanggil kembali dokumen yang dianggap relevan ter[1]nali diantara semua dokumen yang relevan. Persamaan *recall* seperti pada persamaan sebagai berikut[15]:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

d) F-Measure

F-Measure merupakan nilai perbandingan rata-rata *precision* dan *recall* yang dibobotkan. *F-Measure* dibutuhkan jika ingin mendapatkan nilai yang seimbang antara *precision* dan *recall*. Perhtiungan *F-Measure* sebagai berikut [15]:

$$\text{F-Measure} = \frac{2(\text{R} \times \text{P})}{\text{R+P}}$$

Keterangan :
TP = *True Positif*
FN = *False Negatif*
FP = *False Positif*
TN = *True Negatif*
R = *Recall*
P = *Precision*

2.14 Python

Python adalah suatu bahasa pemrograman yang interpretatif multiguna dengan menggunakan filosofi perancangan yang berfokus pada tingkat keterbacaan suatu kode. *Python* berbentuk pemrograman beorientasi objek, pemrograman imperatif, dan pemrograman fungsional [16]. *Python* telah digunakan di berbagai keperluan pengembangan perangkat lunak seperti *internet scripting*, sistem pemrograman, *user interface*, pemrograman komputasi dll. Python memiliki fitur sebagai berikut [17].

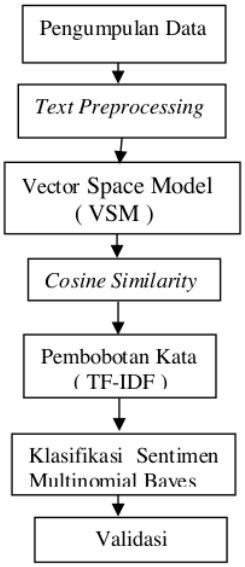
2.15 Jupiter Notebook

Jupyter Notebook merupakan IDE berbasis browser yang dengan kata lain harus memiliki interface untuk menjalankannya. Dengan menggunakan Jupyter Notebook maka proses yang seharusnya menggunakan terminal dapat dijalankan dengan visualisasi di browser.

3. METODOLOGI

3.1 Rancangan Penelitian

Rancangan penelitian yang akan digunakan dalam penelitian ini antara lain mengumpulkan data, text preprocessing, vector space model, cosine similarity, pembobotan kata, klasifikasi, validasi dan evaluasi.



Gambar 1. Alur Penelitian

Jurnal Smart Teknologi
Vol. 3, No. 1, November 2021, Halaman 32 – 40
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

3.2 Pengumpulan Data

Pengumpulan data merupakan suatu hal yang bertujuan untuk memperoleh suatu data yang akan digunakan untuk objek penelitian. Pengumpulan data dilakukan dengan cara crawling data menggunakan website TAGS v6.1. Tahapan crawling data pada Twitter dilakukan dengan cara mengakses API (Application Programming Interface) Twitter yang didapatkan melalui Google Developers. Data yang akan digunakan merupakan data tweet yang berasal dari Twitter dan di dapatkan dari keyword #by.U dan #MPWR dengan waktu pengambilan di bulan Desember 2020. Data yang didapatkan dari tahap crawling data disimpan dengan format CSV (Comma Separated Value).

3.3 Text Processing

Text processing dilakukan untuk mempersiapkan kata pada data narasi sehingga bersih dari noise sebelum dilakukan pembobotan. Proses dari text processing antara lain case folding, tokenizing, normalisasi, stopword removal dan stemming.

3.4 Vector Space Model

Teknik *Vector Space Model* atau VSM memperlihatkan Kata Kunci dan dokumen sebagai suatu vektor pada ruangan multi dimensi. Teknik VSM dapat dihitung dengan pendekatan rumus cosine similarity [11].

3.5 Cosine Similarity

Hasil dari algoritma Cosine Similarity

Rank	D1	D2	D3	D4	Di
Q1	0.50	0.00	0.00	0.00	…
Q2	0.00	1.00	0.00	0.00	…
Q3	0.00	0.00	0.82	0.00	…
Q4	0.00	0.00	0.00	0.71	…
Result	F1	F2	F3	F4	…

Pada Tabel 3.6 Nilai hasil perhitungan Cosine Similarity, merupakan nilai hasil cosine similarity antar Kata Kunci yang terdiri dari Q1,Q2,Q3 dan Q4 dengan Dokumen atau komentar tweet yang terdiri dari D1,D2,D3,D4 dan Di. Pada data tweet komentar D1 termasuk fitur jaringan, data tweet komentar D2 termasuk fitur harga, data tweet komentar D3 termasuk fitur layanan dan data tweet komentar D4 termasuk fitur produk. Nilai cosine similarity menjadi acuan data komentar masuk kategori fitur yang sesuai dengan nilai terbesar yang didapat dari perhitungan cosine similarity.

3.6 Pembobotan Kata TFIDF

Teknik pembobotan digunakan untuk menghitung nilai bobot suatu kata dalam dokumen dengan menggunakan algoritma TF-IDF. Langkah pertama dalam mencari nilai bobot suatu kata adalah dengan menghitung term frequency pada setiap data narasi. Langkah kedua adalah menghitung nilai inverse document frequency (IDF). Langkah terakhir adalah menghitung nilai bobot TF-IDF dengan mengkalikan nilai TF dengan IDF.

3.7 Klasifikasi

Proses klasifikasi yang digunakan pada penelitian ini menggunakan metode Multinomial Naïve Bayes (MNB). Pada tahap klasifikasi akan menentukkan kelas pada data uji.

3.8 Validasi dan Evaluasi

Proses validasi dilakukan pada kumpulan dataset berita hoax dan benar. Untuk mendapatkan hasil akurasi yang optimal maka akan dilakukan proses validasi dengan menggunakan K-Fold Cross Validation. Pada proses validasi tersebut akan membagi dataset menjadi beberapa bagian yang terdiri dari data latih dan data uji. Pada Penelitian ini ditentukan nilai fold K yang digunakan bernilai 2, 4, 5 dan 10. Karena tidak ada aturan formal dalam pemilihan nilai pada K-Fold Cross Validation (Kuhn & Johnson, 2013), maka pemilihan nilai fold K tersebut diambil nilai yang habis dibagi atau tidak menyisahkan nilai, sehingga pada setiap partisi akan memiliki nilai yang seimbang.

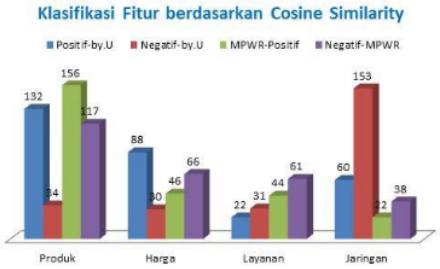
Proses terakhir pada penelitian ini adalah melakukan evaluasi terhadap hasil klasifikasi yang telah dilakukan. Proses evaluasi ini dilakukan untuk mendapatkan nilai akurasi dari penggunaan beberapa algoritma yang telah digunakan sebelumnya. Perhitungan nilai

Jurnal Smart Teknologi
Vol. 3, No. 1, November 2021, Halaman 32 – 40
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

akurasi pada evaluasi merupakan hasil dari perhitungan K-Fold Cross Validation yang memunculkan beberapa nilai akurasi dari beberapa banyak pengujian pada Fold-K. Dari hasil nilai yang didapatkan akan dipilih nilai akurasi yang paling optimum dari beberapa hasil pengujian.

4. HASIL DAN PEMBAHASAN

Data yang akan digunakan berjumlah 550 data by.U dan 550 data MPWR. Dari 550 data tersebut selanjutnya akan dilakukan text preprocessing. Selanjutnya data di klasifikasi berdasarkan aspek menggunakan algoritma Cosine Similarity dengan 4 kategori yaitu harga, jaringan, produk dan layanan. Hasil dari *Cosine Similarity* sebagai berikut



Gambar 2. Klasifikasi Fitur

Untuk dataset by.u didapatkan hasil sebagai berikut :

- 118 meng-tweet terkait harga, disertai 88 tweet memberikan komentar positif dan 30 tweet memberikan komentar negatif
- 166 meng-tweet terkait produk, disertai 132 tweet memberikan komentar positf dan 34 tweet memberikan komentar negatif
- 213 meng-tweet terkait jaringan, disertai 60 tweet memberikan komentar positif dan 153 tweet memberikan komentar negatif
- 53 meng-tweet terkait layanan , disertai 22 tweet memberikan komentar positif dan 31 tweet memberikan komentar negatif.
- Sedangkan untuk dataset MPWR didapatkan hasil sebagai berikut :
- 112 meng-tweet terkait harga, disertai 46 tweet memberikan komentar positif dan 66 tweet memberikan komentar negatif.
- 273 meng-tweet terkait produk, disertai 156 tweet memberikan komentar positif dan 117 tweet memberikan komentar negatif.
- 60 meng-tweet terkait jaringan, disertai 22 tweet memberikan komentar positif dan 38 tweet memberikan komentar negatif.
- 105 meng-tweet terkait layanan, disertai 44 tweet memberikan komentar positif dan 61 tweet memberikan komentar negatif.

Selanjutnya proses klasifikasi diawali dengan pembobotan kata TFIDF dan pembagian data atau partisi data dengan menggunakan *K Fold Cross Validation* untuk mendapatkan model terbaik yang kemudian akan diujikan ke dalam data uji baru. Penggunaan *K Fold* pada penelitian ini adalah 2, 4, 5 dan 10. Setelah dilakukan pembagian data maka masuk kedalam tahap klasifikasi menggunakan algoritma *Multinomial Naïve Bayes (MNB)*. Berikut hasil terbaik yang didapatkan dari hasil klasifikasi algoritma *Multinomial Naïve Bayes (MNB) provider by.U* ditampilkan Tabel 2

Tabel 1. Hasil Klasifikasi *MNB dataset by.U*

KFOLD	AKURASI	PRESISI	RECALL
2	83.94%	98.53%	69.07%
4	87.50%	91%	78.05%
5	88.31%	96.55%	77.78%
10	94.87%	87.50%	100%

Tabel 2. Hasil Klasifikasi *MNB dataset MPWR*

KFOLD	AKURASI	PRESISI	RECALL
2	81.77%	91.25%	72.28%
4	88.54%	92%	87.27%
5	88.31%	81.58%	93.94%
10	92.11%	89.47%	94%

Dari hasil klasifikasi yang didapatkan pada kedua dataset, akan diambil model terbaiknya. Untuk *Multinomial Naïve Bayes, Bernoulli* dan *Rocchio* akan diambil model pada *fold k=10* karena memiliki hasil yang terbaik dari semua *fold k* yang digunakan. Selanjutnya model akan digunakan sebagai data training untuk pengujian terhadap data uji baru. Data yang digunakan sebagai pengujian berjumlah 165. Hasil dari pengujian data baru

terhadap model terbaik ditampilkan pada Tabel 3.

Tabel 3. Hasil Uji Data Baru

DATASET	AKURASI	PRESISI	RECALL
By.U	82.35%	85.00%	75.00%
MPWR	87.50%	91%	90.91%

Berdasarkan Hasil klasifikasi dengan menggunakan data baru didapatkan hasil akurasi, presisi dan *recall* pada dataset by.U menggunakan algoritma *Multinomial Naïve Bayes (MNB)* sebesar 82.35% untuk akurasi, 85.00% untuk presisi dan 75.00% untuk *recall*. Sedangkan pada dataset MPWR algoritma *Multinomial Naïve Bayes (MNB)* mendapatkan hasil sebesar 87.50% untuk akurasi, 91% untuk presisi dan 90.91% untuk *recall*. Hasil yang didapatkan dari validasi data mengalami penurunan, hal ini disebabkan karena adanya karakteristik baru yang ada pada data validasi yang tidak dimiliki pada data model yang menyebabkan penurunan hasil dari segi akurasi, presisi dan *recall*.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan yang dapat diambil dari hasil penelitian yang telah dilakukan adalah:

1. Klasifikasi tweet dengan menggunakan algoritma Cosine Similarity pada dataset provider by.U menghasilkan konsumen menuliskan tweet sebesar 30% dengan kategori produk yang berisi 132 tweet positif dan 34 tweet negatif, 22% dengan kategori harga yang berisi 88 tweet positif dan 30 tweet negatif, 10% dengan kategori layanan yang berisi 22 tweet positif dan 31 tweet negative dan 38% dengan kategori jaringan yang berisi 60 tweet positif dan 153 tweet negatif.
2. Klasifikasi tweet dengan menggunakan algoritma Cosine Similarity pada dataset provider MPWR menghasilkan konsumen menulis tweet sebesar 50% dengan kategori produk yang berisi 156 tweet positif dan 117 tweet negatif, 20 % dengan kategori harga yang berisi 46 tweet positif dan 66 tweet negatif, 19% dengan kategori layanan berisi 44 tweet positif dan 61 tweet negative dan 11% dengan kategori jaringan yang berisi 22 tweet positif dan 38 tweet negatif.
3. Metode Multinomial Naïve Bayes pada dataset mengenai produk by.U rmemberikan hasil akurasi 82,35 %, presisi 60 % dan recall 75 %, sedangkan dataset mengenai produk MPWR memberikan nilai akurasi 87.50 %, presisi 90.91 % dan recall 90.91 %.
4. Secara keseluruhan, preferensi kon-sumen untuk fitur harga pada provider by.U sangat positif dan MPWR sangat positf di kualitas produk yang disajikan.

5.2 Saran

Berdasarkan penelitian yang telah dijabarkan di atas, adapun beberapa saran yang diberikan untuk penelitian selanjutnya adalah sebagai berikut:

1. Menggunakan metode atau algoritma lainnya dalam menentukan klasifikasi aspek fitur pada setiap tweet
2. Menggunakan metode sentimen lainnya seperti KNN, SVM dan NN sebagai perbandingan.
3. Menambahkan kosakata normalisasi

6. DAFTAR PUSTAKA

[1]Cahyono, Y. (2017). Analisis Sentiment pada Sosial Media Twitter Menggunakan Naïve Bayes Classifier dengan Feature Selection Particle Swarm Optimization dan Term Frequency. Jurnal Informatika UniversitasPamulang,2(1),14.https://doi.org/10 .32493/informatika.v2i1.1500

[2]Haqqizar, N., & Larasyanti, T. N. (2019). Analisis Sentimen Terhadap Layanan Provider Telekomunikasi Telkomsel Di Twitter Dengan Metode Naïve Bayes. Prosiding TAU SNAR-TEK 2019 Seminar Nasional Rekayasa Dan Teknologi, 10(2), 1–15.

[3]Wibisono, L. E. (2011). Pengaruh Kualitas Produk, Kualitas Layanan dan Persepsi Harga Terhadap Kepusan Pelanggan 4G XL Di Yogyakarta. Journal of Physics A: Mathematical and Theoretical, 44(8), 51.

[4]Gumelar, T., (2016). "Implementasi Metode Naive Bayes Classifier Untuk Klasifikasi dan Analisis Sentimen pada Sistem Pengaduan RSUD Majalengka". Perpustakaan UNIKOM

[5]Syaifudin, Y. W., & Irawan, R. A. (2018). Implementasi Analisis Clustering Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode K-Means. Jurnal Informatika Polinema, 4(3), 189. https://doi.org/10.33795/jip.v4i3.205

[6]Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya, 2(11), 4704–4713.

[7]Fransiska, S., & Gufroni, A. I. (2020). Sentiment Analysis Provider by . U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method. 7(2), 203–212.

[8]Telkomsel, 2020. "Telkomsel luncurkan by.U, layanan selular prabayar digital end-to-end pertama di Indonesia" . https://www.telkomsel.com/about-us/news/telkomsel-luncurkan-byu-layanan-selular-prabayar-digital-end-end-pertama-di-indonesia. Diakses pada 26 February 2021.

[9]Kevin Rizky Pratama, 2020. "MPWR, Operator Seluler Digital dari Indosat Resmi Meluncur" . https://tekno.kompas.com/read/2020/12/01/16010067/mpwr-operator-seluler-digital-dari-indosat-resmi-meluncur?page=all. Diakses pada

[10]Habibi, R., Setyohadi, D. B., & Wati, E. (2016). Analisis Sentimen Pada Twitter Mahasiswa Menggunakan Metode Backpropagation. Jurnal Informatika, 12(1). https://doi.org/10.21460/inf.2016.121.462

[11]Bania Amburika, Chrisnanto, Y. H., & Uriawan, W. (2016). Teknik Vector Space Model (VSM) dalam Penentuan Penanganan Dampak Game Online Pada Anak. Prosiding SNST Ke-7 Tahun 2016, 1(1), 10–27. http://cogsys.imm.dtu.dk/thor/projects/multimedia/textmining/node5.html

[12]Putra Nuansa, E. (2017). Analisis Sentimen Pengguna Twitter Terhadap Pemilihan Gubernur Dki Jakarta Dengan Metode Naïve Bayesian Classification Dan Support Vector Machine. Institut Teknologi Sepuluh Nopember Surabaya, 1–101.

[13]Sya'bani, M. M., & Umilasari, R. (2018). Penerapan Metode Cosine Similarity dan Pembobotan TF / IDF pada Sistem Klasifikasi Sinopsis Buku di Perpustakaan Kejaksaan Negeri Jember. Justindo(Jurnal Sistem & Teknologi Indonesia), 3(1), 31–42

[14]Shofiya,Feni (2020) Perbandingan Algoritma Support Vector Machine (Svm) Dan Multinomial Naive Bayes (Mnb) Dalam Klasifikasi Abstrak Tugas Akhir (Studi Kasus: Fakultas Teknik Universitas Muhammadiyah Jember).Undergraduate thesis, Universitas Muhammadiyah Jember.

[15]Makhmudah, Umroh. 2019. Analisis Sentimen Terhadap Tweet Kaum Homoseksual Indonesia Menggunakan Metode Support Vector Machine. Fakultas Ilmu Komputer. Jember: Universitas Jember.

[16]Prasetyo, Bagus. (2020). Citra Merek Adidas Pada Suporter dengan Branded Fashion. Fakultas Ilmu Sosial dan Politik. Malang: Universitas Muhammadiyah.

[17]Suhesti, Tyan. 2014. Bahasa Pemrograman Python. Ilmuti, 2013.

Preferensi Konsumen Terhadap Produk By.U Dan MPWR Dengan Analisis Sentimen Berbasis Multinomial Naïve Bayes

ORIGINALITY REPORT

11%	12%	4%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.uii.ac.id Internet Source	1%
2	docplayer.info Internet Source	1%
3	www.suarantb.com Internet Source	1%
4	repository.uin-suska.ac.id Internet Source	1%
5	jurnal.uns.ac.id Internet Source	1%
6	repositori.usu.ac.id Internet Source	1%
7	media.neliti.com Internet Source	1%
8	tekno.kompas.com Internet Source	1%
9	pdfs.semanticscholar.org Internet Source	1%

10	repository.its.ac.id Internet Source	1 %
11	scholar.google.co.id Internet Source	1 %
12	repository.usd.ac.id Internet Source	1 %

Exclude quotes	On	Exclude matches	< 20 words
Exclude bibliography	On		