

ANALISIS SENTIMEN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER PADA ANGKET MAHASISWA

Sunardi ¹⁾, Abdul Fadlil ²⁾, Suprianto ^{2,3)}

¹⁾ Program Studi Teknik Elektro, Universitas Ahmad Dahlan

²⁾ Magister Teknik Informatika, Universitas Ahmad Dahlan

³⁾ Program Studi Sistem Informasi, STMIK PPKIA Tarakanita Rahmawati

Correspondence Author: supri@ppkia.ac.id

ABSTRACT

This study aims to perform a sentiment analysis in the form of student questionnaires to find out student satisfaction in the education process. Sentiment analysis is the process of document classification and in this study is divided into three parts, namely positive, negative and neutral class. Classification is very important in searching documents for users. The classification process begins by dividing the document collection into training data and test data. Training data used naïve bayes classifier method to obtain classification model for class determination on test data. The naïve bayes classifier method is a probabilistic and bayesian theorem. This method is used to classify the results of student opinion data written on the form of student questionnaire satisfaction of lecturers so as to produce the desired classification automatically. The results of this study are used to determine the classification of student questionnaires data so that the data easily legible. The classification of this method has a 75% precision rate, 75% recall and 80% accuracy.

Keyword: Sentiment Analysis, Naïve Bayes Classifier, Student Questionnaire

INTISARI

Penelitian ini bertujuan untuk melakukan analisis sentimen berupa data angket mahasiswa untuk mengetahui kepuasan mahasiswa dalam proses pendidikan. Analisis sentimen adalah proses klasifikasi dokumen dan dalam penelitian ini dibagi menjadi tiga bagian, yaitu kelas positif, negatif dan netral. Klasifikasi sangat penting dalam pencarian dokumen bagi pengguna. Proses klasifikasi diawali dengan membagi koleksi dokumen menjadi data latih dan data uji. Data latih digunakan metode naïve bayes classifier sehingga diperoleh model klasifikasi untuk penentuan kelas pada data uji. Metode naïve bayes classifier adalah metode yang berdasarkan probabilitas dan teorema bayesian. Metode ini digunakan untuk mengklasifikasikan hasil data opini mahasiswa yang dituliskan pada form kepuasan angket mahasiswa terhadap dosen sehingga menghasilkan klasifikasi secara otomatis yang diinginkan. Hasil penelitian ini digunakan untuk menentukan klasifikasi dari data angket mahasiswa sehingga data mudah terbaca. Klasifikasi metode ini mempunyai tingkat precision 75%, recall 75% dan akurasi 80%.

Kata Kunci: Analisis Sentimen, Naïve Bayes Classifier, Angket Mahasiswa

1. PENDAHULUAN

Seiring dengan berkembangnya zaman, manusia dituntut untuk selalu mengembangkan ilmu yang dimilikinya. Sumber daya manusia dalam bidang pendidikan harus meningkatkan ilmu pengetahuan dan cara mengajar yang baik dan efektif khususnya di perguruan tinggi. Hal ini dilakukan untuk menciptakan generasi bangsa yang tumbuh baik sesuai program studi yang diambilnya. Pengajaran yang baik akan membantu mahasiswa untuk mencapai pembelajaran yang maksimal. Kualitas pengajaran dan standarisasi akademik perlu dilakukan evaluasi dan ditingkatkan setiap saat untuk menghasilkan mahasiswa yang berkualitas. Pengajaran yang baik adalah sesuatu yang sulit dan tidak mudah, tergantung kepada para pengajar yang berkaitan langsung dengan proses perkuliahan. Bimbingan dan pelatihan untuk para pengajar sangat diperlukan guna menambah wawasan dan cara masa kini dalam memberikan pengajaran, maka evaluasi kepada para pengajar sangat diperlukan untuk mengidentifikasi kontribusi yang sudah dilakukan dalam pencapaian tujuan pengajaran.

Pada setiap akhir semester, setiap perguruan tinggi umumnya melakukan proses penilaian kinerja dosen untuk mengetahui pencapaian tujuan pengajaran. Sebelum ujian akhir semester, mahasiswa diwajibkan untuk menilai kinerja dosen pada setiap pengajaran yang dilakukan selama satu semester. Ada tiga sub penilaian yang dilakukan, yaitu penilaian dalam hal pengajaran dosen, pelayanan kampus, dan pelayanan asisten laboratorium.

Analisis sentimen merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan informasi (Bo Pang dan Lilian, 2008). Sebuah pengklasifikasian yang sangat diperlukan untuk mengatasi masalah yang terjadi dalam hal perekapan angket mahasiswa.

Penelitian akan melakukan klasifikasi pada data angket mahasiswa menjadi ke dalam tiga kelas, yaitu kelas positif, kelas negatif dan kelas netral. Sistem adalah kumpulan elemen-elemen yang saling berinteraksi dan berinterpedensi dalam lingkungan yang dinamis untuk mencapai tujuan tertentu (Soyusiawaty, Umar, & Mantofani, 2007). Penelitian dilakukan di STMIK PPKIA Tarakanita Rahmawati Tarakan menggunakan aplikasi yang telah dibangun untuk memperoleh data yang akan dijadikan data latih dan data uji pada penelitian ini.

2. LANDASAN TEORI

2.1 Klasifikasi Dokumen

Klasifikasi dokumen adalah bidang penelitian dalam perolehan informasi yang mengembangkan metode untuk menentukan atau mengkategorikan suatu dokumen ke dalam satu atau lebih kelompok yang telah dikenal sebelumnya secara otomatis berdasarkan isi dokumen (Herny Februariyanti dan Eri Zuliarso, 2012). Klasifikasi mempunyai 2 (dua) proses yaitu data latih dan data uji. Data latih digunakan untuk menetapkan standar klasifikasi yang akan digunakan pada data uji. Model klasifikasi yang sudah dibentuk akan dilakukan pengujian untuk mengetahui akurasi dari data uji yang sudah dibentuk.

2.2 Text Mining dan Analisis Sentimen

Text mining mengacu pada proses mengambil informasi berkualitas tinggi dari teks. Informasi yang diambil biasanya mengacu ke beberapa kombinasi *relevansi*, kebaruan, dan *interestingness* (Saraswati, 2011). Analisis sentimen atau yang biasa disebut dengan *Opinion mining* adalah riset *komputasional* dari opini, sentimen, dan emosi yang dituangkan secara tekstual lalu diklasifikasikan menjadi kelompok sentimen positif dan negatif (Feizar, Indrianti, & Yudistira, 2014).

2.3 Naïve Bayes Classifier

Naïve bayes classifier adalah metode klasifikasi yang berdasarkan *probabilitas* dan *Teorema Bayesian* dengan asumsi bahwa setiap variabel X bersifat bebas atau berdiri sendiri dan tidak ada kaitannya dengan variabel lainnya. Metode NBC menempuh dua tahap dalam

proses klasifikasi teks, yaitu tahap pelatihan dan tahap klasifikasi (Amir Hamzah, 2012). Probabilitas adalah kemungkinan terjadinya suatu peristiwa antara 0 s/d 1 (wahyudi & fadlil, 2013). Klasifikasi Gaussian Naive Bayes dapat digunakan untuk memproses atribut numerik pada layanan jaringan komputer (Fadlil, Riadi, & Aji, 2017). Pada tahap pelatihan dilakukan proses analisis terhadap sampel data yang dapat menjadi representasi dokumen. Perhitungan perbandingan antara *term* pada data *testing* dengan setiap kelas yang ada dapat dilakukan dengan persamaan 1 (Eric Meisner, 2010).

$$P(a_j|v_j) = \frac{nc+mp}{n+m} \quad (1)$$

keterangan :

n : jumlah *term* pada data latih dimana $v = v_j$

nc : jumlah *term* dimana $v = v_j$ dan $a = a_j$

p : *probabilitas* setiap kelas dalam data latih

m : jumlah *term* pada data uji

perhitungan untuk menentukan klasifikasi pada data uji dilakukan dengan persamaan 2.

$$V_{nb} = \operatorname{argmax}_{v_j \in V} P(v_j) \prod P(a_i|v_j) \quad (2)$$

2.4 Confusion Matrix

Evaluasi dilakukan untuk mengetahui kinerja dan akurasi dari metode klasifikasi yang telah diterapkan. Confusion matrix merupakan tabel yang terdiri dari banyaknya baris data uji yang diprediksi benar dan salah oleh model klasifikasi yang kemudian digunakan untuk menentukan kinerja dan akurasi dari model klasifikasi seperti ditunjukkan Tabel 1 (Puspitasari, A., Santoso, E., & Indriati, 2017).

Tabel 1. Confusion Matrix

Kategori x	Predicted	
Actual	True Positive	False Positive
	True Negative	False Negative

Keterangan :

TP (True Positive) menunjukkan jumlah data uji yang diklasifikasikan sistem ke dalam kategori x, dan semua data tersebut memang benar termasuk kategori x.

FP (False Positive) menunjukkan jumlah data uji yang tidak diklasifikasikan sistem ke dalam kategori x, tetapi seharusnya semua data tersebut termasuk kategori x.

FN (False Negative) menunjukkan jumlah data uji yang diklasifikasikan sistem ke dalam kategori x, tetapi seharusnya semua data tersebut bukan termasuk kategori x.

TN (True Negative) menunjukkan jumlah data uji yang tidak diklasifikasikan sistem ke dalam kategori x, dan semua data tersebut memang bukan termasuk kategori x

2.5 Precision, Recall dan Accuracy

Precision adalah sebuah tingkat keberhasilan antara informasi yang diminta dengan jawaban yang diberikan. Recall adalah sebuah tingkat keberhasilan dalam menemukan kembali sebuah informasi, sedangkan accuracy adalah sebuah tingkat kedekatan

antara nilai prediksi dengan nilai sebenarnya. Perhitungan akurasi dinyatakan dalam persamaan 3,4 dan 5 (Puspitasari, A., Santoso, E., & Indriati, 2017).

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

3. METODE PENELITIAN

3.1 Objek Penelitian

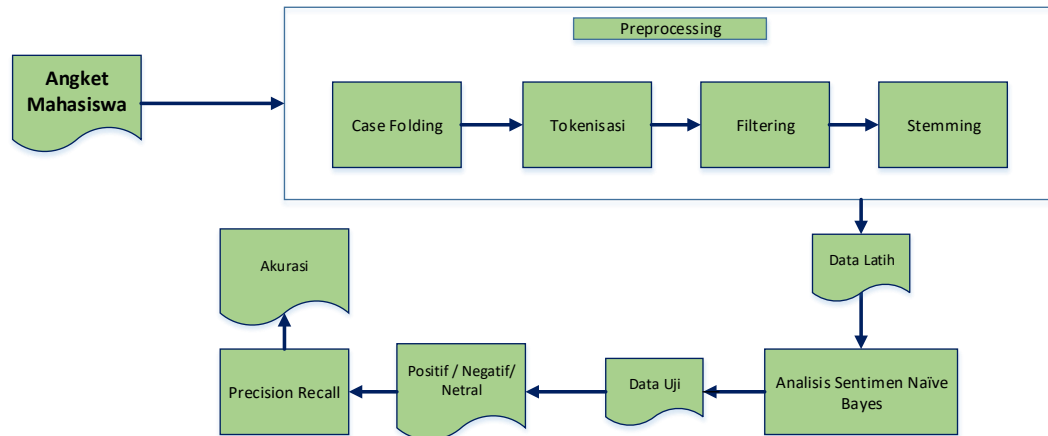
Penelitian ini menggunakan *metode naïve bayes classifier* dengan mengklasifikasikan menjadi tiga kelas, yaitu kelas positif, kelas negatif dan kelas netral. Data yang dibutuhkan dalam penelitian ini adalah data angket mahasiswa yang berisikan komentar mahasiswa pada STMIK PPKIA Tarakanita Rahmawati Tarakan. Terdapat sekitar 800 mahasiswa yang mengisi angket mahasiswa. Sampel yang digunakan pada penelitian ini adalah 100 data angket mahasiswa. Jenis file yang digunakan untuk *input* program adalah file Microsoft Excel (.xls atau .xlsx) kemudian data di *import* ke *database* mysql, selanjutnya data-data akan diolah menggunakan program yang berbasis PHP dan MySQL yang sudah ada dan akan ditambahkan untuk form klasifikasinya. Koleksi data angket selengkapnya dapat dilihat pada Tabel 2 berdasarkan data angket yang diujikan.

Tabel 2. Koleksi Data Angket

No.	Data Angket	Kelas
D1	Metode mengajar sudah baik	Positif
D2	Tidak ada!	Netral
D3	Dosen sering jam tambahan	Negatif
D4	Dosen dalam mengajar terlalu cepat	Negatif
D5	Pertahankan cara mengajar	Positif

Pada Tabel 2 terdapat lima contoh data angket mahasiswa yang telah mempunyai kelas masing-masing, yaitu ada dua data angket pada kelas “positif”, dua pada kelas “negatif” dan satu pada kelas “netral”.

Dalam klasifikasi ini, ada beberapa alur yang dilakukan dimulai dari *input* data, proses, sampai pada tahap akhir yaitu hasil yang diharapkan, seperti pada Gambar 1.



Gambar 1. Alur pengklasifikasian data angket mahasiswa

Tahap pertama dimulai dengan melakukan *pre-processing* data angket mahasiswa dari *case folding* yaitu mengubah semua huruf dalam dokumen menjadi huruf kecil dan menghilangkan tanda baca yang terdapat dalam dokumen, selanjutnya dilakukan Tahap *tokenizing / parsing* adalah tahap pemotongan *string input* berdasarkan tiap kata yang menyusunnya (Triawati, 2009) dan *filtering* untuk mengambil kata-kata penting setelah dilakukan *tokenisasi*.

Proses ini dapat dilakukan dengan algoritma *stoplist* untuk membuang kata yang kurang penting dan *wordlist* untuk menyimpan kata penting, misalnya untuk kata yang kurang penting adalah “di,” “yang” dan lain sebagainya. Tahap berikutnya dilakukan *stemming* untuk menghilangkan tanda awalan atau imbuhan sehingga menjadi kata dasar. Data tersebut kemudian dibuatkan sebagai data latih untuk menentukan nilai *probabilitas* klasifikasi yang dibuat dan digunakan untuk klasifikasi data uji. Tahap terakhir yaitu dilakukan pengukuran persentase ketepatan dalam menentukan hasil dari data uji.

4. HASIL DAN ANALISA

4.1 Pre-Processing

Pre-Processing dilakukan kepada sekumpulan data angket mahasiswa yang telah memiliki kelas masing-masing yang dilakukan secara manual. Pada Tabel 2 terdapat 6 contoh data angket mahasiswa yang mempunyai kelas masing-masing yaitu ada 2 data angket mahasiswa yang mempunyai kelas “positif”, 1 data angket mahasiswa yang mempunyai kelas “netral” dan 3 data angket mahasiswa yang mempunyai kelas “negatif”. Dari data tabel 3.1, dilakukan tahapan selanjutnya dalam *pre-processing* yaitu tokenisasi. Hasil tokenisasi ditunjukkan pada Tabel 3.

Tabel 3. Hasil *Tokenisasi* Data Angket Mahasiswa

No.	Data Angket
D1	metode mengajar sudah baik
D2	tidak ada
D3	dosen sering jam tambahan
D4	dosen dalam mengajar terlalu cepat
D5	pertahankan cara mengajar

Tabel 3 merupakan data angket mahasiswa yang telah melewati tahap pertama dalam proses *pre-processing* yaitu *tokenisasi*. Data dipecah menjadi *token (term)*, huruf pada *term* diubah menjadi huruf kecil dan semua tanda baca dihilangkan. Selanjutnya yaitu masuk

kedalam tahapan kedua dalam *pre-processing* yaitu penghapusan *stopwords*. Bentuk hasil penghapusan *stopwords* ditunjukkan pada Tabel 4.

Tabel 4. Hasil Penghapusan *Stopwords* Data Angket Mahasiswa

No.	Data Angket
D1	metode mengajar baik
D2	
D3	dosen jam tambahan
D4	dosen mengajar cepat
D5	pertahankan mengajar

Tabel 4 merupakan data forum yang telah melewati tahap kedua dalam proses *pre-processing* yaitu penghapusan *stopwords*. Ada beberapa kata yang dihilangkan, contohnya saja pada data forum yang ke-D1, kata “sudah dihapus. Langkah terakhir yaitu tahapan stemming yaitu merubah term menjadi bentuk kata dasar dengan menghilangkan awalan dan akhiran dari *term*. Bentuk hasil *stemming* data angket mahasiswa ditunjukkan pada Tabel 5.

Tabel 5. Hasil *Stemming* Data Angket Mahasiswa

No.	Data Angket
D1	metode ajar baik
D2	
D3	dosen jam tambah
D4	dosen ajar cepat
D5	tahan ajar

Pada Tabel 5, diperoleh kumpulan kata yang telah dilakukan proses *stemming*. Ada beberapa kata yang mengalami proses *stemming*, contohnya pada data forum ke-D1 yaitu kata “mengajar” berubah bentuk menjadi “ajar” dengan menghilangkan awalan “meng”.

4.2 Naïve Bayes Classifier (NBC)

Tahapan awal pada proses NBC yaitu dengan menghitung *probabilitas* masing-masing kelas terhadap keseluruhan data latih. Probabilitas bisa juga disebut dengan ketidakpastian (Yuwono, Fadlil, & Sunardi, 2017). Pada Tabel 1, diketahui bahwa data latih sebanyak 5 data dengan jumlah kelas sebanyak 3 kategori yaitu “Positif”, “Negatif”, dan “Netral”. Berikut adalah perhitungan *probabilitas* untuk kelas “Positif”.

Probabilitas disimbolkan sebagai p .

$$p(\text{Positif}) = \frac{\text{Jumlah kelas positif}}{\text{Jumlah data latih}}$$

$$p(\text{Positif}) = \frac{2}{5} = 0,4$$

Probabilitas untuk kelas “positif” dihitung dan memperoleh nilai yaitu 0,4. Perhitungan juga dilakukan untuk kelas “negatif” dan kelas “netral” dengan metode perhitungan yang sama dengan kelas “positif”. Dari perhitungan tersebut, diperoleh nilai untuk masing-masing kelas yaitu, untuk kelas “negatif” sebesar 0,4 dan kelas “netral” sebesar 0,2.

Dari perhitungan *probabilitas* yang diperoleh, maka langkah selanjutnya adalah melakukan klasifikasi dengan data uji. Berikut contoh data angket untuk diklasifikasikan menggunakan metode *naïve bayes classifier* secara otomatis pada Tabel 6.

Tabel 6. Data uji angket mahasiswa

No.	Data Angket	Kelas
D6	Dosen sangat sering terlambat mengajar!!!	?
D7	Materi yang diajarkan sudah sangat baik, terimakasih	?

Tabel 7 merupakan tabel data uji yang akan diklasifikasikan dengan metode *naïve bayes classifier*, tahap pertama yaitu dengan menentukan nilai n , nc , p , dan m untuk setiap kelas yang ada dan juga pada setiap *term* pada data uji.

Tabel 7. Nilai untuk kelas “positif”

Term	n	nc	p	m
Dosen	5	0	0.4	3
Lambat	5	0	0.4	3
Ajar	5	2	0.4	3

Dari nilai yang telah diperoleh untuk kelas “positif”, maka selanjutnya akan dilakukan perhitungan dengan menggunakan persamaan 1 sebagai berikut :

$$P(\text{Positif} \mid \text{Dosen}) = \frac{0 + 3 \cdot 0.4}{5 + 3} = \frac{1.2}{8} = 0.15$$

Pada perhitungan dengan *term* “lambat” dan “ajar”, dilakukan proses yang sama seperti perhitungan $P(\text{Positif} \mid \text{Dosen})$ dengan menggunakan persamaan 1. Selanjutnya langkah yang dilakukan yaitu mencari nilai maksimal dari hasil perkalian nilai *probabilitas* dan nilai P untuk setiap kelas dengan menggunakan persamaan 2, sebagai berikut :

$$V(\text{Positif}) = 0,4 \cdot 0,15 \cdot 0,15 \cdot 2,15 = 0,01935$$

$$V(\text{Negatif}) = 0,4 \cdot 2,133 \cdot 0,133 \cdot 1,133 = 0,12895$$

$$V(\text{Netral}) = 0,2 \cdot 0,2 \cdot 0,2 \cdot 0,2 = 0,0016$$

$$V_{nb} = \text{argmax} (V(\text{Positif}) \mid V(\text{Negatif}) \mid V(\text{Netral}))$$

$$V_{nb} = \text{argmax} (0.01935 \mid 0.12895 \mid 0.0016)$$

$$V_{nb} = 0.12895$$

Dari perhitungan dengan menggunakan persamaan 2, diperoleh nilai maksimal yaitu 0.12895 adalah nilai v untuk kelas “negatif”. Jadi kesimpulannya untuk data uji D6 termasuk kelas “negatif”. Pada data uji D7 dilakukan proses perhitungan yang sama dengan data uji D6.

4.3 Precision, recall dan Accuracy

Precision, *recall* dan *Accuracy* digunakan untuk pengukuran efektifitas klasifikasi yang dilakukan. Perhitungan akurasi dilakukan dengan persamaan 3,4 dan 5. Hasil klasifikasi dengan metode *naïve bayes classifier* untuk 10 data yang diujikan terdapat pada Tabel 8.

Tabel 8. Hasil Klasifikasi Data Uji

No	Kelas Sebenarnya	Kelas perhitungan NBC
D6	Negatif	Negatif
D7	Positif	Positif
D8	Positif	Positif
D9	Negatif	Negatif
D10	Netral	Positif
D11	Negatif	Negatif
D12	Negatif	Positif
D13	Netral	Netral
D14	Positif	Positif
D15	Negatif	Negatif

Dari Tabel 8 diperoleh nilai TP = 3 , TN = 5 , FP = 1 , FN = 1 . Langkah selanjutnya akan dilakukan perhitungan *precision*, *recall* dan *accuracy* dengan persamaan 3,4 dan 5.

$$Precision = \frac{3}{3 + 1} \times 100\% = \frac{3}{4} = 75\%$$

$$Recall = \frac{3}{3 + 1} \times 100\% = \frac{3}{4} = 75\%$$

$$Accuracy = \frac{3 + 5}{3 + 5 + 1 + 1} \times 100\% = \frac{8}{10} = 80\%$$

Kesimpulan yang diperoleh dari hasil perhitungan data uji sebanyak 10 data yaitu, *precision*=75%, *recall* = 75% dan *accuracy* = 80%.

5. KESIMPULAN

Berdasarkan pembahasan pada bab sebelumnya sesuai dengan harapan yang diinginkan, dapat disimpulkan bahwa dengan menggunakan metode *naïve bayes classifier*, kita dapat melakukan klasifikasi data secara otomatis dari angket mahasiswa menjadi kelas positif, negatif dan netral dengan tingkat akurasi yang tinggi yaitu *precision* 75%, *recall* 75% dan *accuracy* 80%. Pengembangan lebih lanjut, dapat dilakukan penambahan proses pencarian data angket mahasiswa secara otomatis dengan menggunakan *query*, hal ini dilakukan agar tidak semua data angket mahasiswa dilakukan klasifikasi, tetapi hanya yang dibutuhkan oleh user saja.

DAFTAR PUSTAKA

- A Fadlil, I Riadi, S Aji., 2017, DDoS Attacks Classification using Numeric Attribute-based Gaussian Naive Bayes, Japan, *International journal of advanced computer science and applications* 8 (8), 42-50.
- Amir Hamzah. 2012. Klasifikasi Teks dengan Naïve Bayes Classifier (NBC) untuk Pengelompokan Teks Berita dan Abstract Akademis. *Prosiding Seminar Nasional Apikasi Sains & Teknologi (SNAST) Periode III*, p.B269-B277. Yogyakarta.
- DT Yuwono, A Fadlil, S Sunardi, 2017, penerapan metode forward chaining dan certainty factor pada sistem pakar diagnosa hama anggrek coelogyne pandurata, Banjarmasin, *klik-kumpulan jurnal ilmu komputer* 4 (2), 136-145.
- Eric Meisner. 2010. Naïve Bayes Classifier Example.
- Feizar, F. H., Indrianti, & Yudistira, N. (2014). Analisis Sentimen Opini Film Berbahasa Indonesia Berbasis Kamus Menggunakan Metode Neighbor-Weighted K-Nearest Neighbor. Universitas Brawijaya, Teknik Informatika. Malang: Universitas Brawijaya.
- Herny Februariyanti dan Eri Zuliarso. 2012. Klasifikasi Dokumen Berita Teks Bahasa Indonesia menggunakan Ontologi. *Jurnal Teknologi Informasi DINAMIK* Volume 17, No. 1, pp.14-23.
- MJ Wahyudi, A Fadlil, 2013, Sistem pakar untuk mengidentifikasi penyakit udang galah dengan metode theorema bayes, Yogyakarta, *Jurnal Sarjana Teknik Informatika*, 1 (1).
- Pang, B., & Lee, L. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2), 1-135.
- Puspitasari, A., Santoso, E., & Indriati, I. Klasifikasi Dokumen Tumbuhan Obat Menggunakan Metode Improved K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 2, p. 486-492, agu. 2017. ISSN 2548-964X
- Soyusiawaty, Umar, & Mantofani, 2007, sistem informasi geografis objek wisata propinsi kepulauan bangka belitung berbasis web, Yogyakarta, *Prosiding SNATI*, K-17 | vol: | issue : | 2007
- Sumartini Saraswati, N. W. (2011). Text Mining dengan Metode Naive Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis. Denpasar, Bali, Indonesia.