

Implementasi *Particle Swarm Optimization (PSO)* pada *Analysis Sentiment Review Aplikasi Trafi* menggunakan *Algoritma Naive Bayes (NB)*

Eka Rini Yulia¹, Kusmayanti Solecha²

¹Sistem Informasi, STMIK Nusa Mandiri

²Fakultas Teknologi Informasi, Universitas Bina Sarana Informatika

¹e-mail: eka.erl@nusamandiri.ac.id

²e-mail: kusmayanti.ksc@bsi.ac.id

Diterima	Direvisi	Disetujui
29-09-2020	02-11-2020	23-01-2021

Abstrak - Pengembangan Aplikasi Transportasi saat ini semakin besar sehingga banyak sekali vendor-vendor bersaing bisnis dalam menciptakan aplikasi moda transportasi, mulai dari kualitas dan kuantitasnya sehingga sering dipertanyakan. Dengan ini peneliti mengadakan pemutaran aplikasi transportasi yang bernama Trafi untuk mendapatkan pendapat atau komentar aplikasi Trafi dari masyarakat yang telah menggunakan aplikasi dan dituangkan ke dalam media online. Dari sekian banyak komentar yang telah ditinjau untuk mendapatkan satu set data bentuk positif dan negatif dari teks yang akan peneliti olah. Untuk data klasifikasi menggunakan Naive Bayes (NB), NB salah satu algoritma paling populer untuk pengenalan pola. Selain kesederhanaan, Naive Bayes classifier adalah teknik pembelajaran mesin populer untuk klasifikasi teks, Particle Swarm Optimization (PSO) yang dikombinasikan dengan Naive klasifikasi Bayes bersangkutan untuk meningkatkan kinerja. Sebelum digunakan optimasi dengan PSO dalam akurasi set data yang diperoleh 69,50% dan setelah dikombinasikan antara Naive Bayes dan akurasi PSO adalah 72,34%. Gunakan PSO dan Naive Bayes sesuai dengan konsep text mining yang bertujuan untuk menemukan pola-pola yang ada dalam teks, kegiatan yang dilakukan oleh text mining di sini adalah klasifikasi teks.

Kata Kunci: Sentimen Analisis, Review Produk Appstore Android, Algoritma Naive Bayes

Abstract - The development of transportation applications is now getting bigger so that many vendors compete for business in creating transportation mode applications, starting from the quality and quantity so that it is often questioned. With this, the researcher held a transportation application called Trafi to get opinions or comments on applications from people who had used the application and poured it into online media. Of the many comments reviewed to obtain a set of positive and negative forms of data from the text that the researcher will process. For classification data using Naive Bayes (NB), NB is one of the most popular algorithms for pattern recognition. Apart from simplicity, the Naive Bayes classifier is a popular machine learning technique for text classification, Particle Swarm Optimization (PSO) which combines with the Naive Bayes classification to improve performance. Before use, optimization with PSO in the data set accuracy obtained was 69.50% and after the combination of Naive Bayes and PSO accuracy was 72.34%. Use PSO and Naive Bayes according to the concept of text mining which aims to find patterns that exist in text, the activity carried out by text mining here is text classification.

Keywords: Sentiment Analysis, Android Appstore Product Review, Naive Bayes Algorithm

PENDAHULUAN

Berdasarkan penelitian yang sudah dilakukan dalam klasifikasi sentimen terhadap review produk yang tersedia secara *online* diantaranya, analisa sentiment aplikasi smartphone dengan membandingkan metode Support Vector Machine (SVM) dan Naive Bayes (Zhang dkk). Kategorisasi

teks merupakan solusi yang tepat untuk mengelola informasi yang saat ini berkembang dengan sangat cepat dan melimpah. Kategorisasi teks membuat pengelolaan informasi tersebut menjadi efektif dan efisien. Dengan menggunakan kategorisasi teks, dapat dilakukan penyusunan dokumen menurut kategorinya, penyaringan terhadap email *spam*, melakukan penggalan opini (*opinion mining*) dan

analisis sentimen. Algoritma kategorisasi teks saat ini telah banyak berkembang, antara lain: *Support Vector Machines* (SVM), *Naive Bayesian* (NB), pohon keputusan, *K-Nearest Neighbour* (k-NN). Untuk itu pada penelitian ini digunakan metode *Naive Bayes* (NB) karena

Menurut Liu (Liu, 2012) Analisis sentiment atau opinion mining adalah studi komputasional dari opini-opini orang, *appraisal* dan emosi melalui entitas, *event*, dan atribut yang dimiliki.

Pengklasifikasi *Naive Bayes* sangat sederhana dan efisien, (Chen et al., 2009). Di samping kesederhanaannya, pengklasifikasi *Naive Bayes* adalah teknik *machine learning* yang populer untuk klasifikasi teks, dan memiliki performa yang baik pada banyak domain (Ye, Zhang, & Law, 2009).

Particle swarm optimization (PSO) merupakan teknik komputasi evolusioner yang mampu menghasilkan solusi secara global optimal dalam ruang pencarian melalui interaksi individu dalam segerombolan partikel. Setiap partikel menyampaikan informasi berupa posisi terbaiknya kepada partikel yang lain dan menyesuaikan posisi dan kecepatan masing-masing berdasarkan informasi yang diterima mengenai posisi yang terbaik tersebut (Shuzhou & Bo, 2011). PSO banyak digunakan untuk memecahkan masalah optimasi serta pada seleksi fitur (Liu, et al., 2011).

Dengan demikian pada penelitian ini pengklasifikasi *Naive Bayes* dengan pengoptimalan *fitur selection Particle Swarm Optimization* (PSO) akan diterapkan pada sentimen analisis klasifikasi teks pada *review* Aplikasi Trafi bertujuan meningkatkan nilai akurasi.

1. Analisa Sentiment

Analisis sentiment adalah jenis pengolahan bahasa alami untuk melacak mood masyarakat tentang produk tertentu atau topik. Analisis sentimen, yang juga disebut tambang pendapat, melibatkan dalam membangun system untuk mengumpulkan dan meneliti pendapat tentang produk yang dibuat dalam posting blog, komentar, ulasan atau tweet. Analisis sentiment dapat berguna dalam beberapa cara. Misalnya, dalam pemasaran membantu injudging keberhasilan kampanye iklan atau peluncuran produk baru, menentukan versi produk atau jasa yang populer dan bahkan mengidentifikasi demografi suka atau tidak suka terhadap fitur tertentu (Vinodhini, Chandrasekaran, 2012).

2. Pre-processing

Proses pengubahan bentuk bentuk menjadi data yang terstruktur sesuai kebutuhannya untuk proses dalam data mining, yang biasanya akan menjadi nilai-nilai numerik, proses ini sering disebut dengan *text processing* (Ronen Feldman, 2007). Setelah data menjadi data terstruktur dan berupa nilai numerik maka data dapat disajikan sebagai sumber data yang dapat diolah lebih lanjut.

Beberapa proses yang dilakukan adalah sebagai berikut:

1. Tokenizazion

Peneliti menggunakan *Tokenize* untuk memisahkan kata atau huruf dari tanda baca dan simbol.

2. Stopwards Removal

kata yang dianggap tidak perlu dalam pengolahan data sentimen review, sebagai contoh *if, the, of, or, etc.*

3. Steamming

Proses pengubahan bentuk kata menjadi kata dasar. Metode pengubahan bentuk kata menjadi kata dasar ini menyesuaikan struktur bahasa yang digunakan dalam proses stemming (Langgeni et al., 2010).

3. Algoritma Naive Bayes (NB)

Naive Bayes adalah pendekatan yang mengacu pada teorema Bayes, dengan mengkombinasikan pengetahuan sebelumnya dengan pengetahuan baru. Sehingga merupakan salah satu algoritma klasifikasi yang sederhana namun memiliki akurasi tinggi. (Rodiyansyah & Winarko, 2012).

4. Particle swarm optimization (PSO)

Particle Swarm Optimization dapat diasumsikan sebagai kelompok burung secara mencari makanan disuatu daerah. Burung tersebut tidak tahu dimana makanan tersebut berada, tapi mereka tahu seberapa jauh makanan itu berada, strategi terbaik untuk menemukan makanan tersebut adalah dengan mengikuti burung yang terdekat dari makanan tersebut (Salappa, Doumpos, & Zopounidis, 2007).

5. Validasi dan Evaluasi Text Mining

Dari beberapa metode yang digunakan untuk memvalidasi suatu model berdasarkan data yang didapat, seperti *holdout*, *cross validation*, *random sub-sampling*, dan lain-lain. *Confusion matrix* memberikan keputusan yang diperoleh dalam training dan testing, *Confusion matrix* memberikan penilaian performan cek klasifikasi berdasarkan objek dengan benar atau salah (Gorunescu, 2011). *Confusion matrix* berisi informasi aktual (*actual*) dan prediksi (*predicted*) pada system klasifikasi.

Berikut adalah persamaan model Confution Matrix:

1. Nilai *Accuracy* adalah proporsi jumlah prediksi yang benar. Dapat dihitung dengan menggunakan persamaan:

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

METODOLOGI PENELITIAN

Dalam penelitian ada kegiatan penyelidikan (*investigation*), yaitu mencari fakta secara teliti dan teratur menurut kaidah tertentu untuk menjawab suatu pertanyaan. Jadi penyelidikan dilakukan untuk menjelaskan sesuatu. Secara umum metode penelitian diartikan sebagai cara ilmiah untuk mendapatkan data dengan tujuan dan kegunaan tertentu.

Metode penelitian yang penulis lakukan adalah metode penelitian eksperimen, dengan tahapan sebagai berikut:

1. Pengumpulan data set, untuk kemudian diseleksi dari data yang tidak sesuai.
2. Pengolahan data awal, dipilih berdasarkan kesesuaian data dengan metode yang paling baik dari beberapa metode pengklasifikasian teks yang sudah digunakan oleh beberapa peneliti sebelumnya. Model yang digunakan adalah algoritma-Nearest Neighbors (k-NN).
3. Metode Yang diusulkan, untuk meningkatkan akurasi dari Algoritma-Nearest Neighbors (k-NN), maka dilakukan penambahan dengan menggabungkan metode pemilihan fitur filter dan wrapper, yaitu Information gain dan Genetic algorithm.
4. Eksperimen dan Pengujian Metode, dalam eksperimen data penelitian, penulis menggunakan RapidMiner 5 untuk mengolah data. Sedangkan untuk pengujian metode, penulis membuat aplikasi menggunakan bahasa pemrograman PHP dan HTML.

Evaluasi dan Validasi Hasil Evaluasi, dilakukan untuk mengetahui akurasi dari model algoritma-Nearest Neighbors (k-NN). Proses validasi digunakan untuk melihat perbandingan hasil akurasi dari model yang digunakan dengan hasil yang telah ada sebelumnya. Teknik validasi yang digunakan adalah Cross Validation.

Data *review* aplikasi Trafi yang dikumpulkan dari situs <https://play.google.com/store/apps/details?id=com.t.rafid.android.tr&hl=in> Terdiri dari 100 *review positif* dan 100 *review negatif*.

Pengolahan awal data merupakan tindak lanjut dari pengumpulan data, untuk mengurangi waktu pengolahan data. Sample yang digunakan sebanyak 100 *review positif* dan 100 *review negatif*. Dalam tahap *Preprocessing* harus melalui 3 proses, yaitu:

- a. *Tokenization*
- b. *Stopwords Removal*
- c. *Stemming*

HASIL DAN PEMBAHASAN

Pada hasil penelitian kali ini, peneliti menjabarkan hasil dari pengujian algoritma dan optimasinya yang menghasilkan nilai akurasi dan AUC. Ada 2 pengujian yang diterapkan yaitu pertama menggunakan metode Naive Bayes, dan yang kedua metode Naive Bayes yang dioptimalkan dengan *Particle Swarm Optimization*(PSO).

Data training yang digunakan dalam pengklasifikasian *text* terdiri atas 100 *review positif* pada aplikasi Trafi dan 100 *review negatif* pada aplikasi traifi. Data *review* yang akan diolah masih berupa sekumpulan *text* yang dipisah dalam bentuk dokumen. Sebelum diklasifikasikan, data tersebut harus melewati proses tahapan agar data dapat diolah dengan baik.

Hasil Eksperimen yang pertama pengujian metode Naive Bayes ditunjukkan pada tabel dibawah yang nilai akurasinya 69.50% dan AUC 0.488. Berikut Hasil Akurasi pada metode Naive Bayes (NB):

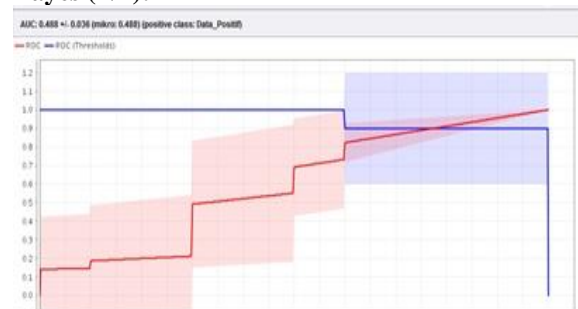
Tabel 1 Hasil Akurasi menggunakan NB

accuracy: 69.50% +/- 11.00% (mikro: 69.50%)			
	True Data_Negatif	True Data_Positif	class precision
pred_Data_Negatif	71	32	69.93%
pred_Data_Positif	29	68	70.10%
class recall	71.00%	68.00%	

Sumber: Hasil Penelitian (2020)

Berdasarkan tabel diatas, menunjukkan bahwa nilai akurasi menggunakan Naive Bayes adalah 69.50%.

Berikut Gambar Hasil ROC pada metode Naive Bayes (NB):



Sumber: Hasil Penelitian (2020)

Gambar 1 Hasil ROC Pengujian Naive Bayes

Kurva ROC pada gambar 4.1 mengekspresikan *confusion matrik* dari tabel 4.4 Diartikan dalam gambar untuk garis horizontal adalah *false positive* dan garis vertikal merupakan garis *true positive*.

Dilihat dari gambar kurva ROC diatas, nilai AUC (*Area Under Curve*) sebesar 0.488 termasuk kedalam *fair Classification*.

Sedangkan hasil eksperimen yang kedua pengujian metode Naive Bayes yang dioptimalkan dengan PSO yang nilai akurasinya mencapai 76.24% dan

AUC 0.636. Berikut hasil akurasi pada metode Naive Bayes (NB):

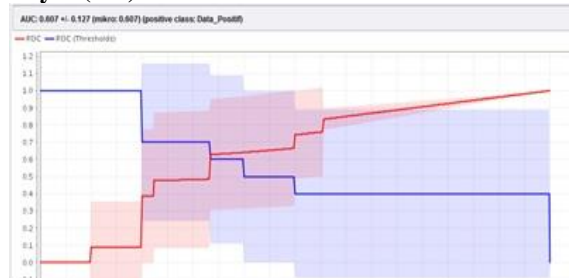
Tabel 2 Hasil Akurasi menggunakan NB - PSO

accuracy 72.34% +/- 0.23% (mikro: 72.31%)			
	True Data_Negatif	True Data_Positif	class precision
pred Data_Negatif	68	26	72.34%
pred Data_Positif	28	73	72.28%
class recall	79.82%	73.74%	

Sumber: Hasil Penelitian (2020)

Berdasarkan tabel diatas, menunjukkan bahwa nilai akurasi menggunakan Naive Bayes berbasis Particle Swarm Optimization adalah 72.34%.

Berikut Gambar HasilROC pada metode Naive Bayes (NB) - PSO:



Sumber: Hasil Penelitian (2020)

Gambar 2 Hasil ROC PengujianNB-PSO

Kurva ROC pada gambar 2. mengekspresikan *confusion matrik* dari tabel 1. Diartikan dalam gambar untuk garis horizontal adalah *false positive* dan garis vertikal merupakan garis *true positive*. Dilihat dari gambar kurva ROC diatas, nilai AUC (*Area Under Curve*) sebesar 0.607 termasuk kedalam *Good Classification*.

Hasil Analisis Evaluasi Hasil dan Validasi Model

Dari hasil pengujian yang peneliti lakukan dari awal pembahasan, pengukuran akurasi menggunakan confusion matrix dan kurva ROC membuktikan bahwa hasil pengujian algoritma Naive Bayes, peneliti menilai belum begitu maksimal dengan sebesar 69.50%. Dan setelah dioptimalkan dengan *Particle swarm optimization*(PSO) hasil akurasinya meningkat 2.84%, yaitu dengan akurasi 72.34%. Dengan dilihat dari hasil penelitian tersebut maka Optimasi PSO dapat meningkatkan nilai akurasi pada algoritma Naive Bayes.

Berdasarkan tabel 2 menunjukkan tingkat akurasi yang cukup baik (72.34%), dan dilihat dari hasil yang muncul pada hasil tabel diatas bisa dicari nilai accuracy, dijabarkan pada perhitungan berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{68+73}{68+28+73+26} = \frac{141}{195} = 0.7614 = 72.31\%$$

Keterangan :

- TP: True Positif
- TN: True Negatif
- FP: False Positif
- FN: False Negatif

Kesimpulan pengujian diatas adalah algoritma Naive Bayes dapat meningkatkan nilai akurasi yang merupakan solusi yang baik dalam permasalahan pada klasifikasi sentimen *review* aplikasi pada android.

KESIMPULAN

penelitian kali ini mencakup aspek yang terkait di antaranya:

1. Implikasi terhadap aspek sistem Hasil evaluasi menunjukkan penerapan Algoritma Naive Bayes (NB) merupakan metode yang cukup baik dalam mengklasifikasi teks *review* Appstore for Android. Metode ini dapat membantu para calon pengguna android dalam menentukan aplikasi apa saja yang layak mereka *install*, supaya tidak lagi asal menginstall aplikasi yang banyak tersedia diberberapa media online, yang dampak buruknya bisa mengakibatkan memory dan kualitas mobile pengguna melemah kualitasnya, dan mengurangi waktu dalam membaca *review* dan komentar mengenai aplikasi pada android.
2. Implikasi terhadap aspek penelitian lanjutan Penelitian selanjutnya bisa menggunakan metode pemilihan fitur ataupun dataset dari domain yang berbeda, seperti *review* hotel, *review* restoran, dan banyak lainnya yang bisa dicari dalam bidang pengembangannya.Pada bagian ini, dijelaskan hasil penelitian dan pada saat yang sama diberikan pembahasan yang komprehensif. Hasil dapat disajikan dalam angka, grafik, tabel, dan lain-lain yang membuat pembaca memahami dengan mudah. Pada bagian ini ditekankan nilai baru dari penelitian yang memuat inovasi, serta implikasinya. Pembahasan dapat dibuat dalam beberapa sub-bab.

REFERENSI

Chen, J., Huang, H., Tian, S., & Qu, Y. (2009). *Feature selection for text classification with Naive Bayes*. *Expert Systems with Applications*, 36(3), 5432–5435.

Feldman, Ronen and Sanger, James. (2007). *The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, New York.

Gorunescu. (2011). *Data Mining Concepts, Models and Techniques*. Romania: Springer-Verlag Berlin Heidelberg

- He, Y., & Zhou, D. (2011). *Self-training from labeled features for sentiment analysis*. *Information Processing & Management*, 47(4), 606–616.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies*, 5(May), 1–167.
- Moraes, R., Valiati, J. F., & Gavião Neto, W. P. (2013). *Document-level sentiment classification: An empirical comparison between SVM and ANN*. *Expert Systems with Applications*, 40(2), 621–633.
- Popescu, A. M., Etzioni, O.: *Extracting Product Features and Opinions from Reviews*, In *Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing*, Vancouver, British Columbia, 2005, 339–346.
- Rodiyansyah, S. F., & Winarko, E. (2012). *Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification*. *FMIPA UGM*, 6(1), 91–100.
- Salappa, A., Doumpos, M., & Zopounidis, C. (2007). *Feature Selection Algorithms in Classification Problems: An Experimental Evaluation*. *Systems Analysis, Optimization and Data Mining in Biomedicine*, 199–212.
- Songbo Tan, Jin Zhang, “An empirical study of sentiment analysis for chinese documents”, *Expert Systems with Applications* 34 (2008) 2622–2629.
- Vinodhini.G, Chandrasekaran.RM. (2012). *Sentiment Analysis and Opinion Mining: A Survey*. *International Journal of Advanced Research in Computer Science and Software Engineering*, ISSN: 2277 128X, Vol 2.
- Wadyono, Agus dan Sudarma S. (2012). *Tip Trik Android untuk Pengguna Tablet & Handphone*. Media Kita, Jakarta.
- Ye, Q., Zhang, Z., & Law, R. (2009). *Sentiment classification of online reviews to travel destinations by supervised machine learning approaches*. *Expert Systems with Applications*, 36(3), 6527–6535.