
KLASIFIKASI TEKS MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER DENGAN SELEKSI FITUR FORWARD SELECTION

Wachid Darmawan, Ari Putra Wibowo, Bambang Ismanto

STMIK Widya Pratama Pekalongan

Jl. Patriot 25 Pekalongan Telp (0285) 427816

wachid.dw@gmail.com, ariputra.stmikwp@gmail.com, bams.stmikwp@gmail.com

ABSTRAK

Penelitian tentang klasifikasi telah banyak dipelajari secara luas yang digunakan dalam berbagai hal, seperti: data mining, machine learning dan database. Selain itu pencarian information retrieval yang dapat digunakan untuk menentukan klasifikasi dokumen, konten berita, target pemasaran serta diagnosis medis. Klasifikasi dokumen menjadi daya tarik tersendiri oleh banyak peneliti selama dua dekade terakhir. Meskipun dalam melakukan klasifikasi dokument selalu ada pembaharuan antara metode dan teknik yang digunakan, namun kebutuhannya masih terus berkembang dan tidak pernah berakhir. Kemampuan untuk melakukan klasifikasi dokumen ke dalam kategori tertentu sangat membantu untuk menghadapi informasi yang berlebihan. Klasifikasi dokumen secara otomatis dikembangkan karena pekerjaan manual tidak lagi efektif. Pada penelitian ini akan dibahas bagaimana algoritma naïve bayes dan forward selection digunakan untuk mengklasifikasi judul skripsi dan tugas akhir yang ada di Prodi TI, SI, MI dan KA yang ada di STMIK Widya Pratama Pekalongan. Naïve bayes akan digunakan untuk menghitung probabilitas kemunculan kata yang mempresentasikan dokumen teks dari judul skripsi dan tugas akhir. Berdasarkan penelitian yang dilakukan diperoleh hasil rata-rata akurasi adalah 67.67% menggunakan algoritma naïve bayes, sedangkan nilai rata-rata akurasi sebesar 92.57% didapatkan pada saat menggabungkan algoritma naïve bayes dengan forward selection. Untuk mengetahui tingkat efektifitas lainnya disarankan untuk melakukan pembobotan lain seperti Backward Selection pada dokument teks yang dapat digunakan untuk penelitian eksperimen lanjutan.

Kata Kunci: text mining, klasifikasi, naïve bayes, forward selection

1. PENDAHULUAN

Text mining adalah proses untuk menghasilkan informasi yang *high-quality* dari dataset yang bersumber data teks meliputi situs website, buku, email, ulasan, dan artikel (Hotho, Nürnberger and Paaß, 2005). Informasi yang *high-quality* ini diperoleh dengan melakukan pemodelan menggunakan berbagai algoritma (Aliwy and Ameer, 2017). Secara umum *text mining* digunakan sebagai klasifikasi (kategori) teks, pengelompokan teks, analisis sentimen, peringkasan dokumen dan pemodelan relasi entitas (Feldman and Sanger, 2007). Salah satu bidang yang paling populer dari proses *text mining* adalah klasifikasi teks. klasifikasi teks adalah interpretasi model yang dapat digunakan untuk mengklasifikasi dokumen baru ke dalam kelas yang sudah ditetapkan sebelumnya (Manning D., Raghavan and Schutze, 2008).

Algoritma klasifikasi teks yang paling banyak digunakan untuk menyelesaikan

klasifikasi teks adalah algoritma *Naïve Bayes* (Wongso *et al.*, 2017). Proses klasifikasi pada algoritma *Naïve Bayes* didasarkan pada peluang bersyarat dari karakteristik salah satu kelas yang ada (Zhang *et al.*, 2011). Salah satu permasalahan pada pengklasifikasian teks adalah banyaknya jumlah atribut yang terdapat pada dataset (Wang *et al.*, 2011). Pada umumnya, jumlah atribut yang digunakan pada proses klasifikasi teks sangat besar dan apabila semua atribut yang ada digunakan akan mempengaruhi kinerja dari pengklasifikasian (Wang *et al.*, 2013). Penggunaan jumlah atribut yang banyak akan membuat akurasi menjadi rendah. Untuk memperoleh akurasi yang lebih baik, harus dilakukan pemilihan atribut dengan menggunakan algoritma yang tepat (Xu, Peng and Cheng, 2012). *Feature selection* merupakan suatu metode analisa data untuk mengoptimalkan kinerja dari algoritma klasifikasi (Wang *et al.*, 2011). Penggunaan *feature selection* berdasarkan pada

pengurangan fitur atau atribut, dengan mengeliminasi atau menghilangkan fitur yang tidak atau kurang relevan (Koncz and Paralic, 2011). Penerapan algoritma *feature selection* yang tepat dapat meningkatkan nilai akurasi (Xu, Peng and Cheng, 2012).

Pada penelitian ini akan dilakukan penggabungan algoritma *naïve bayes classification* dengan *feature selection*. Algoritma *feature selection* yang digunakan adalah *forward selection*. *Forward selection* akan melakukan eliminasi atau menghilangkan fitur yang tidak relevan (Han, Kamber and Pei, 2012). Prinsip kinerja *forward selection* didasarkan pada model *regresi linear* (Noori et al., 2011).

2. LANDASAN TEORI

2.1. Naïve Bayes Classification

Naïve Bayes Classification adalah salah satu algoritma klasifikasi yang menggunakan metode klasifikasi probabilitas dan statistik yang ditemukan oleh Thomas Bayes asal negara Inggris. Algoritma *Naïve Bayes Classification* banyak digunakan untuk memprediksi apa yang akan terjadi di masa yang akan datang berdasarkan pengalaman-pengalaman yang sudah pernah ada atau bisa juga disebut sebagai *Teorema Bayes*. Klasifikasi menggunakan Algoritma *Naïve Bayes Classification* sangat kuat terhadap independensi dari kejadian atau kondisi di masa lalu. Algoritma *Naïve Bayes Classification* merupakan salah satu algoritma terbaik untuk melakukan klasifikasi dibandingkan algoritma lain (Xhemali, et al., 2009). Dengan adanya penelitian tersebut maka *Naïve Bayes Classification* banyak keuntungannya, sehingga banyak digunakan dalam melakukan klasifikasi data.

Keuntungan menggunakan *Naïve Bayes Classification* adalah jumlah data training yang digunakan lebih sedikit dibandingkan menggunakan algoritma lain untuk menentukan nilai estimasi parameter pada saat melakukan klasifikasi. Karena pada Algoritma *Naïve Bayes Classification* mengasumsikan sebagai variabel data independen, jadi *Naïve Bayes Classification* hanya menentukan satu variabel data dalam satu kelas untuk melakukan klasifikasi, bukan keseluruhan data yang ada dari matriks kovarians (Xhemali, et al., 2009).

2.2. Text Preprocessing

Text Preprocessing merupakan bagian dari *Text Mining*, *Text Mining* sendiri merupakan bagian dari *Data Mining*. *Text Mining* dikembangkan karena banyak data dari jenis teks yang sifatnya tidak terstruktur atau tidak terbaca dengan baik. Untuk menstrukturkan teks agar menjadi data yang terstruktur dan memudahkan kita untuk mengolahnya sesuai dengan kebutuhan, untuk melakukan proses mining disebut juga sebagai *Text Preprocessing*. Untuk melakukan proses *Text Preprocessing* belum ada aturan baku untuk melakukannya. Perlakuan *Text Preprocessing* tergantung dari informasi yang ada (dokument teks) serta hasilnya mau seperti apa yang di inginkan. Akan tetapi ada beberapa tahapan untuk memulai melakukan *Text Preprocessing* yang sering digunakan oleh peneliti, seperti: *Normalize Case*, *Tokenizing*, *Filtering* dan *Stemming* (Silge & Robinson, 2017).

2.3. Forward Selection

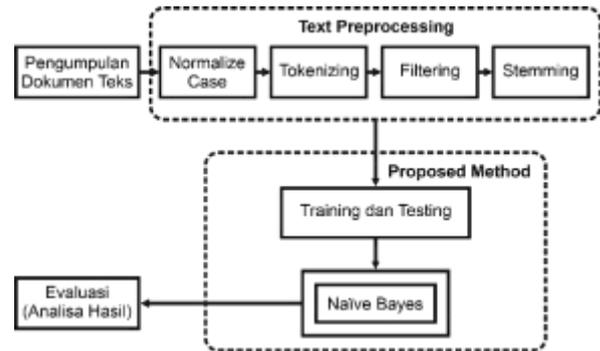
Forward Selection adalah salah satu algoritma yang dapat digunakan untuk menghilangkan atribut-atribut yang tidak relevan (Han, et al., 2012). Algoritma *Forward Selection* dibuat berdasarkan pengembangan dari model regresi linier (Noori, et al., 2011). Penggunaan metode *Forward Selection* adalah memasukkan secara bertahap *prediktor* berdasarkan korelasi parsial terbesar. *Prediktor* yang sudah dimasukkan tidak bisa diganti, dan jika tidak ada pengaruh secara *signifikan* (*sig* dibawah 0,5) kepada variabel respon maka proses *prediktor* akan dihentikan. Dengan adanya hal tersebut metode *Forward Selection* menjadi salah satu pilihan terbaik dalam melakukan perhitungan regresi untuk melakukan eliminasi variabel bebas (Larose, 2006).

3. METODE PENELITIAN

Pada penelitian ini metode yang dipilih adalah penelitian eksperimen, yang mana pada penelitian ini akan menggunakan dokument teks. Dokument teks ini digunakan untuk melihat nilai akurasi terbaik dari beberapa kali percobaan yang dilakukan. Berikut gambaran tahapan penelitiannya, seperti pada gambar 1 berikut ini:



Gambar 1 Tahapan Penelitian



Gambar 2 Kerangka Pemikiran

3.1. Pengumpulan Data

Data yang digunakan pada penelitian ini diperoleh dari judul Skripsi atau Tugas Akhir mahasiswa STMIK Widya Pratama Pekalongan pada kurun waktu 2018-2020. Dari kurun waktu tersebut diambil sebanyak 200 judul Skripsi atau Tugas Akhir dengan rincian, 50 judul untuk Prodi TI, 50 judul untuk Prodi SI, 50 judul untuk Prodi MI dan 50 judul untuk Prodi KA. Pengambilan 50 judul di setiap Prodi dimaksudkan agar data yang digunakan biar *balance* dalam melakukan proses klastering.

3.2. Text Preprocessing

Setelah dokumen teks berupa judul Skripsi dan Tugas Akhir di dapatkan, tahap selanjutnya adalah *text preprocessing* yang meliputi :

- 1) *Normalize Case*: tahapan untuk melakukan konveksi teks yang ada di dokumen teks.
- 2) *Tokenizing*: tahapan untuk melakukan pemecahan teks yang ada di dokumen teks.
- 3) *Filtering*: tahapan untuk menghapus kata-kata konjungsi yang ada di dokumen teks.
- 4) *Stemming*: tahapan untuk menentukan kata dasar di setiap kata yang ada di dokumen teks.

3.3. Eksperimen

Untuk melakukan pengujian model pada penelitian eksperimen yang dilakukan dengan menggunakan aplikasi bantu Rapid Miner. Dari algoritma yang sudah ditentukan, selanjutnya dataset yang sudah ada akan diolah sehingga menghasilkan model yang akan diinginkan. Kerangka pikir dalam penelitian eksperimental yang digunakan adalah sebagai berikut:

3.4. Evaluasi dan Hasil

Setelah melakukan pengujian berdasarkan kerangka pikir yang ada dan telah melakukan eksperimen terhadap semua dokumen teks menggunakan model yang ada. Setelah itu akandi dapatkan nilai akurasi, dari nilai akurasi yang dihasilkan akan dilakukan analisa dan evaluasi. Dari analisa dan evaluasi yang dilakukan akan ditarik kesimpulan akhirnya dari penelitian ini.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Judul Skripsi dan Tugas Akhir dari masing-masing Prodi yang ada di STMIK Widya Pratama di ambil sebanyak 50 judul di setiap Prodi yang dijadikan sebagai Dokumen Teks dalam melakukan penelitian eksperimen berikut. Berikut Dokument Teks yang sudah di kumpulkan dari masing-masing Prodi, diantaranya:

Tabel 1 Dokumen Teks Prodi TI

No	Dokumen Teks Prodi TI
1	Media Informasi Puskesmas Warungasem Kabupaten Batang
2	Pengenalan Budaya Nusantara Berbasis Game Edukasi
3	Virtual Tour Objek Wisata Bandar Ecopark
4	Implementasi Sistem Informasi Geografis Bengkel Dan Dealer Berbasis Mobile Android
5	Media Informasi Kolam Renang Banyu Biru Di Kabupaten Pekalongan Berbasis Multimedia
6	Media Pembelajaran Elektronika Dasar Kelas X Berbasis Augmented

	Reality Pada Jurusan Teknik Audio Video Di Sekolah Menengah Kejuruan Negeri 01 Kandeman
7	Pengenalan Kondisi Fisik Wilayah Indonesia Berbasis Game Edukasi
8	Pengenalan Alat-Alat Farmasi Menggunakan Augmented Reality (Studi Kasus Kelas 10 Jurusan Farmasi Di Sekolah Menengah Kejuruan Islam Nusantara Comal)
9	Media Informasi Wisata Purwahamba Indah Kabupaten Tegal Berbasis Multimedia
10	Media Bantu Budidaya Tanaman Jambu Citra Berbasis Multimedia Pada Kebun Benih Hortikultura Karanganyar
...	...
50	Penerapan Augmented Reality Pada Katalog Meubel Di Cv Rozak Putra

Tabel 2 Dokumen Teks Prodi SI

No	Dokumen Teks Prodi SI
1	Sistem Informasi Pengaduan (E-Complaint) Pada Smp Negeri 16 Pekalongan Berbasis Web
2	E-Commerce Pada Usaha Kerajinan Kulit Pemalang
3	Sistem Informasi Pengaduan Pada Dinas Perhubungan Kabupaten Pekalongan Berbasis Web
4	E - Marketplace Pada Grosir Comal
5	E-Commerce Pada Toko Boneka Di Ammilla Collection Pekalongan
6	E-Commerce Produksi Kain Kasa Pt. Multitex Di Sapugarut
7	E- Commerce Pada Badan Usaha Milik Desa Mataram Sejahtera Desa Kalipucang Wetan Kabupaten Batang
8	Sistem Informasi Pengaduan Masyarakat Pada Satuan Polisi Pamong Praja (Satpol Pp) Kabupaten Batang
9	Implementasi E-Kelurahan Pada Kelurahan Kedungwuni Barat
10	E-Commerce Produk Kerajinan Pada Koperasi Usaha Kecil Dan Menengah Mitra Pelangi Pekalongan
...	...

50	Sistem Pendukung Keputusan Penentu Status Gizi Balita Di Posyandu Taman Asri Dengan Algoritma K-Nearest Neighbor
----	--

Tabel 3. Dokumen Teks Prodi MI

No	Dokumen Teks Prodi MI
1	Sistem Informasi Pembayaran Sumbangan Pembinaan Pendidikan Berbasis Web Pada Sekolah Menengah Kejuruan Muhammadiyah Batang
2	Sistem Informasi Penjualan Pada Enditha Froyzen Batang Berbasis Web
3	Sistem Informasi Penjualan Pada Toko Sheila Store
4	Sistem Informasi Penjualan Pada Karimah Hijab Store Berbasis Web
5	Sistem Informasi Penjualan Di Toko Family
6	Sistem Administrasi Berbasis Web Cv Andhini Perkasa
7	Media Pembelajaran Pengenalan Huruf Hiragana Di Lembaga Pendidikan Dan Keterampilan (Lpk) Wakashio Gakkou Batang
8	Sistem Informasi Booking Foto Di Ucupers Studio
9	Sistem Informasi Administrasi Lpk Aquarius Pekalongan
10	Sistem Informasi Penjualan Berbasis Web Pada Batik Safana
...	...
50	Sistem Pakar Diagnosa Kerusakan Mesin Toyota Avanza Dengan Metode Forward Chaining

Tabel 4. Dokumen Teks Prodi KA

No	Dokumen Teks Prodi MI
1	Rancang Bangun Sistem Informasi Penggajian Pada Honda Cv 54 Motor Batang
2	Komputerisasi Sistem Akuntansi Penjualan Pada Batik Roro Syifa Binagria
3	Rancang Bangun Sistem Penentu Harga Pokok Produksi Pada Pt Unggul Jaya Sejahtera

4	Sistem Akuntansi Persediaan Pada Butik Alesha Fashion Losari
5	Sistem Akuntansi Penjualan Tunai Pada Kevin Phone Cell
6	Sistem Akuntansi Penjualan Di Apotek Pandowo Karanganyar Pekalongan
7	Sistem Akuntansi Penjualan Pada Koperasi Smp 13 Pekalongan
8	Sistem Akuntansi Persediaan Barang Di Batik Kupu-Kupu Pekalongan
9	Sistem Informasi Akuntansi Penjualan Pada Apotek Surya Farma Batang
10	Sistem Akuntansi Persediaan Obat Di Puskesmas Talun Kabupaten Pekalongan
...	...
50	Sistem Akuntansi Pembelian Tunai Tempo Dan Penjualan Tunai Tempo Pada Tb. Cahaya Jati

4.2. Text Preprocessing

Setelah bahan-bahan yang dikumpulkan sudah sesuai, tahap selanjutnya adalah mencari kata-kata (bobot) yang mewakili setiap kategori yang diperlukan melalui beberapa tahapan dalam *Teks Processing*, diantaranya:

4.2.1. Normalize Case

Pada tahap *Normalize Case* akan melakukan perubahan teks pada Dokumen Teks, untuk perubahan dokumen teks yang dilakukan perubahan adalah melakukan konversi teks dari dokumen teks yang terdapat huruf besarnya dirubah menjadi huruf kecil semuanya, contoh:

Sebelum Dilakukan *Normalize Case*:

Sistem Akuntansi Persediaan Barang Di Batik Kupu-Kupu Pekalongan

Setelah Dilakukan *Normize Case*:

sistem akuntansi persediaan barang di batik kupu-kupu pekalongan

4.2.2. Tokenizing

Pada tahap *tokenizing* akan dilakukan pembagian kalimat menjadi kata-kata tersendiri, contoh:

Sebelum Dilakukan *Tokenizing*:

sistem akuntansi persediaan barang di batik kupu-kupu pekalongan

Setelah Dilakukan *Tokenizing*:

sistem
akuntansi
persediaan
barang
di
batik
kupu-kupu
pekalongan

4.2.3. Filtering (Stoping)

Pada tahap *filtering* melakukan proses menghilangkan/menghapus kata-kata konjungsi/penghubung yang ada, seperti pada tabel 5.

Tabel 5. Contoh Kamus Kata Stopping

akhir	bagian	cara
akulah	bakal	cukupkah
anda	baru	dapat
apa	beginian	diakhiri
apakah	belum	dilalui
asal	berjumlah	diucapkan
atau	berkata	dong
...

4.2.4. Stemming

Pada tahap *stemming* dilakukan proses menghapus awalan dan akhiran setiap kata-kata untuk memperoleh kata dasar dari kata yang digunakan.

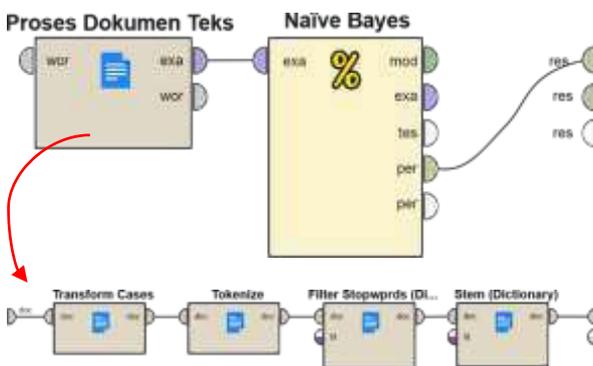
Tabel 6. Contoh Hasil Stemming

Sebelum stemming	Setelah stemming
terancam	ancam
berbahaya	bahaya
kebobolan	bobol
pencetak	cetak
diganti	ganti
berhitung	hitung
dikasari	kasar
...	...

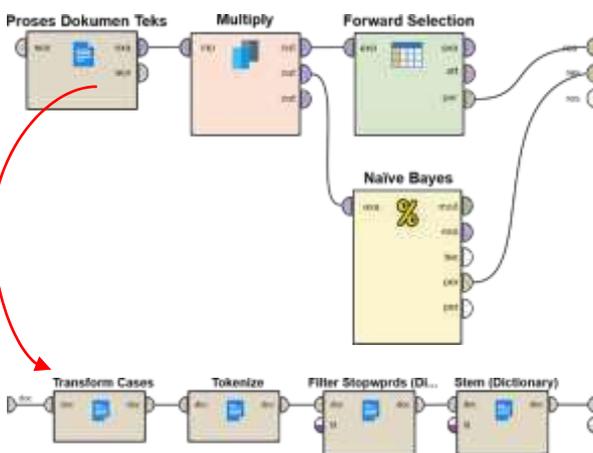
4.3. Eksperimen

Setelah proses *Teks Processing* selesai tahap selanjutnya adalah melakukan eksperimen sebanyak dua kali, yaitu: menggunakan algoritma

Naive Bayes Classification dilanjutkan menggunakan algoritma *Naive Bayes Classification & Forward Selection*. Untuk menunjang akurasi eksperimen yang dilakukan, pada penelitian ini menggunakan alat bantu aplikasi rapid miner. Pada eksperimen ini dilakukan pengujian dengan pembagian dataset untuk menghitung tingkat akurasi menggunakan algoritma *Naive Bayes Classification* dilanjutkan menggunakan algoritma *Naive Bayes Classification & Forward Selection*. Berikut gambaran tahapan penggunaan aplikasi rapid miner tersebut, seperti pada gambar 3 dan 4:



Gambar 3 Eksperimen Naive Bayes



Gambar 4 Eksperimen Naive Bayes dan Forward Selection

Dari hasil eksperimen yang sudah dilakukan seperti terlihat di gambar 3 dan 4. Didapatkan hasil sebagai berikut, yang terlihat pada tabel 7 dan 8.

Tabel 7. Hasil Eksperimen Naive Bayes
accuracy: 67.67% +/- 17.20% (micro average: 67.82%)

	true TI	true SI	true MI	true KA	class presisi on
pred. TI	39	3	9	3	72.22%
pred. SI	6	37	7	5	67.27%
pred. MI	3	4	25	6	65.79%
pred. KA	2	7	10	36	65.45%
class recall	78.00 %	72.55 %	49.02 %	72.00 %	

Berdasarkan tabel 7 dapat di simpulkan, bahwa penelitian eksperimen tentang klasifikasi judul skripsi dan tugas akhir yang menggunakan algoritma *Naive Bayes Classification*, dengan menggunakan dokumen teks sebanyak 50 judul skripsi dan tugas akhir yang diambil dari Prodi TI, SI, KA dan MI di dapatkan nilai *accuracy* rata-rata sebesar 67.67%.

Tabel 8. Hasil Eksperimen Naive Bayes dan Forward Selection

accuracy: 92.57% +/- 6.34% (micro average: 92.57%)

	true TI	true SI	true MI	true KA	class presisi on
pred. TI	45	3	1	0	91.84%
pred. SI	1	47	2	0	94.00%
pred. MI	4	1	47	2	87.04%
pred. KA	0	0	1	48	97.96%
class recall	90.00 %	92.16 %	92.16 %	96.00 %	

Berdasarkan tabel 8 dapat di simpulkan, bahwa penelitian eksperimen tentang klasifikasi judul skripsi dan tugas akhir yang menggunakan algoritma *Naive Bayes Classification* dan *Forward Selection*, dengan menggunakan dokumen teks sebanyak 50 judul skripsi dan tugas akhir yang diambil dari Prodi TI, SI, KA dan MI di dapatkan nilai *accuracy* rata-rata sebesar 92.57%.

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan penelitian eksperimen dengan melakukan komparasi yang sudah dilakukan terlihat pada tabel 7 dan 8 dapat ditarik kesimpulan bahwa nilai rata-rata *accuracy* yang dihasilkan dari menggabungkan algoritma *Naïve Bayes* dan *Forward Selection* didapatkan hasil yang lebih baik yaitu nilai rata-rata *accuracy* sebesar: 92.57%.

5.2. Saran

Pada penelitian eksperimen ini sudah melakukan pembobotan pada dokumen teks yang digunakan, yaitu *Forward Selection*. Untuk penelitian ke depan bisa juga melakukan algoritma pembobotan lain, seperti *Backward Selection* pada dokumen teks yang digunakan atau menggunakan dokumen teks lainnya. Dengan menggunakan algoritma pembobotan yang lain, dengan harapan dapat meningkatkan akurasi dari pengklasifikasian menggunakan algoritma *Naïve Bayes* dan *Forward Selection* atau bisa juga melakukan komparasi menggunakan algoritma klasifikasi lain.

DAFTAR PUSTAKA

- Aliwy, A. H. and Ameer, E. H. A., 2017. Comparative study of five text classification algorithms with their improvements. *International Journal of Applied Engineering Research*, 12(14), pp. 4309–4319. doi: 10.113/I.0973-4562.
- Feldman, R. and Sanger, J., 2007. *The text mining handbook*, Cambridge University Press. New York.
- Han, J., Kamber, M. & Pei, J., 2012. *Data Mining Concepts and Techniques*. United States of America: Morgan Kaufmann is an imprint of Elsevier.
- Hotho, A., Nürnberger, A. and Paaß, G., 2005. A brief survey of text mining. In *Ldv Forum*, 20(1), pp. 19–62.
- Koncz, P. and Paralic, J., 2011. *An approach to feature selection for sentiment analysis*. pp. 357–362.
- Larose, D. T., 2006. *Data Mining Methods and Models*. New Jersey: John Wiley & Sons, Inc.
- Manning D., M., Raghavan, P. and Schütze, H., 2008. Introduction to Information Retrieval, *IEEE Transactions on Knowledge and Data Engineering*. New York: Cambridge University Press.
- Noori, R. et al., 2011. Assessment of Input Variables Determination on the SVM Model Performance Using PCA, Gamma Test, and Forward Selection Techniques for Monthly Stream Flow Prediction. *Journal of Hydrology*, pp. 177-189.
- Silge, J. & Robinson, D., 2017. *Text Mining with R: A Tidy Approach*. New York: O'Reilly Media.
- Wang, S. et al., 2011. A feature selection method based on improved fisher's discriminant ratio for text sentiment classification, *Expert Systems with Applications*. Elsevier Ltd, 38(7), pp. 8696–8702. doi: 10.1016/j.eswa.2011.01.077.
- Wang, S. et al., 2013. Sample cutting method for imbalanced text sentiment classification based on BRC, *Knowledge-Based Systems*. Elsevier B.V., 37, pp. 451–461. doi: 10.1016/j.knosys.2012.09.003.
- Wongso, R. et al., 2017. News Article Text Classification in Indonesian Language, *Procedia Computer Science*. Elsevier B.V., 116, pp. 137–143. doi: 10.1016/j.procs.2017.10.039.
- Xhemali, D., Hinde, C. J. & Stone, R. G., 2009. Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages. *IJCSI International Journal of Computer Science Issues*, pp. 16-23.
- Xu, T., Peng, Q. and Cheng, Y. 2012. Identifying the semantic orientation of terms using S-HAL for sentiment analysis, *Knowledge-Based Systems*. Elsevier B.V., 35, pp. 279–289. doi: 10.1016/j.knosys.2012.04.011.
- Zhang, Ziqiong et al., 2011. Sentiment classification of Internet restaurant reviews written in Cantonese, *Expert Systems with Applications*. Elsevier Ltd, 38(6), pp. 7674–7682. doi: 10.1016/j.eswa.2010.12.147.