

ANALISIS SENTIMEN BERITA SAHAM MENGGUNAKAN ALGORITMA NAIVE BAYES

Skripsi



oleh
THEDA YUWANA SAPUTRA
71150082

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2020

ANALISIS SENTIMEN BERITA SAHAM MENGGUNAKAN ALGORITMA NAIVE BAYES

Skripsi



Diajukan kepada Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

THEDA YUWANA SAPUTRA
71150082

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA 2020

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Kristen Duta Wacana, saya yang bertanda tangan di bawah ini:

Nama : Theda Yuwana Saputra
NIM : 71150082
Program studi : Informatika
Fakultas : Teknik Informasi
Jenis Karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Kristen Duta Wacana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

“ANALISIS SENTIMEN BERITA SAHAM MENGGUNAKAN ALGORITMA NAÏVE BAYES”

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Kristen Duta Wacana berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama kami sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Yogyakarta
Pada Tanggal : 12 April 2020

Yang menyatakan



(Theda Yuwana Saputra)
NIM.71150082

PERNYATAAN KEASLIAN SKRIPSI

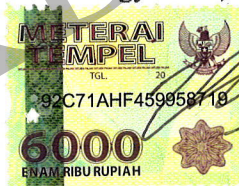
Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

ANALISIS SENTIMEN BERITA SAHAM MENGGUNAKAN ALGORITMA NAIVE BAYES

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 6 Juli 2020



THEDA YUWANA SAPUTRA

71150082

HALAMAN PERSETUJUAN

Judul Skripsi : ANALISIS SENTIMEN BERITA SAHAM
MENGGUNAKAN ALGORITMA NAIVE BAYES
Nama Mahasiswa : THEDA YUWANA SAPUTRA
N I M : 71150082
Matakuliah : Skripsi (Tugas Akhir)
Kode : TIW276
Semester : Genap
Tahun Akademik : 2019/2020

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 6 Juli 2020

Dosen Pembimbing I



Yuan Lukito, S.Kom., M.Cs.

Dosen Pembimbing II

Ditandatangani secara digital
09/07/2020
dikomunikasikan melalui email

Danny Sebastian, S.Kom., M.M., M.T.

HALAMAN PENGESAHAN

ANALISIS SENTIMEN BERITA SAHAM MENGGUNAKAN ALGORITMA NAIVE BAYES

Oleh: THEDA YUWANA SAPUTRA / 71150082

Dipertahankan di depan dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana – Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu

Syarat memperoleh gelar

Sarjana Komputer

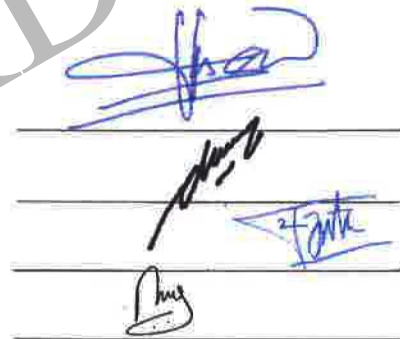
pada tanggal 17 Juni 2020

Yogyakarta, 6 Juli 2020

Mengesahkan,

Dewan Penguji:

1. Yuan Lukito, S.Kom., M.Cs.
2. Danny Sebastian, S.Kom., M.M., M.T.
3. Antonius Rachmat C., S.Kom., M.Cs
4. Nugroho Agus Haryono., S.Si., M.Si



Dekan



(Restyandito, S.Kom., MSIS., Ph.D.)

Ketua Program Studi



(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Penyusunan skripsi ini tidak terlepas dari dukungan berbagai pihak. Peneliti secara khusus mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu. Penulis banyak menerima bimbingan, petunjuk, dan bantuan serta dorongan dari berbagai pihak yang bersifat moral maupun material. Pada kesempatan ini penulis ingin menyampaikan rasa terima kasih kepada:

1. Tuhan Y.M.E. atas berkat, rahmat, dan karunia-Nya sehingga penulis dapat menyelesaikan penelitian ini.
2. Kedua orang tua yang selalu memberikan kasih sayang, semangat, dan mendoakan penulis.
3. Bapak Yuan Lukito, S.Kom., M.Cs dan Bapak Danny Sebastian, S.Kom, M.M., M.T. selaku dosen pembimbing atas waktu, saran, dan bimbingan yang diberikan kepada penulis sehingga laporan dan program tugas akhir dapat selesai dengan baik.
4. Bapak Restyandito, S.Kom, MSIS., Ph.D selaku Dekan Fakultas Teknologi Informasi Universitas Kristen Duta Wacana.
5. Segenap staf dan karyawan yang telah bekerja keras memberikan pelayanan kepada mahasiswa / mahasiswi Universitas Kristen Duta Wacana
6. Kevin Antariksa yang sudah membantu proses labelling data penulis.
7. Teman – teman lain yang sudah memberikan berbagai bantuan dalam pengerjaan tugas akhir ini.

Serta masih banyak lagi pihak-pihak yang tidak bisa penulis sebutkan satu persatu yang memberikan berbagai bantuan kepada penulis. Semoga Tuhan senantiasa membalas semua kebaikan yang telah diberikan. Semoga penelitian ini dapat bermanfaat bagi penulis, maupun bagi orang lain.

Yogyakarta, 6 Juli 2020



Theda Yuwana Saputra

INTISARI

ANALISIS SENTIMEN BERITA SAHAM MENGGUNAKAN ALGORITMA NAIVE BAYES

Banyak cara yang digunakan untuk dapat menginvestasikan harta kita, salah satunya cara adalah dengan berinvestasi di saham. Secara sederhana saham dapat diartikan sebagai tanda penyertaan atau bukti kepemilikan seseorang atau badan dalam suatu perusahaan. Setiap perusahaan yang mengeluarkan saham pasti memiliki trend masing-masing yang memiliki sentimen tertentu yang tentunya akan berpengaruh terhadap pergerakan harga saham. Setiap hal yang mempengaruhi pergerakan saham tentunya akan dimuat dalam situse berita. Tujuan dari penelitian ini adalah untuk membangun sistem yang memberikan informasi terhadap sentimen berita untuk suatu perusahaan tertentu. Sehingga investor dapat terbantu dalam pengambilan keputusan untuk berinvestasi di perusahaan tersebut.

Pada penelitian ini, penulis menggunakan data dari situs berita saham kontan.com. Data diambil menggunakan grabber, kemudian dilakukan pemberian label secara manual. Selanjutnya, dilakukan proses preprocessing seperti *case folding*, dan *stemming*, serta konfigurasi lain yang dibutuhkan seperti penghapusan kata singkatan, penghapusan kata penghubung, dan angka. Proses klasifikasi dilakukan menggunakan metode Naive Bayes.

Pengujian dengan fitur TF-IDF menunjukkan rata-rata akurasi yang dihasilkan sebesar 76,79%. Pemilihan fitur sebesar 100% dari fitur teratas memberikan nilai accuracy tertinggi. Sedangkan pengujian dengan fitur TF-IDF Bigram menghasilkan nilai rata-rata akurasi sebesar 77.13%

Kata Kunci: [klasifikasi teks, naive bayes, berita saham, analisa sentimen]

DAFTAR ISI

PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	v
HALAMAN PENGESAHAN	vi
UCAPAN TERIMA KASIH.....	vii
INTISARI	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL.....	xiii
DAFTAR LISTING	xv
Bab I.....	1
Pendahuluan.....	1
1.1 Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah.....	2
1.4. Tujuan Penelitian.....	2
1.5. Manfaat Penelitian.....	2
1.6. Metode Penelitian.....	3
1.7. Sistematika Penelitian	4
BAB II.....	5
LANDASAN TEORI.....	5
2.1. Tinjauan Pustaka	5
2.2 Landasan Teori.....	8
BAB III	19
PERANCANGAN SISTEM	19
3.1 Spesifikasi Sistem	19
3.1.1 Kebutuhan Fungsional.....	19

3.1.2 Kebutuhan non Fungsional	19
3.1.3 Kebutuhan Perangkat Keras	19
3.1.4 Kebutuhan Perangkat Lunak	19
3.1.5 Kebutuhan <i>Environment</i> , <i>Library</i> ,dan <i>Module</i>	20
3.2. Perancangan Sistem	21
3.2.1. Blok Diagram Sistem.....	21
3.2.2. Arsitektur Sistem	22
3.2.3 Perancangan <i>Grabber</i> Data	23
3.2.4 Perancangan Sistem Klasifikasi.....	24
3.3 Skema Basis Data	31
3.4 Profil Data.....	32
3.5 Kamus Data.....	33
3.6 Langkah Pengujian.....	37
BAB IV	38
IMPLEMENTASI SISTEM.....	38
4.1. Implementasi Sistem	38
4.1.1 Implementasi Proses <i>Labelling</i> Data	38
4.1.2. Implementasi <i>Grabber</i>	39
4.1.3 Implementasi Sistem Klasifikasi	40
4.1.3.3 Implementasi <i>Feature Selection</i>	46
4.1.3.4 Implementasi Klasifikasi	47
4.2 Hasil Pengujian Sistem	50
4.3 Analisa Hasil Pengujian Sistem	69
BAB V	76
KESIMPULAN DAN SARAN.....	76
5.1 Kesimpulan.....	76
5.2 Saran	76

DAFTAR PUSTAKA	77
LAMPIRAN A.....	79

©UKDW

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi KFold 10 Folds.....	18
Gambar 3.1 Blok Diagram Sistem	21
Gambar 3.2 Arsitektur Sistem.....	22
Gambar 3.3 Flowchart alur kerja Grabber Data.....	23
Gambar 3.4 Halaman Utama Web	25
Gambar 3.5 Halaman Berita	25
Gambar 3.6 Halaman Update Emiten	26
Gambar 3.7 Halaman Pengujian Sistem Klasifikasi	27
Gambar 3.8 Flowchart Preprocessing	28
Gambar 3.9 Flowchart Feature Selection.....	29
Gambar 3.10 Flowchart Proses Klasifikasi.....	30
Gambar 3.11 Skema ERD Database Sistem	31
Gambar 3.12 Profil Data	32
Gambar 4.1 Interface PHPmyAdmin.....	38
Gambar 4.2 Halaman Utama Aplikasi.....	40
Gambar 4.3 Tampilan Halaman Berita.....	41
Gambar 4.4 Tampilan Halaman Breakdown.....	41
Gambar 4.5 Tampilan Halaman Emiten List.....	42
Gambar 4.6 Tampilan Halaman Add Emiten.....	42
Gambar 4.7 Tampilan Halaman Test.....	44
Gambar 4.8 Contoh Hasil Klasifikasi Fitur 10%.....	65
Gambar 4.9 Contoh Hasil Klasifikasi Fitur 25%.....	66
Gambar 4.10 Contoh Hasil Klasifikasi Fitur 50%.....	67
Gambar 4.11 Contoh Hasil Klasifikasi Fitur 100%.....	68
Gambar 4.12 Contoh Fitur Nilai Teratas.....	69
Gambar 4.13 Contoh Fitur Nilai Terbawah.....	69
Gambar 4.14 Grafik Precision.....	71
Gambar 4.15 Grafik Accuracy.....	72
Gambar 4.16 Grafik Recall.....	72
Gambar 4.17 Grafik Precision Bigram.....	73
Gambar 4.18 Grafik Recall Bigram.....	73
Gambar 4.19 Grafik Accuracy Bigram.....	74

DAFTAR TABEL

Tabel 2.1 Tabel Contoh Kasus yang Akan Diselesaikan dengan Algoritma Naive Bayes.....	13
Tabel 2.2 Tabel Visualisasi Confusion Matrix.....	16
Tabel 2.3 Tabel Contoh Perhitungan Menggunakan Confusion Matrix.....	16
Tabel 3.1 Tabel News_Data.....	33
Tabel 3.2 Tabel Emiten_Dict.....	33
Tabel 3.3 Tabel tbp_table.....	34
Tabel 3.4 Tabel label_test.....	35
Tabel 3.5 Tabel sentencesplit.....	36
Tabel 3.6 Tabel split.....	36
Tabel 4.1 Tabel Confusion Matrix Data Train 10%.....	50
Tabel 4.2 Tabel Confusion Matrix Data Train 25%.....	51
Tabel 4.3 Tabel Confusion Matrix Data Train 50%.....	51
Tabel 4.4 Tabel Confusion Matrix Data Train 100%.....	52
Tabel 4.5 Tabel Confusion Matrix Data Test 10%.....	52
Tabel 4.6 Tabel Confusion Matrix Data Test 25%.....	53
Tabel 4.7 Tabel Confusion Matrix Data Test 50%.....	53
Tabel 4.8 Tabel Confusion Matrix Data Test 100%.....	54
Tabel 4.9 Tabel Confusion Matrix KFold 10%.....	54
Tabel 4.10 Tabel Confusion Matrix KFold 25%.....	55
Tabel 4.11 Tabel Confusion Matrix KFold 50%.....	55
Tabel 4.12 Tabel Confusion Matrix KFold 100%.....	56
Tabel 4.13 Tabel Rata-rata Nilai Precision pada Semua Percobaan.....	56
Tabel 4.14 Tabel Rata-rata Nilai Recall pada Semua Percobaan.....	57

Tabel 4.15 Tabel Rata-rata Nilai Accuracy pada Semua Percobaan.....	57
Tabel 4.16 Tabel Confusion Matrix Data Train Bigram 10%.....	58
Tabel 4.17 Tabel Confusion Matrix Data Train Bigram 25%.....	58
Tabel 4.18 Tabel Confusion Matrix Data Train Bigram 50%.....	59
Tabel 4.19 Tabel Confusion Matrix Data Train Bigram 100%.....	59
Tabel 4.20 Tabel Confusion Matrix Data Test Bigram 10%.....	60
Tabel 4.21 Tabel Confusion Matrix Data Test Bigram 25%.....	60
Tabel 4.22 Tabel Confusion Matrix Data Test Bigram 50%.....	61
Tabel 4.23 Tabel Confusion Matrix Data Test Bigram 100%.....	61
Tabel 4.24 Tabel Confusion Matrix Data K-Fold Bigram 10%.....	62
Tabel 4.25 Tabel Confusion Matrix Data K-Fold Bigram 25%.....	62
Tabel 4.26 Tabel Confusion Matrix Data K-Fold Bigram 50%.....	63
Tabel 4.27 Tabel Confusion Matrix Data K-Fold Bigram 100%.....	63
Tabel 4.28 Tabel Rata-rata Nilai Precision pada Percobaan dengan Data Bigram.....	64
Tabel 4.29 Tabel Rata-rata Nilai Recall pada Percobaan dengan Data Bigram.....	64
Tabel 4.30 Tabel Rata-rata Nilai Accuracy pada Percobaan dengan Data Bigram.....	65

DAFTAR LISTING

Listing 4.1 Pseudocode Algoritma Grabber.....	39
Listing 4.2 Pseudocode Algoritma Preprocessing	45
Listing 4.3 Pseudocode Feature Selection	46
Listing 4.4 Pseudocode Training Klasifikasi	47
Listing 4.5 Pseudocode Testing Klasifikasi.....	49

©UKDW

Bab I

Pendahuluan

1.1 Latar Belakang

Investasi mempunyai pengertian sebagai suatu keputusan (commitment) yang diambil oleh manajemen untuk mengalokasikan sumber daya perusahaan yang didasarkan atas pertimbangan yang sejalan dengan sasaran jangka panjang (kebijakan manajemen) dengan tujuan memaksimalkan nilai atau kekayaan perusahaan dimasa mendatang (Wisnubroto dan Suyanto, 2019).

Banyak cara yang digunakan untuk dapat menginvestasikan harta kita, salah satunya cara adalah dengan berinvestasi di saham. Secara sederhana saham dapat diartikan sebagai alat kepemilikan seseorang atau badan dalam suatu perusahaan yang berwujud selembar kertas yang menerangkan bahwa pemilik kertas tersebut adalah pemilik perusahaan yang menerbitkan kertas tersebut. Setiap perusahaan yang menerbitkan kertas berharga tersebut dengan tujuan mendapatkan modal atau dana tambahan disebut emiten. Harga saham setiap emiten pun mengalami pergerakan naik turun yang disebabkan berbagai faktor yang menyebabkan investor kesulitan untuk membeli saham sesuai untuk kebutuhan investasi mereka.

Berdasarkan hal tersebut, penulis akan melakukan penelitian untuk menganalisa sentimen dari berita saham berbahasa Indonesia menggunakan metode *Naive Bayes* yang akan digunakan untuk membangun sistem yang memberikan informasi terhadap sentimen berita untuk suatu perusahaan tertentu. Sehingga investor dapat terbantu dalam pengambilan keputusan untuk berinvestasi di perusahaan tersebut.

Terdapat beberapa penelitian sebelumnya yang menggunakan metode *Naive Bayes* sebagai alat untuk klasifikasi salah satunya yaitu penelitian Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes Classifier (Mahardhika dan Zuliardo, 2018). Penulis menerapkan algoritma untuk mendeteksi sentimen dengan tahapan analisis data untuk penentuan rasio data latih dan data test. Data preprocessing meliputi *case folding*, *normalisasi fitur*, *tokenizing* dan *stopword removal* dengan klasifikasi menggunakan *Naive Bayes Classifier* dan menghasilkan *output*.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah diatas, maka masalah yang dibahas dalam penelitian ini dapat dirumuskan sebagai berikut :

1. Bagaimana menerapkan *Naive Bayes* untuk mengklasifikasi sentimen berita dalam kurun waktu tertentu?
2. Bagaimana hasil evaluasi dari kinerja sistem ?

1.3. Batasan Masalah

Berdasarkan permasalahan yang sudah dijabarkan diatas maka batasan yang ditetapkan untuk penelitian adalah sebagai berikut :

1. Data berita hanya dari situs investasi kontan.
2. Data berita hanya berbahasa Indonesia.
3. Data emiten hanya yang terdaftar pada LQ45
4. Menggunakan library Sastrawi

1.4. Tujuan Penelitian

Membangun sebuah sistem yang mampu menganalisa sentimen berita saham dalam kurun waktu tertentu. Dimana sistem dapat membantu investor dalam menyimpulkan suatu sentimen berita saham dalam kurun waktu tertentu, sehingga investor dapat mengetahui gambaran dari perusahaan tersebut, sehingga investor lebih terbantu dalam mengambil kesimpulan sebelum melakukan investasi.

1.5. Manfaat Penelitian

Manfaat yang diharapkan mampu dicapai setelah selesainya penelitian ini adalah :

1. Menambah wawasan serta kemampuan berpikir dalam menerapkan teori yang telah didapat selama perkuliahan kedalam penelitian yang sebenarnya.

2. Hasil penelitian dapat dijadikan sarana dalam pengambilan keputusan saat melakukan *trading* saham. Dengan demikian akan mempermudah *trader* dalam mengambil keputusan.

1.6. Metode Penelitian

1. Studi Literatur

Studi literatur dilakukan dengan mempelajari artikel, jurnal, dan bahan lain terkait dengan topik yang dibahas dengan kata kunci analisis sentimen, algoritma *Naive Bayes*, *text mining*.

2. Perancangan Sistem

Sistem yang dirancang berupa aplikasi berbasis web yang berfungsi mengambil data, menampung data, pelatihan data, dan labelling. Data yang dipilih untuk pelatihan model berada pada jarak rentang waktu tertentu, kemudian diuji dengan data berdasarkan rentang waktu tertentu juga,

3. Pengambilan Data

Data diambil menggunakan grabber yang dibuat dengan bahasa Python. Data akan diambil secara berkala melalui situs-situs terkait berita saham.

4. Pengujian

Pengujian dilakukan dengan pembuatan *confusion matrix* untuk menentukan keakuratan dari hasil analisa sentimen oleh sistem.

5. Analisa dan Evaluasi

Hasil dari pengujian dan evaluasi akan dianalisa kemudian dilaporkan dalam bentuk tugas akhir.

1.7. Sistematika Penelitian

Bagian awal dari penulisan laporan dimulai dengan bab Pendahuluan, Tinjauan Pustaka dan Landasan Teori, serta Perancangan Sistem. Bab Pendahuluan akan membahas tentang latar belakang dari penelitian, masalah yang ditemui, serta tujuan dan manfaat dari klasifikasi berita saham. Tinjauan Pustaka dan Landasan Teori akan membahas tentang dasar-dasar teori yang digunakan sebagai acuan dalam penelitian ini. Bab ketiga, Perancangan Sistem, akan menjelaskan mengenai rancangan sistem yang akan digunakan untuk pengambilan data, pelabelan data, serta pengujian.

Bagian akhir dari penulisan laporan skripsi akan dilanjutkan dengan bab Hasil dan Pembahasan dan ditutup dengan Kesimpulan dan Saran. Hasil dan Pembahasan akan membahas tentang hasil penelitian berupa pengujian data dengan metode *Naive Bayes*. Hasil pengujian akan ditampilkan dalam tabel dan hasil *accuracy*, *precision*, serta *recall*. Kesimpulan dan saran akan menyimpulkan dari hasil pengujian serta memberikan saran untuk penelitian yang berhubungan selanjutnya.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari penelitian ini dapat disimpulkan bahwa metode *Naive Bayes* sudah cukup baik untuk digunakan dalam melakukan klasifikasi pada berita saham. dengan nilai akurasi pada fitur TF-IDF saja sebesar 76,79%, dan fitur TF-IDF disertai bigram sebesar 77,13%.

Penggunaan feature selection sebesar 10% baik pada fitur yang menggunakan bigram dan yang tidak, memiliki hasil *recall* yang paling baik jika dibandingkan dengan 25%, 50%, dan 100%. Semakin banyak fitur yang digunakan, nilai *precision* pada seluruh data semakin meningkat, sedangkan nilai *recall*-nya semakin menurun. Hal ini disebabkan karena semakin banyak fitur yang digunakan, maka semakin banyak juga fitur-fitur yang kurang penting ikut masuk.

Seiring bertambahnya jumlah fitur terpilih maka nilai akurasi mengalami kenaikan dikarenakan adanya fitur tertentu yang dianggap penting untuk suatu berita, data berita yang bervariasi maka kenaikan jumlah fitur yang digunakan untuk mengklasifikasi berita berpengaruh terhadap keakurasian klasifikasi.

Fitur TF-IDF dengan bigram memberikan dampak pada sistem dengan meningkatnya nilai rata-rata akurasi dan sistem tidak mengalami *overfitting*. Dikarenakan semakin banyaknya fitur dan juga adanya pengenalan kata negasi sehingga sistem mengalami peningkatan performa dalam akurasi sebanyak 0,34% dalam mengklasifikasi suatu sentimen berita.

5.2 Saran

Hal yang masih dapat dikembangkan dalam penelitian ini adalah penambahan jumlah data bersih sehingga sistem dapat mengenali berbagai berita dengan gaya penulisan / bahasa yang berbeda dari tiap editor. Rentang waktu pengambilan juga lebih diperluas agar mendapat jumlah data latih yang memiliki ragam fitur yang bervariasi. Tampilan web atau segi antarmuka bisa ditingkatkan untuk lebih memudahkan pengguna dalam mengakses aplikasi web.

DAFTAR PUSTAKA

- Chandra, Denny Nathaniel, Gede Indrawan, and I. Nyoman Sukaraja. "Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram." *Jurnal Ilmiah Teknologi dan Informasia ASIA (JITIKA)*.
- Chrismanto, A. R., & Lukito, Y. (2016). Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan Naive Bayes. *JUI SI, Vol. 02, No. 02*.
- Chrismanto, A. R., & Lukito, Y. (2017). Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes.
- Falahah, & Nur, D. (2015). Pengembangan Aplikasi Sentiment Analysis Menggunakan Metode Naive Bayes (Studi Kasus Analysis Dari Media Twitter). *Seminar Nasional Sistem Informasi Indonesia*.
- Hamzah, A. (2014). Sentiment Analysis Untuk Memanfaatkan Saran Kuesioner Dalam Evaluasi Pembelajaran Dengan Menggunakan Naive Bayes Classifier (NBC). *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)*. Yogyakarta.
- Hidayatullah, A., & Azhari. (2014). Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter.
- Jung, Y. (2018). Multiple predicting K-fold cross-validation for model selection. *Journal of Nonparametric Statistics*, 30(1), 197-215.
- Mahardhika, Y. S., & Zuliarso, E. (2018). ANALISIS SENTIMEN TERHADAP PEMERINTAHAN JOKO WIDODO PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAIVES BAYES CLASSIFIER. *Prosiding SINTAK 2018*.
- Muljono, Artanti, D. P., Syukur, A., Prihandono, A., & Setiadi, D. R. (2018). Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja. *Konferensi Nasional Sistem Informasi (KNSI)*.
- Nasution, M., & Hayaty, M. (2019). Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter. *Jurnal Informatika*, 6(2), 212-218.

- Nugroho. (2011). Sistem Deteksi Plagiarisme Dokumen Teks Dengan Menggunakan Algoritma Rabin-Karpi.
- Nugroho, A. (2018). Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstraksi Fitur N-Gram. *Jurnal Sains Komputer & Informatika (J-SAKTI)*.
- Slamet, C., Andrian, R., Maylawati, D. S., Darmalaksana, W., & Ramdhani, M. A. (2018). Web scraping and Naïve Bayes classification for job search engine. *IOP Conference Series: Materials Science and Engineering (Vol. 288, No. 1, p. 012038)*. IOP Publishing.
- Wati, R. (2016). Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Bayes . *Jurnal Evolusi Volume 4 Nomor 1*.
- Wijaya, A. P., & Santoso, H. A. (2016). Naive Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government. *Journal of Applied Intelligent System, Vol.1, No.1*, 48-55.
- Wisnubroto, P., & Suyanto, M. (2019). ANALISIS CAPITAL BUDGETING SEBAGAI METODE MENILAI KELAYAKAN INVESTASI PERUSAHAAN. *CIEHIS Prosiding*, 1(1), 48-58.