

ABSTRAK

Beberapa tahun belakangan ini kepercayaan masyarakat Indonesia terhadap kepolisian mulai memudar, bersamaan dengan munculnya beberapa kasus yang membuat publik bertanya tentang integritas terhadap personil kepolisian. Masih terdapat beberapa dilema yang berakibat gambaran kepolisian Secara umum masih tampak kabur di mata publik dan media, yang seharusnya menjadi perhatian khusus yang perlu segera disikapi. Kami akan secara otomatis mengklasifikasikan opini publik berbasis media Twitter dalam penelitian ini. Penelitian ini juga dilakukan untuk membuktikan pengaruh seleksi fitur *Chi-Square* terhadap kinerja algoritma *Naïve Bayes Classifier* dalam klasifikasi analisis sentimen. Data penelitian diambil dari Twitter sebanyak 657 data. Pengujian dilakukan dengan menganalisis sentimen dokumen dengan dan tanpa seleksi fitur *Chi-Square*. Dari analisis sentimen tanpa seleksi fitur diperoleh hasil akurasi 80.53% dan *f1-score* 80.60%. Sedangkan dengan seleksi fitur *Chi-Square* (taraf nyata α 0.25) diperoleh hasil akurasi 87.61% dan *f1-score* 88.19%. Dari hasil tersebut, dapat diketahui bahwa seleksi fitur *Chi-Square* mempengaruhi kinerja algoritma *Naïve Bayes* dalam menganalisis sentimen dokumen.

Kata kunci: Analisis sentimen, *Chi-Square*, *Naïve Bayes Classifier*, kinerja kepolisian.

KATA PENGANTAR

Alhamdulillah, Puji dan Syukur kita panjatkan kepada Allah subhanahu wata'ala. Dzat yang hanya kepada-Nya memohon pertolongan, alhamdulillah atas segala pertolongan, rahmat dan kasih sayang-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen Terhadap Kinerja Kepolisian Berdasarkan Klasifikasi Komentar Di Twitter Menggunakan Algoritma Naïve Bayes Dan Seleksi Fitur Chi Square”. Shalawat dan salam kepada Rasulullah Shallahu Alaihi Wasallam yang senantiasa menjadi sumber inspirasi dan teladan terbaik untuk umat manusia.

Kepolisian Republik Indonesia adalah suatu institusi penegak hukum yang ada pada negara Republik Indonesia sebagai pemelihara keamanan dalam negeri melalui upaya penyelenggaraan fungsi kepolisian dalam mencakup pemeliharaan keamanan dan ketertiban masyarakat, penegak hukum, perlindungan, pengayoman, dan pelayanan kepada masyarakat. Kinerja kepolisian merupakan bagian yang mendapat perhatian dari masyarakat. Beberapa tahun belakangan ini kepercayaan masyarakat Indonesia terhadap kepolisian mulai memudar, bersamaan dengan munculnya beberapa kasus yang membuat publik bertanya tentang integritas terhadap personil kepolisian mulai dari beberapa kasus seperti yang dikemukakan oleh portal berita CNN Indonesia. Berdasarkan hal tersebut peneliti tertarik untuk mengetahui bagaimana respon dan opini masyarakat Indonesia terhadap kinerja kepolisian dengan menggunakan data yang bersumber dari media sosial twitter.

Penulis menyadari bahwa dalam menyelesaikan skripsi ini banyak mengalami kesulitan, sehingga terkadang ada rasa putus asa. Namun berkat bantuan, motivasi, bimbingan dan pengarahan yang sangat berharga dari berbagai pihak membuat penulis tetap bersemangat untuk mengerjakan skripsi ini dan akhirnya skripsi dapat diselesaikan.

Penulis menyadari banyak pihak yang memberikan dukungan dan bantuan selama menyelesaikan studi dan tugas akhir ini. Oleh karena itu, sudah sepantasnya penulis dengan penuh hormat mengucapkan terimakasih dan mendoakan semoga Allah memberikan balasan terbaik kepada:

- Bapak Ari Basuki, S.T., M.T selaku Dekan Fakultas Teknik, terima kasih atas dedikasinya kepada Fakultas Teknik dan kepada Prodi Sistem Informasi.
- Bapak Dr. Budi Dwi Satoto, S.T., M.Kom selaku Koordinator Prodi Sistem Informasi yang berjuang keras untuk kemajuan dan kenyamanan Prodi Sistem Informasi.
- Bapak Prof. Dr. Arif Muntasa, S.Si., M.T selaku Dosen Pembimbing 1 yang telah meluangkan waktu dan pikiran dan sudah sangat banyak mengajarkan serta membimbing penulis untuk bisa menyelesaikan skripsi dan meraih gelar S1 nya di Universitas Trunojoyo Madura tercinta ini.
- Bapak Mula'ab, S.Si., M.Kom selaku Dosen Pembimbing 2 yang juga telah meluangkan waktu dan pikiran dan sudah sangat banyak mengajarkan serta membimbing penulis untuk bisa menyelesaikan skripsi dan meraih gelar S1 nya di Universitas Trunojoyo Madura tercinta ini.
- Bapak Dr. Aeri Rachmad, S.T., M.T selaku Dosen Pembimbing Akademik dari Penulis.
- Bapak Husni, S.Kom., M.T selaku Dosen Penguji Skripsi 1 terima kasih atas masukan untuk kesempurnaan skripsi penulis.
- Bapak Firdaus Solihin, S.Kom., M.Kom selaku Dosen Penguji Skripsi 2 terima kasih atas masukan untuk kesempurnaan skripsi penulis.
- Ibu Fifin Ayu Mufarroha, S.Kom., M.Kom selaku Dosen Penguji Skripsi 3 terima kasih atas masukan untuk kesempurnaan skripsi penulis.
- Ibunda Mulyani dan Ayah Agus Sagita yang selalu dengan tulus dan ikhlas mendo'akan penulis, memberikan motivasi, semangat dan kesabaran yang luar biasa untuk mendengarkan keluh kesah penulis agar tetap bersemangat. Terima kasih atas segala kasih sayang Ibu dan Ayah. Do'a kalian selalu sanggup untuk membuat hati penulis mampu menghadapi segala rintangan.
- Terima kasih banyak penulis haturkan kepada seluruh keluarga penulis, nenek sains yang selalu sabar dan ikhlas terus mendo'akan penulis, tante, Om, dan Adik – adik penulis yang selalu mensupport dan mendo'akan penulis.

- Terima kasih kepada Sahabat penulis yang selalu mensupport penulis dan menemani penulis kapanpun dan dimanapun khususnya pada proses pengerjaan Skripsi Penulis
- Terima kasih kepada Tunangan Penulis yang selalu sabar menemani dan mensupport setiap perjalanan penulis untuk mendapatkan gelar Sarjana bersama.
- Terima kasih kepada teman – teman penulis yang telah memberikan wawasan dan berbagi ilmu dalam pengerjaan skripsi.

DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN ORISINALITAS.....	iv
ABSTRAK	v
KATA PENGANTAR	vi
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.2.1 Permasalahan.....	3
1.2.2 Metode Usulan	4
1.2.3 Pertanyaan Penelitian	4
1.3 Tujuan dan Manfaat Penelitian.....	4
1.3.1 Tujuan Penelitian	5
1.3.1 Manfaat Penelitian	5
1.4 Batasan Masalah.....	5
1.5 Sistematika Penulisan.....	6
BAB II KAJIAN PUSTAKA	7
2.1 Analisis Sentimen.....	7
2.2 <i>Text Mining</i>	7
2.3 Twitter	8
2.4 <i>Data Crawling</i>	9
2.5 <i>Text Preprocessing</i>	11
2.6 Seleksi Fitur <i>Chi-Square</i>	12
2.7 Algoritma <i>Naïve Bayes Classifier</i>	13
2.8 <i>Confusion Matrix</i>	14

2.9	Penelitian Terkait.....	15
BAB III METODE USULAN.....		20
3.1	Metode Pengumpulan Data	20
3.2	Pelabelan Pada Data	20
3.3	Data <i>Preprocessing</i>	21
3.4	Seleksi Fitur <i>Chi-Square</i>	23
3.5	Klasifikasi Algoritma <i>Naïve Bayes</i>	23
3.6	Perancangan Sistem.....	25
3.7	Alur Sistem.....	25
3.8	Evaluasi Kinerja Sistem	26
3.9	Skenario Uji Coba	27
BAB IV HASIL DAN PEMBAHASAN		28
4.1	Tahapan Pengumpulan Data.....	28
4.1.1	Pengambilan <i>Tweet</i> pada Twitter	28
4.1.2	Pelabelan pada Data <i>Tweet</i>	31
4.2	Tahapan <i>Preprocessing</i>	34
4.2.1	<i>Remove duplicates</i>	35
4.2.2	<i>Cleaning</i>	37
4.2.3	<i>Transform Cases</i>	39
4.2.4	<i>Tokenization</i>	40
4.2.5	<i>Normalization</i>	42
4.2.6	<i>Filter Token (by Length)</i>	44
4.2.7	<i>Stemming</i>	45
4.2.8	<i>Stop-word Removal</i>	48
4.3	Seleksi Fitur <i>Chi-Square</i>	51
4.4	Klasifikasi <i>Naïve Bayes</i>	57
4.5	Evaluasi Klasifikasi	58
4.6	Skenario Uji Coba	59
BAB V PENUTUP.....		63
5.1	Kesimpulan.....	63
5.2	Saran	63

REFERENSI	65
LAMPIRAN.....	68
DAFTAR RIWAYAT HIDUP.....	110

DAFTAR GAMBAR

Gambar 2.1 Contoh Confussion Matrix	15
Gambar 3.1 Flowchart Preprocessing	21
Gambar 3.2 Flowchart Feature Selection	23
Gambar 3.3 Flowchart Naive Bayes	24
Gambar 3.4 Diagram Alur Sistem.....	26
Gambar 4.1 Kode program tahapan crawling data.....	29
Gambar 4.2 Hasil tahapan crawling data	29
Gambar 4.3 Hasil tahapan crawling data	31
Gambar 4.4 Hasil tahapan crawling data	31
Gambar 4.5 Jumlah data pada masing-masing sentimen	34
Gambar 4.6 Data yang berhasil di load.....	35
Gambar 4.7 Kode untuk menghitung jumlah kata pada dataset.....	35
Gambar 4.8 Bentuk dan jumlah data sebelum dilakukan remove duplicates	36
Gambar 4.9 Code Remove Duplicates	36
Gambar 4.10 Bentuk data sebelum dilakukan tahapan Cleaning.....	37
Gambar 4.11 Code Cleaning	38
Gambar 4.12 Hasil tahapan cleaning	39
Gambar 4.13 Data sebelum di transform case	39
Gambar 4.14 <i>Code Transform Cases</i>	40
Gambar 4.15 Hasil <i>Transform Cases</i> pada data.....	40
Gambar 4.16 Data sebelum dilakukan tahapan <i>Tokenization</i>	41
Gambar 4.17 <i>Code</i> tahapan <i>Tokenization</i>	41
Gambar 4.18 Hasil tahapan <i>Tokenization</i>	41
Gambar 4.19 Bentuk data sebelum dilakukan tahapan <i>Normalization</i>	42
Gambar 4.20 <i>Code</i> untuk melakukan <i>Normalization</i>	43
Gambar 4.21 Dokumen hasil normalisasi	44
Gambar 4.22 Contoh dokumen sebelum di <i>filter token</i>	44
Gambar 4.23 <i>Code Filter Token (by Length)</i>	45
Gambar 4.24 Hasil tahapan <i>filter token</i>	45
Gambar 4.25 Dokumen awal sebelum dilakukan <i>stemming</i>	46

Gambar 4.26 <i>Code Stemming</i>	46
Gambar 4.27 Kamus <i>Stemmer</i>	47
Gambar 4.28 Hasil <i>Stemming</i>	47
Gambar 4.29 Bentuk dokumen sebelum dilakukan <i>stopword removal</i>	48
Gambar 4.30 Membuat kamus kata yang berisi kata-kata <i>stopword</i>	49
Gambar 4.31 <i>Code Stemming</i>	49
Gambar 4.32 Hasil tahapan <i>stopword removal</i>	50
Gambar 4.33 Menghitung jumlah kata setelah data <i>preprocessing</i>	50
Gambar 4.34 Data hasil <i>preprocessing</i>	51
Gambar 4.35 Inisialisasi Fitur dan Target.....	51
Gambar 4.36 Variabel fitur atau X.....	52
Gambar 4.37 Variabel target atau y	52
Gambar 4.38 Observasi fitur	52
Gambar 4.39 Membuat <i>dataframe</i> baru	53
Gambar 4.40 <i>Dataframe</i> dengan nilai kemunculan masing-masing fitur di setiap dokumen.....	53
Gambar 4.41 Kode program untuk menghitung nilai <i>observe frequency</i>	53
Gambar 4.42 Hasil perhitungan nilai <i>observe frequency</i>	54
Gambar 4.43 Perhitungan nilai <i>expected frequency</i>	54
Gambar 4.44 Hasil perhitungan nilai <i>expected frequency</i>	54
Gambar 4.45 <i>Code</i> untuk menghitung nilai <i>chi-score</i>	54
Gambar 4.46 Nilai <i>chi-score</i> masing-masing fitur.....	55
Gambar 4.47 <i>Code</i> seleksi fitur.....	55
Gambar 4.48 Jumlah fitur terpilih setelah di seleksi fitur.....	56
Gambar 4.49 Bentuk dokumen sebelum diseleksi fitur	56
Gambar 4.50 Bentuk dokumen sesudah diseleksi fitur	56
Gambar 4.51 Mengubah dokumen pada <i>dataset</i> menjadi <i>sparse matrix</i>	57
Gambar 4.52 Bentuk <i>sparse matrix</i> dari dokumen	57
Gambar 4.53 <i>Code</i> pembagian data	57
Gambar 4.54 Proporsi jumlah data setelah pembagian data	58
Gambar 4.55 <i>Code fitting</i> model.....	58

Gambar 4.56 <i>Code</i> untuk mendapatkan <i>confusion matrix</i> dari model yang dilatih	59
---	----

DAFTAR TABEL

Tabel 2.1 Penelitian Terkait	17
Tabel 3.1 Contoh data tweet yang telah diberi label.....	20
Tabel 3.2 Contoh hasil penerapan pada Skenario II	27
Tabel 3.3 Contoh Perbandingan Hasil Evaluasi Kinerja Sistem.....	27
Tabel 4.1 Jumlah data tweet berdasarkan kata kunci yang digunakan	30
Tabel 4.2 Contoh data yang didapatkan melalui tahapan <i>crawling</i>	30
Tabel 4.3 Data mentah (hasil crawling) yang telah diberi label	32
Tabel 4.4 Data mentah (hasil crawling) yang telah diberi label	42
Tabel 4.5 Hasil akurasi, presisi, dan recall pada model yang dilatih	59
Tabel 4.6 Confusion matrix dengan model tanpa Chi-Square	59
Tabel 4.7 Hasil evaluasi kinerja sistem tanpa Chi-Square	60
Tabel 4.8 Confusion matrix model dengan taraf nyata 0.25	60
Tabel 4.9 Hasil evaluasi kinerja sistem dengan taraf nyata 0.25	60
Tabel 4.10 Confusion matrix model dengan taraf nyata 0.1	60
Tabel 4.11 Hasil evaluasi kinerja sistem dengan taraf nyata 0.1	61
Tabel 4.12 Confusion matrix model dengan taraf nyata 0.05	61
Tabel 4.13 Hasil evaluasi kinerja sistem dengan taraf nyata 0.05	61
Tabel 4.14 Perbandingan hasil evaluasi kinerja sistem.....	61