

PERBANDINGAN ANALISIS EMOSIONAL PENGGUNA TWITTER PADA PEMINDAHAN IBU KOTA INDONESIA MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE

Shinta Fadhilhsari¹, Widodo², Hamidillah Ajie³

¹ Mahasiswa Prodi Pendidikan Teknik Informatika dan Komputer, Teknik Elektro, FT – UNJ

^{2,3} Dosen Prodi Pendidikan Teknik Informatika dan Komputer, Teknik Elektro, FT – UNJ

¹ shintafadhilhsari@gmail.com, ² widodo@unj.ac.id, ³ hamidillah@unj.ac.id

Abstrak

Pada tahun 2019 pemerintah mewacanakan pemindahan Ibu Kota Republik Indonesia dari Jakarta ke Kalimantan. Hal itu membuat ramai masyarakat memberi opini mereka, yang banyak tertuliskan dalam media sosial Twitter yang biasa disebut tweet. Tweet tersebut dapat diolah untuk menggambarkan bagaimana opini pengguna Twitter akan suatu topik pemindahan Ibu Kota Indonesia. Penelitian ini bertujuan untuk mendapatkan analisis sentimen menggunakan algoritma Naive Bayes Classifier dan Support Vector Machine. Metode penelitian yang digunakan pada penelitian ini adalah menghitung akurasi dan error rate dari masing-masing algoritma, kemudian membandingkannya. Data yang digunakan adalah dataset sebanyak 2,100 tweet. Hasil penelitian dari kedua algoritma, berdasarkan akurasi dari Naive Bayes Classifier adalah 94.96% dan Support Vector Machine adalah 95.19%, jika berdasarkan error rate Naive Bayes Classifier adalah 4.505% dan Support Vector Machine adalah 4.81%. Sehingga dapat disimpulkan bahwa algoritma Support Vector Machine memperoleh nilai akurasi tertinggi dan dianggap sebagai algoritma lebih baik dibandingkan algoritma Naive Bayes pada analisis sentimen.

Kata kunci : *Naive Bayes Classifier, Support Vector Machine, Twitter, Akurasi, Error Rate*

1. Pendahuluan

Upaya pemindahan ibu kota Republik Indonesia dari Jakarta ke Kalimantan sudah mulai dirasakan keramaiannya dimulai pada tahun 2019. Melalui rapat terbatas pemerintahan pada tanggal 29 April 2019 terkait pemindahan ibu kota Indonesia, Presiden Republik Indonesia menawarkan tiga alternatif pemindahannya. Alternatif pertama yakni ibu kota tetap berada di Jakarta. Alternatif kedua yakni mencontoh Malaysia. Ia mengemukakan memindahkan pusat pemerintahan yang dekat dengan Jakarta, yaitu sekitar Jakarta, Bogor, Depok, Tangerang, dan Bekasi (Jabodetabek). Alternatif ketiga yakni memindahkan ibu kota ke luar Pulau Jawa. Dari ketiga alternatif tersebut maka Presiden Republik Indonesia memutuskan pada alternatif ketiga, yaitu memindahkan Ibu Kota Indonesia ke luar Pulau Jawa, khususnya mengarah kepada kawasan timur Indonesia.

Perkembangan media yang begitu pesat, memunculkan banyak media online dari media berita sampai media sosial. Media sosial yang sedang digemari oleh masyarakat sudah begitu banyak, dari Facebook, Twitter, Path, Instagram, LinkedIn dan sebagainya masih banyak lagi (Evadollz, 2014). Kegunaan media sosial belakangan ini tidak hanya untuk bersosialisasi dan menemukan teman baru tetapi untuk kegiatan lain seperti promosi dagangan, penyaluran hobi sampai menyalurkan opininya. Twitter digunakan oleh berbagai kalangan masyarakat, mulai dari remaja hingga orang dewasa. Jumlah pengguna Twitter disebutkan terus meningkat 300.000 user setiap harinya (Yarrow, 2010). Twitter merilis laporan pendataan kuartal keempat perusahaannya. Twitter menyebutkan bahwa terdapat rata-rata 126 juta pengguna aktif setiap harinya selama kuartal keempat (Q4) 2018. Pada tahun 2019 platform Twitter dicatat meningkat 17 persen, ke angka 145 juta pengguna di Indonesia. Biasanya angka ini naik sembilan persen dari tahun ketahun. Ini termasuk pengguna dari situs web resmi atau aplikasi seluler untuk mengakses akun mereka.

Pertumbuhan ini dapat membuat pengguna berbagi kegiatan mereka di media sosial, seperti pendapat dan opini yang bebas diekspresikan terhadap isu-isu yang ada seperti pemindahan ibu kota Indonesia yang akhir-akhir ini diperbincangkan. Berbagai keadaan emosi yang diungkapkan oleh pengguna Twitter dalam bentuk teks. Keambiguan opini pun terjadi pada setiap tweet yang di post oleh pengguna Twitter menimbulkan banyak perbedaan makna teks tersebut.

Pendapat tersebut dikemukakan tanpa adanya keterpaksaan. Perasaan setuju maupun tidak setuju selalu menyertai di setiap tweet. terdapat tweet lainnya yang berupa teks, kemudian akan dilakukan proses ekstraksi

informasi yang terpola (Text Mining). Perilaku ini menjadi kesempatan untuk dilakukannya analisis sentimen. Analisis sentimen akan mengelompokkan polaritas dari teks yang ada dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif atau negatif (Hidayat, 2015).

Pengklasifikasian sentimen dari tweet memiliki beberapa tantangan diantaranya bahasa yang digunakan di Twitter sering tak mempunyai struktur formal dalam kalimat mereka, seperti pemakaian singkatan, perubahan dari huruf ke angka, kurangnya tanda baca dan lain-lain. Oleh karena itu penelitian ini menggunakan metode Naive Bayes dan Support Vector Machine (SVM).

Naive Bayes Classifier merupakan metode yang sederhana dalam proses pengolahan data serta memberikan tingkat akurasi yang baik. Naive bayes terbukti cukup akurat pada permasalahan klasifikasi berbasis teks. Support Vector Machine (SVM) merupakan metode klasifikasi yang mempunyai keunggulan dari metode lain, yaitu selain menggunakan menggunakan jarak sebagai penentunya, metode ini juga menggunakan vektor sebagai syaratnya.

2. Dasar Teori

2.1. Pemindehan Ibu Kota

Pemindehan ibu kota negara merupakan peristiwa yang tidak tabu dan dilaksanakan dengan tujuan memecahkan permasalahan demi kebaikan maupun kemajuan bangsa dan negara. Pemindehan ibu kota di NKRI sangat dimungkinkan karena di dalam Undang-Undang Dasar Republik Indonesia dan Amandemennya tidak diatur secara tegas. Dalam Bab II ayat (2) UUD NKRI tertulis: Majelis Permusyawaratan Rakyat bersidang sedikitnya sekali dalam lima tahun di ibu kota negara. Dalam UUD tersebut tidak ada pasal yang menyebutkan dimana dan bagaimana ibu kota negara diatur. Dengan demikian terdapat fleksibilitas yang tinggi dalam mengatur termasuk memindah ibu kota negara. Dalam pemindehan ibu kota negara, tentu saja diperlukan alasan yang kuat dan mendasar tentang efektifitas fungsinya. (Yahya, 2018)

2.2. Twitter

Twitter merupakan layanan microblogging yang populer di mana pengguna membuat pesan status (disebut "tweets"). Tweet ini terkadang mengungkapkan pendapat tentang berbagai topik. Kami mengusulkan sebuah metode untuk mengekstraksi sentimen secara otomatis (positif atau negatif) dari tweet. Ini sangat berguna karena memungkinkan umpan balik dikumpulkan tanpa intervensi manual. (A Go, R Bhayani, & L Huang, 2009)

2.3. Emosional

Perasaan emosi menunjukkan kegoncangan organisme yang disertai oleh gejala-gejala kesadaran, keperilakuan dan proses fisiologis. Emosi tidak selalu jelek, emosi merupakan pembawa pesan dalam komunikasi interpersonal. (Coleman dan Hammen dalam Rakhmat, 2007:40)

2.4. Analisis Sentimen

Analisis sentimen dilakukan untuk melihat pendapat terhadap sebuah masalah atau dapat juga digunakan untuk identifikasi kecenderungan hal di pasar. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian ataupun aplikasi mengenai analisis sentimen berkembang pesat, bahkan di Amerika kurang lebih 20-30 perusahaan yang memfokuskan pada layanan analisis sentimen. (Liu, 2010).

2.5. Naive Bayes

Naive Bayes Classifier merupakan pengklasifikasian kemungkinan sederhana berdasarkan penerapan teorema Bayes (dari statistik Bayesian) dengan asumsi independen (naif) yang kuat. Sebuah istilah yang lebih deskriptif untuk model kemungkinan yang digaris bawahi merupakan model fitur independen (Susanto, Farisi, 2007). Persamaan dari teorema Bayes adalah:

$$P(H|X) = \frac{P(H|X)P(H)}{P(X)}$$

Keterangan :

- X : Data dengan class yang belum diketahui
- H : Hipotesis data X
- P(H|X) : Probabilitas hipotesis H
- P(H) : Probabilitas hipotesis H (prior probability)
- P(X|H) : Probabilitas X berdasarkan kondisi H
- P(X) : Probabilitas X

2.6. Support Vector Machine

Support Vector Machine adalah sebuah metode untuk klasifikasi baik data linear dan nonlinear. Menggunakan pemetaan nonlinear untuk mengubah data pelatihan asli ke dimensi yang lebih tinggi. Dalam dimensi baru ini, akan mencari linear yang optimal memisahkan hyperplane (yaitu, "batas keputusan" memisahkan tupel dari satu kelas yang lain). Dengan pemetaan nonlinear yang tepat untuk dimensi yang cukup tinggi, data dari dua kelas selalu dapat dipisahkan dengan hyperplane. Support Vector Machine menemukan hyperplane ini menggunakan vektor dukungan ("penting" tupel pelatihan) dan margin (didefinisikan oleh vektor dukungan). (Jiawei Han, dkk., 2012:408).

3. Metodologi

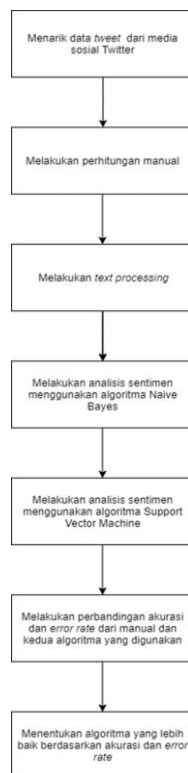
3.1 Tempat dan Waktu Penelitian

Penelitian ini dilakukan di Laboratorium Komputer Gedung L1 Teknik Elektro lantai 2, Fakultas Teknik, Universitas Negeri Jakarta yang berlokasi di Jalan Rawamangun Muka, Jakarta 13220. Penelitian ini dilakukan sejak bulan November 2019 hingga Januari 2020.

3.2 Alat dan Bahan Penelitian

Perangkat keras yang digunakan adalah personal komputer dengan spesifikasi Prosesor Intel Core i5 7200U, Memori DDR3L 1600 MHz SDRAM, 4 GB, Grafis NVIDIA GeForce 930M 2GB DDR3 VRAM, Display 14 inches HD, Hard Drive 1 TB. Adapun perangkat lunak yang digunakan adalah Windows 10 Pro 64-bit, Visual Studio Code, Microsoft Office Professional Plus 2013 dan Google Chrome Versi 79.0.3945.117. Bahan yang digunakan *tweet* dari Twitter.

3.3 Diagram Alir Penelitian



Gambar 1. Diagram Alir Penelitian

4. Hasil dan Analisis

Berikut hasil penelitian dari perhitungan tingkat akurasi, *error rate*, presisi, *recall*, dan *running time* yang dihasilkan dari algoritma Naïve Bayes dan Support Vector Machine dengan 10 kali running

Tabel 1. Hasil Sentimen Menggunakan Algoritma Naïve Bayes

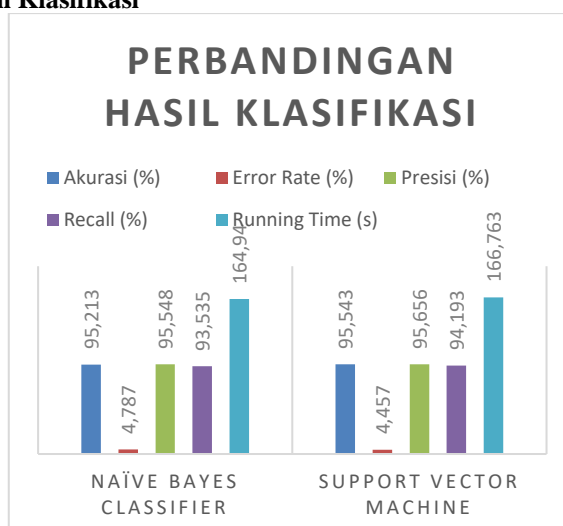
| Ket | Akura si | Error Rate | Presisi | Recall | Running Time |
|----------|-------------|---------------|---------|------------|-----------------|
| Run 1 | 94.19 % | 5.81% | 93.9% | 92.89 % | 252.23 s |

| | | | | | |
|--------|---------|-------|---------|---------|----------|
| Run 2 | 95.57 % | 4.43% | 94.84 % | 95.15 % | 261.64 s |
| Run 3 | 96.05 % | 3.95% | 96.68 % | 94.29 % | 273.27s |
| Run 4 | 95.38 % | 4.62% | 95.31 % | 94.18 % | 172.76 s |
| Run 5 | 95.52 % | 4.48% | 95.93 % | 93.86 % | 121.16 s |
| Run 6 | 94.81 % | 5.19% | 95.25 % | 92.89 % | 124.14 s |
| Run 7 | 95.47 % | 4.53% | 96.43 % | 93.21 % | 130.9 s |
| Run 8 | 94.71 % | 5.29% | 95.54 % | 92.35 % | 106.77 s |
| Run 9 | 95.24 % | 4.76% | 96.31 % | 92.78 % | 99.71 s |
| Run 10 | 95.19 % | 4.81% | 95.29 % | 93.75 % | 106.82 s |

Tabel 2. Hasil Sentimen Menggunakan Support Vector Machine

| Ket | Akurasi (%) | Error Rate | Presisi | Recall | Running Time |
|--------|-------------|------------|---------|---------|--------------|
| Run 1 | 95% | 5% | 95.67 % | 92.89 % | 222.93 s |
| Run 2 | 95.76 % | 4.24% | 95.95 % | 94.4% | 187.17 s |
| Run 3 | 95.14 % | 4.86% | 95.29 % | 93.64 % | 173.58 s |
| Run 4 | 96.24 % | 3.76% | 95.69 % | 95.8% | 172.39 s |
| Run 5 | 95.57 % | 4.43% | 94.84 % | 95.15 % | 137.84 s |
| Run 6 | 95.76 % | 4.24% | 95.45 % | 94.94 % | 147.81 s |
| Run 7 | 94.81 % | 5.19% | 95.65 % | 92.46 % | 161.81 s |
| Run 8 | 95.86 % | 4.14% | 96.25 % | 94.29 % | 140.6 s |
| Run 9 | 95.48 % | 4.52% | 96.12 % | 93.53 % | 167.76 s |
| Run 10 | 95.81 % | 4.19% | 95.65 % | 94.83 % | 155.74 s |

4.4 Perbandingan Hasil Klasifikasi



Gambar 2. Perbandingan Hasil Klasifikasi

Pada algoritma Naïve Bayes mendapatkan nilai akurasi sebesar 95.213% sedangkan pada Support Vector Machine lebih baik sebesar 95.543%. *Error rate* didapatkan pada Naive Bayes sebesar 4.787% dan Support Vector Machine yang lebih rendah yaitu 4.457%. Pada presisi yang lebih baik pada Support Vector Machine sebesar 95.656% dan Naive Bayes sebesar 95.548%. *Recall* terbesar pada Support Vector Machine sebesar 94.193%

sedangka Naive Bayes sebesar 93.535%. Untuk *running time* tercepat pada Naive Bayes sebesar 164.94 detik dan Support Vector Machine sebesar 166.763 detik.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat ditarik kesimpulan bahwa algoritma Support Vector Machine lebih baik dibandingkan Naive Bayes berdasarkan 4 dari 5 parameter yang diuji. Dengan nilai akurasi SVM sebesar, *error rate*, presisi, *recall*, *running time*. Dan nilai akurasi NB sebesar, *error rate*, presisi, *recall*, *running time*.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, saran untuk penelitian serupa dalam sentimen analisis adalah:

1. Pengambilan data menggunakan API Twitter lebih baik saat topik sedang ramai dibicarakan, karena pengambilan data tersebut hanya dalam kurun waktu 7 hari kebelakang.
2. Menambahkan dataset agar mendapatkan hasil yang bervariasi lagi dan juga dengan menambahkan data train sehingga hasil yang didapat semakin akurat.
3. Dapat dilakukan penelitian selanjutnya dengan dikembangkan lagi menggunakan algoritma lain sehingga dapat dibandingkan tingkat akurasi dan *error rate*-nya ataupun parameter lainnya.
4. Hasil penelitian sentimen analisis ini diharapkan menjadi acuan dari pendapat pengguna Twitter dan dapat menjadi suatu kajian studi untuk semua bidang yang memanfaatkannya.

Daftar Pustaka:

- Arief Budi Witarto, dan Dwi Handoko Nugroho Satriyo Anto, *Support Vector Machine- Teori dan Implementasinya dalam Bioinformatika.*, 2003.
- Fachrurrozi., Novi, Y., 2015. *Analisis Sentimen Pengguna Jejaring Sosial Menggunakan Support Vector Machine.* Palembang: Universitas Sriwijaya.
- Fauziah, A., 2017. *Analisa Sentimen Menggunakan Naive Bayes untuk melihat Persepsi Masyarakat terhadap Kenaikan Harga Jual Rokok pada Media Sosial Twitter.* Surakarta: Universitas Muhammadiyah Surakarta.
- Ghulam A. 2017. *Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter.* Ponorogo: Universitas Muhammadiyah Ponorogo.
- Ismail, Sunni., Dwi, H., 2012. *Analisis Sentimen dan Ekstrasi Topik Penentu Sentimen Pada Opini terhadap Tokoh Publik.* Bandung: Jurnal Institut Teknologi Bandung.
- Kamus Besar Bahasa Indonesia (KBBI). <https://www.kbbi.web.id/>. Diakses pada tanggal 22 September 2019.
- Kohavi, Ron and Foster Provost, eds. 1998. *Glossary of Terms. Special Issue on Applications of Machine Learning and the Knowledge Discovery Process, Machine Learning*, vol. 30, pp. 271–274. Kluwer Academic Publishers.
- Listari., Munaffidzul, I., Eky, R., & Edy, W., 2018. *Analisis Sentimen Twitter terhadap Bom Bunuh Diri di Surabaya 13 Mei 2018 menggunakan Pendekatan Support Vector Machine.* Yogyakarta: Seminar Nasional Matematika.
- Mohammad., Wayan, F., 2015. *Analisis Twitter untuk mengetahui karakter seseorang menggunakan algoritma Naive Bayes.* Malang: Seminar Nasional Sistem Informasi Indonesia.
- Moch, Ali., 2018. *Uji Pengaruh Karakteristik Dataset Pada Peforma Algoritma Klasifikasi.* Magelang: STMIK Bina Patria.
- Muhammad, Rifqi., 2018. *Content Analysis on Twiter Users Interaction within First 100 Days of Jakarta's New Government by Using Text Mining.* Yogyakarta: Jurnal Pekomnas.
- Narkhede, Sarang. 2018. *Understanding Confusion Matrix. Towards Data Science, via Medium.*

- Paroubek, Patrick. 2010. *Twitter as a corpus for Sentiment Analysis and Opinion Mining*. Perancis: French National Centre for Scientific Research.
- Petrix, N. 2015. *Sentiment Analysis menggunakan Support Vector Machine(SVM)*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Rachmat, Antonius., 2010. *Algoritma dan Pemrograman dengan Bahasa C – Konsep, Teori, & Implementasi*. Yogyakarta: Penerbit ANDI.
- Rezki, S., Rachmansyah., *Analisis Sentimen pada Sosial Media Twitter terhadap Politik di Indonesia menggunakan Text Mining dengan Metode Naïve Bayes Classifier*. Palembang: STMIK GI MDP.
- Rina, Diani., Untari, Novia., Annisa, Aditsania., 2017. *Analisis Pengaruh Kernel Support Vector Machine (SVM) pada Klasifikasi Data Microarray untuk Deteksi Kanker*. Bandung: Ind. Journal on Computing.
- T. Wilson, J. Wiebe, and P. Ho@mann. *Recognizing contextual polarity in phrase-level sentiment analysis*. In *Proceedings of Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*, Vancouver, CA, 2005.
- X. Hu and H. Liu, “Text Analytics in Social Media”, pada *Mining Text Data*, pp.385-414, Springer, 2012.