

Sentiment Analysis of the Job Creation Law on Twitter Data Using Support Vector Machine

Dhevina Dewantari¹, Ega Javier Harwenda², Mugi Rohimah³,
Muhammad Ridho Kurniawan Pratama⁴, Yova Ruldeviyani⁵
{dhevina.dewantari@ui.ac.id¹, ega.javier@ui.ac.id², mugi.rohimah@ui.ac.id³}

Fakultas Ilmu Komputer, Universitas Indonesia, Depok, Indonesia¹²³

Abstract. Twitter is a microblogging service that allows users to share their opinions, especially on the latest public issues. The Job Creation Law (*UU Ciptaker*) or known as the Job Creation Omnibus Law became a public discussion on Twitter before and after its ratification. This Law amended several previous laws to create employment and increase foreign and domestic investment. Sentiment analysis was needed to analyze public opinion regarding the UU Ciptaker on Twitter, namely by classifying opinions into several classes. The aim was to provide insight to the public regarding the public's reaction to the UU Ciptaker and its effects on public opinion. Support Vector Machine (SVM) was used to classify the data. Based on the classification results of 868 tweets, 217 tweets (25%) were labeled as positive. The 651 tweets (75%) were labeled as negative. This showed that the majority of Twitter users rejected the passage of the UU Ciptaker.

Keywords: public opinion, sentiment analysis, SVM, Twitter, UU Ciptaker

Analisis Sentimen terhadap Undang-Undang Cipta Kerja pada Data Twitter Menggunakan *Support Vector Machine*

Abstrak. Twitter merupakan layanan *microblogging* yang memungkinkan pengguna untuk berbagi opini, terutama tentang berbagai isu publik terbaru. Undang-Undang Cipta Kerja (UU Ciptaker) atau yang dikenal sebagai Omnibus Law Cipta Kerja menjadi perbincangan publik di Twitter menjelang dan setelah pengesahannya. Undang-undang ini mengubah beberapa undang-undang sebelumnya untuk menciptakan lapangan kerja serta meningkatkan investasi asing dan domestik. Analisis sentimen diperlukan untuk menganalisis opini publik terkait UU Ciptaker di Twitter, yakni dengan mengklasifikasikan opini ke dalam beberapa kelas. Tujuannya untuk memberikan wawasan kepada publik mengenai reaksi masyarakat terhadap UU Ciptaker dan pengaruhnya terhadap opini publik. *Support Vector Machine* (SVM) digunakan untuk mengklasifikasikan data. Berdasarkan hasil pengklasifikasian 868 cuitan, diperoleh 217 cuitan (25%) yang bersentimen positif. Adapun 651 cuitan (75%) bersentimen negatif. Ini menunjukkan bahwa sebagian besar pengguna Twitter menolak pengesahan UU Ciptaker.

Kata kunci: analisis sentimen, opini publik, SVM, Twitter, UU Ciptaker

1 Pendahuluan

Layanan *microblogging* Twitter diluncurkan pada tahun 2006 dengan lebih dari 20 juta pengunjung unik setiap bulan [1]. Memanfaatkan fitur cuitan, seseorang dapat memublikasikan informasi berupa teks, gambar, maupun video. Twitter juga dapat menampilkan bermacam cuitan yang sedang populer dari lokasi di sekitar pengguna, serta menggunakan *hashtag* untuk mengelompokkan cuitan sejenis. Dengan menyisipkan *hashtag* pada cuitannya, pengguna dapat memperluas penyampaian informasi. Hal tersebut menjadikan cuitan sebuah wadah yang cukup efektif untuk menyampaikan opini.

Setiap pengguna Twitter dapat menuangkan opini pribadi maupun memberikan komentar terhadap opini orang lain. Opini-opini tersebut dapat dikategorikan sebagai opini dengan sentimen positif dan negatif. Opini dengan sentimen positif dapat dideteksi melalui kata-kata yang mengapresiasi, menyetujui, dan mendukung suatu tema pembahasan. Adapun opini dengan sentimen negatif dapat dideteksi melalui kata-kata yang menolak, menyampaikan ujaran kebencian, atau mengkritik tema pembahasan tersebut.

Pada tanggal 5 Oktober 2020, DPR RI melakukan pengesahan Undang-Undang Cipta Kerja (UU Ciptaker) atau yang lebih dikenal luas sebagai “Omnibus Law Cipta Kerja.” Sebelum disahkan, UU Ciptaker sudah menjadi tren yang dibahas di seluruh Indonesia. Berbagai macam pro dan kontra disampaikan melalui media informasi seperti portal berita, media sosial seperti Twitter, Facebook, maupun media sosial berbentuk jaringan pribadi seperti Whatsapp dan Line. Beberapa hal terkait UU Ciptaker yang memunculkan kontroversi antara lain pengesahan yang dianggap terburu-buru, penyusunan yang tidak transparan, dan adanya pasal-pasal yang dianggap merugikan masyarakat.

Sejak tanggal 14 September 2020 hingga 12 Oktober 2020, terdapat 1.368 cuitan terkait pengesahan UU Ciptaker. Dengan cukup banyaknya cuitan terkait UU Ciptaker yang dikemukakan oleh masyarakat, topik ini pun sempat menjadi tren di Twitter. Penelitian ini bertujuan melakukan analisis sentimen terhadap cuitan-cuitan terkait UU Ciptaker tersebut. Analisis sentimen yang diterapkan menggunakan algoritma SVM. Algoritma ini dipandang memiliki akurasi lebih tinggi dibandingkan dengan metode lain dan sudah pernah digunakan dalam penelitian terdahulu terkait pengesahan undang-undang di negara lain. Analisis sentimen dapat digunakan untuk melihat topik apa yang membuat masyarakat menolak atau menyetujui disahkannya hal tersebut. Selanjutnya, dari hasil analisis sentimen tersebut, dapat dilakukan penelitian lebih lanjut sebagai masukan kepada pemerintah dalam menyikapi hal-hal yang menjadi fokus masyarakat terkait pengesahan UU Ciptaker.

2 Kajian Literatur

Analisis sentimen merupakan metode yang tepat untuk melakukan analisis konten media sosial [2]. Analisis sentimen atau *opinion mining* merupakan metode yang dilakukan untuk mengelompokkan opini atau *review* ke dalam kelas positif atau negatif. Ini penting untuk pembuatan keputusan karena opini telah diekstrak, diidentifikasi, dan dievaluasi dari *review online* [3]. Analisis sentimen diawali dengan menentukan domain himpunan data dan melakukan kegiatan pra-pemrosesan data [4]. Pra-pemrosesan data dilakukan dengan tokenisasi, *stopwords removal*, dan *stemming*. Selanjutnya, dilakukan *transformation* yang merupakan representasi dari jumlah data tekstual [3].

Analisis sentimen merupakan area riset yang membantu mengidentifikasi sentimen dalam data berbentuk teks [5]. Analisis ini berpengaruh terhadap opini, sikap, tindakan, dan reaksi publik [6] sehingga dapat dilakukan untuk mengetahui kecenderungan publik [5]. Kecenderungan publik dapat dikelompokkan sebagai opini dengan sentimen yang bersifat positif atau negatif [7].

Twitter meraih kepopuleran dibandingkan media lain karena luasnya pengguna dan cepatnya pertumbuhan pengguna [5]. Sejak 2016, Twitter telah memiliki 313 juta pengguna dan terus bertambah setiap bulannya. Saat ini Twitter digunakan juga oleh politikus untuk berkampanye, sehingga mendapatkan reaksi publik yang bermacam-macam [8]. Oleh karena itu, Twitter juga kerap digunakan untuk melihat sentimen masyarakat. Banyak riset yang sudah melakukan analisis terhadap perilaku pengguna selama masa sebelum pemilu [9] untuk melihat preferensi publik terhadap sebuah institusi [10]. Twitter digunakan oleh politikus untuk memberikan informasi dan terkadang sebagai media promosi aktif untuk agenda tertentu [11].

Untuk melihat sentimen masyarakat yang terekspresikan lewat Twitter, salah satu metode yang dapat digunakan adalah SVM (*Support Vector Machines*) yang terbukti cocok untuk *text classifiers* [12]. SVM merupakan metode *supervised learning* yang memiliki dasar teori dan dapat melakukan klasifikasi lebih akurat dibandingkan algoritma lain [3]. Pendekatan SVM cocok untuk *multi-class classification* karena dapat mengeliminasi kebutuhan untuk *feature selection* [12]. SVM populer sebagai metode analisis sentimen yang bagus karena hanya sedikit terpengaruh oleh adanya *noisy terms* [3]. SVM dapat memisahkan himpunan data dengan label pada data latih melalui *hyperplane* dengan jarak terjauh dari sampel positif dan negatif [13].

Untuk pemberian label setiap kalimat positif dan negatif, perhitungan dibuat berdasarkan *hyperplane* dengan rumus persamaan (1):

$$f(\phi(x)) = \text{sign}(w \cdot \phi(x) + b) \quad (1)$$

Di mana $f(\phi(x))$ merupakan hasil dari klasifikasi data tes, w adalah bobot, b adalah bias, dan $\phi(x)$ adalah perhitungan data uji Kernel.

Beberapa penelitian telah dilakukan sebelumnya terkait dengan pengesahan maupun amandemen undang-undang suatu negara. Narang melakukan analisis sentimen amandemen Undang-Undang Kewarganegaraan di India pada cuitan Twitter menggunakan algoritma SVM, dengan hasil akurasi sebesar 77,32% dan lebih baik dibandingkan algoritma klasifikasi yang lain [14]. Prayoga melakukan analisis sentimen amandemen RKUHP dan Undang-Undang Anti Korupsi pada cuitan Twitter menggunakan metode *Lexicon Based and Agglomerative Hierarchical Clustering*, dengan akurasi sebesar 61,6% dari jumlah cuitan sebanyak 2.408 cuitan [15]. Terdapat penelitian tentang *text classification* yang membandingkan metode SVM, *kNN*, dan Naive Bayes dan hasilnya SVM merupakan metode yang lebih unggul secara teori dan praktik karena performa dan kekuatannya dalam menangani klasifikasi non-linear [16].

3 Metode Penelitian

Penelitian ini dilakukan melalui beberapa tahapan, yakni pengumpulan data, pemberian respons kalimat, pra-pemrosesan data, serta implementasi dan pengujian. Pengumpulan data

dilakukan dengan memanfaatkan data dari media sosial Twitter. Dengan Twitter REST API yang tersedia, cuitan opini dapat dicari berdasarkan kueri yang diperlukan. Dalam penelitian ini, kueri yang digunakan untuk tahap pengumpulan data adalah kata “Omnibus Law.” Kata tersebut dianggap memiliki keterkaitan dengan objek penelitian dan cukup banyak cuitan opini yang menggunakannya. Cuitan yang dihimpun didapatkan melalui Twitter REST API pada periode 14 September hingga 12 Oktober 2020. Berdasarkan pencarian dengan menggunakan kueri “Omnibus Law,” diperoleh jumlah cuitan sebanyak 1.368 cuitan.

Pemberian respons kalimat bertujuan untuk menggolongkan cuitan menjadi opini dengan sentimen positif dan negatif. Pemberian respons kalimat dilakukan melalui justifikasi ahli. Justifikasi ahli dilakukan oleh guru bahasa Indonesia sebagai informan dengan memberikan respons terhadap 500 sampel cuitan data latih menjadi opini dengan sentimen positif dan negatif, yang selanjutnya digunakan untuk pelatihan data dan prediksi data uji.

Pra-pemrosesan data bertujuan untuk melakukan pembersihan dan transformasi data yang telah dikumpulkan pada tahap sebelumnya. Pada tahap ini, seperti yang dapat dilihat pada Gambar 1, beberapa langkah dilakukan untuk mendapatkan data akhir yang digunakan dalam melakukan pembobotan kata dan analisis sentimen untuk menilai opini publik terhadap UU Ciptaker. Contoh pra-pemrosesan data dapat dilihat pada **Tabel 1**.



Gambar 1. Tahapan Pra-Pemrosesan Data

Tabel 1. Contoh Pra-Pemrosesan Cuitan

Kalimat Awal	
Sebelumnya, Presiden Jokowi menyebut aksi unjuk rasa menolak UU Cipta Kerja disebabkan disinformasi dan hoaks di media sosial. Namun naskah final UU yang disahkan 5 Oktober lalu itu belum bisa diakses publik di saluran resmi pemerintah maupun DPR https://t.co/UeyTnjm5xi	
Pembersihan Data	
Sebelumnya Presiden Jokowi menyebut aksi unjuk rasa menolak UU Cipta Kerja disebabkan disinformasi dan hoaks di media sosial Namun naskah final UU yang disahkan Oktober itu belum bisa diakses publik di saluran resmi pemerintah dan DPR	
Case Folding	
sebelumnya presiden Jokowi menyebut aksi unjuk rasa menolak UU cipta kerja disebabkan disinformasi dan hoaks di media sosial namun naskah final UU yang disahkan oktober itu belum bisa diakses publik di saluran resmi pemerintah dan dpr	
Tokenisasi	
<i>sebelumnya</i> <i>presiden</i> <i>jokowi</i> <i>menyebut</i> <i>aksi</i>	<i>unjuk</i> <i>rasa</i> <i>menolak</i> <i>uu</i> <i>cipta</i>
Stemming	
Sebelum: - <i>menyebut</i> - <i>menolak</i> - <i>disahkan</i>	Setelah: - <i>sebut</i> - <i>tolak</i> - <i>sah</i>
Stopword Removal	
<i>sebelum, dan, di, namun, yang, lalu, itu</i>	

Pada tahap akhir, setelah dilakukan berbagai macam pra-pemrosesan cuitan, dilakukan proses pengimplementasian algoritma SVM dan klasifikasi cuitan. Algoritma SVM diimplementasikan dengan menggunakan bahasa pemrograman Python dengan pustaka *Scikit-learn*, *pandas*, dan *PySastrawi*. Selanjutnya, pengujian dilakukan untuk mengetahui akurasi hasil implementasi algoritma.

4 Hasil dan Pembahasan

Berdasarkan klasifikasi data uji sebanyak 868 cuitan, diperoleh jumlah cuitan dengan sentimen positif sebanyak 217 cuitan (25%), sedangkan cuitan dengan sentimen negatif sebanyak 651 cuitan (75%). Algoritma SVM melakukan proses klasifikasi data uji tersebut dengan akurasi sebesar 80,4%. Dalam proses pelatihan data dan klasifikasi data uji, dilakukan perhitungan beberapa kata terbanyak pada tiap kelas hasil klasifikasi serta dilakukan justifikasi untuk menyaring kata-kata yang penting. Hasil klasifikasi dijabarkan pada **Tabel 2** untuk jumlah kata pada kelas sentimen positif dan negatif.

Tabel 2. Jumlah kata pada kelas sentimen positif dan negatif

Kata Positif	Jumlah Kata Positif	Kata Negatif	Jumlah Kata Negatif
<i>ungkap</i>	52	<i>tolak</i>	137
<i>halal</i>	49	<i>tidak</i>	123
<i>awal</i>	44	<i>aksi</i>	89
<i>sertifikat</i>	38	<i>demo</i>	68

Contoh cuitan berdasarkan klasifikasi pada kelas sentimen positif dan negatif dapat dilihat pada **Tabel 3**.

Tabel 3. Contoh cuitan yang didapatkan

Kalimat	Kelas
Undang-Undang (UU) Cipta Kerja atau Omnibus Law mengubah sistem penerbitan sertifikat halal. Jika sebelumnya sertifikat halal hanya dikeluarkan oleh Majelis Ulama Indonesia (MUI), kini UU Ciptaker memberi alternatif sertifikat halal dapat diberikan ke Badan Penyelenggara Jaminan Produk Halal	Positif
Masyarakat luas tidak persoalkan komunikasinya saja, tapi lebih kpd isi UU Ciptaker nya. Jokowi Tegur Menteri dan Jajarannya, Komunikasi Omnibus Law UU Cipta Kerja Sangat Jelek	Negatif
Sebelumnya, Presiden Jokowi menyebut aksi unjuk rasa menolak UU Cipta Kerja disebabkan disinformasi dan hoaks di media sosial. Namun naskah final UU yang disahkan 5 Oktober lalu itu belum bisa diakses publik di saluran resmi pemerintah maupun DPR	Negatif
Mestinya berbaik sangka pada para pendemo, mereka sebenarnya berangkat dari rasa kekhawatiran terhadap nasib rakyat (pekerja) di negeri ini yang bisa dihimpit oleh investor asing dibawah payung Omnibus Law	Negatif

Sebagaimana terlihat pada **Tabel 2**, diperoleh empat kata terbanyak dari setiap kelas, baik kelas sentimen positif maupun negatif. Sebagian besar pengguna Twitter menolak UU Ciptaker. Hal ini dibuktikan dengan jumlah kata terbanyak dari gabungan cuitan yang diklasifikasikan, yakni sebesar 137 kata. Dari cuitan tersebut, diperoleh kesimpulan bahwa penetapan kebijakan ini memantik aksi demonstrasi. Hal ini dibuktikan dengan adanya 89 cuitan dengan kata aksi dan 68 cuitan dengan kata demo yang merujuk pada penetapan kebijakan UU Ciptaker. Demo yang dilakukan oleh masyarakat disebabkan karena kekhawatiran nasib masyarakat terhadap lapangan pekerjaan yang dihipit oleh investor asing. Penyebab lain yang menjadi penolakan masyarakat adalah persoalan komunikasi tentang UU Ciptaker yang dinilai jelek; contohnya, dokumen final tidak dicantumkan di *website* resmi pemerintah sehingga sulit bagi masyarakat untuk mengaksesnya. Dalam menyikapi penolakan masyarakat terhadap UU Ciptaker, pemerintah dapat meninjau kembali pasal-pasal yang berkaitan dengan lapangan pekerjaan domestik, serta melakukan publikasi naskah final sebagai bagian dari sosialisasi kepada masyarakat.

Banyaknya sentimen positif yang didapatkan tidak terlalu signifikan dibandingkan dengan banyaknya sentimen negatif. Sebagai contoh, kata “halal” dan “sertifikat.” Sebagian besar cuitan bersentimen positif yang didapatkan adalah yang berhubungan dengan penerbitan sertifikat halal yang kini tidak tersentralisasi pada Majelis Ulama Indonesia (MUI). Ini menunjukkan bahwa perhatian masyarakat yang menyetujui UU Ciptaker adalah pada pembahasan terkait sertifikat halal yang tidak lagi dimonopoli oleh MUI.

5 Kesimpulan

Hasil pengujian terhadap 868 cuitan dengan menggunakan metode SVM menghasilkan nilai akurasi sebesar 80,4% dengan 25% cuitan yang diberi respons positif dan 75% cuitan yang diberi respons negatif. Berdasarkan hasil pengujian ini, bisa disimpulkan bahwa mayoritas pengguna sosial media Twitter menolak adanya UU Ciptaker. Penolakan tersebut umumnya berkaitan dengan pasal lapangan pekerjaan, sehingga perlu adanya peninjauan ulang oleh pemerintah terhadap pasal tersebut. Penyebab lain penolakan adalah komunikasi tentang naskah UU Ciptaker yang dinilai jelek. Untuk mengatasi terulangnya permasalahan serupa, pemerintah dapat melakukan konsultasi publik yang tepat dan komprehensif sebelum menetapkan kebijakan.

Dari hasil yang didapat, akurasi dinilai sudah cukup baik. Selanjutnya, upaya meningkatkan akurasi algoritma SVM dapat dilakukan dengan memperbanyak jumlah data latih. Semakin banyak jumlah data latih yang digunakan dalam pembentukan model, maka semakin baik pula kinerja dari algoritma SVM.

Referensi

- [1] A. Giachanou and F. Crestani, “Like it or not: A survey of Twitter sentiment analysis methods,” *ACM Comput. Surv.*, vol. 49, no. 2, 2016, doi: 10.1145/2938640.
- [2] S. S. Istia and H. D. Purnomo, “Sentiment analysis of law enforcement performance using support vector machine and K-nearest neighbor,” *Proc. - 2018 3rd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2018*, pp. 84–89, 2018, doi: 10.1109/ICITISEE.2018.8720969.
- [3] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, “Document-level sentiment classification: An empirical comparison between SVM and ANN,” *Expert Syst. Appl.*, vol. 40, no. 2, pp. 621–633,

- 2013, doi: 10.1016/j.eswa.2012.07.059.
- [4] D. Zimbra, A. Abbasi, D. Zeng, and H. Chen, "The state-of-the-art in twitter sentiment analysis: A review and benchmark evaluation," *ACM Trans. Manag. Inf. Syst.*, vol. 9, no. 2, 2018, doi: 10.1145/3185045.
 - [5] G. Draghici, "Media, Public Opinion and the Contemporary Communication," *Proc. 11th Int. Conf. Electron. Comput. Artif. Intell. ECAI 2019*, pp. 2–5, 2019, doi: 10.1109/ECAI46879.2019.9042068.
 - [6] Z. Jianqiang and G. Xiaolin, "Comparison research on text pre-processing methods on twitter sentiment analysis," *IEEE Access*, vol. 5, pp. 2870–2879, 2017, doi: 10.1109/ACCESS.2017.2672677.
 - [7] E. Kušen and M. Strembeck, "Politics, sentiments, and misinformation: An analysis of the Twitter discussion on the 2016 Austrian Presidential Elections," *Online Soc. Networks Media*, vol. 5, no. January, pp. 37–50, 2018, doi: 10.1016/j.osnem.2017.12.002.
 - [8] D. Antonakaki, P. Fragopoulou, and S. Ioannidis, "A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks," *Expert Syst. Appl.*, vol. 164, no. September 2020, p. 114006, 2021, doi: 10.1016/j.eswa.2020.114006.
 - [9] J. Golbeck and D. L. Hansen, "Computing political preference among Twitter followers," *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 1105–1108, 2011, doi: 10.1145/1978942.1979106.
 - [10] T. Redek and U. Godnov, "Twitter as a political tool in eu countries during the economic crisis: A comparative text-mining analysis," *Drus. Istraz.*, vol. 27, no. 4, pp. 691–711, 2018, doi: 10.5559/di.27.4.06.
 - [11] Y. E. Soelistio, M. Raditia, and S. Surendra, "Simple Text Mining for Sentiment Analysis of," *Proc. 7th ICTS*, no. May, pp. 99–104, 2013, doi: 10.12962/p9772338185001.a18.
 - [12] T. Joachims, "UNIVERSIT AT DORTMUND Fachbereich Informatik Lehrstuhl VIII K unstliche Intelligenz Making Large-Scale SVM Learning Practical LS { 8 Report 24," no. October 1999, 2018, doi: 10.17877/DE290R-5097.
 - [13] S. L. Lo, R. Chiong, and D. Cornforth, "Using support vector machine ensembles for target audience classification on Twitter," *PLoS One*, vol. 10, no. 4, pp. 1–21, 2015, doi: 10.1371/journal.pone.0122855.
 - [14] A. Narang, "Twitter Sentiment Analysis on Citizenship Amendment Act in India," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 8, pp. 1714–1724, Jul. 2020, doi: 10.22214/ijraset.2020.30636.
 - [15] N. Prayoga et al., "Unsupervised Twitter Sentiment Analysis on The Revision of Indonesian Code Law and the Anti-Corruption Law using Combination Method of Opinion Word and Agglomerative Hierarchical Clustering," vol. 8, pp. 200–220, 2020, doi: 10.24003/emitter.v8i1.477.
 - [16] F. Colas and P. Brazdil, "Comparison of SVM and some older classification algorithms in text classification tasks," *IFIP Int. Fed. Inf. Process.*, vol. 217, no. August 2006, pp. 169–178, 2006, doi: 10.1007/978-0-387-34747-9_18.