



Analisis *Text Clustering* Masyarakat Di Twitter Mengenai *Omnibus Law* Menggunakan *Orange Data Mining*

Ratu Aghnia Raffaidy Wiguna¹, Andri Irfan Rifai²

¹Accounting Economics , Universitas Padjadjaran, Bandung, Indonesia

^{2,3}Teknik Sipil, Universitas Internasional Batam, Batam, Indonesia

Email: ¹raffaidyww@outlook.com, ²andrirfan@gmail.com

Abstrak

Sosial media dengan platform twitter menjadi hal menarik untuk diteliti. Trending topik tersebut menghasilkan komentar masyarakat Indonesia yang mengandung opini berupa emosi. *Omnibus law* menjadi salah satu topik hangat yang menarik dibahas. Penelitian ini mencoba menganalisis opini di twitter dengan metode analisis vader yang menghasilkan tweet profiler kemudian visualisasi *distribution*. Penelitian ini menggunakan tools orange data mining dengan mengaplikasikan *Preprocess text* yang meliputi *transformation*, *tokenization*, *normalization*, dan *filtering* yang bertujuan agar *text* bisa dianalisis. Kesimpulan dari penelitian ini yaitu respon masyarakat terhadap UU Cipta Kerja *Omnibus Law* mendapat 6 respon dan yang paling tertinggi responnya adalah masyarakat merasa *surprise*.

Kata Kunci: *Visualisasi Data, Orange Data Mining, OmnibusLaw, Distribution*

1. PENDAHULUAN

Pertumbuhan dunia teknologi tumbuh sangat pesat di dunia tidak terkecuali Indonesia dengan menggapai peringkat ketiga Asia untuk jumlah pengguna internet. Tercatat sebanyak 19, 5 juta pengguna twitter di Indonesia [8]. Platform twitter kini diramalkan oleh isu pengesahan Omnibus Law yang sudah cukup jarang terdengar belakangan ini. Pemberlakuan larangan aktivitas untuk mengendalikan penyebaran Covid- 19 sudah memberikan dampak yang sangat besar terhadap perdagangan serta investasi internasional. Perdagangan global diprediksi jatuh sebesar 13- 32%, sedangkan Penanaman Modal Asing(PMA) dapat terpuruk lebih dalam dengan ditaksir antara 30 serta 40%.

Rantai Nilai Global ataupun Global Value Chains(GVC) tengah diatur ulang supaya kurangi ketergantungan kepada Tiongkok. Asia Tenggara dapat diuntungkan oleh relokasi yang tengah terjalin, namun Indonesia relatif tidak menarik untuk investor asing karena keadaan regulasi yang sangat rumit. Presiden mengajukan RUU Cipta Kerja(Omnibus Law) untuk menyederhanakan keadaan



regulasi di Indonesia, , namun jangkauan RUU ini masih terbatas. Terlebih lagi, diperlukan 400 peraturan pelaksana baru yang wajib terbuat dalam satu bulan serta masih berpotensi menciptakan ribuan peraturan yang lain [3].

Konsep omnibus law yang dikemukakan Presiden Jokowi banyak berkaitan dengan bidang kerja pemerintah di sektor ekonomi. Pada Januari 2020, ada dua omnibus law yang diajukan pemerintah, yaitu Cipta Kerja dan Perpajakan. Secara total, ada 11 klaster yang menjadi pembahasan dalam omnibus law RUU Cipta Kerja, yaitu: Penyederhanaan perizinan tanah, Persyaratan investasi, Ketenagakerjaan, Kemudahan dan perlindungan UMKM, Kemudahan berusaha, Dukungan riset dan inovasi, Administrasi pemerintahan, Pengenaan perintah, Pengendalian lahan, Kemudahan proyek pemerintah dan Kawasan Ekonomi Khusus (KEK)[5].

Terdapat 4 pasal kontroversi omnibus law trending twitter diantaranya pasal 33 RUU mengenai omnibus law mendorong liberalisasi impor secara terang-terangan. Pasal 66 mendukung penuh posisi impor yang disetarakan dengan produksi dalam negeri. Pasal 89 menghapus izin atau cuti khusus untuk keperluan menikah, menikahkan, mengkhitankan hingga adanya anggota satu rumah yang meninggal dunia. Terakhir, pasal 117 menghapus denda minimal praktik monopoli sehingga meringankan hukuman bagi pelaku usaha monopoli [7]. Dari perspektif pemerintah republik Indonesia, RUU Omnibus law cipta kerja dikemukakan dan ditujukan untuk menarik investasi dan memperkuat perekonomian nasional serta memperbaiki iklim investasi yang lebih kondusif disampaikan melalui DPR RI untuk dibahas hingga persetujuan versi konstitusi [6].

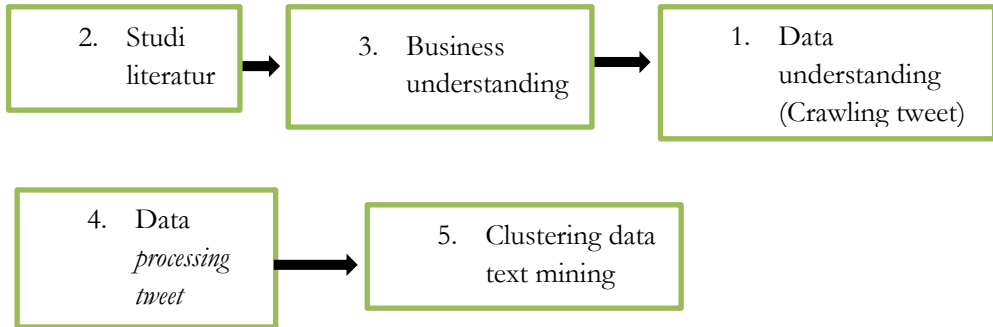
Pengumpulan informasi data dari twitter dapat dilakukan dengan mengintegrasikan Website Scraping serta tools Orange Informasi Mining. Untuk memudahkan mengenali tipe konten dari beberapa informasi tweet, memerlukan proses text mining terhadap data tweet tersebut dengan mengaplikasikan teknik clustering[10]. Untuk mengelompokkan data tekstual bersumber pada kesamaan konten yang dimiliki ke dalam sebagian cluster, sehingga didalam tiap cluster berisi data tekstual dengan konten yang mirip dengan menggunakan teknik clustering.

2. METODOLOGI PENELITIAN

Metode penelitian yang digunakan yaitu metode eksperimen dengan melakukan observasi terhadap variable-variabel objek yang diteliti. Metode eksperimen merupakan suatu penelitian yang kondisi-kondisi tertentu dikendalikan sehingga satu atau beberapa variabel dapat dikontrol.

2.1. Metode Analisis Data

Langkah-langkah yang akan dilakukan dalam penelitian ini terdapat pada desain penelitian di bawah [8] :



Gambar 1. Desain Penelitian

1. Studi literatur
Penelitian sebelumnya serta studi literatur definisi serta hal-hal yang berkaitan dengan text mining.
2. *Business Understanding*
menganalisa masalah dan fakta yang saat ini sedang berkembang di masyarakat.
3. *Data Understanding*, pengumpulan tweet (*crawling tweet*) [11] menggunakan API Twitter dari tanggal 28 Oktober 2020. Pencarian opini menggunakan metode pengumpulan tweets manual yaitu memasukkan berbagai macam kata kunci yang berhubungan dengan topik Omnibus Law pada kolom pencarian twitter dengan menggunakan kata pencarian : omnibus law, rakyat Bersama omnibus law ataupun pencarian dengan hastag contoh #omnibuslaw #rakyatbersamaomnibuslaw
4. *Data preprocess Text (Transformation, Tokenization, Normalization dan Filtering)*. Penentuan class attribute dan Load Dictionary yaitu mencocokkan kata dasar dengan kamus kata sentimen untuk mengetahui kandungan sentimen (positif, netral, negatif). Seluruh data tweet dilabelkan menurut kelas, kelas yang akan dipakai pada penelitian ini ada 3, yaitu kelas positif, kelas negatif dan kelas netral. Proses labelling tweet dilakukan secara manual.
5. *Clustering data text mining* dengan orange data mining menggunakan visualisasi *distribution* yang memvisualisasikan data text mining dengan emosi pengguna tweet yang telah diproses tersebut.

2.2. Metode Pengolahan Data

1. Web scrapping

Metode pengambilan data digunakan untuk mengumpulkan data pendukung yang dibutuhkan dalam pembuatan aplikasi. Untuk dataset yang digunakan diambil melalui Twitter API dengan proses crawling menggunakan *tools orange datamining*. Total dataset yang digunakan pada aplikasi analisis sentimen opini terhadap RUU Omnibus Law Indonesia pada twitter sebanyak 1000 tweet menggunakan 2 hashtag yaitu #OmnibusLaw dan #RakyatBersamaOmnibusLaw yang masing-masing hashtag periode pengumpulan 28 October 2020.

2. Orange data mining

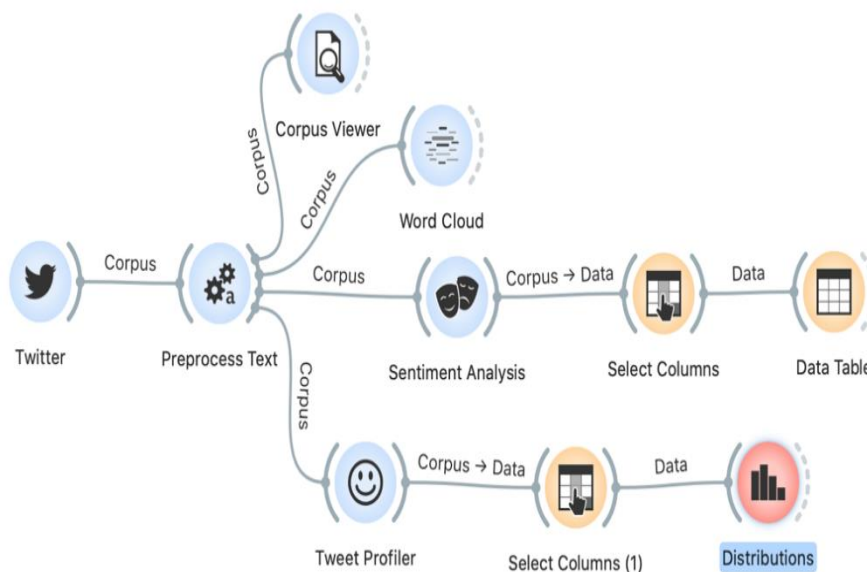
Orange adalah teknologi pembelajaran mesin open source atau perangkat lunak penambangan data. Orange dapat digunakan untuk analisis dan visualisasi data eksploratif. Ini memberikan platform untuk pemilihan eksperimen, pemodelan prediktif, dan sistem rekomendasi dan dapat digunakan untuk penelitian genomik, biomedis, bioinformatika, dan pengajaran. Orange selalu disukai jika faktor inovasi, kualitas, atau keandalannya terlibat. Orange mempermudah pemakai bermain dengan data *open source* serta melaksanakan proses data analytics secara intuitif [3]. Pada permasalahan riset ini Orange Data Mining menunjukkan sebagian widget untuk mencari data informasi kata yang dominan timbul dari konten status serta pendapat/komentar account twitter yang hendak menciptakan tampilan word cloud dari widget Orange Data Mining.

Teknik data mining membantu dalam menemukan pengetahuan tersembunyi dalam tim data yang dapat digunakan untuk menganalisis dan memprediksi perilaku di masa depan. Klasifikasi adalah salah satu metode penambangan catatan yang menetapkan label kelas ke sekumpulan kasus yang tidak diklasifikasikan.

3. HASIL DAN PEMBAHASAN

3.1 Skenario Penelitian

Penggunaan *Orange Data Mining* menampilkan *Design Widget Text Clustering* yang disajikan dalam alur proses Seperti gambar dibawah ini:



Gambar 2. *Design widget text clustering*

Data yang telah dicrawling di twitter akan diinput dan dianalisa satu persatu berdasarkan objek. Selanjutnya, akan dihubungkan terhadap widget yang diperlukan untuk penelitian sehingga menghasilkan desain widget seperti gambar diatas.

3.2 Data Crawling (twitter)

Pada penelitian ini yang menjadi data penelitian adalah trending topik tweet masyarakat Indonesia di *twitter*. Tweet diambil sebanyak 1000 tweet dengan hashtag #OmnibusLaw dan #RakyatBersamaOmnibusLaw dari API dan key token twitter. Dimana hasil akan dianalisa menjadi cluster serta di rekapitulasi sehingga menciptakan kata yang dominan timbul dari status serta pendapat/komentar tersebut. Dalam Orange Data Mining data inputan bisa disebut dengan corpus. Widget corpus yaitu kumpulan dokumen yang bisa menyajikan jumlah baris kalimat, serta memastikan mana fitur yang akan dan tidak akan diinput untuk keperluan analisis [10].

Gambar 3. Data text mining (twitter)

3.2.1 *Preprocess Text*

Sebelum melaksanakan analisis text, pelaksanaan text mining untuk mengolah text dalam perihal ini merupakan preprocess text. Text dipisahkan menjadi unit yang lebih kecil (token), kemudian transformation, tokenization, normalization, serta filtering. Langkah- langkah dalam analisis yang diterapkan secara berurutan dapat diaktifkan serta di non- aktifkan pada Orange Data Mining dalam widget Preprocess Text [10]. Dibawah ini tahapan dalam preprocess text dilakukan sebelum text dianalisis memakai Orange Data Mining:

1. **Transformation**

Pertama yang dilakukan dari *preprocess text* adalah *transformation* merupakan proses mengubah data *input* untuk transformasi huruf kecil secara *default*.

- a. *Lowercase* berfungsi untuk mengubah semua teks menjadi huruf kecil.
- b. *Remove Accents* untuk menghilangkan seluruh aksen didalam teks.
- c. *Parse html* untuk mengetahui tag html dan mem-parsing teks.
- d. *Remove url* untuk menghapus url dari teks.

2. **Tokenization**

Metode memecah teks menjadi komponen yang lebih kecil (kata, kalimat, *bigrams*). a. *Word & Punctuation* akan membagi teks dengan kata per kata dan membiarkan juga simbol tanda baca. *This example = (This), (example), (.)* Proses ini dilakukan setelah *transformation*.

3. **Normalization**

Proses selanjutnya adalah *normalization* yang berlaku *stemming* dan *lemmatization text*. *Text* yang telah dipisahkan kata demi kata akan menghasilkan sebuah *text* yang berdiri sendiri dalam sebuah kalimat. Konten status serta pendapat umumnya memiliki penyusunan yang

kurang sempurna(typo). Diharapkan dengan proses normalisasi ini text tersebut akan dikenal maknanya dengan memakai WordNet Lemmatizer mempraktikkan jaringan persamaan kata(sinonim) kognitif buat token(kata) bersumber pada pada basis informasi lexicon(kamus) bahasa Indonesia yang besar dari NLTK(Alami Language Toolkit).

4. Filtering

Menghapus ataupun menaruh opsi kata. Disini merupakan proses dimana proses filtering kata– kata, simbol yang tidak dibutuhkan dalam proses berikutnya(analisis sentimen).

- a) Stopwords, menghapus stopwords dari bacaan(misalnya, menghapus and, or, this...). ini bisa dicoba dengan memilah bahasa yang akan disaring(filter). NLTK server memberikan stopwords yang bisa diunduh untuk keperluan bahasa. Tetapi dalam riset ini stopwords secara default memakai bahasa inggris. Untuk menyaring kata-kata kunci yang disediakan.
- b) *Regexp* menghapus perkata yang sesuai dengan ekspresi reguler. Default diatur untuk menghapus tanda baca.
- c) *Most Frequent Token*, ialah text yang dipisahkan kata demi kata. Dalam fitur ini hendak memastikan berapa jumlah token paling banyak yang akan timbul serta hendak dianalisis dalam suatu dokumen(corpus). Dalam riset ini menetapkan 1, 000, 000 bagaikan *Most Frequent Token*.

Tahapan preprocess text telah dilakukan, kemudian data telah berbentuk text–text terpisah dan dapat dilihat dalam bentuk word cloud pada Orange Data Mining.

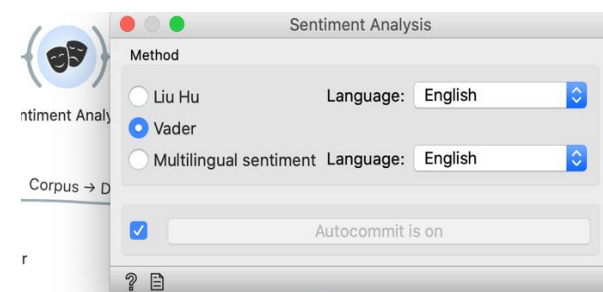


Gambar 4. World cloud omnibuslaw

Word cloud bisa digunakan untuk mengetahui frekuensi kemunculan kata paling banyak. Word cloud membagikan tampilan semacam gambar diatas setelah melewati proses preprocess text, . Dalam hasil tampilan gambar diatas, word cloud dengan frekuensi jumlah kata paling banyak. Jika semakin banyak frekuensi kemunculan kata, dimensi hurufnya pada word cloud menjadi besar pula. Word cloud ialah alterasi atau variasi untuk menunjukkan hasil dari tahapan preprocess text. Corak dari kata membuat tampilan lebih menarik serta mudah dimengerti, foto diatas merupakan hasil tahapan preprocess text dari informasi– informasi yang sebelumnya berisi baris kalimat komentar tweets omnibus law sepanjang bulan oktober 2020.

3.3. Analisis Sentimen

Proses analisis menggunakan algoritma vader untuk mengklasifikasikan polarity(class sentiment) ialah positif, negatif, serta neutral dengan skor total ialah compound.



Gambar 4. Widget Sentiment Analysis Orange Data Mining

Vader akan mengkategorikan, dan berikan skor text bersumber pada nilai pada tiap kata– kata yang terdaftar dalam lexicon vader. Hasil akhir dari evaluasi merupakan skor total ialah compound. Skor total inilah yang hendak direkapitulasi serta dibanding hasilnya. Dalam menganalisis terdapat sebagian atribut variable yang hendak difokuskan dimana dari fitur yang jadi atribut yang digunakan untuk dianalisis(used features) dari widget corpus ialah kolom Texttrans serta ialah text hasil translate yang ada pada tiap informasi kementar, yang tujuannya buat memperoleh hasil berbentuk atribut positif negatif, neutral, serta skor total(compound).

3.4 Tabel Data

Tabel data juga merupakan reaksi hasil untuk menampilkan atribut mana saja yang dipilih untuk ditampilkan sebagai *output* dalam *widget select column* pada

langkah sebelumnya. Dengan bantuan data crawling twitter menggunakan API hasil analisis pada tabel data hasil analisis sentimen akan dihitung seberapa *positif*, *negatif*, dan *neutral* dengan melihat dari skor total (*Compound*) dengan formula perhitungan format data *numerik*.

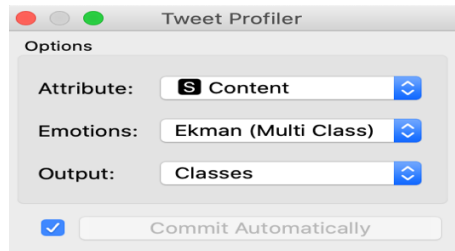
Data Table						
Content True	pos	umber of Retweet	neg	Number of Likes	neu	compound
Kata Bijak Jokowi soal Omn...	0	4	0	7	1	0
Industri dan keberlanjutan ...	0	0	0	0	1	0
Ada Pasal Omnibus Law Cl...	0	36	0	112	1	0
#3 Isu Lingkungan Hidup D...	0	1	0	1	1	0
Isi-Isi Hidup buruh/Hidup ...	0	102	0	297	1	0
yang paling utama adalah ...	0	9	0	15	1	0
ya mau cari kerja ataupun ...	0	0	0	0	1	0
waspada provokasi omnibu...	0	0	0	0	1	0
wah makin mantap aja nih ...	0	0	0	0	1	0
terdapat lebih dari 7 juta or...	0	9	0	13	1	0
solusi hambatan transform...	0	0	0	0	1	0
raga dan nyawa merajut se...	0	0	0	3	1	0
perekonomian para buruh ...	0	0	0	0	1	0
perekonomian para buruh ...	0	0	0	0	1	0
perekonomian para buruh ...	0	0	0	2	1	0
perekonomian para buruh ...	0	0	0	0	1	0
perekonomian para buruh ...	0	0	0	0	1	0
perekonomian para buruh ...	0	0	0	0	1	0
perekonomian para buruh ...	0	0	0	0	1	0
para mafia yg kebakaran ku...	0	0	0	0	1	0
para generasi milenial haru...	0	0	0	0	1	0
omnibus law solusi transfor...	0	0	0	0	1	0
omnibus law meningkatkan...	0	0	0	0	1	0
omnibus law mendorong m...	0	0	0	0	1	0
omnibus law memudahkan ...	0	0	0	0	1	0
omnibus law membuka ban...	0	0	0	0	1	0
omnibus law memberikan k...	0	0	0	0	1	0
omnibus law memberikan ...	0	0	0	0	1	0
omnibus law memberikan ...	0	0	0	0	1	0
omnibus law memberikan ...	0	0	0	1	1	0
omnibus law memberikan ...	0	0	0	0	1	0
omnibus law memberikan ...	0	0	0	0	1	0
omnibus law klaster ketena...	0	0	0	0	1	0
omnibus law dan generasi ...	0	1	0	0	1	0
omnibus law dan generasi ...	0	0	0	0	1	0
omnibus law dan generasi ...	0	0	0	0	1	0
omnibus law adalah kebutu...	0	0	0	0	1	0
milenial solid mendukung o...	0.11	0	0	0	0.89	0.1531
milenial dukung omnibus la...	0	0	0	0	1	0
mantap betul nih omnibus l...	0	0	0	0	1	0
manifest omnibus law bag...	0	0	0	0	1	0
lembaga dunia ini ternyata ...	0	0	0	0	1	0
keberhasilan di dalam keada...	0	0	0	0	1	0

Gambar 6. Tabel Data

3.5 Tweet Profiler

Tweet Profiler mengambil data tentang sentimen dari server untuk setiap tweet(ataupun dokumen) yang diberikan. Widget mengirimkan informasi ke server, tempat model menghitung probabilitas serta/ ataupun skor emosi. Widget menunjang 3 klasifikasi emosi, ialah Ekman, Plutchik serta Profile of Mood States(POMS). Klasifikasi emosi terdapat 3 kategori yaitu Ekman, Plutchik ataupun Profile of Mood States. Klasifikasi kelas jamak akan menghasilkan satu emosi yang sangat bisa jadi per dokumen, sedangkan banyak label akan menciptakan nilai dalam kolom untuk tiap emosi. Penelitian ini akan menggunakan atribut Konten untuk analisis, klasifikasi emosi Ekman dengan opsi multi-kelas dan memilih untuk mengamati variabel Emosi yang telah dikelompokkan dengan orange data mining. Pada penelitian ini menggunakan data 1000 tweets mengenai omnibus law 2020. Data yang telah di crawling

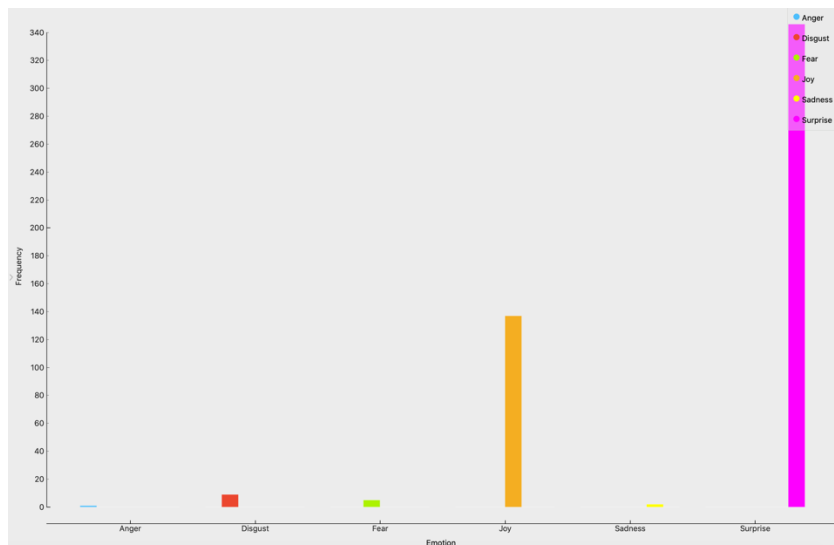
menggunakan widget dari orange data mining dengan Corpus dan dihubungkan ke Tweet Profiler.



Gambar 7. Tweet Profiler

3.6 Distribution

Widget Distribusi menampilkan distribusi nilai atribut diskrit atau kontinu. Jika data berisi variabel kelas, distribusi dapat dikondisikan di kelas. Setelah melakukan tweet profiler pada widget Langkah selanjutnya menghubungkan corpus pada distribution. Hasil akan terlihat 6 bentuk emosi dari data twitter yang telah di input. Pada penelitian ini widget menampilkan emosi para pengguna twitter dengan hashtag #omnibuslaw dan #RakyatBersamaOmnibuslaw. Dari hasil 6 emosi ini data menunjukkan bahwa respon dari twitter per tanggal 28 oktober adalah Joy dan Surprise.



Gambar 8. Distribution

4. KESIMPULAN

Berdasarkan hasil uraian analisis yang dilakukan diatas, dapat ditarik kesimpulan yaitu:

1. Beberapa trending topik yang menjadi fokus pembicaraan mengenai omnibus law di twitter adalah mengenai *burub maju Bersama omnibus law, Bersama pulih ekonomi bangsa, UU Cipta Kerja, waktunya milenial berbisnis, omnibus sejahtera*, dimana yang sekiranya menjadi hal penting bagi pemerintah untuk diperhatikan keluhan rakyat.
2. Metode analisis menggunakan tweet profiler dapat mengetahui mood atau emosi para pengguna twitter dengan trending topik yang sedang terjadi di Indonesia khususnya mengenai UU Cipta Kerja Omnibus Law.
3. Dengan melakukan analisis clustering distribution kita dapat mengetahui klasifikasi pengguna twitter dengan visualisasi emosi yang telah diinput ke dalam setiap corpus dalam orange data mining.

5. DAFTAR PUSTAKA

- [1] "orange biolab." [Online]. <https://orange.biolab.si/workflows/>.
- [2] N. Anggraini and H. Suroyo, "Comparison of Sentiment Analysis against Digital Payment 'T-cash and Go-pay' in Social Media Using Orange Data Mining," *J. Inf. Syst. Informatics*, vol. 1, no. 2, pp. 152–163, 2019.
- [3] antara, "Indonesia Peringkat Lima Pengguna Twitter," *kominfo*, 2012. .
- [4] N. and J. D. Colnerič, "Emotion Recognition on Twitter: Comparative Study and Training a Unison Model," *In IEEE Transactions on Affective Computing*, 2018. .
- [5] Dandy Bayu Bramasta, "Secepat Kilat, Berikut Fakta soal Omnibus Law UU Cipta Kerja," *kompas*, jakarta, p. 1, Oct-2020.
- [6] H. Harahap and A. Hamid, "Analysis of The Importance of Omnibus Law "Cipta Karya" in Indonesia," *Int. J. Sci. Res. Manag.*, vol. 8, no. 08, pp. 236–250, 2020.

-
- [7] M. Hidayat, “Jegal Omnibus Law Trending Twitter, Simak 4 Pasal Kontroversi RUU Cipta Kerja,” *tempo.co*, jakarta, p. 1, Aug-2020.
- [8] S. Juanita, “Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes,” *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 552, 2020.
- [9] S. Kodati and R. Vivekanandam, “Analysis of Heart Disease using in Data Mining Tools Orange and Weka Sri Satya Sai University Analysis of Heart Disease using in Data Mining Tools Orange and Weka,” *Glob. J. Comput. Sci. Technol.*, vol. 18, no. 1, 2018.
- [10] A. Patunru, A. Surianta, and A. Tenggara, “Menarik Penanaman Modal Asing (PMA),” no. 4, pp. 1–9, 2020.
- [11] A. Sentiya, H. Suroyo, F. I. Komputer, and U. B. Darma, “Analisis Text Clustering Akun Fanpage Shopee Indonesia Dengan Komentar Followers Menggunakan Tools Orange Data Mining,” *Bina Darma Conf. Comput. Sci.*, pp. 1055–1067.