[~~https://bisnis.tempo.co/read/1574076/sederet-pernyataan-kontroversial-mendag-soal-kelangkaan-minyak-goreng/full&view=ok~~](https://bisnis.tempo.co/read/1574076/sederet-pernyataan-kontroversial-mendag-soal-kelangkaan-minyak-goreng/full&view=ok)

~~Aturan Harga Eceran Tertinggi (HET) yang ditetapkan oleh Kementerian Perdagangan Republik Indonesia pada Peraturan Menteri Perdagangan Republik Indonesia Nomor 6 Tahun 2022 memberikan dampak yang cukup signifikan pada kelangkaan minyak goreng di pasar Indonesia.~~

~~Pada Penelitian (Afsharizadeh, Maryam & Mohsennia, Mohsen, 2021) minyak goreng merupakan salah satu bahan dalam memproduksi bahan bakar biodiesel. Biodiesel adalah bahan bakar alternatif yang digunakan sebagai bahan bakar mesin diesel. Sedangkan, menurut Satuan Kerja Khusus Pelaksana Kegiatan Usaha hulu migas dan gas bumi (SKK Migas) terdapat ketidakseimbangan antara produksi dan permintaan minyak mentah di Indonesia yang menyebabkan pemerintah harus mengimpor minyak.~~

~~https://ekonomi.bisnis.com/read/20220224/44/1504514/waduh-defisit-minyak-indonesia-harus-impor-500-ribu-barel-minyak#:~:text=Saat%20ini%20produksi%20minyak%20mentah,hingga%201%2C5%20juta%20bph.~~

~~Invasi yang dilakukan Rusia ke Ukraina mengganggu stabilitas kegiatan impor bahan baku minyak goreng Indonesia. Bahan baku yang terganggu terutama pada menipisnya bahan komoditas minyak dan gas (migas). Peningkatan harga bahan baku minyak goreng menyebabkan pasokan minyak goreng diselundupkan dan diekspor secara illegal oleh pihak yang tidak bertanggung jawab demi mencari keuntungan. Akibat tindakan illegal tersebut mengakibatkan terjadinya kelangkaan minyak goreng di pasar Indonesia.~~

Dengan dimikian dibutuhkan suatu analisis penilaian dari pendapat atau pandangan (sentiment) masyarakat terhadap kebijakan pemerintah mengenai undang-undang perlindungan data pribadi yang baru disahkan. Sentiment Analysis dilakukan dengan memanfaatkan suatu pemrosesan model machine learning untuk mendeteksi cuitan masyarakat terkait kinerja pemerintah pada media sosial twitter. Diharapkan dengan adanya penelitian ini, dapat membantu pegawai pemerintah atau aparatur sipil negara (ASN) dalam menganalisa ketepatan suatu kebijakan yang telah dibuat dan dilaksanakan.

Pada penelitian ini dilakukan perbandingan model algoritma multinomial naïve bayes dan support vector machine untuk mengetahui analisis sentiment opini publik terkait undang-undang perlindungan data pribadi. Metode yang dilakukan adalah klasifikasi dengan evaluasi performa menggunakan confusion matrix.

===

Kinerja pemerintah menjadi perhatian penting dalam peningkatan mutu kebijakan dan pegawai ASN. Seperti opini masyarakat terkait dampak dari kebijakan yang telah dibuat dan berjalan. (Laksono, Rachmawan Adi et.al, 2019) menggunakan sentiment analisis untuk menganalisa review customer dan mencari 10 restoran terbaik di Surabaya.

Kemudian pada penelitian (Sutoyo, Edi & Almaarif, Ahmad, 2020) memanfaatkan analisis sentiment untuk mengetahui opini publik tentang wacana pemindahan ibu kota dari Jakarta ke Penajam Paser Utara dan Kutai Kartanegara.

Selain itu, sentimen analisis juga dapat digunakan untuk melihat emosi seputar pengalaman penderitaan publik bersama dari pandemi COVID-19

dan termasuk isolasi sosial skala besar serta hilangnya nyawa manusia dalam penelitian (Lwin, May Oo et.al, 2020).

===

Tweet diambil sejak tanggal 19 Oktober 2022 hingga 28 Oktober 2022 terkait cuitan masyarakat mengenai undang-undang perlindungan data pribadi dengan jumlah data sebesar 380 tweets.

Data diambil menggunakan teknik crawling pada Social Media Twitter dengan melakukan filter pada retweet dan keywordnya adalah “uu pdp” dan “perlindungan data pribadi”

Setelah itu, data yang telah dicrawling disimpan ke dalam csv

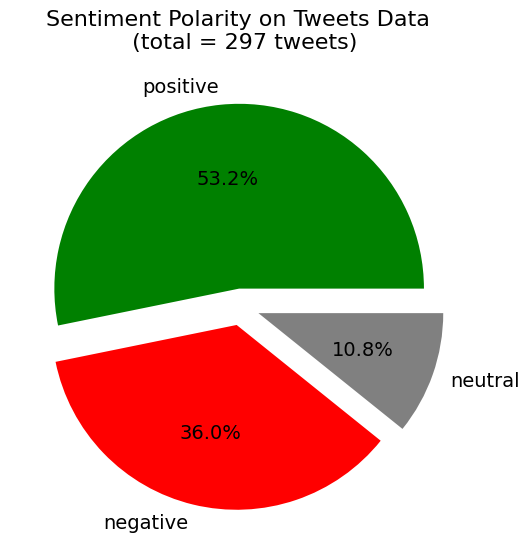
Kemudian, data dilakukan preprocessing untuk pembersihan data dengan beberapa tahapan sebagai berikut

1. Menghapus duplikasi
2. Menghapus mentions
3. Menghapus hash tag
4. Menghapus underscore
5. Menghapus beberapa tanda dan symbol
6. Menghapus retweet
7. Menghapus hyperlink
8. Menghapus quotes
9. Menormalisasi huruf menjadi huruf kecil
10. Melakukan tokenizing
11. Memfilter teks ke dalam Bahasa Indonesia
12. Merubah teks dengan stemming menggunakan Sastrawi

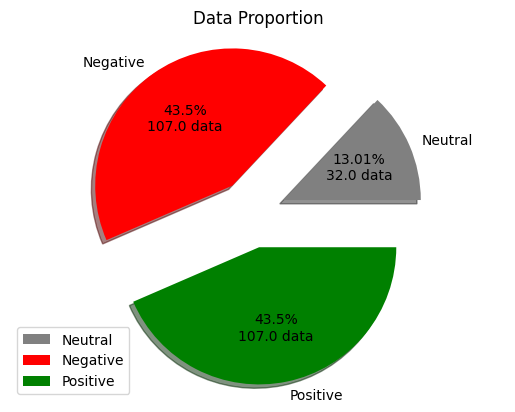
Pada Proses labeling, data diproses menggunakan sentiment polarity dengan kamus Indonesia sentiment lexicon yang dilakukan beberapa tahapan sebagai berikut

1. Setiap kata akan dilakukan pengecekan ke kamusnya yang negative dan positif.
2. Setiap kata yang ada didalam kamus akan diberikan poin sesuai dengan bobot yang berada di kamus.
3. Setiap kata akan digabung menjadi kalimat utuh dan dihitung bobot poinnya untuk diklasifikasikan.
4. Jika kalimat tersebut mendapatkan bobot poin lebih dari 0 maka akan diklasifikasikan sebagai positif, jika kurang dari 0 maka akan diklasifikasikan sebagai negative, dan untuk kalimat berbobot 0 diklasifikasikan sebagai neutral.

Setelah proses preprocessing dan labelling, data akhir keseluruhan berjumlah 297 data dengan tiga kelas, yaitu positif sebesar 158 data, negative sebesar 107 data, dan netral sebesar 32 data. Proporsi data cenderung positif dengan proporsi sebesar 53.2%



Karena data ini imbalance, maka dilakukan Teknik undersampling dengan mengurangi data positif sehingga memiliki proporsi sama dengan data negative. Sehingga didapatkan data sebagai berikut.



Karena data netral tidak digunakan dalam membuat model maka kita dapat menghapusnya.

Chart, pie chart

Description automatically generated

Pada tahapan pemodelan, sebanyak 214 data dibagi ke dalam dua kategori, yaitu data training untuk melatih model dan data test untuk mengevaluasi model. Data dibagi sebesar 90% sebesar 192 data untuk data training dan 10% sebesar 22 data untuk data test.

Setelah data dibagi, data train dan test dipelajari kosa-katanya dan diubah menjadi matriks menggunakan teknik transform pada vectorizer.

Kemudian data train yang sudah ditransform menggunakan vectorizer dilatih menggunakan model multinomial naïve bayes yang dibandingkan dengan support vector machine. Pada evaluasinya, data diuji menggunakan data test yang telah ditransform menggunakan vectorizer juga.

Hasil akurasi model training dari model multinomial naïve bayes didapatkan sebesar 93.22% dengan akurasi test cross-validation sebesar 77.27%.

Sedangkan model support vector machine hanya mendapatkan akurasi training sebesar 95.83% dengan akurasi test sebesar 72,72%.

Terakhir dalam mengukur performa evaluasi, penelitian ini menggunakan confusion matrix untuk mengetahui nilai overall accuracy, precision, recall, dan f1-scorenya. Model multinomial naïve bayes mendapatkan nilai overall accuracy sebesar 77.27% dengan weighted avg pada precision 75%, recall 81.81%, dan f1-score 78.26%.

Sedangkan model support vector machine mendapatkan nilai overall accuracy sebesar 72.72% dengan weighted avg pada precision 77.77%, recall 63.63%, dan f1-score 70%.