

Analisis Sentimen Masyarakat terhadap Vaksin Covid-19 pada Twitter dengan Naïve Bayes

(Pasca Vaksinasi Perdana)

Sur Hidayat, Kusri, Ferry Wahyu Wibowo

Magister Teknik Informatika

Universitas Amikom Yogyakarta

Jl. Ring Road Utara, Sleman, Yogyakarta

sur.1274@students.amikom.ac.id, kusri@amikom.ac.id, ferry.w@ amikom.ac.id

Abstrak— Percepatan dalam menanggulangi penyebaran Covid-19 di Indonesia dilakukan oleh Pemerintah dengan memberikan vaksin pada masyarakat. Pemerintah telah melakukan vaksinasi perdana pada Januari 2021 oleh presiden dan pejabat pemerintahan. Informasi tentang vaksin Covid-19 telah digulirkan sejak November 2020 dan menimbulkan berbagai pendapat di masyarakat, mendukung dan menolak vaksin. Setelah vaksinasi perdana, tetap ada pendapat yang positif dan negatif. Pendapat ini bisa menjadi penilaian tanggapan masyarakat terhadap vaksinasi. Twitter bisa dimanfaatkan untuk analisis sentimen masyarakat terhadap isu tertentu. Data dari Twitter diambil dan dikumpulkan menggunakan API twitter yang tersedia, sehingga diperoleh pendapat berupa teks komentar berdasarkan hastag yang diinput. Penelitian ini bertujuan melakukan analisis tanggapan masyarakat terhadap vaksin covid-19 dengan mengklasifikasikan tweet dalam kategori positif, netral dan negatif. Pengelompokan tanggapan masyarakat menggunakan metode Naïve Bayes untuk mengetahui jumlah opini negative, netral dan positif. Crawling data Twitter dengan hastag #vaksinasi, #vaksinacovid19, #vaksinacovid, #VaksinUntukKita, #vaksinsinovac dengan jumlah 1000 tweet dalam bahasa Indonesia. Google Colab untuk melakukan crawling, analisis data dan kalsifikasi Naïve Bayes, maka diperoleh hasil pendapat positif sebanyak 32,7%. Sedangkan pendapat negatif adalah 14,4%. Akurasi menggunakan metode klasifikasi Naïve Bayes mencapai 92,2%. Validasi dilakukan dengan metode confusion matrix dan diperoleh nilai akurasi 92,2%, presisi 92,6% dan recall 91,8%.

Kata kunci— analisis sentimen; vaksin; twitter; naïve bayes

I. PENDAHULUAN

Pandemi yang disebabkan oleh virus corona atau Covid-19 ditetapkan secara resmi oleh *World Health Organization* (WHO) mulai Maret 2020 [1]. Penyebaran virus corona berawal dari Kota Wuhan, China, pada akhir 2019 dan telah menyebar ke dunia dengan jumlah pasien mencapai 95,612,831 orang dan yang meninggal 2,066,176 orang sampai dengan 20 Januari 2021 [2]. Pada Maret 2020 pertama kali ditemukan kasus Covid-19 di Indonesia, yang menjangkiti dua orang di Depok Jawa Barat [3]. Penularan Covid-19 di Indonesia terus bertambah bahkan meningkat, pada tanggal 20 Januari 2021 terdapat

jumlah 951.651 kasus dengan kematian 27.203 orang[4]. Pandemi memberikan dampak langsung di seluruh negara dunia, sebagai akibat dari diberlakukannya protokol kesehatan yang harus ditetapkan pada seluruh aspek kegiatan, mulai dari pembatasan sosial hingga *lockdown* sehingga menghambat aktivitas manusia. Efek lebih lanjut, berpotensi masalah besar terhadap kesehatan manusia dan efek buruk pada ekonomi global. Untuk itu penyebaran virus harus bisa dikendalikan dengan efektif[5].

Pesatnya penyebaran COVID-19 dan bahaya darurat kesehatan jika tidak segera diatasi, salah satu cara yang dilakukan untuk mencegah penyebaran virus ini adalah vaksinasi [5]. Fungsi dari vaksin adalah melindungi dan meningkatkan deteksi virus yang masuk tubuh. Vaksinasi dapat mengurangi dan menghambat penyebaran penyakit di dalam kelompok masyarakat [6]. Sangat penting vaksin dikembangkan dengan spesifikasi aman dan efektif sehingga dapat mencegah penyebaran penyakit lebih luas dan berlangsung lama [5].

Pemerintah Indonesia secara aktif dan cepat akan melakukan kegiatan vaksinasi yang diberikan kepada masyarakat. Untuk mendukung proses vaksinasi yang cepat dan lancar maka Presiden Joko Widodo mengeluarkan Peraturan Presiden Republik Indonesia Nomor 99 Tahun 2020 pada tanggal 5 Oktober 2020. Peraturan Presiden ini berisi tentang Pengadaan Vaksin dan Pelaksanaan Vaksinasi Dalam Rangka Penanggulangan Pandemi COVID-19. Isi dari Perpres ini mengatur kewenangan pemerintah, kementerian atau lembaga pemerintah dalam rencana kegiatan vaksinasi [7]. Perpres ini menjadi dasar untuk melakukan kerjasama untuk pengadaan vaksin di Indonesia dengan mitra internasional [8].

Publikasi vaksinasi dilakukan dengan masif pada semua aspek yaitu dari kelayakan penggunaan vaksin, kemungkinan risiko pasca vaksinasi, sampai tahapan dan prosedur pemberian vaksin untuk masyarakat. Masyarakat edukasi melalui kampanye yang berisi tentang pentingnya pengendalian penyebaran Covid-19 di Indonesia. Proses vaksinasi Covid-19 juga telah dilaksanakan dengan diawali pemberian Vaksin

Sinovac kepada Presiden Joko Widodo dan para pejabat Pemerintah pada tanggal 15 Januari 2021.

Tanggapan dan opini publik disampaikan melalui berbagai media. Media sosial merupakan media yang banyak digunakan untuk menyampaikan pendapat. Menurut statistik digital global pada Juli 2020 dalam We Are Social 2020, jumlah pengguna media sosial di Indonesia pada 2020 melebihi 150 juta. Data tentang pengguna media sosial di Indonesia menunjukkan bahwa twitter merupakan media sosial yang populer digunakan dengan angka mencapai 52% dari keseluruhan data penggunaan media sosial di Indonesia [12]. Data pada twitter merupakan bagian dari big data yang bersifat stream dan bisa digunakan sebagai sumber data analisis terkait komentar pengguna di media sosial. Sumber data pada twitter ini sebagai alternatif dan terobosan dalam mengumpulkan data di internet dalam proses survey tradisional. Dengan mempertimbangkan tahapan proses yang diperlukan, maka prosesnya sangat panjang, dan dianggap sulit untuk menggunakan survei tradisional secara langsung untuk mengumpulkan data tentang tanggapan dan opini publik.

Berbeda dengan harus melakukan survei tradisional, pengumpulan data melalui media sosial diyakini akan memberikan efisiensi di segala aspek. Efisiensi ini mencakup biaya untuk memperoleh data dalam jumlah minimum, dapat memperoleh data secara real time, dan menghasilkan data yang berisi informasi yang lebih detail untuk menggambarkan opini publik yang sebenarnya [13]. Banyak studi sebelumnya juga telah melakukan aktivitas yang menggunakan data dari media sosial twitter untuk menganalisis reaksi dan opini publik. Misalnya, studi dilakukan untuk memahami opini publik tentang kebijakan ganjil dan genap India [14], dan untuk memahami layanan LRT India. Bagaimana dengan opini publik. Los Angeles [15], Chicago [13] dan KRL commuter line di Indonesia [16].

Berdasarkan latar belakang di atas, peneliti menggunakan data dari twitter untuk mendapatkan data tanggapan dan pendapat masyarakat terhadap Vaksin Covid-19. Data yang diperoleh dari twitter akan diolah untuk melakukan analisis sentimen dengan kategori positif, netral dan negatif. Metode Naïve Bayes digunakan untuk mengklasifikasikan tanggapan terhadap vaksin Covid-19. Untuk proses crawling data dan kalsifikasi menggunakan aplikasi *Google Colab*. Pada hasil akhir diharapkan akan memperoleh besaran nilai kelompok tanggapan positif dan negatif terhadap pelaksanaan vaksinasi Covid-19.

I. METODE

A. Analisis Sentimen

Riset analisis sentimen yang dikembangkan sejak tahun 2003 dan merupakan bagian dari text mining merupakan riset komputasi yang didasarkan pada setiap ekspresi emosi, emoji, opini, komentar dan ekspresi teks[17]. Analisis sentimen atau analisis sentimen dalam bahasa Indonesia adalah teknik atau metode yang digunakan untuk mengidentifikasi bagaimana menggunakan teks untuk mengekspresikan emosi dan membagi emosi menjadi emosi positif atau negatif [18]. Dalam penggunaan media sosial khususnya dan pengguna internet secara umum akan banyak menulis tentang pengalaman, opini, dan hal-hal lain yang berkaitan dengan mereka.

Mengekspresikan perasaan dengan perasaan positif, netral dan negatif dapat diungkapkan dengan cara yang sangat kompleks [20].

B. Twitter

Twitter adalah layanan jejaring sosial yang memungkinkan pengguna untuk saling berkomunikasi. Pengguna twitter mengirimkan pesan dengan batasan sampai 140 karakter. Pesan dalam bentuk huruf, angka dan symbol ini disebut tweet. Hashtag digunakan untuk menandai topik di Twitter. Kata-kata pada tag topik dimulai dengan "#", diikuti dengan topik yang sedang dibahas. Tag sering digunakan untuk meningkatkan visibilitas tweet pengguna.

C. Naïve Bayes Classifier

Klasifikasi Naïve Bayes merupakan metode klasifikasi statistik yang dapat memprediksi probabilitas anggota kelas, seperti probabilitas sampel tertentu yang termasuk dalam kelas tertentu [21]. Penelitian ini akan menggunakan metode ini untuk mengklasifikasikan data yang diekstrak dari Twitter dan membagi data tersebut menjadi kategori positif, negatif dan netral. Metode Naïve Bayes digunakan dalam penelitian klasifikasi sentimen ini karena memiliki beberapa keunggulan diantaranya kesederhanaan, kecepatan dan akurasi yang tinggi. Klasifikasi menggunakan Naïve Bayes memiliki formula perhitungan (1) sebagai berikut [22]:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

X = data dengan kelas tidak dikenal

H = hipotesis data X adalah kelas khusus

P(H|X) = probabilitas hipotesis H didasarkan pada kondisi X

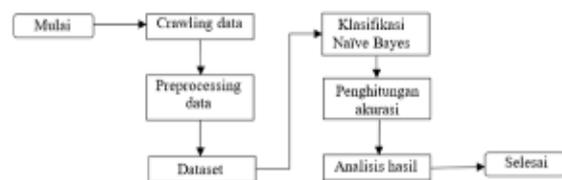
P(H) = probabilitas hipotesis H

P(X|H) = probabilitas hipotesis X didasarkan pada kondisi H

P(X) = probabilitas X

D. Crawling data

Sistem yang dibangun pada penelitian ini digunakan untuk memprediksi setiap tweet tentang pendapat positif, negative atau netral tentang vaksinasi covid-19 di Indonesia dan dalam bahasa Indonesia. Untuk memperoleh data tweet menggunakan application programming interface (API) Twitter. API Twitter diperoleh dengan registrasi sebagai developer dan di setuju oleh pihak Twitter. Proses selanjutnya adalah dilakukan pre-processing data untuk membersihkan data dari karakter yang berbeda dengan angka dan huruf. Setelah data bersih, pilih data latih dan data uji. Gambar 1 menggambarkan sistem yang digunakan.



Gambar 1. Diagram alir penelitian

II. HASIL DAN PEMBAHASAN

A. Proses Crawling data

Pengumpulan data atau proses perayapan data Twitter ini menggunakan kode API Twitter yang telah diberikan, dan bahasa pemrograman *python* dalam web aplikasi *Google Colab* membantu proses pengambilan data Twitter. Kode API Twitter adalah antarmuka pemrograman aplikasi di web twitter ini. Kunci layanan API ini berisi sekumpulan fungsi, protokol, komponen, dan perintah, untuk menyederhanakan pemrograman saat membangun sistem informasi atau aplikasi. Kunci API Twitter merupakan aplikasi yang berisi *access key*, *access secret*, *consumer keys* dan *consumer secret*. *Consumer keys*, *consumer secret*, *access key*, dan *access secret* ini digunakan untuk mengakses dan mengambil data tweet sesuai dengan parameter yang digunakan dan dalam batas yang diijinkan oleh twitter. Selain membutuhkan kunci API Twitter, untuk proses mengambil data menggunakan *Google Colab*, dengan library dari *python-3*. Proses crawling dilakukan dengan memasukkan kunci API, kemudian dilakukan pembatasan tanggal pengambilan data tweet yang sesuai dengan hastag, seperti ditunjukkan pada Gambar 2.

```
auth = OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)

def scrapetweets(search_words, data_since, data_until):
    today = datetime.today().strftime("%Y-%m-%d")
    last_week = datetime.today() - timedelta(7)
    last_week = last_week.strftime("%Y-%m-%d")

    search_words = "#vaksinasi OR #vaksinCOVID19 OR #vaksinCOVID OR #VaksinUntukKita"
    date_since = last_week
    date_until = today
```

Gambar 2. Penggunaan Code API untuk Crawling

Selanjutnya code yang digunakan pada Gambar 2 adalah Pengambilan data dari Twitter menggunakan parameter hastag yang berjumlah 5. Hastag saat proses crawling data menjadi topic Twitter atau menjadi pembahasan pada tweet. Hastag yang digunakan adalah *#vaksinasi*, *#vaksinCOVID19*, *#vaksinCOVID*, *#VaksinUntukKita* dan *#vaksinSinovac*. Data crawling dengan batasan jumlah yaitu 1000 tweet pada tanggal 14 Januari 2021 sampai dengan 20 Januari 2021.

B. Preprocessing data

Dalam preprocessing terdapat beberapa langkah untuk mengubah kalimat-kalimat dalam tweets ke dalam bahasa standar, karena tweets tidak sepenuhnya dikumpulkan menggunakan kata-kata standar. Pre-processing sangat berguna untuk menghilangkan beberapa notasi yang tidak digunakan. Preprocessing dilakukan dengan menggunakan library bahasa pemrograman *python-3* yang diinstall pada *Google colab*. Preprocessing juga digunakan untuk mendapatkan hasil dataset yang maksimal. Dalam proses tokenisasi, sangat berguna untuk memisahkan kata, simbol, frasa, dan entitas dalam teks. Dalam proses ini juga menggunakan library *nlTK* untuk melakukan preprocessing.

Pada Gambar 3 Menunjukkan proses pembersihan kata dalam bahasa Indonesia, kemudian dilakukan proses tokenize data untuk mengelompokkan kata dan huruf menjadi kode yang bisa dibaca oleh algoritma.

```
def preprocessing_data(x):
    return p.clean(x)

def tokenize_data(x):
    return p.tokenize(x)

data['tweet_clean'] = data['text'].apply(preprocessing_data)
data['tweet_clean'] = data['tweet_clean'].apply(tokenize_data)
data = data.drop_duplicates()
```

Gambar 3. Code preprocessing dan tokenize

C. Stemming Data

Dalam proses stemming dapat digunakan untuk menghilangkan imbuhan tiap kata, termasuk imbuhan sebelum atau sesudah kata. Dalam proses stemming dilakukan dengan bantuan library *google translator* yang termasuk dalam fasilitas *Google colab*. Dalam proses ini, menggunakan translator untuk mempercepat proses stemming dengan mengubah kata menjadi bahasa inggris. Proses ini ditunjukkan pada Gambar 4 sebagai berikut.

```
from google_trans_new import google_translator
translator = google_translator()

def convert_eng(tweet):
    return translator.translate(tweet, lang_tgt='en')

data['tweet_english'] = data['tweet_clean'].apply(convert_eng)
```

Gambar 4. Proses stemming data

Proses stemming dengan code porter stemmer dilakukan untuk memproses data pada stemming bahasa inggris yang ditunjukkan pada Gambar 5.

```
ps = PorterStemmer()

def stemming_data(x):
    return ps.stem(x)

data['tweet_english'] = data['tweet_english'].apply(stemming_data)
```

Gambar 5. Porter stemmer

D. Menggunakan Naive Bayes

Pada proses ekstraksi fitur dan proses pengklasifikasian Naive Bayes yang nantinya akan di compress menjadi satu class pipeline vectorizer kemudian transformer dilanjutkan classifier. Pengklasifikasian tweet tersebut dilakukan dengan bantuan library *python-3* yang mempunyai nama library *scikit-learn* untuk proses klasifikasi naive bayes. Untuk pembacaan data, *python -3* memiliki fasilitas library *numpy* dan juga *pandas*. Untuk library *scikit-learn* disini yang digunakan adalah Pipeline, CountVectorizer, Naive Bayes classifier dan akurasi. Untuk langkah awal pengerjaan proses ekstraksi fitur dan klasifikasi adalah dilakukan proses menginstall library yang diperlukan. Proses pengklasifikasian data ini dilakukan dengan menggunakan perhitungan probabilitas dari setiap kelas, maka diperoleh hasil jelas dari prediksi data yang di-input. Tahapan akhir setelah melakukan semua proses pengklasifikasian, maka menghitung dari akurasi dari algoritma yang digunakan.

```

from textblob.classifiers import NaiveBayesClassifier
cl = NaiveBayesClassifier(train_set)
print('Akurasi Test:', cl.accuracy(dataset))

Akurasi Test: 0.922

```

Gambar 6. Code running proses Naïve bayes

Pada Gambar 6 diperlihatkan proses running untuk menganalisa data dengan algoritma naïve bayes dan diperoleh hasil test akurasi sebesar 0,922. Nilai akurasi yang didapatkan dari pengujian dengan seluruhnya jumlah 1000 data.

```

status = pd.DataFrame({'klasifikasi_bayes': status})
data['klasifikasi_bayes'] = status

print(f'\nHasil Analisis Data:\nPositif = {total_positif}\nNetral = {total_netral}')
print(f'\nTotal Data : {total}')

Hasil Analisis Data:
Positif = 327
Netral = 538
Negatif = 143

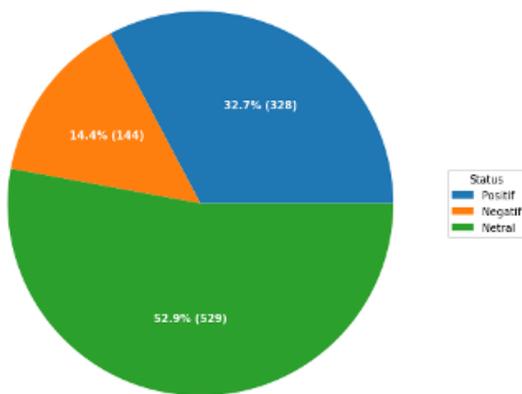
Total Data : 1000

```

Gambar 7. Code running untuk training dan test data dengan Naïve Bayes

Setelah dilakukan proses klasifikasi dengan algoritma Naïve Bayes diperoleh hasil analisis data dari total 1000 tweet dengan pendapat positif berjumlah 327 tweet, jumlah pendapat negatif adalah 143 tweet (lihat Gambar 7). Sedangkan untuk pendapat yang netral sebanyak 530 tweet. Tingkat akurasi dari analisis yang dihasilkan adalah 92,2% dengan Naïve Bayes. Untuk menguji hasil yang valid dilakukan pengujian dengan metode confusion matrix dan diperoleh nilai akurasi 92,2%, presisi 92,6% dan recall 91,8%.

Selanjutnya Gambar 8 menunjukkan hasil analisis ditampilkan dengan visualisasi grafik pie.



Gambar 8. Grafik Pie Hasil klasifikasi positif, negative dan netral

III. KESIMPULAN

Berdasarkan hasil analisis yang dilakukan terkait sentimen masyarakat terhadap vaksin covid-19 setelah dilaksanakan vaksinasi perdana, dengan sumber data pada tanggal tanggal 14 – 20 Januari 2021, dengan menggunakan metode Naïve Bayes di peroleh angka positif sebesar 32,7% dan sentimen negatif sebesar 14,4%. dengan akurasi sebesar 92,2%. Untuk menguji

hasil yang valid dilakukan pengujian dengan metode confusion matrix dan diperoleh nilai akurasi 92,2%, presisi 92,6% dan recall 91,8%. Penjelasan dari angka yang diperoleh menunjukkan bahwa tingkat sentimen positif lebih besar dibandingkan dengan sentimen negative pada media sosial twitter. Dengan hasil penelitian ini bisa dilihat bahwa masyarakat menerima dan mendukung vaksin covid-19 berdasarkan data dari Twitter. Namun penelitian ini masih terdapat beberapa kekurangan bisa dilihat pada angka pendapat netral sebesar 52,9%, menunjukkan bahwa banyak pendapat yang umum. Hal tersebut bisa terjadi karena proses stemming menggunakan tweet English dan library berdasarkan bahasa inggris. Pada penelitian selanjutnya bisa digunakan library satrawi pada proses stemming serta menambahkan data training dan data uji dengan membandingkan hasil akurasi klasifikasi algoritma lainnya seperti support vector machine (SVM).

DAFTAR PUSTAKA

- [1] WHO.Question and answers hub (who.int), <https://covid19.who.int/> , 2021.
- [1] WHO. WHO Coronavirus Disease (COVID-19) Dashboard | WHO Coronavirus Disease (COVID-19) Dashboard, <https://covid19.who.int/> , 2021.
- [2] Nuraini R. Kasus Covid-19 Pertama, Masyarakat Jangan Panik _ Indonesia. Indonesia.go.id [Internet].; <https://indonesia.go.id/narasi/indonesia-dalamangka/ekonomi/kasus-covid-19-pertama-masyarakat-jangan-panik/> , 2020
- [3] Peta Sebaran COVID-19 | Satgas Penanganan COVID-19, <https://covid19.go.id/> , 2021
- [4] Liu C, Zhou Q, Li Y, Garner L V, Watkins SP, Carter LJ, et al. Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. 2020.
- [5] Sari I.P., Sriwidodo. Perkembangan Teknologi Terkini dalam Mempercepat Produksi Vaksin Covid-19. 2020;5(5):204–17.
- [6] Peraturan Presiden Republik Indonesia; 2020 p. 1–13.
- [7] Hakim RN. Menlu Retno dan Menteri BUMN Akan ke Inggris dan Swiss Amankan Stok Vaksin Covid-19. kompas.com [Internet]. 2020; Available from: https://nasional.kompas.com/read/2020/10/12/09_074911/menlu-retno-dan-menteri-bumn-akan-ke-inggris-dan-swiss-amankan-stok-vaksin
- [8] Hastuti RK. Mohon Doanya! Bulan Depan Indonesia Mulai Vaksinasi Covid-19. cnbcindonesia.com [Internet]. 2020; <https://www.cnbcindonesia.com/news>
- [9] Anwar F. Program Vaksin COVID-19 Mulai November, Apa Itu Emergency Use Authorization? detik.com [Internet]. 2020; <https://health.detik.com/berita-detikhealth>
- [10] Artanti A ayu. Kabar Gembira, Pemerintah Mulai Program Vaksin November 2020 , <https://www.medcom.id/ekonomi/bisnis/> , 2020
- [11] Digital 2020 - We Are Social. <https://wearesocial.com/blog/2020/07/digital-use-around-the-world-in-july-2020>, 2020.
- [12] Collins C, Hasan S, Ukkusuri S V., A Novel Transit Rider Satisfaction Metric : Rider Sentiments Measured From Online Social Media Data, Journal Public Transportation, 2013.
- [13] Basu R, Khatua A, Jana A, Ghosh S. Harnessing, Twitter Data for Analyzing Public Reactions to Transportation Policies : Evidences from The Odd Even Policy in Delhi India. 2017.
- [14] Luong TTB, Houston D. Public Opinions Of Light Rail Service In Los Angeles : An Analysis Using Twitter Data. iConference 2015 Proceeding. 2015.
- [15] Pratama MO, Satyawan W, Jannati R, Pamungkas B, Raspiani, Syahputra ME, The Sentiment Analysis of Indonesia Commuter Line Using Machine Learning Based on Twitter Data. Journal Physic Conference Ser. 2019.

- [16] Pramana S, Yuniarto B, Mariyah S, Santoso I, Nooraeni R., Data Mining dengan R Konsep Setara Implementasi Pertama. Bogor In Media, 2018.
- [17] Nasukawa, T., Yi, J., Sentiment Analysis: Capturing Favorability Using Natural Language Processing. In Proceedings of The 2nd International Conference on Knowledge Capture. pp. 70–77. 2003.
- [18] Cvijikj I.P., Michahelles, F., Understanding Social Media Marketing: A Case Study on Topics, Categories and Sentiment on a Facebook Brand Page. 2011.
- [19] Coletta, L.F.S. et al., Combining Classification and Clustering for Tweet Sentiment Analysis. Brazilian Conference on Intelligent Systems, IEEE, pp. 210–215. 2014.
- [20] Troussas, C. et al., Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning. In IISA 2013. IEEE, pp. 1–6. 2013
- [21] Leung, K. M., Naive Bayesian Classifier. Polytechnic University Department of Computer Science/Finance and Risk Engineering. 2017.