

Analisis Sentimen Twitter Terhadap Bitcoin dan Cryptocurrency Berbasis Python TextBlob

Rizky Parlika¹, Sunu Ilham Pradika², Amir Muhammad Hakim³, Kholilul Rachman N.M⁴

¹²³⁴Informatika, UPN "Veteran" Jawa Timur
¹rizkyparlika.if@upnjatim.ac.id

²17081010045@student.upnjatim.ac.id

³17081010051@student.upnjatim.ac.id

⁴17081010055@student.upnjatim.ac.id

Abstrak — Teknologi *Cryptocurrency* saat ini sangat berkembang ditandai dengan banyaknya orang yang mau berpartisipasi untuk menjadi bagian di dalamnya. Salah satu *Cryptocurrency* yang terkenal adalah *Bitcoin*. Namun, bagi sebagian orang tentu tidak mengerti apa itu *Bitcoin* dan bagaimana cara kerjanya. Sehingga, hal itu membuat mereka ragu untuk ikut turut serta menjadi bagian dalam teknologi ini. Sehingga perlu dibuat sebuah analisis sentimen mengenai teknologi ini sehingga dapat memberi pengetahuan apa itu sesungguhnya *Bitcoin*. Melalui *Twitter*, analisis sentimen dapat dilakukan guna mencari tahu bagaimana pendapat banyak orang mengenai *Bitcoin*. Hal ini dapat dilakukan dengan memanfaatkan *API* yang disediakan oleh *Twitter* untuk mengumpulkan data berupa *tweets*. Kemudian, melakukan analisis sentimen menggunakan *library Python* untuk mengklasifikasikan suatu *tweet* menjadi positif, negatif, atau netral. Dari hasil klasifikasi menggunakan dataset yang berjumlah 3433 data didapatkan bahwa 41,3% diklasifikasikan positif, 44,9% netral, dan 13,7% negatif. Lalu, untuk visualisasi dari keseluruhan pendapat orang-orang mengenai *Bitcoin* dapat dibentuk dengan menggunakan *Word Cloud* sehingga hanya akan muncul kata-kata penting yang berhubungan dengan *Bitcoin*. Diharapkan setelah penelitian ini dilakukan dapat dilihat bagaimana pendapat orang-orang khususnya pengguna *Twitter* mengenai *Bitcoin*.

Kata Kunci — *Bitcoin*, *Textblob*, *Cryptocurrency*, *Word Cloud*, *NLP*

I. PENDAHULUAN

Pada era digital seperti sekarang data adalah sumber daya yang memenuhi kehidupan. data yang banyak ini kemudian semakin menumpuk setiap harinya dan apabila dibiarkan tentunya hanya akan menjadi sampah data. Salah satunya adalah data *tweets* yang ada pada *Twitter*. Data *tweet* tersebut semakin bertambah banyak seiring dengan bertambahnya hari dan apabila dibiarkan akan menjadi sampah data. Belakangan ini populer sebuah teknik untuk menggali informasi dari tumpukan data khususnya pada *Twitter* yakni *Text Mining* atau sering juga disebut dengan analisis sentimen. Analisis sentimen merupakan proses memahami, mengekstrak dan mengolah data teksual secara otomatis untuk mendapatkan informasi dengan tujuan untuk melihat kecenderungan opini terhadap sebuah objek, apakah cenderung beropini positif atau

negatif atau netral [1]. Analisis sentimen umumnya dilakukan dengan menggunakan *Natural Language Processing* (NLP). Dalam beberapa tahun terakhir harga *Bitcoin* terus meningkat dan telah menjadi investasi bagi para investor *Bitcoin*. *Bitcoin* merupakan salah satu *Cryptocurrency* yang bersifat terdesentralisasi dan tidak diatur dan dijamin oleh otoritas pusat. *Bitcoin* bersifat sangat fluktuatif dan sering kali membuat resah para investor *Bitcoin* [2]. Oleh karena itu untuk memaksimalkan keuntungan dan meminimalisir resiko, maka diperlukan sebuah cara bagi investor agar dapat memprediksi harga *Bitcoin* khususnya dengan melakukan analisis sentimen *Bitcoin*.

Beberapa penelitian yang terkait dengan sentimen analisis untuk melakukan prediksi suatu harga yakni penelitian yang dilakukan oleh Kalyani Joshi dkk (2016) yang mampu menerapkan analisis sentimen untuk memprediksi harga saham dengan artikel berita dan harga saham sebelumnya dengan menggunakan 3 metode, yakni *Support Vector Machine*, *Random Forest* dan *Naive Bayes Classifier* [3].

Disisi lain, pada penelitian yang dilakukan oleh Jessica (2019) yang mampu menerapkan analisis sentimen pada lima akun official news di *Twitter* untuk memprediksi harga saham *Facebook* dengan menganalisis *Moving Average* saja [4]. Lalu, pada penelitian yang dilakukan oleh Yancong Xie dkk (2017) yang mampu menerapkan *Text Mining* menggunakan data dari artikel finansial online di China dengan metode *Support Vector Machine* untuk memprediksi harga saham di pasar saham China [5].

Penggunaan akses *Application Programming Interface* (*API*) juga dimanfaatkan Parlika, R., Pratama, A., (2019) untuk menampilkan berita dari web menggunakan bot telegram, dimana web penyedia berita menyediakan akses berita melalui share *API*. Selain dari web, telegram menyediakan share *API* yang cukup untuk membangun aplikasi yang bermanfaat melalui teknologi bot nya, dan ini dimanfaatkan oleh Parlika, R., dkk (2019) untuk membuat aplikasi login terenkripsi. Seperti halnya telegram, twitter menyediakan share *API* yang cukup lengkap untuk membangun aplikasi tambahan [14]. Penelitian ini hendak melakukan analisis sentimen *tweets* mengenai *Bitcoin* untuk melihat pengaruh sentimen terhadap harga *Bitcoin*. Berbeda dengan metode yang digunakan oleh penelitian-penelitian terdahulu untuk melakukan prediksi harga, pada penelitian ini metode yang digunakan adalah

dengan menggunakan *library Python* yaitu *TextBlob* [15] untuk melakukan klasifikasi lalu melakukan visualisasi dengan menggunakan *Word Cloud*.

II. TINJAUAN PUSTAKA

A. Library Python TextBlob

Python TextBlob merupakan *library Python* (versi 2 dan 3) yang digunakan untuk memproses data textual. *TextBlob* menyediakan API yang dapat digunakan untuk pemrosesan bahasa alami (NLP) seperti ekstraksi frase kata benda, analisis sentimen, klasifikasi, penerjemahan, dll. [6]. Hasil objek *TextBlob* digunakan untuk memproses pembelajaran bahasa alami dan *library TextBlob* hanya dapat mengenali Bahasa Inggris.

B. Natural Language Processing

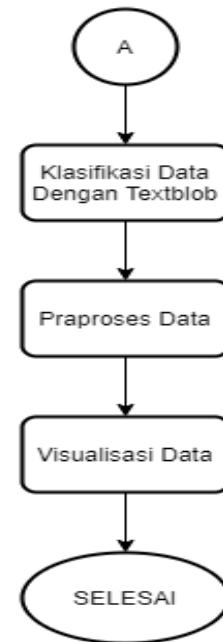
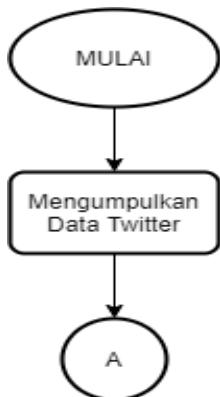
Bahasa alami (*Natural Language*) adalah bahasa yang biasa digunakan oleh manusia untuk berkomunikasi [9]. Istilah *Natural Language Processing* (NLP) biasanya digunakan untuk mendeskripsikan fungsi dari komponen perangkat lunak atau perangkat keras pada sistem komputer yang dapat menganalisis atau mensintesis bahasa alami, baik lisan maupun tulisan (teks) [10]. Tujuan utama dari NLP adalah membuat mesin dapat mengerti dan memahami makna bahasa manusia lalu memberikan respon yang sesuai [11]. NLP merupakan salah satu cabang kecerdasan buatan yang berfokus pada pemrosesan bahasa alami yang digunakan untuk mengidentifikasi teks positif, negatif dan netral [7].

C. Word Cloud

Word Cloud merupakan gambaran visual dari frekuensi kemunculan kata-kata pada suatu kumpulan teks. Ukuran huruf menentukan frekuensi kemunculan sebuah kata, maka semakin besar ukuran huruf maka semakin besar kemunculan kata tersebut. Sebaliknya, semakin kecil ukuran huruf maka semakin kecil frekuensi kemunculan kata tersebut [8].

III. METODE

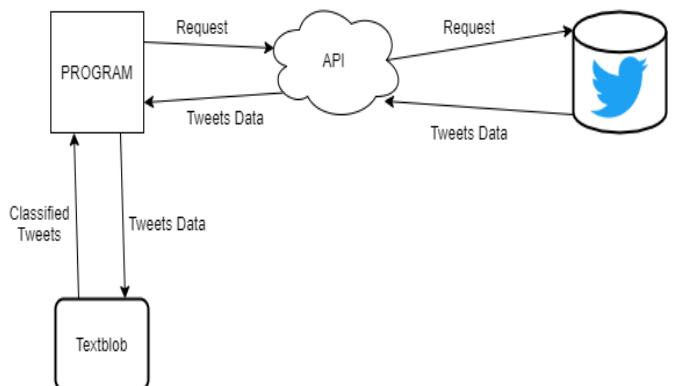
Metode penelitian yang digunakan pada penelitian ini terdiri dari 4 tahapan yang dimulai dari pengumpulan data *tweets* hingga visualisasinya, seperti yang tampak pada diagram alir berikut:



Gbr. 1 Diagram alir metode penelitian yang digunakan pada penelitian ini.

Penelitian ini dimulai dengan mengumpulkan data *tweets* dari *Twitter* dengan menggunakan *API* yang sudah disediakan oleh *Twitter*. Pada tahap ini juga dilakukan klasifikasi sentimen menggunakan *TextBlob* yang sebelumnya dilakukan pra proses secara sederhana sebelum dihitung tingkat polaritas dan subjektivitasnya oleh *TextBlob* dengan menggunakan data yang sudah diambil. Kemudian, sebelum melakukan visualisasi, data *tweets* terlebih dahulu dilakukan pra proses sehingga menjadi terstandar. Lalu, visualisasi data dilakukan dengan menggunakan *Word Cloud*.

A. Pengumpulan Data dan Klasifikasi



Gbr. 2 Skema pengumpulan data.

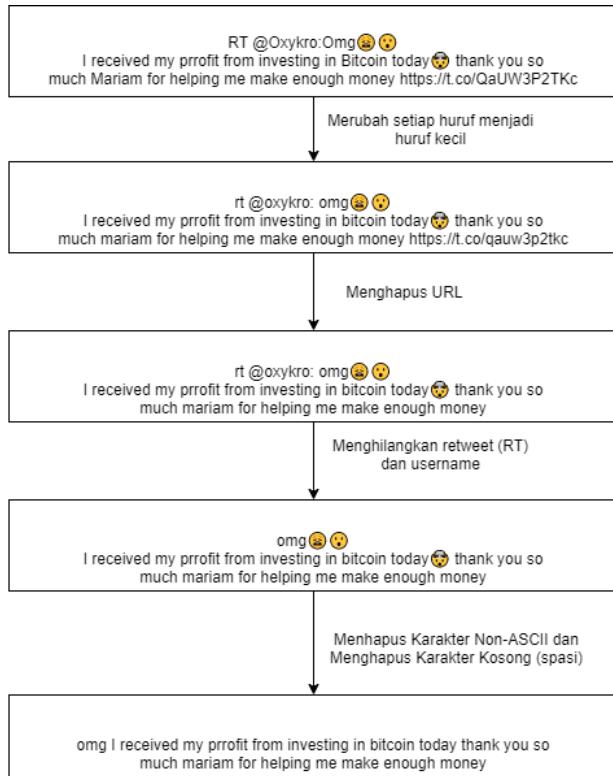
Pada Gbr 2 tampak skema kerja program dalam pengumpulan data. Program akan melakukan *request* ke *API* dengan menggunakan *API key* dan token yang sebelumnya sudah diperoleh. Kemudian *API* akan melakukan *request* ke *database Twitter* dan ketika *database* mendapatkan *request* tersebut maka *database* akan mengirimkan kembalikan berupa data *tweets*. Data *tweets* tersebut berjumlah maksimal 100 dalam sekali pengambilan atau bisa kurang. Data tersebut kemudian sebelum disimpan ke dalam program akan terlebih

dahulu dan dikenai pra proses sederhana dan selanjutnya diklasifikasikan menggunakan *TextBlob* yang akan mengirimkan sebuah kembalian. Kembalian yang didapatkan saat klasifikasi adalah positif yang berarti nilai polaritas antara 0.0 sampai 1.0, kemudian netral apabila nilai polaritasnya adalah 0.0, dan terakhir negatif apabila nilai polaritasnya antara 0.0 sampai -1.0. Nilai polaritas dari tiap kata sendiri sudah didefinisikan pada *TextBlob* sehingga tidak perlu melakukan perhitungan sendiri. Pengumpulan data dilaksanakan pada tanggal 7 maret - 15 maret 2020 dan mendapatkan sebanyak 3343 data. Contoh data yang sudah terklasifikasi adalah sebagaimana pada Tabel 1.

TABEL I
DATA YANG BERHASIL DIKUMPULKAN

Tweets	Sentiment
Omg I received my profit from investing in Bitcoin today thank you so much Mariam for helping me make enough mo... https://t.co/QaUW3P2TKc	Positif
RT @BitcoinMemeHub: "The prosecutor was so dumb that they did not even know that Bitcoin.con wasn't the official website of Bitcoin & they a...	Negatif
RT @GrayscaleInvest: The #EthereumClassic network was designed to expand upon the use cases afforded by #Bitcoin and serve as a decentraliz...	Netral

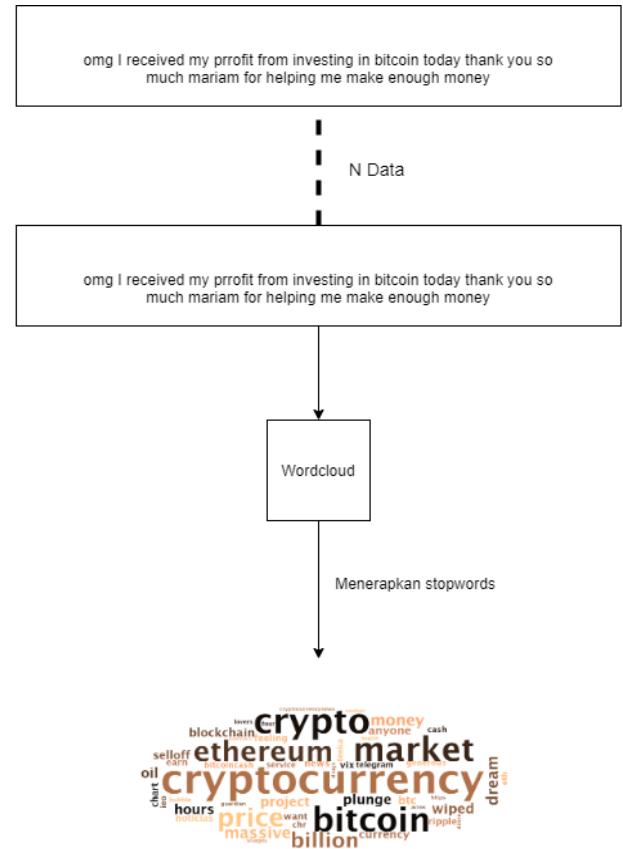
B. Pengumpulan Data dan Klasifikasi



Gbr. 3 Pra proses data.

Pada Gambar 3 tampak langkah-langkah pra proses yang dilakukan pada penelitian ini untuk mewujudkan data yang terstandarisasi. Data yang sudah disimpan dalam program akan dilakukan pra proses yang pertama dengan merubah setiap huruf menjadi huruf kecil. Lalu kemudian dilakukan penghapusan *URL*. Setelah itu, menghilangkan kata *retweet* (RT) yang dalam hal ini sudah dikecilkan menjadi "rt" dan juga *username* dari pengguna yang dikenai *retweet*. Terakhir akan dilakukan penghapusan pada semua karakter Non-ASCII dan karakter kosong pada satu data tersebut sehingga pra proses selesai. Data juga dilakukan pengecekan duplikasi data, sehingga apabila ada data yang sama maka akan dihapus agar menjadi data yg unik. Hasil dari pra proses ini adalah data terstandarisasi yang siap untuk dilakukan visualisasi.

C. Pengumpulan Data dan Klasifikasi



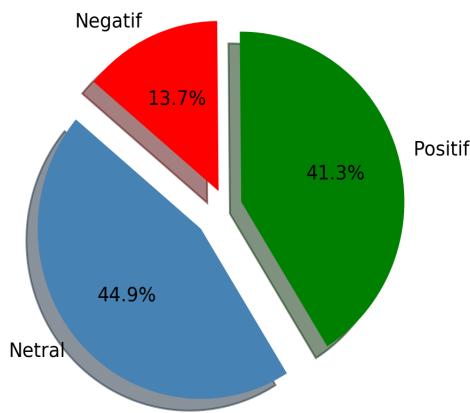
Gbr. 4 Visualisasi data.

Data tweets yang sudah dilakukan praproses kemudian akan dipisahkan berdasarkan masing-masing sentimen yaitu negatif, positif, dan netral. Lalu, setelah dipisahkan maka akan diterapkan *stopwords* atau kata umum yang tidak memiliki makna seperti yang, di, adalah. Sehingga, hanya akan tersisa kata yang memiliki makna seperti kata kerja, kata benda, dan lain sebagainya. Setelah kata bersih dari *stopwords* maka akan dilakukan visualisasi dengan menggunakan *Word Cloud*.

IV. HASIL DAN PEMBAHASAN

Setelah dilakukan klasifikasi data dan praproses maka didapatkan hasil yang akan dilakukan visualisasi menjadi beberapa jenis, yakni pie chart dan word cloud. Berikut selengkapnya:

A. Pie Chart



Gbr. 5 Hasil klasifikasi data tweets bitcoin dan cryptocurrency

Seluruh data tweets yang berjumlah 3433 data telah diklasifikasikan menggunakan library python textblob dan didapatkan hasil seperti pada Gbr 5 dimana 41,3% tweets adalah positif, 44,9% adalah netral, dan 13,7% adalah negatif. Dari hasil tersebut menunjukkan bahwa pengguna twitter masih lebih mengungkapkan sentimen yg positif atau netral daripada sentimen negatif. Hal ini menunjukkan bahwa eksistensi bitcoin diterima oleh banyak kalangan. Walaupun tidak sedikit juga yang memberikan sentimen negatif. Namun, secara keseluruhan sentimen pengguna twitter adalah netral ditunjukkan dengan jumlah paling banyak yaitu sekitar 1541 tweets.

B. Word Cloud

Pada hasil visualisasi dengan menggunakan word cloud maka terdapat 3 word cloud, yaitu word cloud positif, negatif, dan netral. Ketiga word cloud tersebut akan menampilkan kata yang paling sering disebut oleh pengguna twitter yang memudahkan untuk dianalisa daripada harus melihat tweet satu per satu. Word cloud tersebut terdapat pada Gbr 6, Gbr 7, dan Gbr 8.



Gbr. 6 Word cloud positif

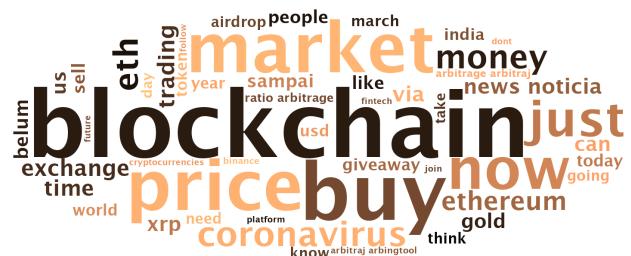
Pada Gbr 6 adalah hasil visualisasi word cloud positif. Dapat dianalisa bagaimana pengguna twitter terdapat 4

kata dengan bentuk paling besar yakni blockchain, free, market, dan now. Hal ini artinya pengguna twitter banyak sekali menyebutkan 4 kata tersebut di postingan mereka. Kemudian terdapat kata money atau uang dan juga buy atau beli. Maka dapat ditarik sebuah pengertian bahwa terdapat aksi beli pada bitcoin dan cryptocurrency. Terdapat juga kata price atau harga yang berhubungan dengan beli, karena pada saat membeli sesuatu tentu terdapat harga yang ditawarkan. Dan uang sendiri dapat berarti bahwa bitcoin adalah sebuah "uang digital" pada masa kini. Lalu, terdapat kata buy atau beli yang juga ada kata now atau sekarang. Kedua kata tersebut bisa saling berhubungan karena pada saat data diambil harga bitcoin sedang jatuh.



Gbr. 7 Word cloud negatif

Pada Gbr 7 adalah hasil visualisasi word cloud negatif. Terdapat 2 kata besar di sana yaitu market atau pasar dan price atau harga. Dapat disimpulkan bahwa sentimen negatif pengguna twitter pada bitcoin berfokus pada harga dan pasar. Tentu saja hal ini akan memberikan sebuah kesimpulan bahwa terjadi sesuatu dengan harga bitcoin dalam rentang waktu pengambilan data. Pada word cloud Gbr 7 juga terdapat kata price probability atau perkiraan harga. Banyak orang yang sedang membicarakan perkiraan harga yang mungkin akan jatuh sehingga digolongkan ke dalam sentimen negatif.



Gbr. 8 Word cloud netral

Dan terakhir pada Gbr 8 adalah hasil visualisasi word cloud netral. Terdapat 6 kata yang bisa dihubungkan menjadi satu kesatuan yang saling terhubung yaitu market, blockchain, price, buy, now, dan coronavirus. Coronavirus juga tidak kalah hangat dalam pembahasan sentimen netral karena pengguna twitter tentu akan membicarakan apakah coronavirus ini mempengaruhi harga sekarang pada bitcoin atau tidak. Bisa ditarik sebuah interpretasi bahwa sentimen pengguna twitter sedang berfokus pada harga pada market yang mengindikasikan untuk melakukan pembelian tanpa adanya sebuah ekspresi positif maupun negatif pada waktu sekarang.

V. KESIMPULAN

Setelah melakukan serangkaian metode penelitian didapatkan hasil bahwa

1. Pengklasifikasian sentimen menggunakan library *python textblob* bisa dilakukan dengan hasil yang cukup baik.
2. Visualisasi menggunakan *word cloud* sangat memudahkan dalam hal analisa untuk mendapatkan sebuah pengetahuan dari topik yang sedang diperbincangkan oleh pengguna twitter berikut sentimennya.
3. Pada penelitian ini, seluruh data tweets yang berjumlah 3433 data telah diklasifikasikan menggunakan library *python textblob* dan didapatkan hasil dimana 41,3% tweets adalah positif, 44,9% adalah netral, dan 13,7% adalah negatif.
4. Pada penelitian ini, pada hasil visualisasi dengan menggunakan word cloud maka bisa ditarik sebuah interpretasi bahwa sentimen pengguna twitter sedang berfokus pada harga pada market yang mengindikasikan untuk melakukan pembelian tanpa adanya sebuah ekspresi positif maupun negatif pada waktu sekarang.

Tentu saja, penelitian ini masih bisa dikembangkan dengan merubah sistem klasifikasi untuk mendapatkan hasil yang lebih baik lagi.

REFERENSI

- [1] I. F. Rozi, S. H. Pramono, and E. A. Dahlan, “Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi,” Electr. Power, Electron. Commun. Control. Informatics Semin., vol. 6, no. 1, pp. 37–43, 2012.
- [2] R. A. Juanda, Jondri, and A. A. Rohmawati, “Prediksi Harga Bitcoin Dengan Menggunakan Recurrent Neural Network,” E-Proceeding Eng., vol. 5, no. 2, pp. 3682–3690, 2018.
- [3] Joshi, Kalyani & N, Bharathi & Rao, Jyothi. (2016). Stock Trend Prediction Using News Sentiment Analysis. International Journal of Computer Science and Information Technology. 8. 67-76. 10.5121/ijcsit.2016.8306.
- [4] Jessica, Jessica (2019) Analisis Sentimen Akun Official News Media Twitter dalam Prediksi Saham Facebook. Bachelor Thesis thesis, Universitas Multimedia Nusantara.
- [5] Y. Xie, “Stock Market Forecasting Based on Text Mining Technology: A Support Vector Machine Method,” J. Comput., vol. 12, no. 6, pp. 500–510, 2017, doi: 10.17706/jcp.12.6.500-510.
- [6] TextBlob, “TextBlob Documentation,” TextBlob, p. 1, 2020.
- [7] Duncan, B., & Zhang, Y. (2015). Neural networks for sentiment analysis on Twitter. Proceedings of 2015 IEEE 14th International Conference on Cognitive Informatics and Cognitive Computing, ICCI*CC 2015. <https://doi.org/10.1109/ICCI-CC.2015.7259397>.
- [8] Bashri, Moh Fajar Ainul. 2017. “Analisis Sentimen Menggunakan Latent Dirichlet Allocation dan Visualisasi Topic Polarity Wordcloud,” Universitas Diponegoro.
- [9] Purnama, I. K. E., & Zaini, A. (2009). Pengembangan Agent Antarmuka Cerdas Berbasis Bahasa Alami untuk Bahasa Indonesia yang Diterapkan Pada Game Edukasi Kecakapan Hidup (Life Skill). ITS Digital Repository. Diakses pada 10 Oktober 2013 dari <http://digilib.its.ac.id/public/ITS-Research-11468-132137894-Conclusion.pdf>
- [10] C. I. Ratnasari, S. Kusumadewi, and L. Rosita, “Model Natural Language Processing untuk Perumusan Keluhan Pasien,” Semin. Nas. Inform. Medis V, pp. 11–18, 2014.
- [11] R. Alamanda, C. Suhery, and Y. Brianorman, “Aplikasi Pendekripsi Plagiat Terhadap Karya Tulis Berbasis Web Menggunakan Natural Language Processing dan Algoritma Knuth-Morris-Pratt,” J. Coding, Sist. Komput. Untan, vol. 04, no. 1, pp. 33–44, 2016.
- [12] Parlika, R., Pratama, A., 2019, “Aplikasi Penampil Berita Realtime Berbasis Bot Telegram menggunakan API Web (APBR Versi 1.0)”, Seminar Nasional “Santika 2019” dengan Tema : “Revolusi Data dalam Pengembangan Industri Digital”, Rabu, 4 September 2019
- [13] Parlika, R., Lugito, M., I., P., Hadiansya, R., P., Vinza , H., S., Faris, H., P., 2019, “Membangun Login Terenkripsi menggunakan Bot Telegram dan Basisdata Mysql”, Seminar Nasional “Santika 2019” dengan Tema : “Revolusi Data dalam Pengembangan Industri Digital”, Rabu, 4 September 2019
- [14] API reference index, <https://developer.twitter.com/en/docs/api-reference-index>, diakses 24 maret 2020 1:13 PM
- [15] TextBlob: Simplified Text Processing, Release v0.15.2. (Changelog) , <https://textblob.readthedocs.io/en/dev/>, diakses 24 maret 2020 1:22 PM