

PENGELASAN SENTIMEN POLITIK DI MEDIA SOSIAL MENGUNAKAN PEMBELAJARAN MESIN

Muhammad Anas Farhan Abdul Razak
Azuraliza Binti Abu Bakar

Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia

ABSTRAK

Isu politik di Malaysia semasa pandemik *CoVid-19* menjadi semakin panas apabila terdapat pelbagai isu yang dikaitkan dengan ahli politik. Ianya terbukti benar kerana orang ramai sering memperkatakan isu-isu ini di dalam media sosial seperti *Twitter*. Saban hari, jumlah *tweets* yang terkumpul semakin banyak sehingga mencecah ratusan ribu. Hal ini telah menyukarkan sesetengah pihak yang ingin mendapatkan gambaran keseluruhan mengenai sesuatu isu yang diperkatakan. Oleh itu, matlamat kajian adalah mengelaskan *tweets* berasaskan politik berdasarkan sentimen seperti positif, neutral dan negatif bagi membantu dan memudahkan pihak berkepentingan atau kerajaan membuat keputusan penting dalam pentadbiran. Kajian ini menggunakan bahasa pengaturcaraan *Python* dan beberapa jenis model pengelasan pembelajaran mesin seperti *Naïve Bayes*, *Support Vector Machine*, *Random Forest* dan *Logistic Regression* berpandukan metodologi *CRISP-DM*. Hasil analisis tersebut memberikan peratusan sentimen keseluruhan bagi setiap topik dan boleh dirujuk apabila sesetengah pihak berkepentingan memerlukan panduan dalam isu-isu yang tersenarai.

1 PENGENALAN

Media sosial merupakan sebahagian daripada kehidupan masyarakat yang berperanan sebagai sebuah platform untuk berkomunikasi antara satu sama lain tanpa batasan. Perkataan media sosial ialah gabungan daripada media dan sosial. Media merupakan alat atau perantara komunikasi dalam perhubungan manakala sosial pula adalah berkaitan dengan persahabatan, pergaulan dan aktiviti masa lapang (Kamus Dewan dan Pustaka Edisi Keempat 2014). Media sosial meliputi pelbagai aplikasi seperti *Twitter*, *Facebook*, *Instagram* dan banyak lagi. Pada era yang semakin berkembang teknologinya, kebanyakan pengguna media sosial menggunakan media sosial sebagai platform untuk meluahkan perasaan dan juga pendapat. Sejarar dengan

keadaan semasa, pengguna media sosial sering meluahkan pendapat dan perasaan mereka terhadap isu politik yang sedang goyah di saat pandemik *CoVid-19* melanda negara ini. Hal ini telah menyebabkan pertambahan sentimen yang ada di dalam media sosial.

Menurut laporan harian, perebutan kuasa yang berlaku ketika Malaysia diserang oleh wabak pandemik *Covid-19* terus menular dan tiada penghujungnya. Ahli-ahli politik di Malaysia di mana mereka seperti tidak takut mati ketika seluruh dunia sedang berperang dengan wabak koronavirus ini. Tidak kiralah pemimpin kerajaan atau pembangkang, mereka lebih mementingkan wang, kuasa dan kedudukan (Sinar Harian 2020). Hal ini menyebabkan orang ramai sering melemparkan pendapat serta perasaan mereka di media sosial berkenaan isu politik yang melanda negara ini. Isu politik sering dibincangkan oleh pengguna media sosial dan kadang-kadang menjadi trending di sesetengah platform media sosial. Hal ini sekali gus telah meningkatkan jumlah sentimen yang berasakan politik dan menyebabkan orang ramai menjadi keliru dengan status perkembangan mengenai politik ini.

Secara kesimpulannya, kajian ini dicadangkan kerana terdapat cabaran yang dihadapi oleh pengkaji untuk menganalisis sentimen politik dalam era pandemik *CoVid-19* ini. Pendekatan secara analisis sentimen dicadangkan untuk menganalisis *tweets* di *Twitter* bagi mengkaji pandangan awam terhadap isu-isu politik yang berlaku semenjak wabak pandemik melanda negara ini. Sentimen-sentimen ini dianalisis dan dikelaskan kepada beberapa kumpulan sentimen yang berbeza iaitu positif, neutral dan negatif. Hasil akhir kajian ini boleh membantu pihak yang berkepentingan atau kerajaan dalam membuat keputusan penting dalam urusan pentadbiran.

2 PERNYATAAN MASALAH

Melalui pemerhatian yang dijalankan, terdapat ribuan sentimen *tweets* yang berasakan politik di dalam media sosial *Twitter*. Semenjak pergelutan kuasa antara parti-parti politik di Malaysia berlaku, isu politik makin kerap menjadi perbincangan orang ramai. Hal ini telah menyebabkan peningkatan jumlah *tweets* sentimen yang sangat banyak. Apabila terdapat begitu banyak pendapat yang dilontarkan oleh orang ramai, sentimen positif, neutral dan juga negatif telah bercampur-aduk. Ini telah mendatangkan cabaran kepada seseorang pengkaji yang ingin menganalisis sentimen politik pada era pandemik *CoVid-19* ini. Perkara ini juga telah

menyukarkan semua pihak seperti orang ramai, pihak berkepentingan dan kerajaan dari mendapat gambaran keseluruhan keputusan bagi setiap isu politik yang dibincangkan.

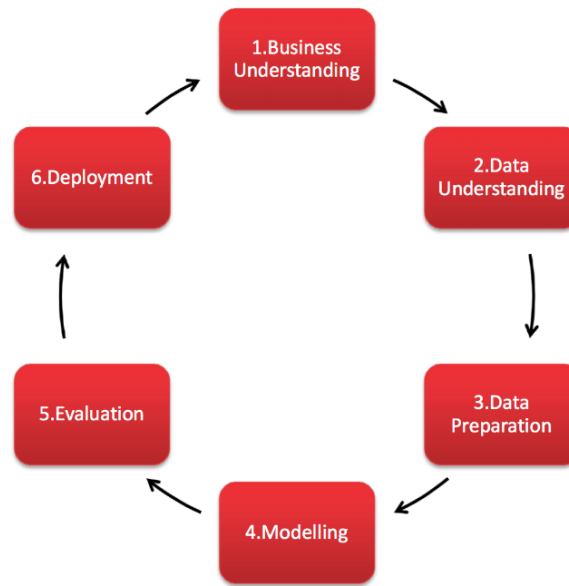
3 OBJEKTIF KAJIAN

Objektif utama dalam projek Pengelasan Sentimen Politik di Media Sosial Menggunakan Pembelajaran Mesin ini ialah:

- i. Mereka bentuk dan membangunkan sebuah model yang dapat mengelaskan sentimen politik yang terdapat di dalam media sosial menggunakan pendekatan algoritma pembelajaran mesin.
- ii. Membangunkan sistem visualisasi pengelasan sentimen politik secara atas talian.

4 METOD KAJIAN

Metodologi menunjukkan proses penting yang berlaku dalam memastikan sesuatu projek yang dijalankan berjalan dengan lancar mengikut fasa yang telah ditetapkan. Untuk projek Pengelasan Sentimen Politik di Media Sosial Menggunakan Pembelajaran Mesin ini, metodologi yang digunakan sepanjang tempoh pelaksanaan ialah *CRISP-DM*. Nama *CRISP-DM* adalah singkatan bagi proses merentas industri untuk perlombongan data (*Cross-industry process for data mining*). Terdapat 6 fasa di dalam *CRISP-DM* ini iaitu Pemahaman Perniagaan (*Business Understanding*), Pengetahuan Data (*Data Understanding*), Penyediaan Data (*Data Preparation*), Pemodelan (*Modelling*), Penilaian (*Evaluation*) dan Penggunaan (*Deployment*) seperti yang ditunjukkan pada Rajah 1.

Rajah 1 Fasa *CRISP-DM*

4.1 Fasa Pemahaman Perniagaan

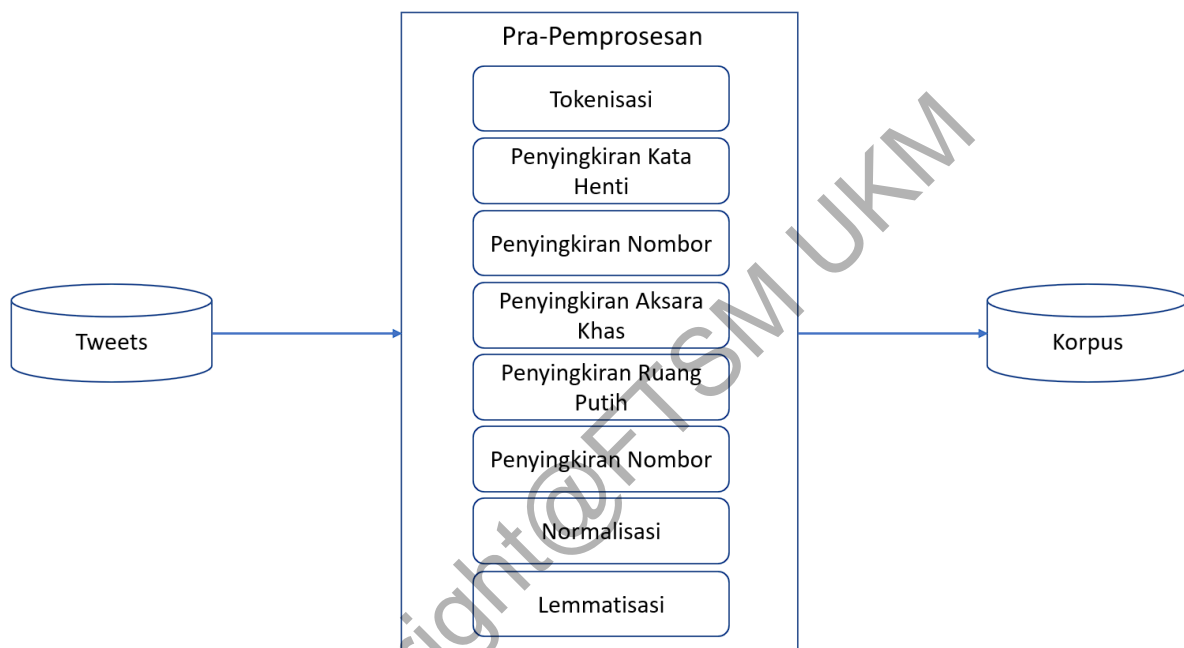
Fasa pemahaman perniagaan adalah fasa di mana objektif utama kajian dikenalpasti. Dalam projek ini, objektif utamanya adalah untuk membangunkan sebuah modal pengelasan sentimen dan juga sebuah sistem visualisasi pengelasan. Projek ini akan menggunakan *Python* sebagai platform utama dan akan menggunakan beberapa algoritma pembelajaran mesin yang terdiri daripada *Naïve Bayes*, *SVM*, *Random Forest* dan *Logistic Regression*. Dalam fasa ini juga, kriteria yang akan digunakan untuk menentukan sama ada projek ini berjaya dari sudut perniagaan perlu dikenalpasti. Projek ini dianggap berjaya sekiranya data-data ini berjaya dikelaskan mengikut sentiment masing-masing iaitu positif, neutral atau negatif.

4.2 Fasa Pengetahuan Data

Fasa pengetahuan data adalah fasa di mana data-data yang telah dirangkak keluar dari media sosial disenaraikan dari mana sumber datangnya dan juga kaedah yang digunakan untuk mendapatkan data-data ini. Dalam projek ini, data sentimen yang dirangkak keluar dari *Twitter* menggunakan perisian *Twitter API* direkodkan untuk rujukan masa hadapan. Fasa ini juga memerlukan data untuk diterangkan dari segi format, kuantiti dan ciri permukaan lain yang ditemui. Selain itu, data perlu diperiksa untuk memastikan semua data adalah lengkap, tiada nilai yang kosong dan sesuai untuk digunakan.

4.3 Fasa Penyediaan Data

Semasa fasa penyediaan data, data yang dikumpul melalui pra-pemprosesan, seperti pembersihan data, transformasi data dan pengurangan data. Proses pra-pemprosesan ini penting kerana ianya proses yang mengubah data mentah kepada format yang boleh difahami oleh mesin dan membantu mengubah bunyi dari ciri dimensi tinggi ke ruang dimensi rendah untuk mendapatkan seberapa banyak maklumat yang tepat dari teks. Teknik ini dapat membantu meningkatkan ketepatan model pengelas.



Rajah 2 Proses Fasa Pra-Pemprosesan

4.4 Fasa Pemodelan

Dalam fasa pemodelan ini, sebuah model Pengelasan Sentimen Politik di Media Sosial Menggunakan Pembelajaran Mesin akan dibangunkan. Dengan menggunakan *Python* sebagai platform utama serta disokong dengan algoritma pembelajaran mesin, model ini dapat mengelaskan setiap sentimen yang telah diekstrak mengikut sentimen masing-masing iaitu positif atau negatif. Sebelum model ini diberikan set ujian, ianya terlebih dahulu dilatih dengan menggunakan set latihan. Hal ini kerana model pengelas perlu dilatih bagi mendapatkan ketepatan yang lebih tinggi.

4.5 Fasa Penilaian

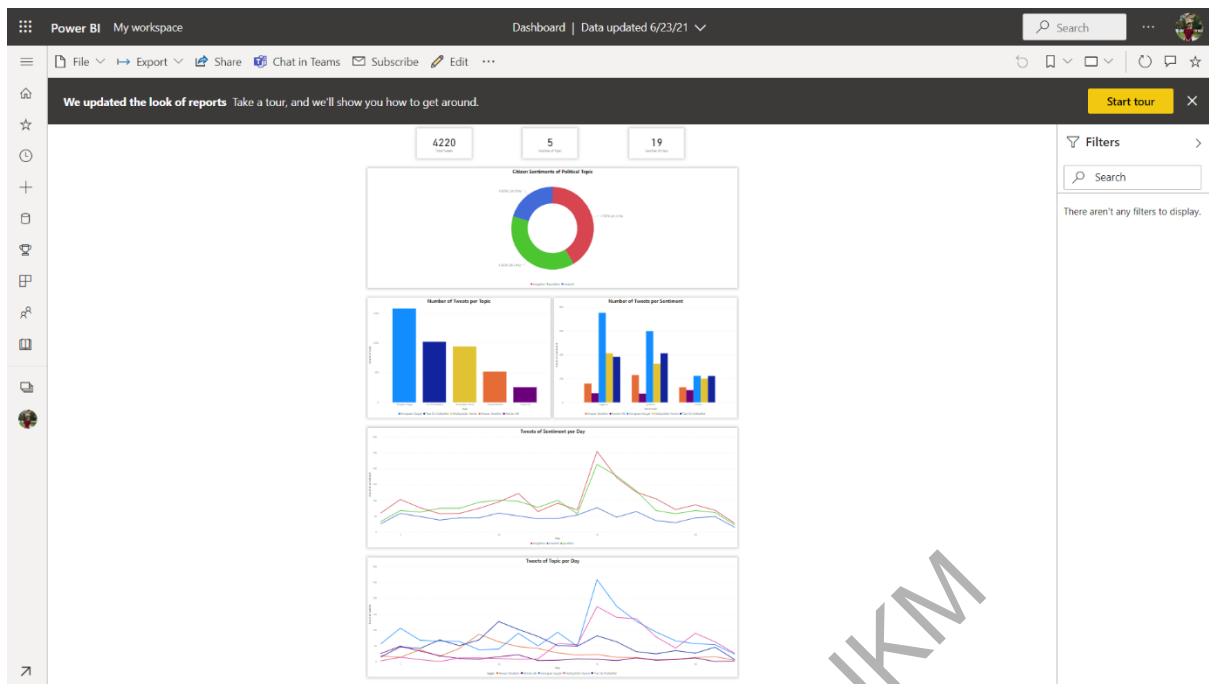
Semasa fasa penilaian, keputusan yang diperolehi dari model yang dibangunkan akan direkodkan dan disimpan. Keputusan ini akan dianalisis untuk menentukan sama ada projek ini telah memenuhi objektif perniagaan awal iaitu berjaya mengelaskan sentimen berasaskan politik di media sosial. Seterusnya, ulasan proses dibuat untuk membuat ringkasan tinjauan keseluruhan proses dan mencatat sebarang proses yang harus diulangi semula. Setelah itu, sebarang kemungkinan tindakan perlu disenaraikan dan keputusan akan dibuat sama ada keputusan projek ini sudah memadai atau tidak.

4.6 Fasa Penggunaan

Fasa penggunaan adalah fasa terakhir di dalam metodologi *CRISP-DM*. Di dalam fasa ini, model yang dibangunkan dan berjaya memenuhi objektif akan digunakan. Selain itu, laporan akhir yang mengandungi segala penglibatan data yang merangkumi semua hasil sebelumnya, meringkas dan mengatur hasilnya perlu direkodkan. Bagi pembentangan akhir, keputusan yang diperolehi dari fasa penilaian akan dibentangkan dalam bentuk visual bagi memudahkan orang ramai faham dan melihatnya. Akhir sekali, pengalaman dokumentasi akan dihasilkan bagi merekod segala pengalaman penting yang diperolehi semasa projek.

5 HASIL KAJIAN

Bahagian ini membincangkan hasil daripada proses pembangunan model Pengelasan Sentimen Politik Di Media Sosial Menggunakan Pembelajaran Mesin. Terdapat beberapa perisian yang digunakan untuk membangunkan sistem hasil keputusan analisis yang telah dilakukan. Dalam kajian ini, Microsoft Power BI digunakan bagi menghasilkan visual untuk papan pemuka. Power BI merupakan satu aplikasi visualisasi dan analisis data interaktif yang membolehkan pengguna menjana papan pemuka dengan mudah hanya dengan memasukkan data. Selepas data dimasukkan, aplikasi ini akan memberi beberapa pilihan visualisasi untuk dipilih oleh pengguna. Papan pemuka dihasilkan bagi memudahkan penganalisis memahami keputusan yang dihasilkan. Rajah dibawah menunjukkan visual bagi papan pemuka untuk kajian ini.



Rajah 3 Visual Yang Dihasilkan Menggunakan Power BI

Selain itu, Tableau Public juga digunakan untuk menghasilkan papan pemuka untuk memaparkan keputusan data. Tableau adalah mesin analisis visual yang menjadikannya lebih mudah untuk membuat analisis visual interaktif dalam bentuk papan pemuka. Papan pemuka ini memudahkan penganalisis bukan teknikal dan pengguna akhir menukar data menjadi grafik interaktif yang dapat difahami. Sebelum itu, data yang telah diproses perlu dimuat naik ke Google Spreadsheet terlebih dahulu sebelum menukarnya menjadi visual. Dengan menggunakan bantuan Google API, data berjaya dimuat naik ke spreadsheet dengan mudah.

```
scope = ['https://spreadsheets.google.com/feeds', 'https://www.googleapis.com/auth/drive']
credentials = ServiceAccountCredentials.from_json_keyfile_name('myfile.json', scopes=scope)
gc = gspread.authorize(credentials)
spreadsheet_key = '10TRJgPR5w0p98cMrawkiCw37SKXbdR4q85ePSTU1HgA'
```

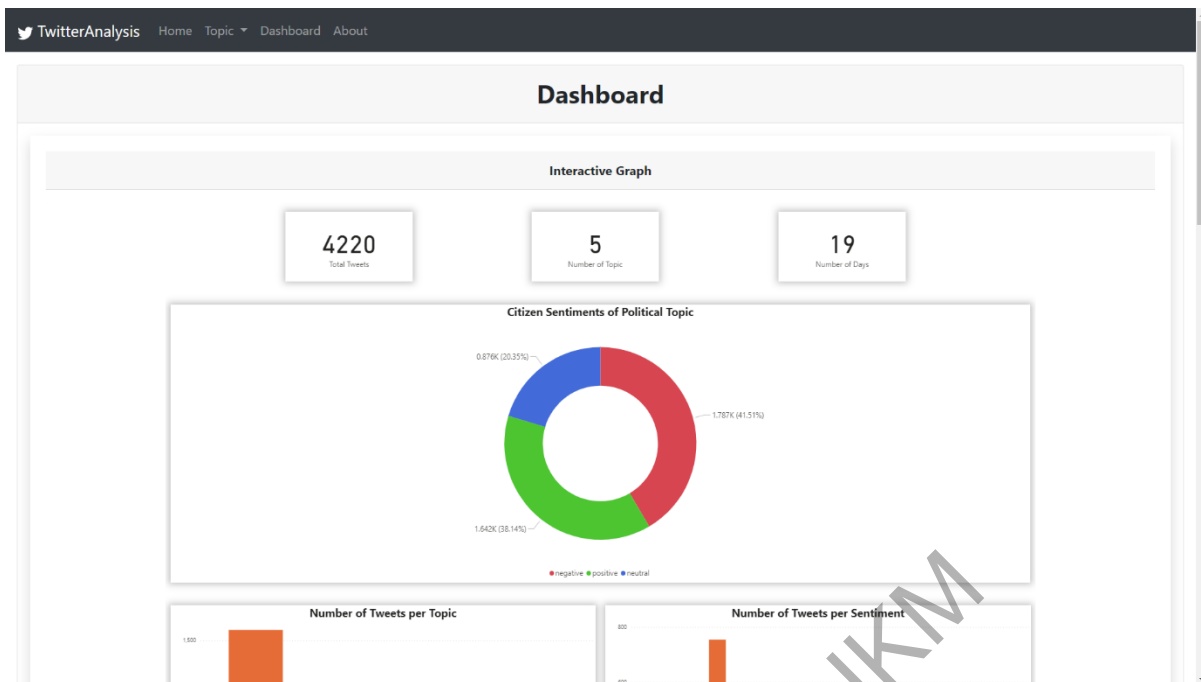
Rajah 4 Penyambung Antara Pythin Jupyter Dan Google Spreadsheet

Setelah data dimuat naik, kini ianya sudah bersedia untuk divisualkan. Dengan menggunakan aplikasi Tableau Public, hasil visual daripada analisis data dapat dipaparkan pada papan pemuka ini. Papan pemuka yang telah siap dipaparkan pada profil pengkaji di Tableau Public seperti gambar rajah di bawah.



Rajah 5 Visual Yang Dihasilkan Menggunakan Tableau Public

Bagi membangunkan sebuah laman web, Visual Studio Code digunakan bagi memudahkan pembangunan sistem untuk tujuan memaparkan hasil kajian yang diperolehi secara atas talian. Visual Studio Code adalah penyunting kod yang diselaraskan dengan sokongan untuk operasi pengembangan seperti penyahpejatan, menjalankan tugas, dan kontrol versi. Ini bertujuan untuk menyediakan hanya alat yang diperlukan oleh pembangun untuk kitaran kod-debug kod cepat dan meninggalkan aliran kerja yang lebih kompleks kepada IDE yang lebih lengkap, seperti Visual Studio IDE. Berikut adalah rajah antara muka bagi papan pemuka yang memaparkan hasil keputusan yang diperolehi dalam kajian ini.



Rajah 6 Hasil Antara Muka Bagi Papan Pemuka

6 KESIMPULAN

Sistem Pengelasan Sentimen Politik di Media Sosial Menggunakan Pembelajaran Mesin ini dijangka dapat membantu dan memudahkan pihak berkepentingan atau kerajaan membuat keputusan penting dalam pentadbiran. Sistem ini juga memainkan peranan penting di Malaysia kerana terdapat peningkatan jumlah sentimen tweets berasaskan politik di media sosial berikutan orang ramai yang kini mudah untuk mendapatkan akses ke media sosial dan juga permasalahan kerajaan yang gagal dalam menangani peningkatan dan penularan kes CoVid-19.

Hasil daripada analisis sentimen yang dilakukan telah divisualisasikan dalam bentuk graf bagi memudahkan orang ramai dan juga pihak berkepentingan untuk memahami hasil yang diperolehi. Penggunaan perisian Power BI dan Tableau Public dalam projek ini telah membantu memudahkan kerja visualisasi hasil yang diperolehi. Fungsi interaktif yang ditawarkan oleh Power BI telah menjadikan hasil visualisasi semakin cantik dan menarik. Walaupun terdapat kekangan yang dihadapi semasa melakukan kajian ini, ianya telah diselesaikan dengan menggunakan teknik yang sesuai.

7 RUJUKAN

Anonymous. Evolusi Media Sosial.

http://evolusimediasosial.blogspot.com/2015/12/sejarah-dan-evolusi-media-sosial_16.html [17 Disember 2015]

Azhar, Muhammad. Hafidz, Noor. Rudianto, Biktra. Gata, Windu. 2014. *Marketplace Sentiment Analysis Using Naïve Bayes and Support Vector Machine*. Universiti Islam 45 Bekasi. [14 September 2020]

Bannister, Kristian. *Understanding Sentiment Analysis: What It Is & Why It's Used*.

Brandwatch. <https://www.brandwatch.com/blog/understanding-sentiment-analysis/> [26 Februari 2018]

Brownlee, Jason. *Naïve Bayes for Machine Learning. Machine Learning Mastery*.

<https://machinelearningmastery.com/naive-bayes-for-machine-learning/> [11 April 2016]

Davydova, Olga. Data Monsters. *Text Pre-processing in Python: Steps, Tools, and*

Examples. *Medium*. <https://medium.com/@datamonsters/text-preprocessing-in-python-steps-tools-and-examples-bf025f872908> [16 Oktober 2018]

Elgabry, Omar. (Mar 2019). *The Ultimate Guide to Data Cleaning. Towards Data*

Science. <https://towardsdatascience.com/the-ultimate-guide-to-data-cleaning-3969843991d4> [1 Mac 2019]

Gupta, Rishabh. (May 2017). *Advantage of CSV file. GitHub*.

<https://github.com/prakhargvp/csvWithCplusplus/wiki/Advantages-of-CSV-file> [13 Mei 2017]

Lum Choi Kian. 2019. Analisis Sentimen Dalam Bitcoin Tweets. Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia.

<http://www.ftsm.ukm.my/cybersecurity/file/research/technicalreport/PTA-FTSM-2019-013.pdf> [3 Januari 2019]

Malik, Usman. *Implementing SVM and Kernel SVM with Python's Scikit-Learn. Stack*

Abuse. <https://stackabuse.com/implementing-svm-and-kernel-svm-with-pythons-scikit-learn> [10 Mei 2019]

- Merak Jalanan. 2020. Rebut kuasa ketika pandemic Covid-19 terus menular. Berita Harian. <https://www.sinarharian.com.my/article/106035/SUARA-SINAR/Pojok/Rebut-kuasa-ketika-pandemik-Covid-19-terus-menular> [20 Oktober 2020]
- Nations, Daniel. 2018. *What Is Social Media?. Lifewire*. <https://www.lifewire.com/what-is-social-media-explaining-the-big-trend-3486616> [26 Januari 2021]
- Pandey, Pranjal. *Data Pre-processing: Concept. Towards Data Science*. <https://towardsdatascience.com/data-preprocessing-concepts-fa946d11c825> [25 November 2019]
- Pandey, Parul. *Data Visualisation with Tableau. Datacamp*. <https://www.datacamp.com/community/tutorials/data-visualisation-tableau> [25 September 2018]
- Qian, Vicky. *Step-By-Step Twitter Sentiment Analysis: Visualizing Multiple Airlines' PR Crises [Updated For 2020]. iPullRank*. <https://ipullrank.com/step-step-twitter-sentiment-analysis-visualizing-united-airlines-pr-crisis> [26 April 2017]
- Rodrigues, Israel. *CRISP-DM methodology leader in data mining and big data. Towards Data Science*. <https://towardsdatascience.com/crisp-dm-methodology-leader-in-data-mining-and-big-data-467efd3d3781> [17 Februari 2020]
- Shaikh, Raheel. *Feature Selection Techniques in Machine Learning with Python. Towards Data Science*. <https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e> [28 Oktober 2018]
- Smith, Robert. *The Key Differences Between Rule-Based AI and Machine Learning. Medium*. <https://becominghuman.ai/the-key-differences-between-rule-based-ai-and-machine-learning-8792e545e6> [14 Julai 2020]
- Sunilkumar, Devipriya. *From Analysis to Dashboarding-Connecting notebooks to Tableau. Towards Data Science*. <https://towardsdatascience.com/from-analysis-to-dashboarding-connecting-notebooks-to-tableau-483fa373f3a4> [30 April 2020]
- White, Bryan. *Sentiment Analysis: VADER or TextBlob?. Towards Data Science*. <https://towardsdatascience.com/sentiment-analysis-vader-or-textblob-ff25514ac540> [27 Mei 2020]