

ANALISIS SENTIMEN TERHADAP INISIATIF “ONE-BELT ONE-ROAD” CHINA DI TWITTER

CHOONG HENG HUI

PROF. DR. SALWANI ABDULLAH

Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia

ABSTRAK

Inisiatif “*One-Belt One-Road*” (OBOR) yang juga dikenali sebagai “*Belt and Road Initiative*” (BRI) merupakan antara topik hangat yang berterusan di Malaysia. Isu ini banyak dibincangkan dan dipertikai disebabkan ianya merupakan salah satu projek yang boleh memberi kesan utama terhadap pembangunan Malaysia. Oleh yang demikian, kertas ini membentangkan kajian yang menjalankan analisis sentimen terhadap inisiatif “*One-Belt One-Road*” China di Twitter. Proses analisis berasaskan kaedah leksikon ini mengandungi lima langkah iaitu pengumpulan data, pra-pemprosesan data, analisis leksikon, analisis statistik dan pembentangan hasil analisis. *Tweets* berkaitan OBOR telah dikumpulkan sebagai data untuk analisis. Langkah-langkah pra-pemprosesan dilakukan terhadap data mentah yang dikumpulkan untuk meningkatkan ketepatan analisis. Analisis leksikon dijalankan terhadap data yang telah dibersihkan menggunakan pustaka *Python*, iaitu VADER dan TextBlob. Dalam langkah analisis statistik, jumlah tag sentimen *Tweets* dan kekerapan perkataan yang seringkali diguna telah dikira. Hasil analisis telah dibentangkan dalam bentuk rajah *wordcloud*, carta pai dan graf siri masa. Sejumlah 1860 *Tweets* telah dikumpulkan manakala hanya 874 data *Tweets* yang tertinggal selepas langkah pra-pemprosesan digunakan dalam analisis. Keputusan analisis sentimen menunjukkan bahawa majoriti *Tweets* (49.89%) berkaitan inisiatif “*One-Belt One-Road*” adalah bersifat positif. Sentimen neutral dan negatif pula masing-masing merangkumi 36.96% dan 13.18%. Frasa hangat yang diperolehi termasuk “BRI”, “Forest City. Johor”, “China”, “ECRL”, “Malaysia” dan “project”. Penambahbaikan yang boleh dilakukan terhadap projek ini termasuk pemilihan frasa penuh sahaja sebagai kata kunci untuk pengumpulan data dan pembinaan langkah pra-pemprosesan yang lebih lengkap.

1 PENGENALAN

Inisiatif “*One-Belt One-Road*” (OBOR) yang juga dikenali sebagai “*Belt Road Initiative*” (BRI) merupakan satu topik hangat yang dibicarakan oleh rakyat Malaysia berikutan perjanjian Malaysia dengan China pada tahun 2017. Dua projek yang terkandung ialah “*East Coast Rail Link*” (ECRL) dan “*Malaysia-China Kuantan Industrial Park*” (MCKIP). Walaupun projek ECRL pernah dibatalkan oleh Tun Dr. Mahathir selepas beliau dipilih menjadi Perdana Menteri Malaysia ketujuh, projek ini telah berjaya diteruskan selepas perbincangan Malaysia dengan China untuk mengurangkan kos pembayaran.

Walaupun negara China berjanji untuk membantu meningkatkan tahap ekonomi negara yang terlibat dalam inisiatif ini (Ferdinand, 2016), terdapat juga komen-komen dan analisis negatif tentangnya. Kebimbangan yang merisaukan pelbagai pihak telah disuarakan: 1) kos dan risiko strategik dan kewangan yang timbul disebabkan pelaburan BRI dan China terhadap infrastruktur sensitif; 2) kebangkitan China dari segi ekonomi dan kemungkinan penyelewengan pengaruh ekonominya terhadap negara serantau demi menyesuaikan diri dengan agenda China (Smith, 2018).

Oleh yang demikian, terdapat sokongan idea yang berbeza di laman sosial media terutamanya Facebook dan Twitter dalam kalangan rakyat Malaysia. Pendapat dan hujah pihak yang berlainan telah pun mengelirukan rakyat biasa yang bukan pakar ekonomi. Pahlawan papan kekunci, atau biasanya dikenali sebagai “*keyboard warrior*” pula merumitkan keadaan dengan memperkenalkan unsur-unsur tidak berkaitan dalam komen atau status sosial media mereka seperti muslihat, pakatan sulit, malah perkauman yang merupakan isu sensitif di Malaysia. Tindakan tidak bermanfaat ini seterusnya memudaratkan cubaan kerajaan untuk mengambil tindakan yang sepatutnya.

Di sini, analisis sentimen terhadap komen-komen di laman sosial media memainkan peranan penting dengan mengenal pasti pandangan pengguna (Ramanathan, et al. 2019). Pengenalan jenis komen tersebut boleh memudahkan penganalisis dan penyelidik untuk menentukan kesahihan dan mengasingkan pandangan sebenar oleh rakyat daripada kekacauan “*keyboard warrior*”.

2 PENYATAAN MASALAH

Algoritma untuk menganalisis sentimen terhadap inisiatif “*One-Belt One-Road*” di Twitter tidak wujud. Justeru, usaha untuk mengkaji dan memahami pandangan pengguna Twitter terhadap inisiatif ini tidak dapat dijalankan. Topik berkenaan inisiatif ini yang kerap disebut dalam *Tweets* juga tidak dapat dikenal pasti.

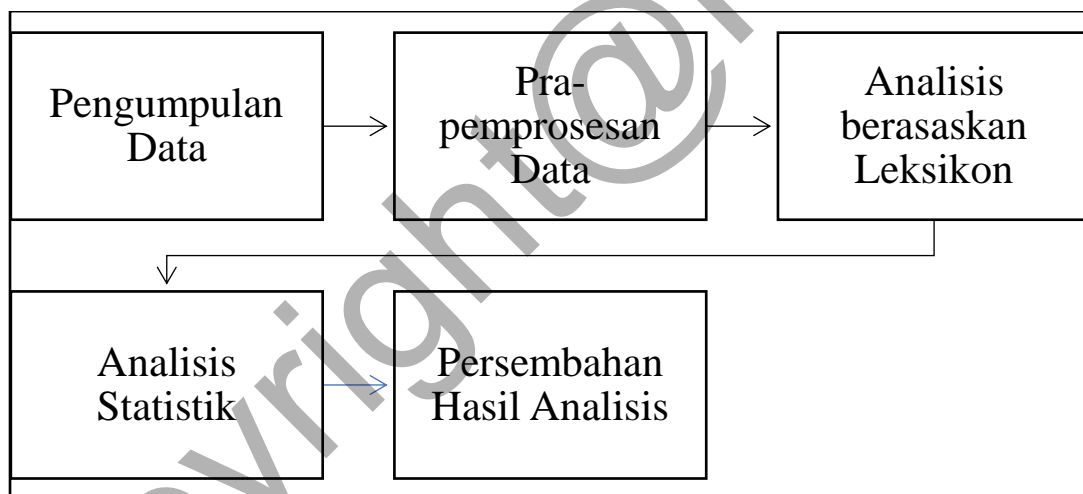
3 OBJEKTIF KAJIAN

1. Untuk menghasilkan algoritma yang mampu menganalisis sentimen terhadap komen di Twitter.
2. Untuk mengkaji pandangan majoriti terhadap inisiatif “*One-Belt One-Road*” melalui komen di Twitter.
3. Untuk mengenal pasti frasa hangat berkaitan dengan inisiatif “*One-Belt One-Road*” di Twitter.

4 METOD KAJIAN

Kajian ini bersifat kualitatif dan melibatkan pengumpulan data dari laman sosial media Twitter. Bahasa pengaturcaraan yang akan digunakan dalam kajian ini merupakan *Python* sahaja. Shahnawaz dan Parmanand mencadangkan 5 langkah utama untuk analisis sentimen dalam kajian mereka pada tahun 2017. Rangka tersebut telah diubah suai dan diguna pakai dalam kajian ini seperti yang ditunjukkan dalam rajah 1.

Data teks dalam *Tweets* akan dikumpulkan terlebih dahulu menggunakan kod yang sedia ada (Henrique, 2018). *Tweets* yang dikumpulkan akan dijalankan pra-pemrosesan dahulu untuk tujuan pembersihan data. Kemudian, analisis sentimen akan dijalankan terhadap data yang telah dibersihkan melalui kaedah leksikon menjadi positif, negatif atau neutral. Pengiraan frasa yang paling kerap disebut dalam *Tweets* juga akan dilakukan bersama pengiraan hasil pengelasan bagi analisis leksikon secara umum dalam langkah analisis statistik. Pembentangan hasil analisis pada tahap terakhir analisis akan mempersembahkan data berkaitan dalam bentuk grafik.



Rajah 1 Carta aliran proses analisis

4.1 Fasa Perancangan

Fasa perancangan kajian ini dimulakan dengan pengesahan masalah, objektif dan skop kajian. Selepas itu, kajian literasi dijalankan untuk mendapatkan ilham bagi mencapai objektif yang ditetapkan. Kaedah kajian disahkan selepas pemahaman topik dan jalan penyelesaian melalui kajian literasi. Pseudokod algoritma dihasilkan dahulu untuk memastikan kebolehlaksanaan algoritma dan penepatan objektif kajian dari segi teori. Data *Tweets* yang dikumpulkan dan diguna dalam kajian juga difahami terlebih dahulu. Jadual 1 menunjukkan butiran data *Tweets* yang berkaitan. Perisian *Jupyter Notebook* terpilih sebagai persekitaran pembangunan bersepadu untuk pembangunan algoritma dalam bahasa pengaturcaraan *Python*.

Jadual 1 Nama, definisi dan jenis objek data yang dikumpulkan

Nama Data	Definisi Data	Jenis Objek (<i>datatype</i>)
<i>Date</i>	Tarikh dan masa <i>Tweet</i> diterbitkan	Rentetan (<i>str</i>)
<i>Username</i>	Nama pengguna di <i>Twitter</i> yang menerbitkan <i>Tweet</i>	Rentetan (<i>str</i>)
<i>Retweets</i>	Jumlah penerbitan semula <i>Tweet</i> oleh pengguna lain	Integer (<i>int</i>)
<i>Favourites</i>	Jumlah “ <i>Like</i> ” terhadap <i>Tweet</i> oleh pengguna lain, sebagai tanda penghargaan atau persetujuan terhadap <i>Tweet</i>	Integer (<i>int</i>)
<i>Text</i>	Teks <i>Tweet</i> yang diterbitkan, boleh mengandungi <ol style="list-style-type: none"> 1. pautan laman sesawang dan pautan gambar <i>Twitter</i> serta 2. <i>emoji</i> dan emotikon (simbol atau ikon yang mewakili pemaparan perasaan) 	Rentetan (<i>str</i>)
<i>Mentions</i>	@ <i>Username</i> , panggilan untuk pengguna <i>Twitter</i> yang lain dalam <i>Tweet</i>	Rentetan (<i>str</i>)
<i>Hashtags</i>	Perkataan atau frasa yang mengikuti simbol “#” dalam <i>Tweet</i> , boleh digunakan untuk mencari <i>Tweet</i> lain yang mengandungi kata kunci atau topik yang sama	Rentetan (<i>str</i>)

Sumber definisi: *Twitter* 2019

4.2 Fasa Analisis

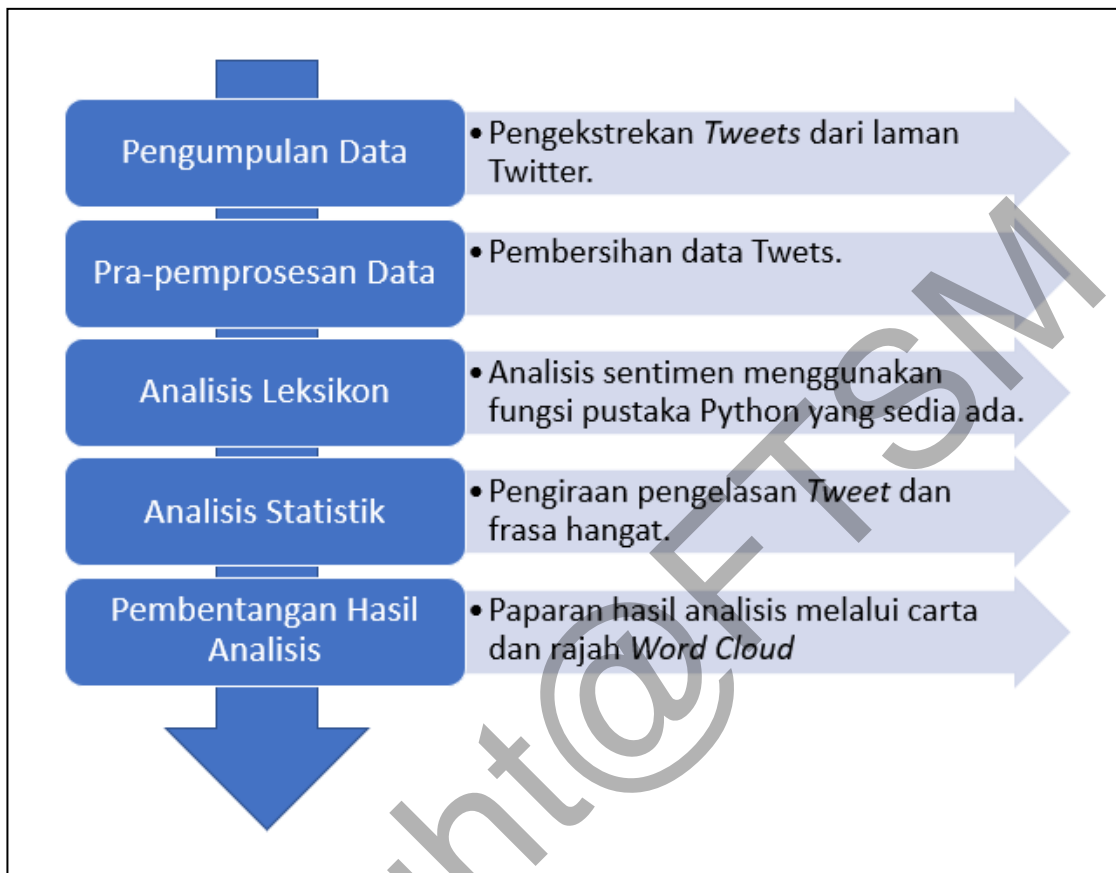
Analisis leksikon kajian ini mengguna pakai pakaj “*vaderSentiment*” dan “*textblob*”. “*vaderSentiment*”, dikenali sebagai *VADER* merupakan peralatan analisis berasaskan leksikon dan peraturan yang direka khas untuk mengkaji sentimen yang dinyatakan di laman media sosial. Walaupun tiada maklumat terperinci tentang pemarkahan sentimen oleh laman sesawang rasmi *TextBlob*, rujukan blog (Schumacher, 2015) mencadangkan bahawa *TextBlob* mempunyai sifat yang merupai *vaderSentiment*, iaitu berasaskan leksikon dan peraturan.

Purata daripada markah “*compound*” daripada *VADER* dan markah “*polarity*” daripada *TextBlob* digunakan sebagai nilai pengelasan tag sentimen. Pengelasan sentimen dibuat berdasarkan nilai ambang yang dicadangkan oleh pencipta *VADER*, iaitu 0.05 dan -0.05 untuk menentukan *Tweet* positif, neutral dan negatif.

4.3 Fasa Reka Bentuk

Algoritma analisis sentimen ini terdiri daripada gabungan beberapa lapisan kod mengikut susunan seperti dalam rajah 2. Fungsi setiap langkah analisis sentimen tersedia di tepi nama lapisan kod dalam rajah 2. Lima fail penting dalam algoritma yang dihasilkan daripada reka bentuk ini dinamakan “*scrapeTweets.py*”, “*preprocessTweets.py*”, “*analyseTweets.py*”, “*analyseTweetsStatistics.py*” dan “*analysisOutput.py*” mengikut setiap langkah dalam rajah 2. Fail “*main.ipynb*”

dihasilkan pula untuk memanggil algoritma penuh atau setiap langkah secara berasingan.



Rajah 2 Butiran carta aliran proses analisis sentimen.

4.4 Fasa Pengujian

4.4.1 Pengumpulan Data

Pengujian algoritma analisis sentimen dijalankan mengikut reka bentuk algoritma. Untuk memudahkan susunan dan pencarian hasil algoritma, beberapa folder telah diwujudkan terlebih dahulu yang dinamakan “TextFile”, “*Tweets*”, “Profiling” dan “Graphs”. Output yang berkaitan akan disimpan dalam folder masing-masing. Fail *txt* yang mengandungi kata kunci tentang inisiatif “*One-Belt One Road*” dinyatakan dalam “TextFile” dan dinamakan “keywords.txt” dahulu. Disebabkan kata kunci yang dinyatakan tidak dapat dibaca sekali gus oleh algoritma, kata kunci telah dibahagikan kepada dua fail yang berbeza dan fail yang kedua dinamakan “keywords2.txt”. Sejumlah 23 kata kunci digunakan untuk pencarian.

Kemudian, fail “main.ipynb” dijalankan untuk memanggil fungsi-fungsi berkaitan dalam fail yang lain. Kod dalam “*scrapeTweets.py*” dijalankan dan mengumpulkan data-data *Tweets* menggunakan kata kunci yang telah dinyatakan. Tarikh pengumpulan *Tweets* bermula daripada 1 Januari 2018 hingga 1 Januari 2020.

Pencarian bahasa yang digunakan dalam *Tweets* ditetapkan sebagai Bahasa Inggeris sahaja manakala lokasi *Tweets* ditetapkan di Malaysia. Empat fail dijana daripada kod tersebut, iaitu “1.*Tweets_part1.csv*”, “2.*Tweets_part2.csv*”, “3.*compiled_Tweets.csv*” dan “i.*Profiling Report.html*”. Ketiga-tiga fail *csv* disimpan dalam folder “*Tweets*” manakala fail “i.*Profiling Report.html*” disimpan dalam folder “*Profiling*”.

Fail pertama dan kedua merupakan hasil pengumpulan *Tweets* melalui dua fail *txt* yang mengandungi kata kunci berbeza. Fail ketiga pula merupakan gabungan dua fail tersebut untuk memudahkan analisis. 1628 *Tweets* telah disimpan dalam fail pertama dan 232 *Tweets* disimpan dalam fail kedua. Sejumlah 1860 *Tweets* telah dikumpulkan sebagai data mentah dan disimpan dalam fail “3.*compiled_Tweets.csv*”. Rajah 3, 4 dan 5 masing-masing menunjukkan lima baris kandungan output fail *csv* daripada kod “*scrapeTweets.py*”.

Date	Username	Retweets	Favorites	Text	Mentions	Hashtags
27/12/2019 16:48	A	0	0	BB/YMCA Golf Fellow	# AndrewC	
27/12/2019 16:45	B	0	0	BB/YMCA Monthly Golf Fellows		
25/12/2019 17:16	C	0	0	@The Bellevue Resort	https://	
25/12/2019 16:22	D	0	0	Forest City part 2	@Forest City J	
25/12/2019 12:33	E	0	0	OMG BRI, HELLO I MISSED YOU!!		

Rajah 3 Sebahagian kandungan fail “1.*Tweets_part1.csv*”.

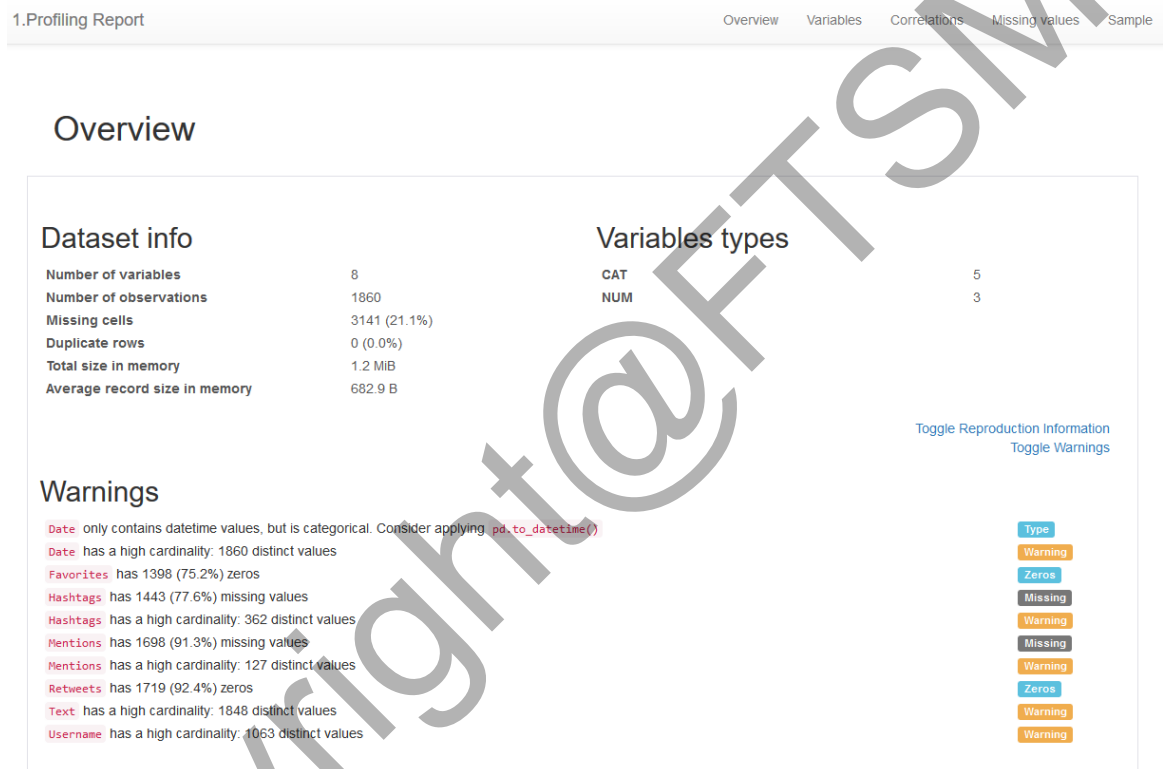
Date	Username	Retweets	Favorites	Text	Mentions	Hashtags
26/12/2019 9:49	F	0	1	Interview with Star r	# IQI # Ban	
23/12/2019 8:33	G	0	0	I'm at CCCC ECRL PROJECT MAN		
21/12/2019 18:23	H	0	2	Kuala Lumpur Bike V	# Malaysia	
17/12/2019 21:41	I	6	22	Bandar Malaysia is integral in fc		
17/12/2019 17:16	J	2	2	# Malaysia today ink	# Malaysia	

Rajah 4 Sebahagian kandungan fail “2.*Tweets_part2.csv*”.

Date	Username	Retweets	Favorites	Text	Mentions	Hashtags
27/12/2019 16:48	A	0	0	BB/YMCA Golf Fellow	# AndrewC	
27/12/2019 16:45	B	0	0	BB/YMCA Monthly Golf Fellows		
26/12/2019 9:49	F	0	1	Interview with Star r	# IQI # Ban	
25/12/2019 17:16	C	0	0	@The Bellevue Resort	https://	
25/12/2019 16:22	D	0	0	Forest City part 2	@Forest City.	

Rajah 5 Sebahagian kandungan fail “3.*compiled_Tweets.csv*”.

Fail “i.Profiling Report.html” merupakan laporan profil data *Tweets* yang mengandung penerangan ringkas tentang fail “3.compiled_*Tweets.csv*”. Maklumat statistik yang ringkas berkenaan data *Tweets* boleh diperoleh daripada fail *HTML* ini. Contohnya dari rajah 6, butiran pemboleh ubah yang ada dalam fail data *Tweets* terpapar termasuk jumlah, nama dan juga jumlah nilai yang hilang dalam setiap pemboleh ubah. Data ini digunakan untuk menentukan langkah pra-pemprosesan yang dilakukan terhadap data *Tweets*.



Rajah 6 Sebahagian kandungan fail “i.Profiling Report.html”.

4.4.2 Pra-pemprosesan Data

Setelah pengumpulan data, proses pembersihan data dijalankan terhadap data mentah tersebut. Kod “preprocess*Tweets.py*” dipanggil dan fail “3.compiled_*Tweets.csv*” dibaca. Tiga fail output dari proses ini adalah “4.preprocessed_*Tweets.csv*”, “text_dictionary.txt” dan “ii.Profiling Report.html”.

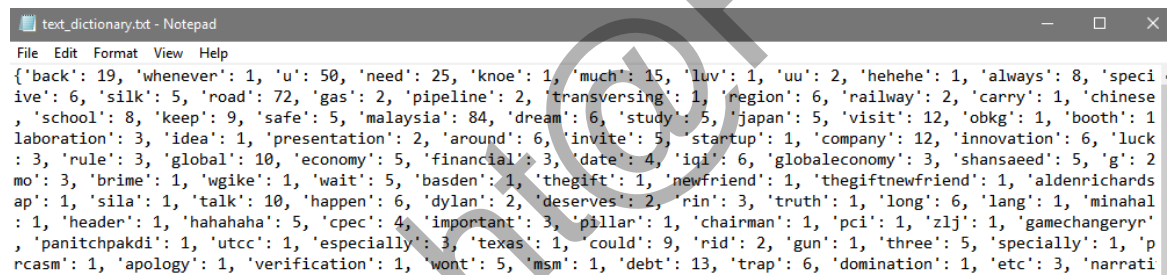
Data *Tweets* yang telah dijalankan pra-pemprosesan disimpan sebagai “4.preprocessed_*Tweets.csv*” dalam folder “*Tweets*”. Semua perkataan akar dan kekerapannya yang diperoleh dari setiap *Tweets* disimpan dalam “text_dictionary.txt” di folder “TextFile”. Fail ini akan digunakan dalam proses analisis kemudian. Fail “ii.Profiling Report.html” yang merupakan laporan profil tentang

“4.preprocessed_Tweets.csv” juga dijana ke dalam folder “Profiling”. 874 Tweets tertinggal setelah langkah pra-pemrosesan ini.

Rajah 7 menunjukkan kandungan fail “4.preprocessed_Tweets.csv”. Rajah 8 menunjukkan kandungan fail “text_dictionary.txt”. Rajah 9 menunjukkan kandungan “ii.Profiling Report.html”.

Date	Username	Retweets	Favorites	Text	Preproce	Tokenized	Mentions	Hashtags
27/12/2019	A	0	0	BB/YMCA	BB YMCA	['BB', 'YMC	None	# Andrew
27/12/2019	B	0	0	BB/YMCA	BB YMCA	['BB', 'YMC	None	None
26/12/2019	F	0	1	Interview	Interview	['Interview	None	# IQI # Bar
25/12/2019	D	0	0	Forest Cit	Forest Cit	['Forest', 'N	None	None
25/12/2019	E	0	0	OMG BRI,	OMG BRI,	['OMG', 'B	None	None

Rajah 7 Sebahagian kandungan fail “4.preprocessed_Tweets.csv”.



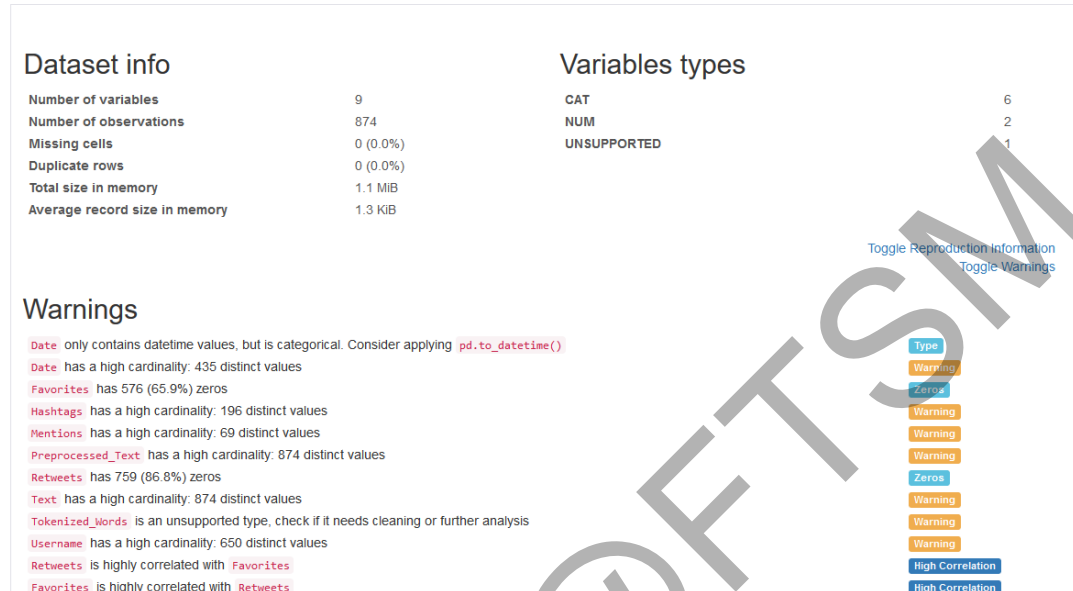
```

File Edit Format View Help
{'back': 19, 'whenever': 1, 'u': 50, 'need': 25, 'knoe': 1, 'much': 15, 'luv': 1, 'uu': 2, 'hehehe': 1, 'always': 8, 'speci
ive': 6, 'silk': 5, 'road': 72, 'gas': 2, 'pipeline': 2, 'transversing': 1, 'region': 6, 'railway': 2, 'carry': 1, 'chinese
, 'school': 8, 'keep': 9, 'safe': 5, 'malaysia': 84, 'dream': 6, 'study': 5, 'japan': 5, 'visit': 12, 'obkg': 1, 'booth': 1
laboration': 3, 'idea': 1, 'presentation': 2, 'around': 6, 'invite': 5, 'startup': 1, 'company': 12, 'innovation': 6, 'luck
: 3, 'rule': 3, 'global': 10, 'economy': 5, 'financial': 3, 'date': 4, 'iqi': 6, 'globaleconomy': 3, 'shansaeed': 5, 'g': 2
mo': 3, 'brime': 1, 'wgike': 1, 'wait': 5, 'basden': 1, 'thegift': 1, 'newfriend': 1, 'thegiftnewfriend': 1, 'aldenrichards
ap': 1, 'sila': 1, 'talk': 10, 'happen': 6, 'dylan': 2, 'deserves': 2, 'rin': 3, 'truth': 1, 'long': 6, 'lang': 1, 'minahal
: 1, 'header': 1, 'hahahaha': 5, 'cpec': 4, 'important': 3, 'pillar': 1, 'chairman': 1, 'pci': 1, 'zlj': 1, 'gamechangeryr'
, 'panitchpakdi': 1, 'utcc': 1, 'especially': 3, 'texas': 1, 'could': 9, 'rid': 2, 'gun': 1, 'three': 5, 'specially': 1, 'p
rcasm': 1, 'apology': 1, 'verification': 1, 'wont': 5, 'msm': 1, 'debt': 13, 'trap': 6, 'domination': 1, 'etc': 3, 'narrati

```

Rajah 8 Sebahagian kandungan fail “text_dictionary.txt”.

Overview



Rajah 9 Sebahagian kandungan fail “ii.Profiling Report.html”.

4.4.3 Analisis berasaskan Leksikon

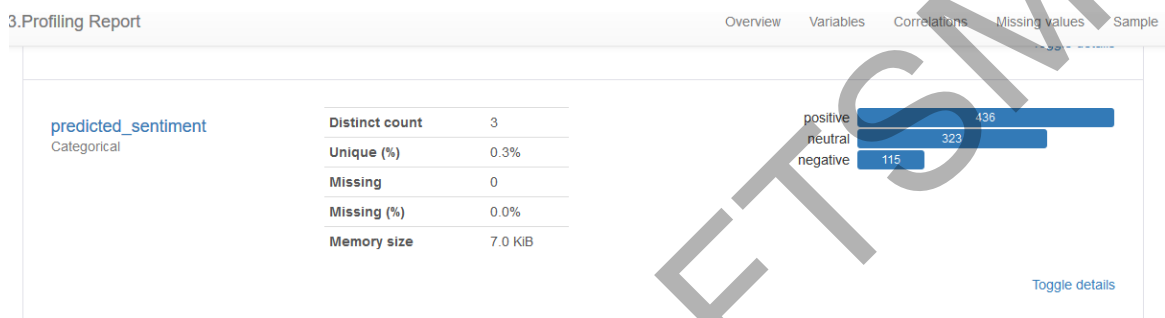
Dalam langkah analisis leksikon, fail “`analyseTweets.py`” dipanggil untuk menganalisis data *Tweets* fail “`4.preprocessed_Tweets.csv`”. Output fail dari proses ini adalah fail “`5.analysed_Tweets.csv`” dan fail “`iii.Profiling Report.html`”.

Fail “`5.analysed_Tweets.csv`” disimpan dalam folder “*Tweets*”. Rajah 10 menunjukkan lima baris kandungan fail “`5.analysed_Tweets.csv`” yang mengandungi data *Tweets* yang telah dibersihkan dan pemarkahan serta pengelasan tag sentimen untuk setiap Tweet.

Fail “`iii.Profiling Report.html`” yang merupakan laporan profil tentang “`5.analysed_Tweets.csv`” disimpan dalam folder “*Profiling*”. Dari laporan ini, jumlah tag sentimen telah diperolehi untuk setiap kelas. Rajah 11 menunjukkan bahawa jumlah tag sentimen positif yang paling banyak adalah positif manakala jumlah tag sentimen negatif adalah paling sedikit. Namun, data ini tidak dapat digunakan secara langsung untuk penghasilan graf. Justeru, langkah seterusnya dilakukan.

Date	Text	Preprocess	tokenized	compound	polarity	average_s	predicted
27/12/2019	BB/YMCA	BB YMCA	['BB', 'YMC	0.6369	0	0.31845	positive
27/12/2019	BB/YMCA	BB YMCA	['BB', 'YMC	0	0	0	neutral
26/12/2019	Interview	Interview	['Interview	0	0	0	neutral
25/12/2019	Forest Cit	Forest Cit	['Forest', '	0	0	0	neutral
25/12/2019	OMG BRI,	OMG BRI,	['OMG', 'B	-0.3595	0	-0.17975	negative

Rajah 10 Sebahagian kandungan fail “5.analysed_Tweets.csv”.



Rajah 11 Sebahagian kandungan fail “iii.Profiling Report.html”.

4.4.4 Analisis Statistik

Dalam langkah analisis statistik, fail “5.analysed_Tweets.csv” digunakan oleh “analyseTweetsStatistics.py” untuk pengiraan jumlah setiap tag sentimen dan frasa hangat. Tiga fail output dalam proses ini adalah “result_statistics.txt”, “filtered_text_dictionary.txt” dan “6.filtered_dictionary_output.csv”.

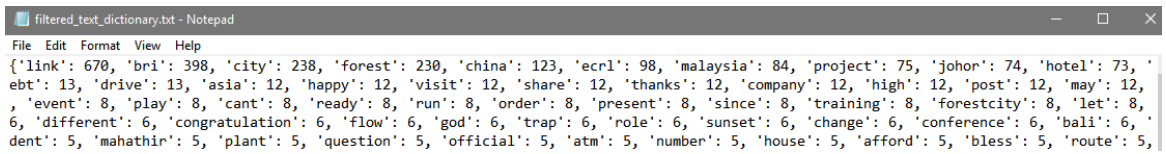
Keputusan pengiraan akan disimpan dalam fail *txt* “result_statistics.txt” di folder “TextFile” seperti Rajah 12. Kemudian, perkataan dalam “text_dictionary.txt” ditapis dan disusun secara tertib menurun mengikut kekerapan perkataan. Perkataan hangat yang telah diatur disimpan dalam “6.filtered_dictionary_output.csv” di folder “Tweets” untuk memudahkan semakan manual dan “filtered_text_dictionary.txt” di folder “TextFile” untuk pembentangan grafik. Rajah 13 dan 14 menunjukkan kandungan kamus yang telah ditapis dalam bentuk *txt* dan *csv*.

```

result_statistics.txt - Notepad
File Edit Format View Help
436
115
323

```

Rajah 12 Kandungan fail “result_statistics.txt”.



```

filtered_text_dictionary.txt - Notepad
File Edit Format View Help
{'link': 670, 'bri': 398, 'city': 238, 'forest': 230, 'china': 123, 'ecrl': 98, 'malaysia': 84, 'project': 75, 'johor': 74, 'hotel': 73, '
ebt': 13, 'drive': 13, 'asia': 12, 'happy': 12, 'visit': 12, 'share': 12, 'thanks': 12, 'company': 12, 'high': 12, 'post': 12, 'may': 12,
, 'event': 8, 'play': 8, 'cant': 8, 'ready': 8, 'run': 8, 'order': 8, 'present': 8, 'since': 8, 'training': 8, 'forestcity': 8, 'let': 8,
6, 'different': 6, 'congratulation': 6, 'flow': 6, 'god': 6, 'trap': 6, 'role': 6, 'sunset': 6, 'change': 6, 'conference': 6, 'bali': 6, '
dent': 5, 'mahathir': 5, 'plant': 5, 'question': 5, 'official': 5, 'atm': 5, 'number': 5, 'house': 5, 'afford': 5, 'bless': 5, 'route': 5,

```

Rajah 13 Sebahagian kandungan fail “filtered_text_dictionary.txt”.

link	670
bri	398
city	238
forest	230
china	123
ecrl	98

Rajah 14 Sebahagian kandungan fail “6.filtered_dictionary_output.csv”.

4.4.5 Pembentangan Hasil Analisis

Akhir sekali, proses pembentangan hasil analisis secara grafik dijalankan oleh fail “analysisOutput.py”. Tujuh fail output daripada langkah ini termasuk satu *Wordcloud*, satu carta pai dan empat graf siri masa. Fail “filtered_text_dictionary.txt” digunakan untuk menjana *Wordcloud* yang menggambarkan taburan kekerapan perkataan; fail “result_statistics.txt” digunakan untuk menghasilkan carta pai yang menggambarkan peratusan tag sentimen *Tweets*; fail “5.analysed_Tweets.csv” digunakan untuk menghasilkan graf siri masa yang menunjukkan taburan tag sentimen *Tweets* sepanjang masa. Semua rajah dan graf disimpan di folder “Graphs” dalam bentuk *png*. Rajah-rajah di bawah menunjukkan output kod “analysisOutput.py”.

Rajah 15 menunjukkan *Wordcloud* yang dijana. Saiz perkataan adalah berkadar dengan kekerapan sesuatu perkataan. Kelihatan perkataan “link”, “bri”, “forest”, “city”, “ecrl” dan “china” mempunyai saiz yang lebih besar berbanding dengan perkataan-perkataan yang lain. Hal ini disebabkan perkataan-perkataan ini merupakan antara 10 perkataan yang paling kerap disebut dalam *Tweets*.

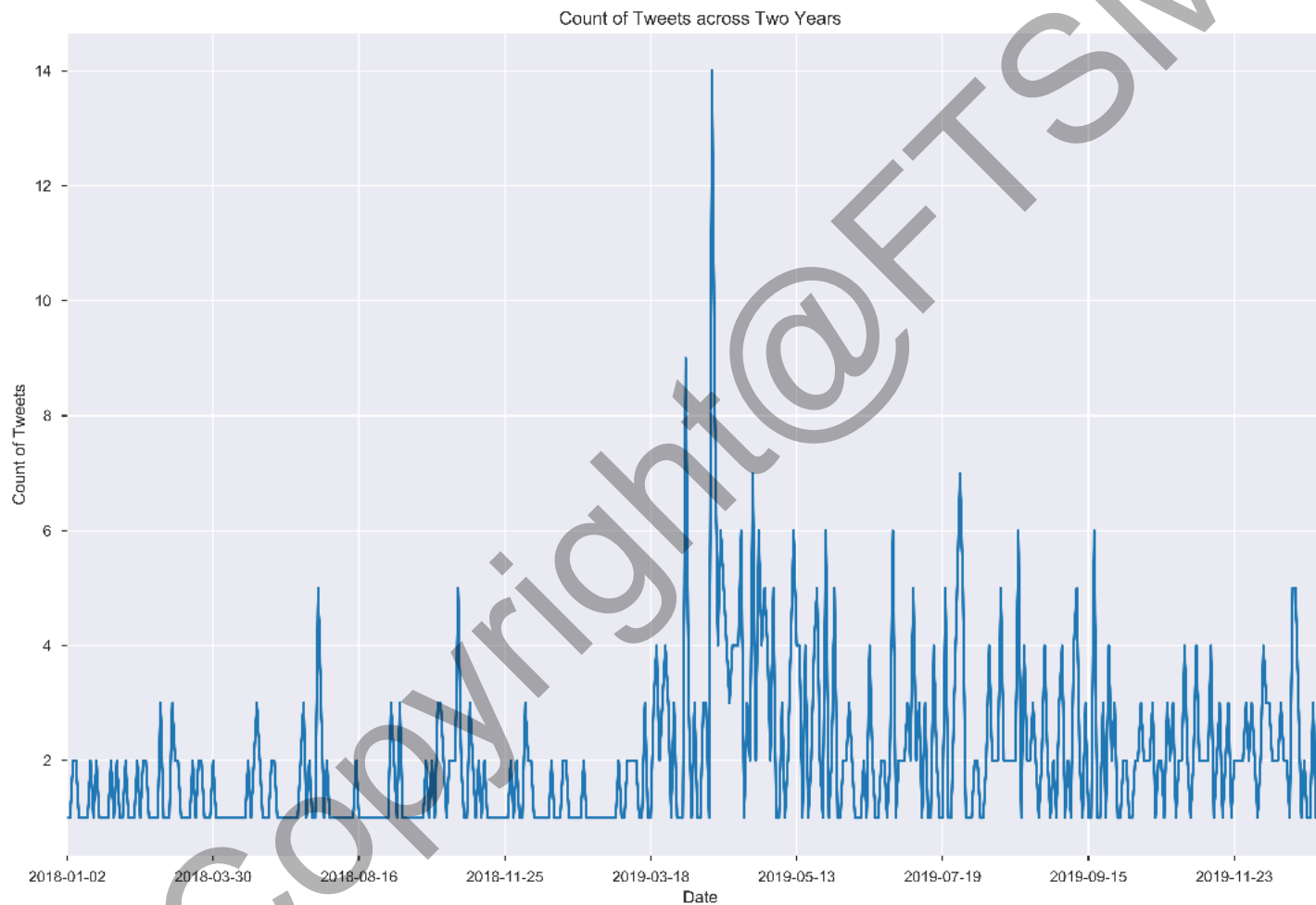
Rajah 17 menunjukkan graf siri masa keseluruhan *Tweets* yang tidak diberikan pengelasan tag sentimen bermula awal tahun 2018 sehingga awal tahun 2020. Kelihatan *Tweets* berkenaan inisiatif “*One-Belt One-Road*” paling kerap dibincangkan pada sekitar April 2019 di mana jumlah *Tweets* mencapai 14 pada suatu hari.

Rajah 18 menunjukkan graf siri masa untuk *Tweets* positif dalam dua tahun. Paling banyak Tweet dikelaskan sebagai positif pada sekitar April 2019 dan mencapai 14 Tweet. Corak ini selaras dengan corak graf siri masa keseluruhan *Tweets*.

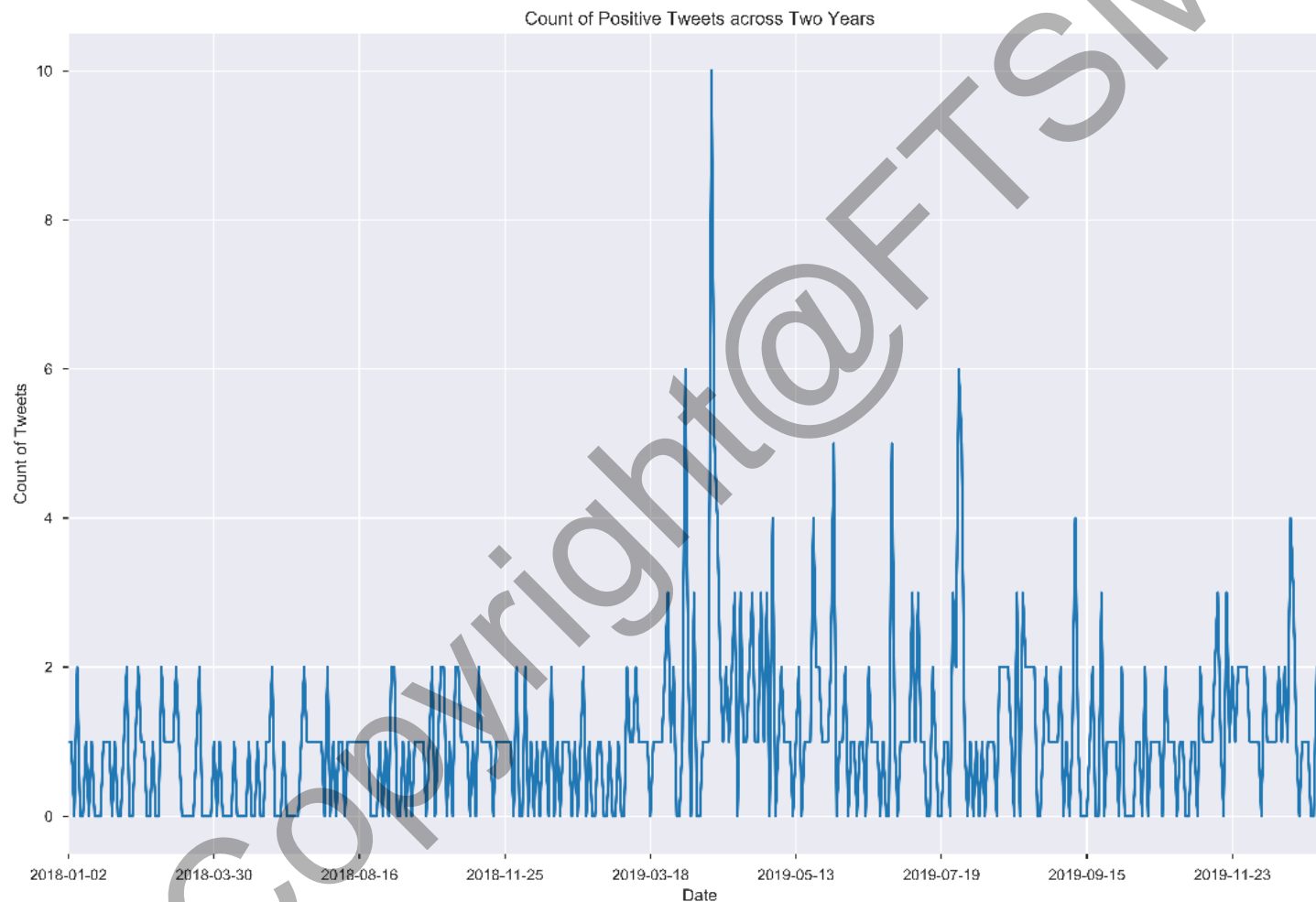
Rajah 19 menunjukkan graf siri masa untuk *Tweets* negatif. Tweet maksimum hanya adalah 5 pada sehari pada sekitar bulan Mei 2019. Rajah 20 menunjukkan graf siri masa untuk *Tweets* neutral. 5 Tweet dikelaskan sebagai neutral sebanyak dua kali dalam April 2019.

Rajah 21 menunjukkan graf siri masa *Tweets* keseluruhan yang telah dikelaskan. Tag positif diwarnakan biru, tag negatif diwarnakan jingga manakala tag neutral diwarnakan hijau.

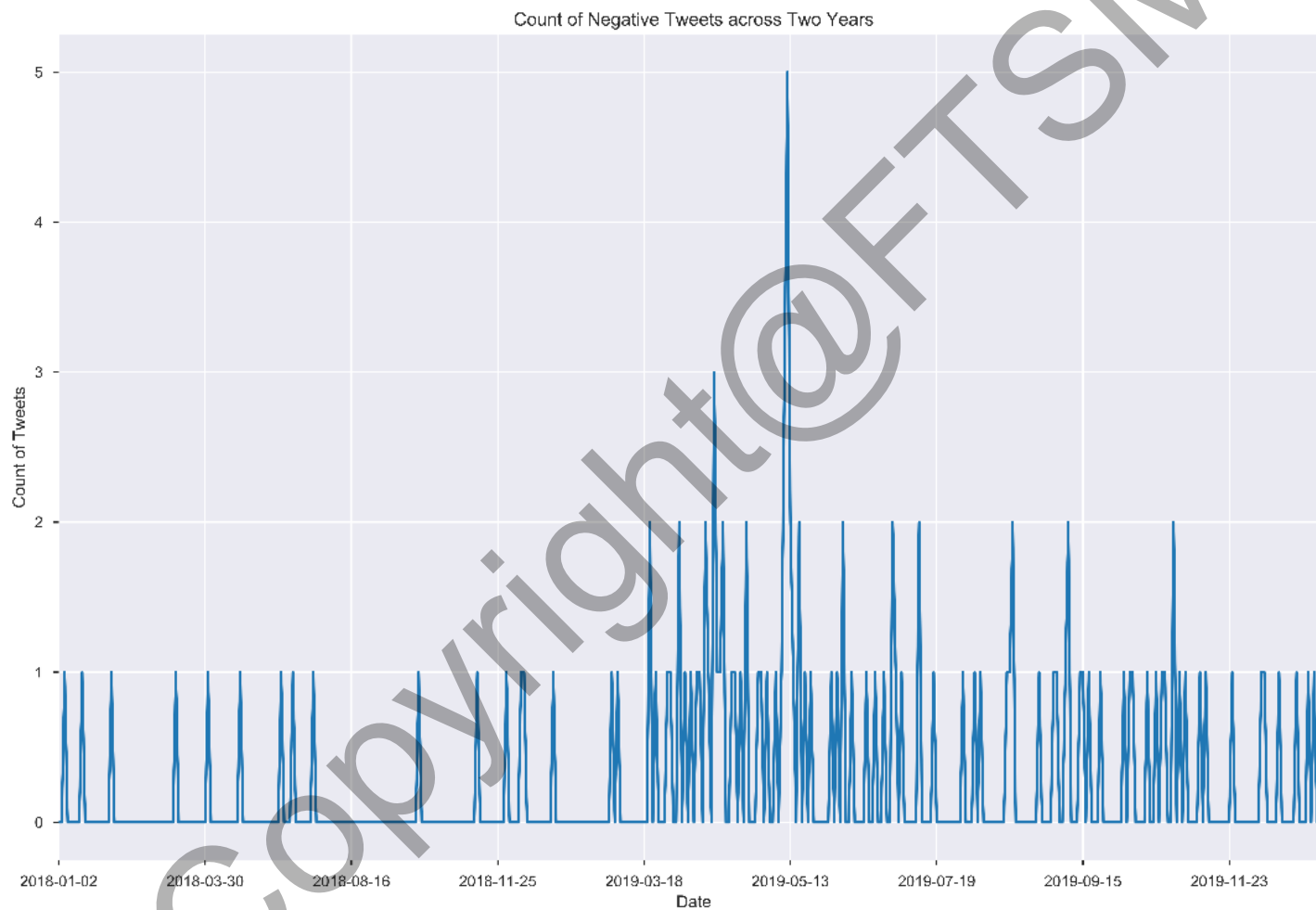
Copyright@FTSM



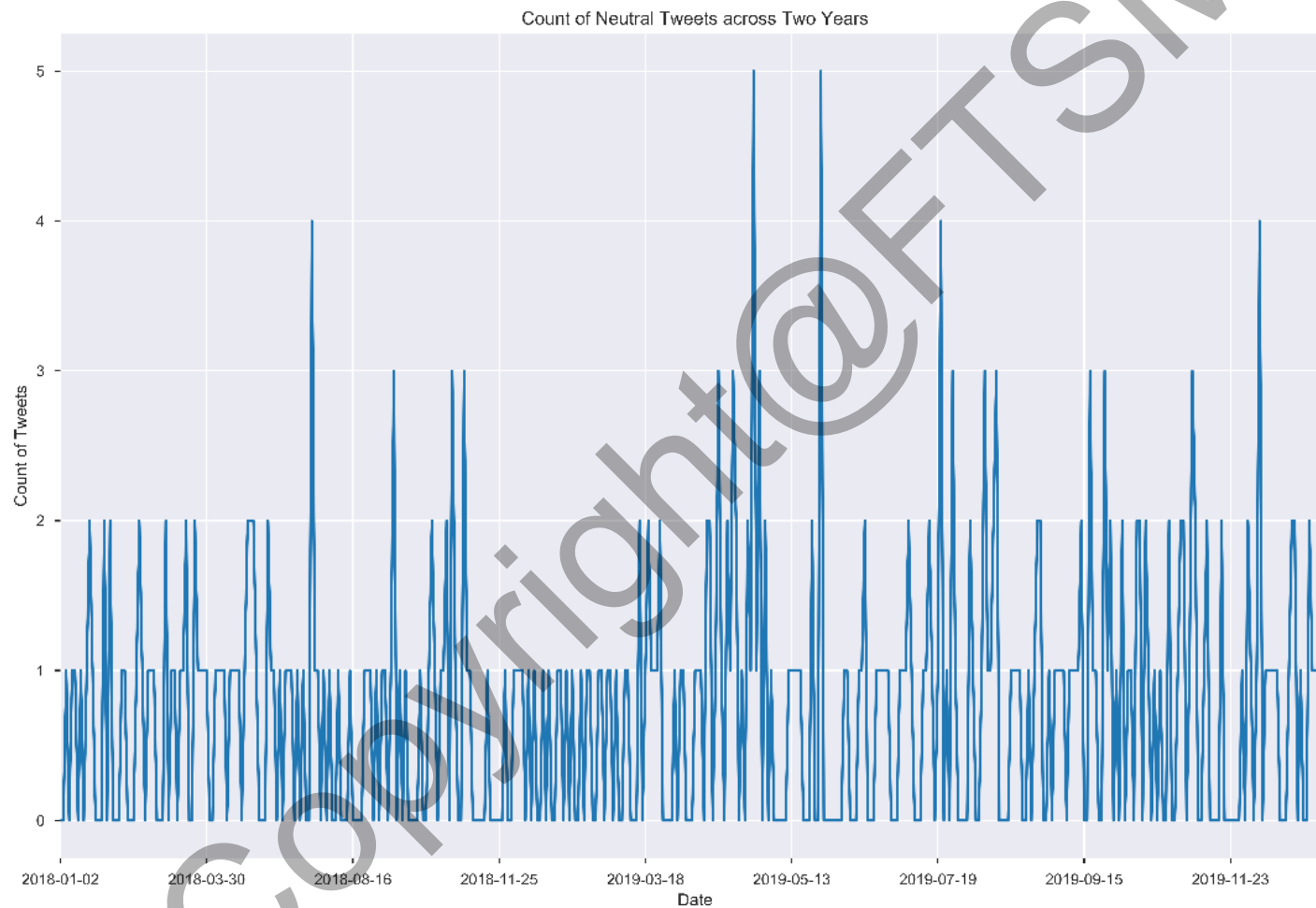
Rajah 17 Graf siri masa keseluruhan yang belum dikelaskan.



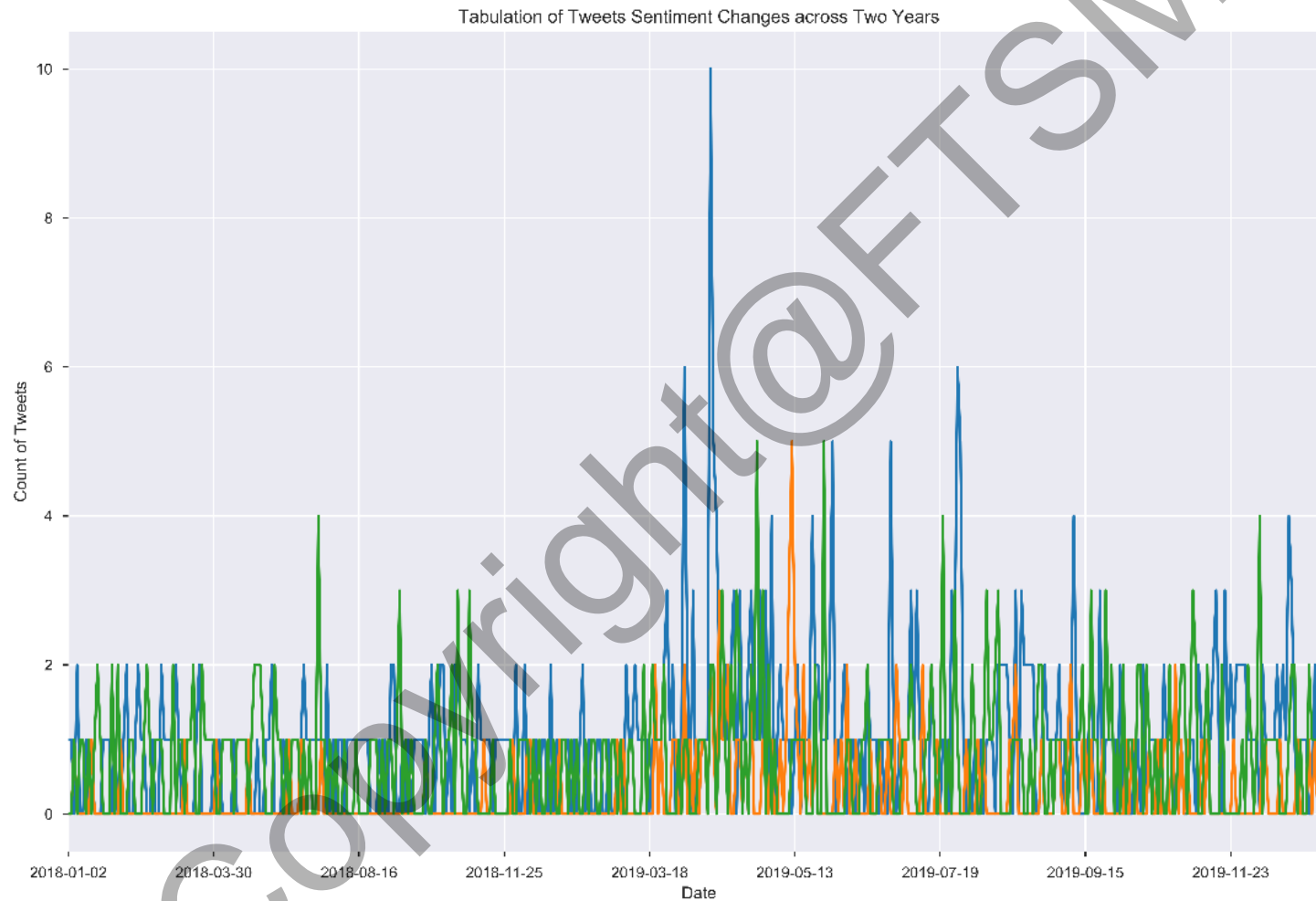
Rajah 18 Graf siri masa untuk *Tweets* yang dikelaskan positif.



Rajah 19 Graf siri masa untuk *Tweets* yang dikelaskan negatif.



Rajah 20 Graf siri masa untuk *Tweets* yang dikelaskan neutral



Rajah 21 Graf siri masa untuk *Tweets* keseluruhan yang telah dikelaskan.

5 HASIL KAJIAN

Hasil kajian ini akan dibentangkan mengikut tiga objektif yang telah dinyatakan dalam Bab 1. Secara keseluruhannya, algoritma yang mampu menganalisis sentimen terhadap komen di Twitter telah berjaya dihasilkan menggunakan bahasa pengaturcaraan *Python*. Lima proses analisis yang dirancang telah berjaya dibina secara bermodul dan berfungsi seperti yang diinginkan.

Secara umumnya, pandangan majoriti terhadap inisiatif “*One-Belt One-Road*” dalam Twitter adalah bersifat positif. Sejumlah 436 daripada 874 *Tweets*, iaitu 49.89% *Tweets* adalah bersifat positif. Jumlah *Tweets* yang bersifat neutral adalah dalam 323 atau 36.96% manakala *Tweets* yang bersifat negatif adalah sejumlah 115 atau 13.18%.

Sepuluh frasa hangat tertinggi yang berkaitan dengan inisiatif ini di Twitter adalah “link”, “bri”, “city”, “forest”, “china”, “ecrl”, “malaysia”, “project”, “johor” dan “hotel”. Susunan perkataan-perkataan tersebut adalah mengikut tertib menurun. Kekekapan perkataan-perkataan hangat tersebut adalah di antara 73 dan 670. Frasa tertentu boleh dihasilkan daripada perkataan-perkataan hangat yang dikumpulkan termasuk “China BRI”, “Forest City. Johor” dan “Malaysia ECRL Project”.

100 *Tweets* yang telah dianalisis telah terpilih secara rawak untuk penyemakan manual tentang ketepatan analisis. 78% *Tweets* telah dikelaskan secara tepat. Namun begitu, hanya 51% *Tweets* didapati berkaitan dengan topik inisiatif “*One-Belt One-Road*” secara langsung. Sejumlah 40 daripada 51 atau 78.43% *Tweets* tersebut telah dikelaskan secara tepat.

Frasa hangat yang diperoleh boleh ditafsirkan bahawa *Forest City* di Johor serta projek ECRL di Malaysia merupakan topik berkenaan “China BRI” yang banyak dibincangkan dalam Twitter. Kemunculan perkataan “hotel” pula mungkin disebabkan oleh lokasi pengguna dimasukkan sekali dalam *Tweets* mereka. Pemeriksaan manual secara ringkas menunjukkan bahawa perkataan “hotel” biasanya muncul bersama “*Forest City*”.

6 KESIMPULAN

Secara konklusi, kesemua objektif kajian ini telah tercapai. Algoritma untuk menganalisis sentimen terhadap inisiatif “*One-Belt One-Road*” telah dihasilkan. Pandangan majoriti terhadap inisiatif ini dalam Twitter adalah bersifat positif. Frasa-frasa hangat yang diperoleh dari data *Tweets* termasuk “China BRI”, “Forest City. Johor” dan “Malaysia ECRL Project”.

Kekurangan kajian ini termasuk pengumpulan data tidak relevan yang banyak serta pembersihan data yang tidak lengkap terhadap bahasa rojak dan singkatan dalam *Tweets*. Penambahbaikan yang boleh dilakukan terhadap projek ini termasuk pemilihan kata kunci penuh sahaja serta penambahbaikan langkah pra-pemprosesan data.

7 RUJUKAN

- Bird, Steven, Edward Loper and Ewan Klein (2009), *Natural Language Processing with Python*. O'Reilly Media Inc.
- Clement, J. 2019. Twitter: number of monthly active users 2010-2019. <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/> [13 Oktober 2019].
- Desai, R. D. 2018. Sentiment Analysis of Twitter Data. *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, (Iciccs), 114–117.
- Ferdinand, P. 2016. Westward ho—the China dream and ‘one belt, one road’: Chinese foreign policy under Xi Jinping. *International Affairs*, 92(4), 941–957.
- Henrique, J. 2018. GetOldTweets-python. <https://github.com/Jefferson-Henrique/GetOldTweets-python> [20 November 2019].
- Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. Ann Arbor, MI, June 2014.
- Ikoru, V., Sharmina, M., Malik, K., & Batista-Navarro, R. 2018. Analyzing Sentiments Expressed on Twitter by UK Energy Company Consumers. *2018 5th International Conference on Social Networks Analysis, Management and Security, SNAMS 2018*, 95–98.
- Kaur, H., Mangat, V. & Nidhi. 2017. A survey of sentiment analysis techniques. *Proceedings of the International Conference on IoT in Social, Mobile, Analytics and Cloud, I-SMAC 2017*, 921–925.
- Kaur, H., Mangat, V., & Nidhi. 2017. A survey of sentiment analysis techniques. *Proceedings of the International Conference on IoT in Social, Mobile, Analytics and Cloud, I-SMAC 2017*, 921–925.
- Mishra, P., Rajnish, R. & Kumar, P. 2016. Sentiment Analysis of Twitter Data: Case Study on Digital India. *2016 International Conference on Information Technology (InCITE) – The Next Generation IT Summit*, 148–153
- Raghupathi, D., Yannou, B., Farell, R. & Poirson, E. 2015. Customer sentiment appraisal from user-generated product reviews: a domain independent heuristic algorithm. *International Journal on Interactive Design and Manufacturing*, 9(3), 201–211.

- Ramanathan, V. & Meyyappan, T. 2019. Twitter text mining for sentiment analysis on people's feedback about Oman tourism. *2019 4th MEC International Conference on Big Data and Smart City, ICBDS 2019*, 1–5.
- Ramanathan, V., Meyyappan, T. 2019. Twitter text mining for sentiment analysis on people's feedback about Oman tourism. *2019 4th MEC International Conference on Big Data and Smart City, ICBDS 2019*, 1–5.
- Shahnawaz, Astya. P. 2017. Sentiment analysis: Approaches and open issues. *Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA 2017, 2017-Janua*, 154–158.
- Smith, J. M. 2018. China's Belt and Road Initiative: Strategic Implications and International Opposition. *Background*, (3331), 1–25.
- Twitter Inc. 2019. Glossary. <https://help.twitter.com/en/glossary> [19 November 2019].
- Vamshi, K. B., Pandey, A. K., & Siva, K. A. P. 2018. Topic Model Based Opinion Mining and Sentiment Analysis. *2018 International Conference on Computer Communication and Informatics, ICCCI 2018*, 1–4.
- Vanaja, S., Belwal, M. 2018. Aspect-Level Sentiment Analysis on E-Commerce Data. *Proceedings of the International Conference on Inventive Research in Computing Applications, ICIRCA 2018, (Icirca)*, 1275–1279.
- Vu, D. 2019. Generating WordClouds in Python. <https://www.datacamp.com/community/tutorials/wordcloud-python> [20 November 2019].
- Wagh, R., Punde, P. 2018. Survey on Sentiment Analysis using Twitter Dataset. *Proceedings of the 2nd International Conference on Electronics, Communication and Aerospace Technology, ICECA 2018, (Iceca)*, 208–211.
- Zadrozny, P. & Kodali, R. 2013. Sentiment Analysis. Dlm. *Big Data Analytics Using Splunk*. Apress, Berkeley, CA 255–282
- Zvarevashe, K., Olugbara, O. O. 2018. A framework for sentiment analysis with opinion mining of hotel reviews. *2018 Conference on Information Communications Technology and Society, ICTAS 2018 - Proceedings*, 1–4.