

TWITTER SENTIMENT ANALYSIS ON SIGNIFICANT EVENTS THAT HAPPENS AROUND UKM, BANGI

Tay Fui Kien

Dr.Lailatul Qadri Zakaria

Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia

ABSTRAK

Analisis sentimen merupakan satu process dalam menentukan nada emosi sama ada positif, neutral atau negatif di sebalik penggunaan bahasa tabii. Twitter adalah sejenis aplikasi rangkaian sosial dan mikroblog. Melalui platform seperti ini, pengguna Twitter dapat membincangkan isu semasa, berkongsi pendapat berkaitan mana-mana topik. Twitter merupakan antara aplikasi yang terpilih untuk pengumpulan data yang penting yang berlaku di sekitar UKM, Bangi. Sebagai pengetahuan, Twitter telah memutuskan bahawa perkembangan had mesej Twitter (*tweets*) kepada 280 aksara dilancarkan secara terbuka pada 7 November 2017. Namun, pengguna Twitter masih mempraktikkan penggunaan mesej yang ringkas dan tidak rasmi seperti akronim, emotikon, bahasa yang bercampur dan melakukan kesalahan ejaan. Analisis sentimen data Twitter bagi mengenal pasti ketepatan emosi para pengguna Twitter berdasarkan sifat-sifat dan emotikon yang terdapat dalam *tweets* menjadi satu tugas yang mencabarkan pengkaji. Perisian Studio R digunakan bagi mengatasi masalah kekompleksan dan kerumitan semasa membuat analisis sentimen Twitter. Fungsi 'gsub' dalam Studio R digunakan bagi menapis data *tweets* seperti menyingkirkan emotikon, @-mention, tanda baca, nombor, URL dan tanda aneh (*bizarre signs*). Proses tokenisasi, penyingkiran *stopwords* dan penandaan POS juga dipraktikkan sepanjang proses analisis sentimen. Teknik n-grams digunakan untuk menangkap teks yang membawa maksud negasi seperti 'not happy' (sentimen negatif), 'not bad' (sentimen positif). Rshiny digunakan bagi memaparkan hasil keputusan kajian dalam bentuk web. Pembangunan sistem analisis sentimen Twitter mengenai peristiwa-peristiwa penting yang berlaku di sekitar UKM, Bangi memerlukan bantuan aplikasi Studio R dan RShiny. Kesimpulannya, jangkaan hasil pengenalanpastian sentimen pengguna Twitter berkaitan

#

peristiwa penting yang berlaku di sekitar UKM, Bangi dapat dianalisis. Semoga pembangunan sistem ini dapat memudahkan khlayak ramai dalam pencapaian maklumat yang berguna dan memanfaatkan pada masa akan datang.

1 PENGENALAN

Analisis sentimen adalah cara perolehan kontekstual teks dalam mengenal pasti dan mengekstrak maklumat subjektif yang terkandung dalam sumber maklumat dengan tujuan mengkaji sentimen sosial dan emosi terhadap perkhidmatan, jenama atau produk (Shashank Gupta 2018). Analisis sentimen juga didefinisikan sebagai satu penggunaan Pemprosesan Bahasa Tabii (NLP), statistik, dan analisis teks bagi mengekstrak di samping mengenal pasti sentimen teks ke dalam kategori positif, negatif, atau neutral (Matt Kiser 2016). Tambahan pula, analisis sentimen merupakan proses memahami sesuatu pendapat suara berkaitan dengan subjek tertentu secara automatik berdasarkan tulisan ataupun secara lisan (MonkeyLearn 2018).

Analisis sentimen sering digunakan untuk mencapai keputusan binari di mana seseorang menerima atau menolak sesuatu, seseorang suka atau tidak suka sesuatu, atau produk itu baik atau buruk (Matt Kiser 2016). Sebagai manusia, kita dapat mengesan perasaan dan emosi seseorang berdasarkan teks dengan berkesan dan biasanya dapat mengesan dalam keadaan yang hampir tidak disedari. Kita mampu menapis semua konteks yang mengandungi perkataan, frasa, objek ataupun senario, mengklasifikasikan bahagian-bahagian yang relevan, membandingkannya dengan pengalaman masa lalu, dan dapat menggunakannya untuk memperdalam pemahaman kita terhadap maklumat yang kita capai (Pradeep Palani 2018). Namun, dalam era teknologi kini, kemajuan dalam Kecerdasan Mesin dan Pemprosesan Bahasa Tabii (NLP) telah memberi impak terhadap analisis semantik melalui algoritma yang canggih, komputer yang berkuasa tinggi, dan membekalkan data bagi melatih kecerdasan komputer dalam aspek ini. Dengan ini, mesin berkeupayaan dalam memahami bahasa tabii termasuk sentimen, penilaian, sikap, dan emosi orang ramai melalui bahasa tulisan (MonkeyLearn 2018).#

#

2 PENYATAAN MASALAH

Laman web mikroblog telah berkembang menjadi sumber maklumat yang pelbagai. Dalam kajian analisis sentimen, Twitter merupakan antara pilihan pelbagai mikroblog yang digunakan dalam membantu pengkaji membina model bagi mengklasifikasikan ‘*tweets*’ kepada sentimen positif, neutral dan negatif. Namun begitu, terdapat beberapa masalah telah dikenal pasti. Antaranya adalah bagaimana cara untuk mengenal pasti emosi para pengguna Twitter terhadap peristiwa yang berlaku di sekitar UKM, Bangi. Seperti yang sedia maklum, Twitter merupakan laman web mikroblog yang digunakan dalam kajian ini sebagai sumber pengumpulan data dan maklumat. Pada zaman moden ini, orang ramai suka berkongsi maklumat, perasaan dan pengamalan diri di laman web mikroblog seperti Twitter.

Di samping itu, Twitter telah menghadkan jumlah perkataan iaitu maxima 280 aksara dalam setiap *tweets*. Sekatan saiz mesej ini menggalakkan penggunaan bahasa yang bijak, menjadikan tweets sangat mudah diimbis, dan juga sangat mencabar bagi pengguna Twitter untuk menulis dengan baik (Paul Gil 2018). Dengan ini, beberapa masalah kajian telah ditimbulkan iaitu bagaimana cara untuk mengenal pasti sifat-sifat yang penting dalam *tweets* seperti emotikon, *parts of speech (POS)* dan *stemwords*. Masalah kajian yang seterusnya adalah bagaimana cara untuk menyingkirkan *hashtags*, *targets*, singkatan, tanda baca, nombor, dan aksara khas.

3 OBJEKTIF KAJIAN

Objektif utama bagi projek ini adalah untuk mengenal pasti emosi orang ramai berdasarkan *tweets* yang disebarkan dalam Twitter berdasarkan peristiwa yang berlaku di kawasan sekitar UKM, Bangi pada tempoh masa tertentu. Objektif yang kedua adalah untuk membangunkan sistem analisis sentimen data Twitter yang mampu menentukan keputusan sentimen bagi peristiwa-peristiwa yang berkaitan.

4 METOD KAJIAN

Bagi memastikan sistem analisis sentimen Twitter dapat dibangunkan dengan sempurna, beberapa fasa pembangunan perlu diikuti. Fasa pembangunan termasuk fasa perancangan,

#

fasa analisis, fasa reka bentuk, fasa pengujian dan dokumentasi. Dokumen spesifikasi keperluan sistem bagi sistem yang bakal dibangunkan adalah penting bagi memastikan kelancaran pelaksanaan sistem tersebut dan juga bagi mengelakkan sebarang masalah yang tidak diingini. Selain itu, keperluan berfungsi dan keperluan tidak berfungsi dijelaskan bagi mencapai objektif pembangunan sistem tersebut. Manakala penggunaan perisian dan perkakasan yang bersesuaian juga penting dalam meningkatkan prestasi sistem yang akan dibangunkan.

4.1 Fasa Perancangan

Fasa perancangan membincangkan proses pengenalpastian masalah, objektif, skop, kekangan, cadangan penyelesaian dan metodologi kajian. Langkah yang seterusnya merujuk kepada perbincangan kajian susastera hasil daripada pembacaan bahan bercetak dan bahan yang diperolehi menerusi media elektronik seperti laporan saintifik, laporan rasmi, dokumen teknikal dan lain-lain. Kajian susastera perlu dilakukan bagi memberikan pendedahan idea dan hala tuju yang jelas mengenai topik yang berkaitan dengan proses analisis sentimen ini. Kajian kesusteraan yang dijalankan juga dapat memberikan gambaran jelas mengenai topik kajian di samping memastikan kajian yang dilaksanakan dapat memenuhi objektif yang digariskan.

4.2 Fasa Analisis

Fasa analisis diperlukan untuk pengkaji mentafsir dan menganalisis maklumat yang dikumpulkan dalam fasa perancangan. Proses analisis ini adalah mustahak bagi mengenalpasti masalah dan teknik yang akan dipraktikkan supaya dapat memastikan kelancaran dalam pembangunan projek dan memenuhi skop dan matlamat kajian yang ditetapkan.

4.3 Fasa Reka Bentuk

Melalui fasa reka bentuk, gambaran mengenai sistem yang dibangunkan dapat diperlihatkan dengan jelas dan menjadi satu panduan semasa proses menghasilkan sistem analisis sentimen yang sistematik. Fasa ini juga membincangkan reka bentuk seni bina

#

(*architecture design*), reka bentuk pangkalan data (*database design*), reka bentuk antara muka (*interface design*) dan reka bentuk algoritma (*algorith design*). Spesifikasi reka bentuk memberi keterangan yang mendalam mengenai proses ‘Analisis Sentimen Twitter Mengenai Peristiwa-Peristiwa Penting Yang Berlaku Di Sekitar UKM, Bangi’. Pembinaan reka bentuk seperti ini memudahkan pembangun dan pengguna untuk memahami proses pembangunan sistem web ini. Di samping itu, ia juga membantu dalam penghasilan sistem yang dapat memenuhi keperluan dan permintaan pengguna.

4.4 Fasa Pengujian

Pengujian sistem dijalankan bagi memastikan sistem yang dibangunkan menepati keperluan dan dapat berfungsi tanpa sebarang masalah yang kritikal. Pengkaji mempraktikkan pengujian berasaskan struktur dengan menggunakan pendekatan reka bentuk kotak putih (*white-box*). Pendekatan ini digunakan bagi menguji keberkesanan berdasarkan komponen struktur dalaman dan seni bina sistem analisis sentimen Twitter.

Proses pengujian akan dijalankan sebanyak lima kali. Pengujian pertama dijalankan dengan memilih 1200 data *tweets* sebagai data latihan dan 300 data *tweets* sebagai data ujian. Pengujian yang seterusnya dilakukan dengan menambah 150 data *tweets* sebagai data latihan dan 50 data *tweets* sebagai data ujian ke dalam dataset yang digunakan sebagai ujian sebelumnya. Butiran komposisi setiap data latihan dan data ujian untuk setiap proses pengujian dapat diperolehi di jadual 5.1.

Jadual 1 Butiran komposisi setiap data latihan dan data ujian untuk setiap proses pengujian.

Nombor pengujian	Bilangan Data Latihan	Bilangan Data Ujian	Pemadanan Sentimen Positif	Pemadanan Sentimen Negatif	Ketepatan (%)
1	900	300	175	53	0.760
2	1050	350	214	66	0.767
3	1200	400	236	74	0.776
4	1350	450	252	93	0.785
5	1500	500	290	107	0.809
Nilai Purata:					0.7794

#

proses pengumpulan data sehingga terkumpul data selain daripada UKM, Bangi. Dengan ini, terdapat beberapa aspek yang dapat diperbaiki dan ditambahbaikkannya supaya dapat menghasilkan sistem yang lebih efisien pada masa akan datang. Terdapat satu cadangan bagi para pengkaji yang bakal menggunakan Studio R semasa dalam proses pengumpulan data. Pengkaji digalakkan menggunakan kata kunci yang spesifik dengan disertai koordinat lokasi bagi meningkatkan lagi ketepatan data yang dikumpulkan.

Antara cadangan penambahbaikan sistem adalah dengan menambahkan fungsi yang membenarkan pengguna memuatnaikkan data yang dikehendaki. Pengguna dapat melakukan analisis sentimen berdasarkan data yang dimuatnaik melalui sistem ini bagi mendapatkan hasil keputusan kajian yang berkait rapat. Selain itu, pengkaji juga mencadangkan supaya analisis sentimen Twitter dapat dilakukan dalam masa nyata iaitu sistem ini perlu dilengkapi dengan fungsi dari proses pengumpulan data hingga paparan hasil keputusan. Cadangan yang seterusnya adalah mempraktikkan teknik pembelajaran mesin bagi menghasilkan keputusan yang lebih bagus dan efektif.

6 KESIMPULAN

Pengkaji telah berjaya memenuhi kedua-dua objektif kajian seperti yang awal dicadangkan. Beberapa kelebihan dan kelemahan sistem analisis sentimen Twitter mengenai peristiwa-peristiwa penting yang berlaku di sekitar UKM, Bangi telah dikenal pasti dan diterangkan dalam bab ini. Selain itu, cadangan penambahbaikan juga diusulkan bagi membantu penghasilan sistem yang lebih sempurna pada masa akan datang.

Secara keseluruhannya, sistem analisis sentimen Twitter mengenai peristiwa-peristiwa penting yang berlaku di sekitar UKM, Bangi berjaya dibangunkan dan menepati objektif kajian. Kajian ini diharapkan dapat membantu pihak UKM bagi mengenal pasti keseluruhan sentimen terhadap peristiwa-peristiwa penting yang berlaku sejak kebelakangan ini. Kajian ini juga merupakan satu alternatif bagi membantu mana-mana pihak agar dapat mendapatkan maklumat yang diperlukan dan mampu mengenal pasti analisis sentimen melalui penggunaan sistem ini.

#

7 RUJUKAN

Matt Kiser. 2016. A Beginner's Guide to Sentiment Analysis. <https://medium.com/@mattkiser/a-beginner-s-guide-to-sentiment-analysis-888390a8085a> [28 September 2018].

MonkeyLearn Inc. 2018. The Basics of Sentiment Analysis. <https://monkeylearn.com/sentiment-analysis/> [14 Oktober 2018].

Pradeep Palani. 2018. Understanding Semantic Analysis (And Why This Title is Totally Meta). <https://zetaglobal.com/blog-posts/understanding-semantic-analysis-title-totally-meta/> [27 September 2018].

Shashank Gupta. 2018. Sentiment Analysis: Concept, Analysis and Applications. <https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17> [14 Oktober 2018].

#

#