

ANALISIS SENTIMEN BAGI SEKTOR PELANCONGAN DI MALAYSIA

GOH JOEN KEN

MOHD RIDZWAN YAAKUB

*Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia, 43600 UKM Bangi,
Selangor Darul Ehsan, Malaysia*

ABSTRAK

Sektor pelancongan di Malaysia telah menjadi penyumbang penting kepada ekonomi negara dan sumber kebanggaan negara selama bertahun-tahun. Kajian ini meneroka analisis sentimen dalam konteks sektor pelancongan di Malaysia. Objektif kajian ini adalah untuk membentangkan keputusan empirik analisis sentimen untuk mengekstrak beberapa aspek perkhidmatan yang berguna untuk memanfaatkan pembangunan industri pelancongan berdasarkan maklum balas pengguna. Ini memberikan alasan untuk mendakwa bahawa berdasarkan teknologi yang dicadangkan adalah mungkin untuk mengesan maklum balas pengguna perkhidmatan pelancong dan untuk merumuskan cadangan untuk firma pelancongan untuk penambahbaikan aktiviti mereka. Ini bertujuan untuk meningkatkan industri pelancongan. Kajian menggunakan data semakan pelanggan daripada tapak web TripAdvisor untuk membangunkan model klasifikasi menggunakan algoritma pembelajaran mesin dan menggunakan teknik pemprosesan bahasa tabii.

PENGENALAN

Sektor pelancongan Malaysia berdiri sebagai tonggak penting ekonomi negara, mempesonakan pelancong dengan warisan budayanya yang kaya. Menurut laporan baru-baru ini, pelancongan menyumbang 2.7% yang ketara kepada KDNK Malaysia, menonjolkan kepentingannya yang tidak dapat dinafikan kepada kemakmuran kewangan negara (Statista, 2023b). Walau bagaimanapun, dalam era ketersambungan digital ini, sentimen pengembara memainkan peranan penting dalam membentuk nasib industri. Dalam konteks ini, penyelidikan kami memulakan penerokaan menyeluruh tentang "Analisis Sentimen untuk Sektor Pelancongan di Malaysia." Dengan peningkatan platform dalam talian dan media sosial, industri pelancongan semakin dipengaruhi oleh suara digital pelancong, baik tempatan mahupun antarabangsa (Deraman et al., 2021). Namun begitu, kekayaan data dalam talian ini sebahagian besarnya masih belum diterokai, menimbulkan jurang penyelidikan yang menarik yang ingin dicapai oleh projek kami. Isu teras yang kami tangani ialah kekurangan analisis sistematik mengenai sentimen orang ramai dalam sektor pelancongan Malaysia. Apabila pengembara berkongsi pengalaman, ulasan dan cadangan mereka dalam talian, sentimen mereka boleh sama ada menguatkan atau menghakis daya tarikan industri. Projek kami

berusaha untuk membedah paduan suara digital ini, mendedahkan arah aliran, sentimen dan pandangan yang boleh memaklumkan pihak berkepentingan, penggubal dasar dan pemain industri. 2 Objektif utama penyelidikan ini merangkumi memahami sentimen pelancong terhadap pelbagai aspek sektor pelancongan Malaysia, memantau persepsi orang ramai yang berkembang dari semasa ke semasa, dan mengenal pasti kawasan yang berpotensi untuk diperbaiki. Kami akan menggunakan teknik Pemprosesan Bahasa Semulajadi (NLP) terkini, memanfaatkan sejumlah besar data yang tersedia pada platform media sosial dan tapak web pelancongan. Kajian ini bukan sekadar latihan akademik, ia mempunyai implikasi yang mendalam kepada pihak berkepentingan sektor pelancongan. Dengan memberikan pandangan tentang sentimen awam, kami menyasarkan untuk memperkasakan pembuat keputusan dan pemain industri dengan pengetahuan yang diperlukan untuk meningkatkan tawaran sektor, strategi pemasaran dan pengalaman pelawat. Akhirnya, penyelidikan kami berhasrat untuk mengukuhkan kedudukan Malaysia sebagai destinasi pelancongan global. Dalam bahagian seterusnya, kami akan mendalami metodologi, sumber data, teknik analisis dan membentangkan penemuan analisis sentimen kami. Melalui penemuan ini, kami berusaha untuk menjelaskan hubungan rumit antara sentimen awam dan sektor pelancongan Malaysia.

METODOLOGI KAJIAN

Spesifikasi keperluan system dalam membangunkan model Analisis Sentiment bagi Sector Pelancongan di Malaysia dengan menggunakan pembelajaran mesin akan dinyatakan dalam bab tersebut. Spesifikasi keperluan system merupakan suatu dokumen yang mengandungi spesifikasi keperluan yang diperlukan untuk membangunkan model tersebut dan dokumen ini akan menjadi sesuatu rujukan yang penting dalam sepanjang tempoh kajian ini berlangsung kerana ianya dapat menerangkan keperluan kepada kajian secara terperinci. Perkakasan dan perisian yang digunakan dalam kajian ini akan dinyatakan dengan jelas dan proses-proses semasa menjalankan kajian ini seperti pra-pemprosesan data, set data, algoritma dan pembelajaran mesin. Hal ini kerana maklumat ini amat penting untuk membangunkan model kepada kajian dan juga dapat menggambarkan keseluruhan projek. Seterusnya, spesifikasi keperluan system merupakan suatu dokumen yang mengandungi spesifikasi keperluan yang diperlukan untuk membangunkan model tersebut dan dokumen ini akan menjadi sesuatu rujukan yang penting dalam sepanjang tempoh kajian ini berlangsung kerana ianya dapat menerangkan keperluan kepada kajian secara terperinci. 21 Dalam bab ini, perkakasan dan perisian yang digunakan dalam kajian ini akan dinyatakan dengan jelas dan proses-proses semasa menjalankan kajian ini seperti prapemprosesan data, set data, algoritma dan pembelajaran mesin. Hal ini kerana maklumat ini amat penting untuk membangunkan model kepada kajian dan juga dapat menggambarkan keseluruhan projek. Spesifikasi reka bentuk bertujuan untuk mengenalpasti dan menentukan proses komponen, modul, antara muka dan pangkalan data untuk sistem supaya dapat melengkapkan spesifikasi keperluan yang telah ditetapkan. Spesifikasi reka bentuk melibatkan pemodelan data yang diperolehi bagi memudahkan pergerakan data dalam sistem. Oleh itu, pembangunan sistem boleh menjadi lebih sistematik dan mengikut sistem yang telah ditetapkan. Bab ini akan mengenalpasti proses reka bentuk supaya kajian tersebut akan dijalankan dengan lancar dan sistematik. Proses reka bentuk adalah reka bentuk senibina, pra-pemprosesan data, pangkalan data, set data, reka bentuk antara muka dan reka bentuk algoritma. Proses reka bentuk tersebut amat

penting untuk dapat menggambarkan keseluruhan kajian dan perlu dirangcangkan supaya dapat keputusan yang amat sesuai untuk membangun model pengelasan analisis sentimen.

Fasa pengumpulan data

Fasa pengumpulan data merupakan fasa yang perlu dilakukan dalam kajian Analisis Sentimen Sektor pelancongan di Malaysia. Dengan menggunakan pengikisan web sebagai kaedah utama untuk mengumpul data, memfokuskan pada mengekstrak maklumat berharga daripada tapak web TripAdvisor. Pengikisan web ialah teknik yang melibatkan pengekstrakan automatik data daripada halaman web, dan dalam konteks ini, ia berfungsi sebagai cara yang cekap untuk mengumpulkan ulasan dan penilaian yang dijana pengguna. Memanfaatkan bahasa pengaturcaraan Python dan perpustakaan yang berkaitan seperti Selenium untuk menavigasi melalui tapak web TripAdvisor, secara sistematik mendapatkan data yang berkaitan dengan ulasan pengguna, penilaian dan butiran lain yang berkaitan dengan entiti minat tertentu. Proses pengumpulan data yang komprehensif daripada TripAdvisor, platform popular untuk ulasan berkaitan perjalanan, meletakkan asas untuk analisis dan cerapan seterusnya tentang sentimen dan pilihan pengguna dalam domain yang dipilih. Inilah kodnya untuk mengikis web bagi tapak web TripAdvisor

Fasa pembersihan data

Fasa Pembersihan Data merupakan fasa yang amat penting dalam kajian ini. Set data yang diekstrak melalui TripAdvisor mempunyai data yang tiada berkaitan dengan kajian, format data yang salah dan tidak lengkap. Terdapat pengguna tapak web menggunakan perkataan yang bersingkat untuk berkongsi perasaan mereka di tapak web seperti 'tmr' as tomorrow, 'b4' as before, 'ady' as already dan lain-lain. Selain itu, terdapat pengguna tapak web juga menyalah ejaan perkataan semasa berkongsi perasaan mereka di komen tapak web. Justeru, proses pembersihan data perlu dijalankan supaya kesalahan ini tidak memberi kesan terhadap keputusan akhir kajian. Data yang mempunyai nombor, aksara khas, symbol yang tidak mempunyai maksud dan tidak berkaitan dengan tajuk kajian seperti emoji yang terdapat dalam data dengan bentuk teks, iaitu '=', '(', '==||', dan lain-lain yang perlu dikeluarkan supaya model dapat mengelaskan sentiment dengan menggunakan masa yang tersingkat.

Fasa pra-pemprosesan teks

Fasa pra-pemprosesan teks adalah fasa yang penting semasa menjalankan kajian ini. Ia terdapat beberapa fasa yang berlangsung semasa proses pra-pemprosesan teks berlaku. Semua fasa yang dilakukan di dalam proses ini bertujuan untuk membersihkan data yang terdapat di komen laman web supaya dapat mengurangkan kerumitannya untuk tugas pengelasan sentiment dan mengurangkan masa yang diambil oleh model klasifikasi sentiment untuk menganalisis data. Pada fasa pertama pra-pemprosesan teks adalah tokenisasi, semua ayat yang berada di dalam set data akan dibahagikan menjadi perkataan tunggal dan perkataan tersebut juga dipanggil sebagai token. Tujuan membahagikan ayat kepada token adalah dapat mengabaikan ruang putih antara 2 perkataan supaya boleh meneruskan fasa seterusnya dengan cepat dan tepat. Kemudian adalah penapisan kata henti. Ini adalah fasa yang amat penting kerana ia dapat menyinkirkan dan membersihkan perkataan yang tidak membawa maksud tetapi mempunyai kekerapan yang tinggi di dalam data seperti perkataan 'a', 'an', 'the', dan lain-lain supaya perkataan tersebut tidak dapat mempengaruhi ketepatan kepada keputusan semasa menjalankan klasifikasi sentiment nanti. Penandaan POS 26 Seterusnya

adalah fasa normalisasi, fasa ini dijalankan untuk mengubahkan perkataan kepada perkataan yang standard. Contohnya, perkataan seperti 'baddddd' atau 'goooooooodddd' perlu digantikan dengan 'bad' dan 'good'. Inilah perkataan yang perlu diperhatikan kerana perkataan tersebut mempunyai maksud yang amat penting dan pengaruh yang kuat kepada keputusan kepada klasifikasi sentiment. Tambahan pula, fasa lemmatisasi perlu dijalankan untuk menukarkan perkataan yang mengandungi imbuhan kepada perkataan asal. Hal ini kerana perkataan yang pelbagai jenis sukar untuk mengelaskan walaupun perkataan mempunyai maksud yang sama dan perkataan yang pelbagai jenis juga akan menyebabkan model klasifikasi sukar untuk memahami semasa menjalankan klasifikasi sentiment. Akhirnya, fasa yang terakhir adalah Penandaan POS, semasa fasa Penandaan POS, setiap perkataan atau dikenali sebagai token akan ditandai dengan penandaan tatabahasa seperti kata nama, kata kerja, kata sifat dan lain-lain supaya dapat mengetahui bahawa sentiment kepada perkataan tersebut.

Fasa klasifikasi sentiment

Dalam fasa klasifikasi sentiment ini, terdapat beberapa jenis algoritma pembelajaran mesin sering digunakan untuk menganalisis data. Dalam analisis sentiment, teknik klasifikasi kepada algoritma pembelajaran mesin yang sering digunakan secara luas adalah Naïve Bayes, Support Vector Machine(SVM), Random Forest, dan logistic Regression. Dalam kajian ini, keempat-empat pengelasan kepada algoritma pembelajaran mesin akan digunakan. Tujuan menggunakan keempat-empat algoritma pembelajaran mesin adalah untuk melakukan perbandingan antara model dari segi ketepatan dan prestasi supaya dapat menentukan model yang paling sesuai dalam kajian tersebut walaupun keempat-empat algoritma biasa digunakan dalam tugas klasifikasi.

KEPUTUSAN DAN PERBINCANGAN

Penyelidik bertujuan untuk membandingkan prestasi semua model Sentimen Analisis Bagi Sektor Pelancongan di Malaysia iaitu Naive Bayes, Hutan Rawak, dan SVM yang digabungkan dengan TFIDF, BOW dan N-gram masing masing dan model ensemble. Untuk menilai model dengan tepat, penyelidik mempertimbangkan nisbah pembahagian data latihan dengan 80-20 lalu menjalankan penyiasatan terhadap kesan penalaan hiperparameter pada prestasi model, membandingkan keputusan mereka selepas penalaan. Bermula dengan model yang lebih ringkas, pengelas Naive Bayes, apabila digabungkan dengan teknik pengekstrakan ciri yang berbeza, menunjukkan hasil yang berbeza-beza. Pendekatan TFIDF, yang mengambil kira kepentingan perkataan dalam konteks set dokumen, mencapai ketepatan 57%. Ini menunjukkan bahawa walaupun TFIDF menangkap perkaitan istilah dengan baik, ia mungkin tidak mencukupi sendiri untuk klasifikasi sentiment dalam konteks ini. Sebaliknya, pendekatan BOW dan Ngram, yang masing-masing memfokuskan pada kekerapan perkataan dan urutan perkataan, kedua-duanya mencapai ketepatan 65%. Ini menunjukkan bahawa kaedah ini lebih baik untuk menangkap teks yang menunjukkan sentiment, walaupun kaedah ini masih kurang daripada model berprestasi tinggi. Apabila beralih ke pengelas Hutan Rawak, terdapat peningkatan yang ketara dalam ketepatan untuk semua teknik pengekstrakan ciri. TFIDF digabungkan dengan Hutan Rawak mencapai ketepatan 67%, mengatasi sedikit

kaedah BOW dan N-gram, yang masing-masing mencapai 66% dan 65%. Peningkatan ini menyerlahkan keupayaan Hutan Rawak untuk mengendalikan set ciri yang pelbagai dan menangkap corak kompleks dalam data, menjadikannya lebih berkesan daripada Naive Bayes untuk tugas ini. Prestasi yang paling ketara diperhatikan dengan pengelas SVM. Kedua-dua kaedah TFIDF dan BOW mencapai ketepatan 72%, menunjukkan peningkatan yang ketara berbanding model sebelumnya. Ini mencerminkan keupayaan SVM untuk mencipta sempadan keputusan yang mantap, memisahkan kelas dengan berkesan walaupun dalam ruang berdimensi tinggi. Terutama, kaedah N-gram yang digabungkan dengan SVM mencapai ketepatan tertinggi antara semua model yang diuji, pada 74%. Ini menunjukkan bahawa pendekatan N-gram, yang menangkap jujukan perkataan dan dengan itu lebih banyak konteks, apabila dipasangkan dengan keupayaan pengelasan SVM yang berkuasa, amat sesuai untuk analisis sentimen dalam domain ini. Akhirnya, model ensemble yang menggabungkan Memori Jangka Pendek Panjang (LSTM), Unit Berulang Berpagar (GRU) dan Rangkaian Neural Konvolusi (CNN) mencapai ketepatan 71%. Walaupun ini adalah prestasi yang kukuh, ia lebih rendah sedikit daripada model SVM yang berprestasi terbaik. Pendekatan ensemble ini memanfaatkan kekuatan seni bina pembelajaran mendalam, menangkap kebergantungan temporal dan ciri dalam teks. Walaupun kerumitan dan keupayaannya untuk mengendalikan corak yang rumit, ia tidak mengatasi ketepatan SVM yang lebih mudah dengan model N-gram. Kesimpulannya, pengelas SVM yang digabungkan dengan kaedah pengekstrakan ciri N-gram muncul sebagai pendekatan paling berkesan untuk analisis sentimen dalam sektor pelancongan Malaysia, mencapai ketepatan tertinggi sebanyak 74%. Ini menekankan kepentingan kedua-dua menangkap jujukan perkataan menggunakan algoritma pengelasan yang berkuasa. Walaupun model pembelajaran mendalam ensemble juga menunjukkan prestasi yang baik, keputusan menunjukkan bahawa untuk tugas khusus ini, gabungan kaedah pembelajaran mesin tradisional dengan perwakilan teks yang sesuai boleh menghasilkan hasil yang lebih baik.

Jadual 4.11 Jadual Perbandingan antara Model

Jadual Perbandingan	
Model	Ketepatan (%)
Naïve Bayes + TFIDF	57
Naïve Bayes + BOW	65
Naïve Bayes + N-Gram	65
Hutan Rawak + TFIDF	67
Hutan Rawak + BOW	66
Hutan Rawak + N-Gram	65
SVM + TFIDF	72
SVM + BOW	72
SVM + N-Gram	74
Model Ensembl (LSTM + GRU + CNN)	71

Kekuatan dan Kekangan

Kekuatan projek ini terletak pada penilaian komprehensif berbilang model pembelajaran mesin dan nisbah pembahagian data latih-uji, memberikan cerapan berharga tentang prestasi pada model ini. Selain itu, projek ini menggunakan teknik visualisasi siri masa, yang membolehkan penerokaan arah aliran dan corak harga yang jelas dan intuitif. Kejayaan SVM digabungkan dengan n-gram dan ensembel model dengan mencapai ketepatan yang tinggi menunjukkan keteguhan mereka dalam menangkap corak yang kompleks dalam bentuk teks. Walau bagaimanapun, batasan tersebut termasuk prestasi rendah model lain, menunjukkan keperluan untuk kaji lanjut untuk meningkatkan keberkesannya. Satu lagi had projek ini ialah masa pengiraan lanjutan yang diperlukan untuk latihan dan penalaan hiperparameter model pembelajaran mesin, terutamanya apabila bergantung pada sumber seperti Google Colab. Disebabkan oleh kekangan kuasa pengiraan dan memori pada platform berasaskan awan, proses itu boleh memakan masa, menghadkan keupayaan untuk meneroka ruang carian hiperparameter yang lebih besar atau ensembel model yang lebih kompleks sepenuhnya.

Cadangan Penambahbaikan

Untuk meningkatkan keteguhan dan ketepatan analisis sentimen terhadap sektor pelancongan di Malaysia, beberapa bidang utama untuk penambahbaikan telah dikenal pasti. Pertama, menangani ketidakseimbangan dalam set data adalah penting. Set data semasa mengandungi bilangan ulasan positif yang tidak seimbang berbanding dengan ulasan neutral dan negatif. Ketidakseimbangan ini boleh memesonkan keputusan, menjadikan analisis sentimen kurang mewakili taburan pendapat sebenar. Kerja masa depan harus bertujuan untuk mengumpul set data yang lebih seimbang dengan menggabungkan ulasan yang lebih neutral dan negatif. Satu lagi bidang untuk penambahbaikan ialah tidak konsisten dalam panjang ulasan dan ketepatan label. Set data ini menunjukkan kecenderungan untuk ulasan positif menjadi jauh lebih panjang daripada yang neutral atau negatif. Percanggahan panjang ini boleh menimbulkan berat sebelah dalam analisis sentimen, kerana ulasan yang lebih panjang mungkin mengandungi maklumat yang lebih bernuansa yang condong ke arah positif. Untuk mengurangkan ini, penyelidikan masa depan harus menyeragamkan panjang ulasan dengan sama ada memotong ulasan yang terlalu panjang atau mengembangkan yang lebih pendek melalui teknik parafrasa. Selain itu, meningkatkan ketepatan label sentimen adalah penting. Pengesahan manual atau kaedah separa automatik yang menggabungkan pertimbangan manusia dengan pembelajaran mesin boleh dilaksanakan untuk memastikan bahawa label menggambarkan dengan tepat sentimen yang dinyatakan dalam ulasan. Akhir sekali, cabaran untuk mengendalikan pelbagai bahasa dan memastikan terjemahan yang tepat mesti ditangani. Proses semasa menterjemah ulasan bahasa asing ke dalam Bahasa Inggeris kadangkala boleh mengubah makna asal, yang membawa kepada salah klasifikasi sentimen. Untuk mengatasinya, kerja masa hadapan harus mempertimbangkan untuk menggunakan model terjemahan yang lebih canggih yang lebih baik dalam memelihara konteks dan makna. Sebagai alternatif, analisis sentimen boleh dijalankan dalam bahasa asal menggunakan model NLP berbilang bahasa yang mampu memproses pelbagai bahasa tanpa memerlukan terjemahan. Pendekatan ini 74 bukan sahaja akan meningkatkan ketepatan tetapi juga mengekalkan nuansa budaya dan konteks teks asal.

KESIMPULAN

Kesimpulannya, projek ini telah berjaya mencapai objektifnya untuk menyiasat Analisis Sentimen terhadap Sektor Pelancongan di Malaysia menggunakan pelbagai model pembelajaran mesin dan meneroka kesan nisbah pembahagian data latih-uji yang berbeza. Penilaian komprehensif model ensembel dan SVM digabungkan dengan ngram memberikan cerapan berharga tentang prestasinya. Selain itu, projek ini juga menyerlahkan cabaran yang ditimbulkan oleh batasan pengiraan dalam platform berasaskan awan dan keperluan untuk penalaan hiperparameter yang meluas. Bagi penyelidik masa depan, projek ini berfungsi sebagai asas akademik yang kukuh untuk Analisis Sentimen terhadap Sektor Pelancongan di Malaysia, menawarkan panduan tentang pemilihan model, dan pengoptimuman hiperparameter untuk meningkatkan keupayaan ramalan.

PENGHARGAAN

Segala puji dan syukur bagi Allah, Yang Maha Pengasih lagi Maha Penyayang. Saya dengan rendah hati menadah tangan tanda kesyukuran, terharu kerana telah menyiapkan laporan usulan saya dengan bimbingan dan rahmat yang dikurniakan kepada saya. Saya ingin merakamkan setinggi-tinggi penghargaan kepada penyelia saya, Assoc. Prof Dr. Mohd Ridzwan Yaakub, atas bimbingan, sokongan dan dorongan sepanjang setahun sebagai penyelia laporan usulan saya. Kepakaran dan pandangan beliau yang tidak ternilai telah memainkan peranan penting dalam membentuk hala tuju dan hasil projek saya. Saya amat berterima kasih atas masa dan usaha yang dia telah laburkan untuk saya dan kerja saya. Saya tidak akan dapat menyiapkan laporan usulan ini tanpa bimbingan beliau. Terima kasih, Assoc. Prof Dr. Mohd Ridzwan Yaakub. Saya juga ingin mengucapkan terima kasih juga kepada para pensyarah dari FTSM yang telah menaburkan ilmu pengetahuan kepada saya sepanjang pengajian saya di Universiti Kebangsaan Malaysia (UKM). Saya juga ingin merakamkan ribuan terima kasih kepada ahli keluarga saya kerana memberi sokongan dan semangat serta sentiasa berdoa terhadap kejayaan saya di universiti. Akhirnya, terima kasih kepada semua yang terlibat secara langsung atau tidak langsung dalam menghasilkan laporan usulan ini. Saya juga ingin memohon maaf sekiranya terdapat kesilapan sepanjang pelaksanaan laporan usulan ini.

RUJUKAN

- Statista. 2023b. Tourism industry direct contribution as share of GDP Malaysia 2013-2022.
- Deraman, N. A., Buja, A. G., Samah, K. a. F. A., Jono, M. N. H. H., Isa, M. a. M. & Saad, S. 2021. A Social Media Mining Using Topic Modeling and Sentiment Analysis on Tourism in Malaysia During Covid19. IOP Conference Series: Earth and Environmental Science 704(1): 012020.
- Mishra, R. K., Urolagin, S., Jothi, J. a. A., Neogi, A. S. & Nawaz, N. 2021. Deep Learning-based Sentiment Analysis and Topic Modeling on Tourism During Covid-19 Pandemic. Frontiers in Computer Science 3(
- Manosso, F. C., & Domareski Ruiz, T. C. 2021. Using sentiment analysis in tourism research: A systematic, bibliometric, and integrative review. *Journal of Tourism, Heritage & Services Marketing*, 7(2)(16-27.
- Puh, K. & Bagić Babac, M. 2023. Predicting sentiment and rating of tourist reviews using machine learning. *Journal of Hospitality and Tourism Insights* 6(3): 1188-1204.
- Tan, C.-Y. & Soon, S.-V. 2023. Tourism demand for Malaysia: Further evidence from panel approaches. *Asia Pacific Management Review* 28(4): 459-469.
- Anon. 2023. Hotel reviews sentiment analysis using machine learning. *International Research Journal of Modernization in Engineering Technology and Science*
- Bahrawi, N. 2019. Sentiment Analysis Using Random Forest Algorithm-Online Social Media Based. *Journal of Information Technology and Its Utilization* 2(2): 29-33.
- Dina, N. Z. 2020. Tourist sentiment analysis on TripAdvisor using text mining: A case study using hotels in Ubud, Bali. *African Journal of Hospitality, Tourism and Leisure* 9(2): 1-10.
- Farisi, A., Sibaroni, Y. & Faraby, S. 2019. Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier. *Journal of Physics: Conference Series* 1192(012024).
- Gupta, G. & Gupta, P. 2019. Twitter Mining for Sentiment Analysis in Tourism Industry. 2019 Third World Conference on Smart Trends in Systems Security and Sustainability (WorldS4), hlm. 302-306.
- Karthika, P., Murugeswari, R. & Manoranjithem, R. 2019. Sentiment Analysis of Social Media Network Using Random Forest Algorithm. 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), hlm. 1-5.
- Laksono, R. A., Sungkono, K. R., Sarno, R. & Wahyuni, C. S. 2019. Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes. 2019 12th International Conference on Information & Communication Technology and System (ICTS), hlm. 49-54.

- Nayoan, R. a. N., Hidayatullah, A. F. & Fudholi, D. H. 2021. Convolutional Neural Networks for Indonesian Aspect-Based Sentiment Analysis Tourism Review. 2021 9th International Conference on Information and Communication Technology (ICoICT), hlm. 60-65.
- Salur, M. U., Ī, A. & Alghrsi, S. A. 2019. SmartSenti: A Twitter-Based Sentiment Analysis System for the Smart Tourism in Turkey. 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), hlm. 1-5.
- Wongkar, M. & Angdresey, A. 2019. Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter. 2019 Fourth International Conference on Informatics and Computing (ICIC), hlm. 1-5.

Goh Joen Ken (A190479)
Assoc. Prof Dr. Mohd Ridzwan Yaakub
Fakulti Teknologi & Sains Maklumat
Universiti Kebangsaan Malaysia

Copyright@FTSM
UKM