

ANALISIS SENTIMEN BERDASARKAN ASPEK DALAM ULASAN FILEM

LEE JIA YEE

SABRINA TIUN

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

ABSTRAK

Pada masa kini, terdapat banyak filem dan siri drama yang tersedia di saluran berbayar. Daripada pelbagai pilihan yang ada, memilih apa yang hendak ditonton menjadi satu tugas yang sukar. Analisis sentimen tradisional sering kali mengabaikan butiran halus mengenai pendapat benar penonton tentang filem. Analisis Sentimen Berdasarkan Aspek (ABSA) muncul sebagai satu penyelesaian kepada masalah ini. Objektif projek ini adalah untuk mengekstrak dan menilai sentimen penonton berdasarkan aspek-aspek tertentu, membolehkan penonton untuk memilih filem dengan memahami aspek tertentu yang disukai atau tidak disukai. Analisis sentimen berdasarkan aspek (ABSA) adalah sejenis teknik Pemprosesan Bahasa Tabii yang digunakan untuk menentukan sentimen berdasarkan aspek-aspek tertentu dalam sesuatu produk, perkhidmatan atau topik dalam teks. Projek ini menggabungkan pengetahuan daripada kedua-dua teknik Pemprosesan Bahasa Tabii (NLP) dan pembelajaran mesin (ML). Teknik web scraping akan digunakan untuk mengumpulkan data ulasan filem bagi memastikan analisis yang komprehensif. Kemudian, teknik pra pemprosesan teks akan digunakan untuk membersihkan data. Teras ABSA adalah mengenal pasti dan mengekstrak aspek atau entiti tertentu dari teks. Aspek merujuk kepada unsur atau ciri yang sedang dibincangkan dalam sesuatu topik. Polariti sentimen (positif atau negatif) akan ditentukan untuk setiap aspek yang disebut dalam ulasan. Model pembelajaran mesin akan digunakan untuk meramalkan sentimen aspek yang diekstrak. Pada akhir projek ini, satu model *ensemble* ABSA yang mampu menganalisis ulasan filem dan memberikan visual yang jelas mengenai aspek-aspek tertentu yang mempengaruhi sentimen penonton akan dibangunkan. Ramalan sentimen ini ditunjukkan melalui antaramuka.

Kata Kunci: Analisis Sentimen Berdasarkan Aspek, Pemrosesan Bahasa Tabii, Pembelajaran Mesin

PENGENALAN

Kemunculan pelbagai platform penstriman filem seperti Amazon Movies, Netflix, dan HBO Max membolehkan pengguna berkongsi pendapat mereka dalam pelbagai format dengan cekap. Ini termasuk ulasan dalam format seperti teks dan sistem penilaian yang menggambarkan pelbagai sentimen dari positif hingga negatif. Laman web penilaian popular seperti IMDB, Rotten Tomatoes dan Metacritic menjadi tempat pengguna berkongsi ulasan dan penilaian mereka (Stegner 2020). Ulasan filem, sama ada daripada pengkritik profesional atau pengguna biasa, berperanan sebagai sumber maklumat penting bagi penonton dalam membentuk pilihan dan jangkaan mereka.

Sentimen ialah perasaan atau pendapat, terutamanya berdasarkan emosi (Oxford Learner University). Analisis sentimen, juga dikenali sebagai perlombongan pendapat adalah proses mengenal pasti dan mengekstrak pendapat dan sentimen daripada data teks. Ia digunakan secara meluas dalam pelbagai domain seperti media sosial, perdagangan elektronik, dan ulasan filem. Analisis sentimen menggabungkan teknik pembelajaran mesin dan pemrosesan bahasa tabii (PBT) (Onalaja et al. 2021). Dalam konteks ulasan filem, analisis sentimen dapat digunakan untuk menilai sentimen keseluruhan ulasan yang disampaikan oleh pengguna terhadap filem. Ia membantu mengenal pasti aspek positif dan negatif filem serta memberikan gambaran keseluruhan tentang filem. Oleh itu, analisis sentimen ulasan filem boleh dianggap sebagai penyelidikan pasaran untuk industri hiburan (Wang et al. 2020).

Kebanyakan kajian lepas hanya memberi tumpuan kepada analisis sentimen secara keseluruhan dalam ulasan filem. Oleh itu, projek ini bertujuan untuk membangunkan model Analisis Sentimen Berasaskan Aspek (ABSA) untuk mengekstrak dan mengategorikan sentimen yang berkaitan dengan aspek tertentu filem. ABSA merupakan satu pendekatan yang lebih halus dalam analisis. Pelaksanaan ABSA dalam ulasan filem membantu penonton mencari filem yang sejajar dengan kegemaran mereka. Hal ini kerana ABSA dapat menganalisis aspek-aspek dalam ulasan filem seperti plot, lakonan, skrip dan muzik secara terpisah. Sebagai contoh, penonton mungkin teruja dengan aksi dalam filem tetapi kecewa dengan perkembangan plotnya. Situasi ini jelas menunjukkan wujudnya sentimen campuran dalam ulasan filem. Dengan menganalisis elemen-elemen ini, penonton memperoleh

pemahaman yang lebih mendalam tentang kelebihan atau kelemahan spesifik dalam filem. Manfaat pelaksanaan ABSA tidak hanya terhad kepada penonton. Malah, pihak industri filem juga dapat memahami aspek mana yang menerima pujian atau kritikan supaya dapat meningkatkan prestasi mereka dalam produksi filem. Perlombongan pendapat dalam ulasan filem telah terbukti sangat penting dan memberi manfaat kepada pengeluar filem dan pengguna (Yadav et al. 2021).

METODOLOGI KAJIAN

Metodologi proses pembangunan yang digunakan dalam projek ini ialah *Machine Learning Life Cycle* atau kitaran hayat pembelajaran mesin. Kaedah ini merupakan satu proses kitaran di mana setiap fasa perlu diikuti apabila menjalankan projek. Metodologi ini merangkumi tujuh fasa iaitu pengumpulan data (*Gathering data*), penyediaan data (*Data Preparation*), pra-pemprosesan data (*Data Wrangling*), analisis data (*Analyse Data*), melatih model (*Train Model*), menguji model (*Test Model*), dan penyebaran model (*Deployment*). Metodologi ini dipilih disebabkan oleh kesesuaian dengan projek ini. Hal ini kerana projek ini fokus pada penggunaan algoritma pembelajaran mesin, di mana kebanyakan fasa berkaitan dengan pemprosesan data dan pembinaan model. Keunggulan metodologi ini dalam mengintegrasikan kitaran hidup pembelajaran mesin dari pengumpulan data hingga penilaian model memberikan landasan yang kukuh bagi keseluruhan pelaksanaan projek ABSA dalam ulasan filem. Selain itu, metodologi ini mudah difahami dan dilaksanakan bagi projek ini. Dalam konteks pembangunan model pembelajaran mesin, kitaran ini penting kerana memberi fokus kepada penggunaan hasil model dan penilaian untuk menambahbaikkan set data. Set data yang berkualiti tinggi adalah cara paling efektif untuk melatih model yang berkualiti tinggi (Hofesmann 2021). Dengan ini, projek yang teratur dan efektif akan dihasilkan.



Rajah 1 Fasa-fasa Metodologi Kitaran Hayat Pembelajaran Mesin (Jaiswal 2021)

Fasa Pengumpulan Data, Penyediaan Data dan Pra-pemrosesan Data

Data akan dikumpulkan dari *Rotten Tomatoes* menggunakan teknik *web scraping*. Semua data yang diekstrak daripada *Rotten Tomatoes* adalah mentah dan belum menjalani sebarang proses teks. Oleh itu, pra-pemrosesan teks perlu dilakukan untuk membersihkan data hingar dan menjadikannya sesuai untuk analisis. Langkah ini termasuk terjemahan, pembersihan, dan normalisasi teks untuk memastikan konsistensi dan kualiti data yang tinggi.

Fasa Analisis Data

Dalam fasa ini, set data ulasan filem akan dianalisis untuk mengenal pasti aspek-aspek relevan seperti plot, acting, directing, music, visual, dan character menggunakan teknik Pengecaman Entiti Bernama (NER) dengan *spaCy*. Proses ini membolehkan pengekstrakan aspek-aspek yang signifikan dari setiap ulasan. Selain itu, pengekstrakan ciri menggunakan *BERT Embedding* akan dilakukan untuk menukarkan teks kepada vektor nombor, yang menangkap makna kontekstual dan semantik teks dengan lebih baik.

Fasa Melatih Model

Dalam fasa ini, model pengelas sentimen aspek akan dilatih untuk setiap aspek yang diekstrak. Dua algoritma pembelajaran mesin dicadangkan: *Artificial Neural Network (ANN)* dan *Support Vector Machine (SVM)*. ANN dipilih kerana keupayaannya untuk menangani kompleksiti dan non-lineariti data, manakala SVM dipilih sebagai perbandingan kerana kecekapan dan ketepatannya dalam pengelas sentimen.

Fasa Menguji Model

Dalam fasa ini, setiap model pengelas sentimen aspek yang telah dilatih akan dinilai menggunakan metrik ketepatan. Seterusnya, model-model ini akan digabungkan menggunakan teknik purata mudah (*simple averaging*) untuk membentuk model *ensemble* ABSA. Model *ensemble* ini kemudian akan digunakan untuk meramal sentimen keseluruhan dan sentimen berasaskan aspek bagi setiap ulasan. Prestasi model *ensemble* ABSA ini akan dinilai menggunakan pelbagai metrik penilaian termasuk ketepatan, kejituan, dapatan semula, dan skor-F1.

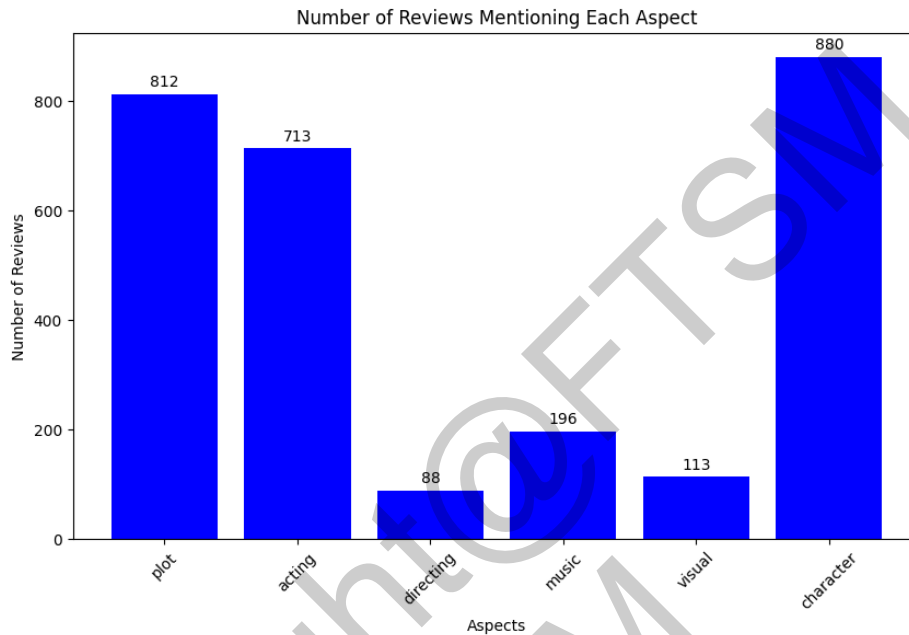
Fasa Penyebaran Model

Setelah model *ensemble* ABSA dinilai dan didapati memenuhi keperluan projek, model ini akan disebar menggunakan antaramuka *Streamlit*. Antaramuka ini membolehkan pengguna memasukkan ulasan filem dan mendapatkan ramalan sentimen bagi setiap aspek serta sentimen keseluruhan. Ini menjadikan model ABSA mudah digunakan dan dapat diakses oleh pengguna tanpa memerlukan pengetahuan teknikal mendalam. Model yang disebar akan membantu pengguna dalam memahami sentimen dalam ulasan filem dengan lebih baik.

KEPUTUSAN DAN PERBINCANGAN

Rajah 2 menunjukkan carta bar yang menunjukkan frekuensi ulasan yang menyebutkan pelbagai aspek berkaitan dengan filem. Paksi-x diberi label "Aspek" dan menyenaraikan enam kategori: *plot*, *acting*, *directing*, *music*, *visual*, *character*. Paksi-y diberi label "Bilangan ulasan," dengan nilai dari 0 hingga 900 dengan peningkatan 100.

Aspek *character* mempunyai jumlah ulasan tertinggi dengan 880 ulasan, diikuti rapat oleh *plot* dengan 812 ulasan. *Directing* mempunyai jumlah ulasan yang paling rendah iaitu sebanyak 88 ulasan sahaja. Data ini menunjukkan kebanyakan penonton cenderung membincangkan *character* dan *plot* secara mendalam manakala aspek *directing* jarang disebutkan.



Rajah 2 Bilangan Ulasan yang Menyebut Setiap Aspek

Model pengelasan sentimen aspek akan dilatih untuk meramalkan sentimen berasaskan enam aspek yang dikenal pasti. Projek ini akan menggunakan ANN untuk melatih model, dan SVM digunakan sebagai perbandingan. Kajian lepas yang dilakukan oleh Borele dan Borikar (2016) tentang analisis sentimen menunjukkan bahawa penggunaan ANN berprestasi tinggi dalam tugas pengelasan. Selain itu, SVM dipilih untuk membuat perbandingan dengan ANN kerana asas teori yang kukuh dan keberkesannya dalam ruang berdimensi tinggi (Dhiraj 2019). Prestasi kedua-dua model ini akan dibandingkan berdasarkan ketepatan masing-masing. Berdasarkan jadual 1, ANN mencapai ketepatan yang lebih tinggi berbanding dengan SVM dalam semua aspek kecuali aspek *directing*. Hal ini mungkin disebabkan oleh jumlah ulasan mengenai aspek *directing* yang terhad iaitu sebanyak 88 ulasan seperti yang ditunjukkan dalam graf sebelum ini. ANN sering memerlukan set data yang lebih besar untuk mencapai prestasi optimum (Donges 2023).

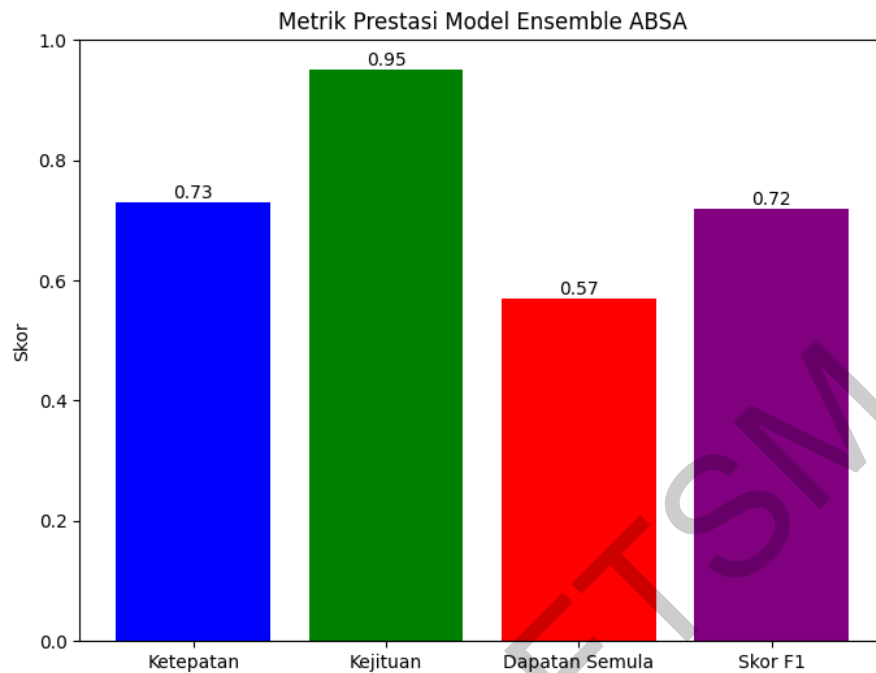
Jadual 1 Perbandingan Ketepatan Model Pengelas Sentimen Aspek

Aspek	Ketepatan Model	
	<i>Artificial Neural Network (ANN)</i>	<i>Support Vector Machine (SVM)</i>
<i>Plot</i>	0.79	0.70
<i>Acting</i>	0.82	0.78
<i>Directing</i>	0.58	0.71
<i>Sound</i>	0.88	0.79
<i>Visual</i>	0.84	0.56
<i>Character</i>	0.70	0.69

Prestasi model *ensemble* ABSA kemudian dinilai menggunakan beberapa metrik klasifikasi iaitu ketepatan, kejitian, dapatan semula dan skor F1 dari modul 'sklearn.metrics' untuk memberikan pemahaman yang komprehensif tentang ketepatan dan kebolehpercayaannya. Berdasarkan Jadual 2, model *ensemble* ABSA telah menunjukkan prestasi yang kukuh dengan ketepatan sebanyak 73.14%, yang menunjukkan bahawa model ini meramal kelas dengan betul sebanyak 73% daripada data ujian. Kejitian yang tinggi sebanyak 95.15% mendedahkan bahawa apabila model meramal kelas positif, ia adalah betul sebanyak 95.15%, menonjolkan keberkesannya dalam mengurangkan positif palsu. Walau bagaimanapun, nilai dapatan sebanyak 57.48% menunjukkan bahawa model ini hanya mengenal pasti 57.48% daripada kes positif sebenar, menunjukkan tahap negatif palsu yang sederhana. Skor F1 sebanyak 71.67% mencerminkan keseimbangan antara kejitian dan dapatan semula, memberikan ringkasan keberkesanan keseluruhan model. Rajah 3 menunjukkan prestasi model *ensemble* ABSA dalam bentuk graf.

Jadual 2 Prestasi Model *Ensemble* ABSA

	Ketepatan	Kejitian	Dapatan Semula	Skor F1
Model <i>ensemble</i> ABSA	0.7314	0.9515	0.5748	0.7167



Rajah 3 Prestasi Model *Ensemble* dalam Graf

Kemudian, antaramuka model ABSA dibangunkan menggunakan perisian *Streamlit* dan bahasa pengaturcaraan *Python*. Rajah 4 menunjukkan antaramuka yang disediakan kepada pengguna untuk memasukkan ulasan filem dalam ruangan teks yang disediakan. Pengguna boleh menaip ulasan filem mereka dalam bentuk beberapa ayat.

Aspect Based Sentiment Analysis of Movie Reviews

Type your review here

Submit Clear

Rajah 4 Antaramuka bagi pengguna memasukkan ulasan filem

Rajah 5 menunjukkan antaramuka yang dipaparkan kepada pengguna selepas mereka menekan butang 'Submit'. Model ABSA akan memproses dan menunjukkan analisis ulasan

filem pada antaramuka ini. Ramalan sentimen keseluruhan mengenai ulasan tertentu akan ditunjukkan (positif atau negatif). Selain itu, antaramuka ini juga menunjukkan satu jadual mengenai analisis ulasan filem secara terperinci. Untuk setiap aspek, teks yang berkaitan dengan aspek tersebut, skor sentimen yang diberikan, dan label sentimennya (positif atau negatif) ditunjukkan. Jika aspek tidak wujud dalam ulasan, simbol '-' akan dipaparkan. Di samping itu, graf turut dipaparkan untuk menunjukkan analisis skor sentimen yang diselaraskan bagi aspek-aspek yang terdapat dalam ulasan. Paksi-x menunjukkan aspek-aspek yang ditemui dalam ulasan, manakala paksi-y menunjukkan skor sentimen yang diselaraskan.

Penggunaan skor yang diselaraskan dalam graf bertujuan untuk memberikan perspektif yang lebih jelas mengenai hasil analisis sentimen. Dengan menyesuaikan skor daripada titik asas neutral 0.5, graf ini berjaya menonjolkan perbezaan dalam sentimen untuk setiap aspek yang dianalisis. Penyesuaian ini membantu membezakan antara sentimen positif dan negatif dengan lebih jelas: Skor yang melebihi 0.5 menunjukkan sentimen positif (diperlihatkan dalam bar hijau), manakala skor yang lebih rendah daripada 0.5 menunjukkan sentimen negatif (diperlihatkan dalam bar merah). Pendekatan ini memberi pengguna representasi visual yang mudah difahami tentang bagaimana sentimen terhadap setiap aspek dibandingkan dengan harapan neutral, memudahkan penafsiran hasil analisis. Dengan itu, penggunaan skor yang diselaraskan dalam graf meningkatkan ketelusan dan kejelasan penilaian sentimen berhubung dengan ulasan filem yang dianalisis.

Aspect Based Sentiment Analysis of Movie Reviews

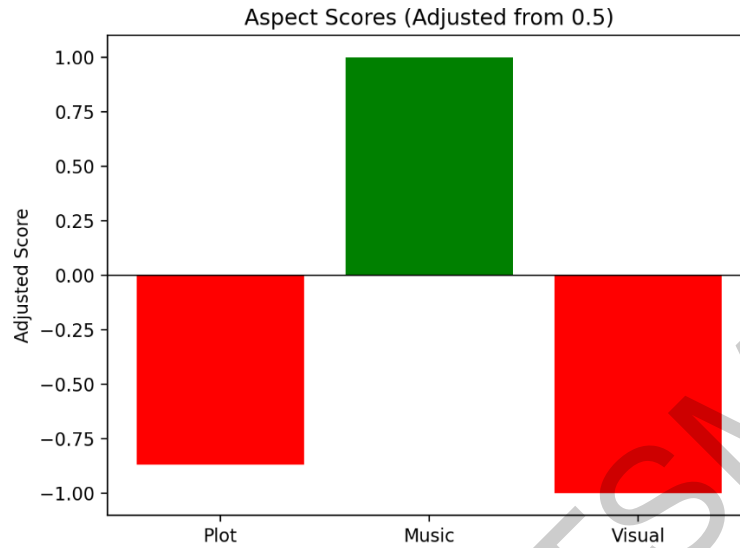
Type your review here

The soundtrack is good. But the story line is terrible. The visual is very bad.

Submit Clear

The review above is negative.

Aspect	Review	Score	Sentiment
Plot	story line terrible	0.0648	Negative
Acting	-	-	-
Director	-	-	-
Music	soundtrack good	1.0	Positive
Visual	visual bad	0.0	Negative
Character	-	-	-



Rajah 5 Antaramuka menunjukkan ramalan sentimen berdasarkan aspek

KESIMPULAN

Projek ini telah berjaya membangunkan model ABSA yang khusus untuk ulasan filem. Model ini menggabungkan teknik NLP dan pembelajaran mesin untuk mengekstrak dan menganalisis sentimen berdasarkan aspek-aspek seperti *plot*, *acting*, *directing*, *music*, *visual*, dan *character*. Data ulasan filem dikumpulkan dari *Rotten Tomatoes* menggunakan teknik web scraping dan melalui pra-pemrosesan data yang teliti. Model ABSA ini kemudian dilatih untuk mengenal pasti sentimen bagi setiap aspek dan menggabungkan ramalan tersebut menggunakan teknik purata mudah untuk menentukan sentimen keseluruhan. Selain itu, satu antaramuka pengguna juga dibangunkan untuk memudahkan pengguna memasukkan ulasan filem dan mendapatkan ramalan sentimen berdasarkan aspek-aspek tersebut.

Model ABSA yang dibangunkan mempunyai beberapa kekuatan utama. Pertama, ia mampu mengekstrak aspek-aspek penting dalam ulasan filem menggunakan teknik NER dan padanan kata kunci. Kedua, model ini dapat meramalkan sentimen bagi setiap aspek supaya memberikan maklumat yang lebih terperinci kepada penonton. Ketiga, penggabungan ramalan menggunakan teknik purata mudah merupakan cara yang cepat dan berkesan untuk menentukan sentimen keseluruhan berdasarkan ulasan filem. Tambahan pula, antaramuka pengguna yang dibangunkan memudahkan interaksi pengguna dengan model ini, menjadikannya lebih praktikal dan mudah digunakan.

Walau bagaimanapun, model ini juga mempunyai beberapa kekangan. Salah satu kekangan utama adalah data yang tidak seimbang yang boleh menjejaskan prestasi model. Pengekstrakan aspek menggunakan NER dan padanan kata kunci mungkin kurang berkesan dalam mengenal pasti aspek yang disebut secara tidak langsung atau tidak jelas dalam ulasan. Senarai kata kunci yang digunakan untuk mengekstrak aspek mungkin tidak cukup komprehensif atau tepat, yang boleh menyebabkan ketidaktepatan dalam pengekstrakan aspek. Selain itu, model ABSA menghadapi kesukaran dalam mengenal pasti sentimen yang kompleks seperti ironi dan sarkasme, yang boleh menjejaskan ketepatan ramalan sentimen.

Bagi meningkatkan keberkesanan dan ketepatan model ABSA, beberapa cadangan penambahbaikan boleh dipertimbangkan untuk kajian masa hadapan. Pertama, penggunaan data ulasan filem yang lebih besar dan lebih pelbagai dari pelbagai genre filem boleh membantu dalam memastikan model ini lebih umum dan berkesan untuk semua jenis ulasan filem. Kedua, penggunaan model pengekstrakan aspek yang lebih maju seperti BERT boleh meningkatkan ketepatan dalam mengenal pasti aspek yang lebih halus dan tersembunyi. Ketiga, penerapan teknik penggabungan model yang lebih kompleks, seperti boosting atau teknik *ensemble learning* lain, boleh memberikan hasil yang lebih baik dalam menentukan sentimen keseluruhan. Dengan mengatasi kekangan yang ada dan melaksanakan cadangan-cadangan ini, model ABSA mampu menghasilkan ramalan aspek dan sentimen yang tepat.

PENGHARGAAN

Pertama sekali saya ingin mengucapkan setinggi-tinggi penghargaan dan terima kasih kepada penyelia Dr. Sabrina binti Tiun. Beliau telah banyak meluangkan masa dalam memberi tunjuk ajar, bantuan, dan nasihat yang begitu berguna dan bernilai sepanjang projek ini. Panduan beliau yang berharga dan teliti telah menginspirasi saya dalam cara yang tidak terhitung jumlahnya. Selain itu, saya juga ingin mengucapkan terima kasih kepada pensyarah-pensyarah Fakulti Teknologi dan Sains Maklumat yang telah banyak mendidik saya sepanjang tempoh masa pembelajaran di Universiti Kebangsaan Malaysia. Keterlibatan mereka telah meningkatkan kematangan intelektual saya yang akan membantu saya pada masa akan datang. Seterusnya, ucapan terima kasih tidak terhingga untuk ahli keluarga yang tercinta dan rakanrakan yang telah banyak memberi sokongan sepanjang projek ini. Akhir sekali, seikhlas tulus kata terima kasih kepada semua pihak yang telah memberi bantuan, nasihat, dan bimbingan secara langsung dan tidak langsung dalam menjayakan projek ini.

RUJUKAN

- Borele, P. & Borikar, D. A. 2016. An approach to sentiment analysis using artificial neural network with comparative analysis of different techniques. *IOSR Journal of Computer Engineering (IOSR-JCE)* 18(2): 64-69.
- Dhiraj, K. 2019. Top 4 advantages and disadvantages of Support Vector Machine or SVM. <https://dhirajkumarblog.medium.com/top-4-advantages-and-disadvantages-of-support-vector-machine-or-svm-a3c06a2b107> [15 Jun 2024]
- Donges, N. 2023. Pro and Cons of Neural Networks. <https://resources.experfy.com/ai-ml/pros-and-cons-of-neural-networks/> [16 Jun 2024]
- Hofesmann, E. 2021. The Machine Learning Lifecycle in 2021. <https://towardsdatascience.com/the-machine-learning-lifecycle-in-2021-473717c633bc> [6 November 2021]
- Jaiswal, S. 2021. Machine Learning Life Cycle. <https://www.javatpoint.com/machine-learning-life-cycle> [6 November 2021].
- Onalaja, S., Romero, E. & Yun, B. 2021. Aspect-based Sentiment Analysis of Movie Reviews. *SMU Data Science Review* 5(3): 10.
- Wang, Y., Shen, G. & Hu, L. 2020. Importance evaluation of movie aspects: Aspect-based sentiment analysis. 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), hlm. 2444-2448.
- Yadav, K., Kumar, N., Maddikunta, P. K. R. & Gadekallu, T. R. 2021. A comprehensive survey on aspect-based sentiment analysis. *International Journal of Engineering Systems Modelling and Simulation* 12(4): 279-290.

Lee Jia Yee (A187581)
Dr. Sabrina Tiun
Fakulti Teknologi & Sains Maklumat,
Universiti Kebangsaan Malaysia