

ANALISIS SENTIMEN ULASAN PRODUK KECANTIKAN DAN PENJAGAAN KULIT

NURUL ATIKAH BINTI MANSOR

ZALINDA BINTI OTHMAN

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

ABSTRAK

Langkah-langkah penjagaan asas muka sangat penting dalam kehidupan harian kita dan perlu dilakukan secara rutin untuk mendapatkan hasil kulit yang sihat. Hal ini, produk kecantikan dan penjagaan kulit telah menjadi keperluan utama bagi semua individu terutamanya wanita. Walau bagaimanapun, tidak semua produk kecantikan dan penjagaan kulit mempunyai kualiti yang bagus mengikut keperluan pengguna. Oleh itu, mereka sudah semestinya akan memilih produk yang mempunyai kualiti dan memberi kesan terbaik terhadap kulit mereka dengan melihat ulasan dari pengguna lain. Penjagaan wajah yang betul akan membantu anda memiliki kulit yang cantik, sihat dan licin. Analisis sentimen merupakan cara untuk menganalisis dan mengklasifikasikan ulasan pengguna kepada pendapat positif, negatif atau neutral berkenaan produk yang dipertikaikan. Selain itu, ia juga penting untuk mengetahui kekerapan perkataan yang sering muncul dalam ulasan pengguna tersebut. Jumlah produk kecantikan dan penjagaan kulit yang berada di pasaran dapat menarik perhatian pengguna untuk membeli. Dengan ini, ia sering digunakan sebagai rujukan untuk produk yang ingin di beli oleh pengguna. Semakin positif ulasan pengguna, semakin besar keyakinan pembeli untuk membeli produk tersebut. Sebaliknya, jika semakin banyak ulasan negatif, semakin kurang keyakinan pembeli untuk membeli. Objektif projek ini dijalankan adalah untuk mengumpul semua data berkenaan produk kecantikan dan penjagaan kulit dari ulasan pengguna dari laman sesawang Sephora.com. Oleh itu, hasil akhir kajian ini adalah penghasilan web papan pemuka dengan menggunakan *Streamlit* untuk memaparkan hasil keputusan analisis sentimen. Kesimpulannya, dengan menggunakan algoritma tersebut, pembeli dapat menambah keyakinan mereka berdasarkan ulasan produk tersebut. Data yang telah dikumpulkan sangat berguna untuk mendapatkan keputusan akhir projek ini.

Kata kunci: Analisis sentimen, Ulasan produk, *Streamlit*

PENGENALAN

Kulit adalah salah satu organ paling besar dan utama pada badan manusia. Hampir semua orang mempunyai kira-kira 2meter persegi kulit pada badan masing-masing. Oleh itu, dengan menjaga kulit anda dengan baik ia mungkin mempunyai kesan kepada kesihatan umum badan anda. Lapisan paling luar kulit berfungsi sebagai penghalang terhadap faktor luaran. Ia dipengaruhi oleh lebih banyak faktor luaran daripada yang anda fikirkan. Antara faktor yang boleh mempengaruhi kulit kita adalah terdedah kepada sinaran UV, terdedah kepada cahaya matahari tanpa perlindungan untuk jangka masa yang terlalu lama, tidak mendapat rehat atau nutrisi yang secukupnya dan penuaan.(Catherine, 2018). Generasi sekarang lebih mementingkan tentang kesihatan mereka, termasuklah tentang penjagaan kulit yang konsisten. Tidak kira apa jenis keadaan kulit sama ada berjerawat, berminyak atau kering yang mungkin dihadapi oleh wajah anda, penjagaan kulit yang baik dan konsisten akan menjamin bahawa kulit anda akan kelihatan cantik dan muda. Dengan menguasai asas penjagaan kulit, anda masih boleh memanjakan diri anda. Pilihan gaya hidup yang sihat dan penjagaan kulit yang baik boleh membantu mencegah dan melambatkan penuan dan masalah jerawat pada wajah anda.

Terdapat beberapa langkah yang boleh anda diambil untuk memastikan anda memiliki kulit yang sihat. Rutin penjagaan kulit asas yang komprehensif terdiri daripada lima langkah penting iaitu membersihkan, menyegarkan, menggunakan serum, melembapkan dan menggunakan pelindung matahari. Pertama, mencuci muka dengan menggunakan pencuci muka yang sesuai dengan keadaan kulit bagi menghilangkan kotoran, minyak berlebihan dan kekotoran dari permukaan kulit. Langkah ini membantu membersihkan permukaan kulit muka untuk produk seterusnya diserap dengan lebih berkesan. Selepas proses membersihkan muka, “*toning*” membantu mengimbangi tahap pH kulit dan menghilangkan kesan kekotoran yang tinggal. Langkah seterusnya adalah penggunaan serum, iaitu formula pekat yang direka untuk menyasarkan masalah kulit tertentu seperti garis halus, bintik gelap atau penghidratan. Serum menyediakan bahan aktif yang kuat yang meresap lebih dalam ke dalam kulit. Selepas serum, pelembapan adalah penting untuk menghidrat dan mengunci kelembapan, menjadikan kulit yang licin dan anjal. Akhir sekali, penggunaan pelindung matahari adalah langkah yang amat penting untuk melindungi kulit daripada sinaran UV yang berbahaya, yang boleh menyebabkan penuaan pramatang dan meningkatkan risiko kanser kulit.(Tolentino, 2022). Langkah-langkah tersebut membentuk penjagaan kulit asas tetapi menyeluruh, ia hendaklah dilakukan secara konsisten agar dapat menyumbang kepada kulit yang lebih sihat, bersih dan dilindungi.

Hal ini, terdapat banyak penyelesaian yang menjanjikan untuk mengembalikan kesan penuaan, mengurangkan kedutan, memutihkan kulit dan banyak lagi. Jadi, anda perlulah jalankan kajian untuk menentukan sama ada sesuatu produk itu sebenarnya penting dan sesuai untuk kesihatan kulit atau ia merupakan produk yang akan memudaratkan lagi kulit anda. Sebagai pengguna, anda boleh membaca ulasan produk untuk menilai bahawa produk tersebut berbaloi untuk dibeli atau tidak. Ini kerana setiap hari ramai pengguna akan menulis ulasan pada sesetengah produk, jadi jumlah ulasan terdiri daripada ratusan hingga ribuan dan mengandungi pelbagai pendapat dari pengguna. Ulasan produk boleh berikan kebaikan untuk pengguna yang baru ingin mencuba produk penjagaan kulit kerana dengan membaca ulasan tersebut, pengguna boleh mendapatkan maklumat tentang kualiti produk penjagaan kulit dan kecantikan berdasarkan pengalaman pengguna lain guna produk itu. Secara umumnya dalam kajian ini, analisis sentimen merupakan teknik untuk menganalisis sentimen pengguna dalam ulasan produk dan dapat mengenal pasti ulasan tersebut positif, negatif dan neutral. Semua sumber data yang telah dikumpulkan dalam projek ini adalah sangat penting bagi kita untuk mengkaji ulasan produk yang terbaik untuk membantu pembeli dalam membuat pilihan terhadap produk kecantikan dan penjagaan kulit.

Terdapat beberapa contoh kajian lepas yang berkaitan dengan topik kajian ini. Kajian pertama dari (Fadly et al. 2022) bertajuk "Analisis Sentimen pada Produk Penjagaan Kulit Asli". Kajian ini mengumpul 10,000 ulasan daripada Sephora.com dan menggunakan lima pengelas iaitu Bayes Naif, KNN, Mesin Vektor Sokongan (SVM), Pohon Keputusan dan pembelajaran mendalam untuk menganalisis data. Pembelajaran mendalam dan pohon keputusan mencapai hampir 80% ketepatan dan 60% nilai F1. Seterusnya adalah kajian daripada (Wardani et al. 2022) yang menerangkan tentang kajian analisis sentimen mengenai ulasan produk kecantikan menggunakan kaedah Modified Balanced Random Forest (MBRF) dan Chi-Square sebagai pemilihan ciri. Set data yang digunakan dalam kajian itu diperolehi daripada laman sesawang "Female Daily", yang menyediakan perkhidmatan maklumat tentang produk kecantikan wanita. Model terbaik daripada kajian menghasilkan ketepatan purata dan purata nilai F1 masing-masing sebanyak 81.75% dan 71.90%. Kajian ini menyimpulkan bahawa kaedah MBRF berkesan dalam mengklasifikasikan data yang tidak seimbang, dan kaedah Chi-Square dapat meningkatkan prestasi sistem. Selain itu, kajian analisis sentimen produk kecantikan dan penjagaan kulit yang dijalankan oleh Nadifa et al. pada tahun 2021 (Putri et al. 2021). Hasil penelitian menunjukkan nilai ketepatan 85.89%, dengan klasifikasi menggunakan algoritma mesin vektor sokongan (SVM) yang mendapat nilai ketepatan 85.98%

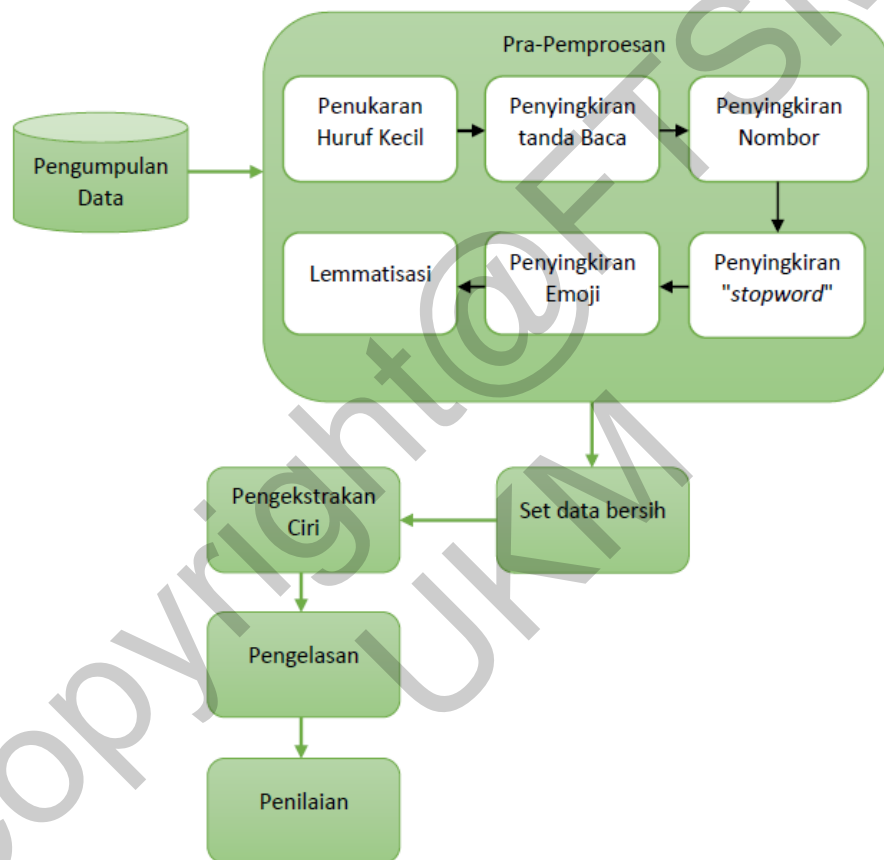
dengan kernel sigmoid. Di samping itu, kajian oleh (Tri Astuti et al. 2022) menggunakan algoritma dan Kaedah Pengoptimuman Kumpulan Partikel (PSO) untuk menganalisis sentimen dari ulasan produk Wardah UV Shield Essential Sunscreen Gel SPF 30 PA+++ . Hasil penilaian menggunakan kaedah penilaian silang 10 k-fold menunjukkan bahawa algoritma PSO mencapai ketepatan sebanyak 78.85%, manakala algoritma Bayes Naif mencapai ketepatan sebanyak 77.96%. Matriks kekeliruan juga menunjukkan bahawa algoritma PSO memiliki nilai ketepatan yang lebih baik berbanding dengan algoritma Bayes Naif. Kajian terakhir adalah daripada (Sari et al. 2020) mengenai peranan penting produk kecantikan dan penjagaan kulit dalam kehidupan harian wanita yang kini dianggap sebagai keperluan utama. Melalui penggunaan RapidMiner, kajian mendapati kaedah Bayes Naif memberikan ketepatan 80.45%.

Terdapat banyak model yang boleh menentukan sentimen sesuatu teks, sama ada positif, negatif atau neutral. Jadi, projek ini bertujuan untuk mengumpulkan data ulasan pengguna pada sesuatu produk kecantikan dan penjagaan kulit dan mengkaji model yang sesuai untuk menganalisis sentimen. Oleh itu, tujuan utama projek ini adalah untuk mengelaskan ulasan produk yang dikaji kepada positif, negatif atau neutral dengan menggunakan analisis sentimen. Selain itu, kajian ini juga akan membangunkan beberapa model pembelajaran mesin untuk menganalisis ramalan yang tepat agar dapat membantu pengguna dalam membuat keputusan. Seterusnya, kajian ini juga akan membangunkan papan pemuka analisis sentimen ulasan produk kecantikan dan penjagaan kulit menggunakan aplikasi web *Streamlit*. Oleh itu, kajian ini amat berguna dalam membantu ramai pembeli mengkaji terlebih dahulu produk tersebut berdasarkan ulasan yang telah dibuat oleh pengguna terdahulu.

Terdapat pelbagai produk kecantikan dan penjagaan kulit yang berbeza di zaman kini. Oleh itu, kajian ini hanya menekankan data daripada beberapa produk sahaja. Skop kajian ini bertujuan untuk memudahkan pembeli membuat keputusan sama ada produk tersebut berbaloi untuk dibeli atau tidak berdasarkan ulasan-ulasan pengguna yang lepas. Selain itu, algoritma pembelajaran mesin seperti Pohon Keputusan, Hutan Rawak, Mesin Vektor Sokongan, Regresi Logistik, K-Jiran Terdekat dan Bayes Naif telah digunakan untuk membangunkan model ramalan yang mampu meramalkan keputusan yang tepat. Akhir sekali, keputusan ulasan produk tersebut akan dipaparkan melalui papan pemuka dengan menggunakan aplikasi web *Streamlit*.

METODOLOGI KAJIAN

Model klasifikasi aspek sentimen adalah model yang dirancang khusus untuk mengidentifikasi fasa-fasa tertentu dari suatu produk atau yang terkandung dalam ulasan dan menentukan sentimen yang terkait dengan fasa tersebut. Oleh yang demikian, untuk mendapatkan hasil keputusan yang tepat dan berkualiti, pembangunan model yang sesuai adalah sangat penting untuk memastikan kajian ini berjalan dengan lancar. Model yang dinyatakan di bawah mudah difahami dan dapat digunakan semasa proses menganalisis sentimen.



Rajah 1 Model Klasifikasi Aspek Sentimen

Fasa Pengumpulan Data

Fasa pengumpulan data merupakan langkah pertama yang perlu dilakukan dalam kajian analisis sentimen ini. Dalam kajian ini, set data yang digunakan bagi projek ini dimuat turun dari laman Kaggle dan diekstrak melalui laman sesawang Sephora.com. Data yang telah dikumpulkan adalah data yang mentah dan mempunyai jumlah yang besar untuk diproses. Proses ini melibatkan pemilihan dataset yang kaya dengan ulasan produk, termasuk maklumat seperti teks ulasan, jenama produk dan atribut lain seperti penilaian produk, harga dan sebagainya

Fasa Pra-pemrosesan Teks

Fasa pra-pemrosesan data adalah salah satu teknik perlombongan data yang merangkumi proses transformasi data menjadi maklumat yang boleh difahami. Data mentah biasanya tidak konsisten, tidak relevan dan mengandungi teks hingar (*noisy text*). Oleh itu, sebarang singkatan perkataan, perkataan henti, tanda baca dan lemmatisasi perlu dikeluarkan sebelum menjalankan analisis sentimen. Fasa ini penting kerana ia boleh menjejaskan kualiti hasil kajian yang tepat. Kemudian, untuk pembersihan data, tokenisasi ialah proses menggantikan data sensitif dengan simbol pengenalan unik yang mengekalkan semua maklumat penting tentang data tanpa menjejaskan keselamatannya. Medium yang akan digunakan untuk membersihkan data ialah *Python*. Akhir sekali, data yang telah dikumpulkan di fasa ini akan disimpan untuk digunakan di fasa seterusnya.

Fasa Pengekstrakan Ciri

Apabila lakukan kajian tentang analisis sentimen, fasa pengekstrakan ciri ialah proses mencari dan mengekstrak butiran atau ciri yang berkaitan daripada teks data yang boleh digunakan untuk menentukan sentimen yang dinyatakan dalam ulasan produk. Ini merupakan fasa penting dalam membangunkan model analisis sentimen. Pertama, model Bag of words adalah salah satu teknik untuk mengekstrak data. Ia dipanggil (BOW) kerana sebarang maklumat tentang susunan atau struktur perkataan dalam data tidak diambil kira dan akan melihat data yang diketahui sahaja. Seterusnya adalah Frekuensi Istilah-Frekuensi Dokumen Terbalik (TF-IDF). Ia adalah ukuran statistik yang menilai kepentingan sesuatu perkataan dalam data berbanding koleksi data (korpus). Ia memberikan pemberat yang lebih tinggi kepada perkataan yang lebih khusus dalam set data sersebut. Di samping itu, Leksikon ialah koleksi perkataan atau frasa yang dikaitkan dengan skor sentimen iaitu positif, negatif dan neutral. Model ini dapat menentukan polariti tanggapan atau pendapat. Model boleh memahami dan menilai sentimen dengan lebih tepat apabila ciri yang paling penting dan relevan dikenal pasti. Ia memberikan pemahaman yang lebih mendalam tentang ulasan pengguna terhadap produk yang dikaji. Mengekstrak ciri ini membolehkan model memahami dan memproses teks dalam format yang boleh diproses oleh algoritma pembelajaran mesin. Oleh itu, memilih kaedah pengekstrakan ciri yang betul boleh menyumbang dengan ketara kepada prestasi model analisis sentimen. (Majumder, 2023).

Fasa Pengelasan

Fasa pengelasan dalam analisis sentimen ulasan produk melibatkan beberapa langkah penting. Pertama, data yang telah diproses dibahagi kepada dua subset utama iaitu data latihan (80%) dan data ujian (20%). Data latihan digunakan untuk melatih model pembelajaran mesin, sementara data ujian digunakan untuk menilai ketepatan model yang telah dilatih. (Kiruthika, 2023). Beberapa model dan algoritma digunakan dalam proses ini, termasuk Pohon Keputusan, Hutan Rawak, Mesin Vektor Sokongan (SVM), Regresi Logistik, dan K-Jiran Terdekat (KNN). Ketepatan model terlatih dinilai menggunakan data ujian untuk menentukan model yang menunjukkan ketepatan tertinggi. Secara keseluruhannya, fasa ini bertujuan untuk mengklasifikasikan sentimen ulasan produk kecantikan dan penjagaan kulit dengan menggunakan pelbagai algoritma pembelajaran mesin untuk menghasilkan ramalan yang tepat.

Fasa Penilaian

Fasa penilaian memberikan pemahaman tentang tahap keupayaan model untuk mengklasifikasikan sentimen dengan tepat dan membantu menilai keberkesanan model analisis sentimen. Fasa ini akan memaparkan keputusan analisis sentimen sama ada ulasan produk yang telah dianalisis tergolong dalam sentimen positif, negatif atau neutral. Selain itu, dapat membantu pembeli membuat keputusan dalam memilih produk yang sesuai berdasarkan hasil data yang telah dikaji. Setresunya, metrik penilaian seperti ketepatan, kepersisan, dapatan semula dan skor-f1 digunakan untuk menilai prestasi model secara kuantitatif, memastikan keseimbangan antara kepersisan dan dapatan semula untuk hasil yang berkesan dalam mengklasifikasikan sentimen ulasan produk. Akhir sekali, hasil keputusan akan dipaparkan dan divisualisasikan dalam bentuk graf yang mudah difahami dan lebih menarik.

KEPUTUSAN DAN PERBINCANGAN

Bagi menentukan model klasifikasi yang terbaik, penilaian ketepatan telah dijalankan untuk mengkaji model yang menunjukkan hasil yang tertinggi. Metrik penilaian juga telah dijalankan bagi model pembelajaran mesin yang telah dibangunkan untuk memilih algoritma yang paling sesuai bagi mengklasifikasikan ulasan produk kecantikan dan penjagaan kulit. Metrik penilaian mengukur seberapa baik model melakukan ramalan atau prediksi dengan membandingkan hasil ramalan dengan nilai sebenar dalam set data.

Dalam analisis sentimen, ulasan atau teks biasanya akan dikategorikan kepada tiga kelas utama iaitu positif, negatif dan neutral. Ini kerana positif menunjukkan ulasan yang menunjukkan kepuasan atau pandangan baik terhadap produk. Bagi negatif pula, ia menunjukkan ulasan yang menunjukkan ketidakpuasan atau pandangan buruk terhadap produk dan akhir sekali neutral iaitu ulasan yang tidak menunjukkan emosi yang kuat atau berada di antara positif dan negatif. Untuk menentukan nilai akhir keputusan bagi setiap model, *weighted average* akan dipilih kerana ia mengambil kira bilangan contoh dalam setiap kelas. Contohnya, jika terdapat ketidakseimbangan dalam jumlah contoh antara kelas, seperti lebih banyak ulasan positif berbanding neutral atau negatif, *weighted average* memberikan gambaran yang lebih tepat tentang prestasi model secara keseluruhan. Ini kerana *weighted average* mengira metrik purata dengan mengambil kira jumlah contoh dalam setiap kelas dan memberikan lebih banyak berat kepada kelas yang lebih besar. Hal ini penting kerana metrik penilaian seperti ketepatan, kejitian, dapatan semula dan F1-skor boleh berbeza-beza antara kelas yang berbeza. Oleh itu, Jadual 1, 2, 3, 4, 5 dan 6 menunjukkan keputusan model pembelajaran mesin berdasarkan dua kaedah iaitu tidak menggunakan kaedah pensampelan semula dan menggunakan kaedah SMOTE (*Oversampling*).

Jadual 1 Keputusan Pohon Keputusan

Metrik Penilaian	Ketepatan	Kejitian	Dapatan semula	F1-skor
Tidak menggunakan kaedah pensampelan semula	0.9073	0.9097	0.9108	0.9097
Menggunakan kaedah SMOTE (<i>Oversampling</i>)	0.9093	0.9126	0.9072	0.9058

Jadual 2 Keputusan Hutan Rawak

Metrik Penilaian	Ketepatan	Kejitian	Dapatan semula	F1-skor
Tidak menggunakan kaedah pensampelan semula	0.9488	0.9469	0.9472	0.9470
Menggunakan kaedah SMOTE (<i>Oversampling</i>)	0.9497	0.9475	0.9478	0.9476

Jadual 3 Keputusan Mesin Vektor Sokongan

Metrik Penilaian	Ketepatan	Kejituan	Dapatan semula	F1-skor
Tidak menggunakan kaedah pensampelan semula	0.9498	0.9497	0.9498	0.9498
Menggunakan kaedah SMOTE (<i>Oversampling</i>)	0.9498	0.9497	0.9498	0.9498

Jadual 4 Keputusan Regresi Logistik

Metrik Penilaian	Ketepatan	Kejituan	Dapatan semula	F1-skor
Tidak menggunakan kaedah pensampelan semula	0.8703	0.8707	0.8703	0.8699
Menggunakan kaedah SMOTE (<i>Oversampling</i>)	0.8707	0.8710	0.8707	0.8703

Jadual 5 Keputusan K-Jiran Terdekat

Metrik Penilaian	Ketepatan	Kejituan	Dapatan semula	F1-skor
Tidak menggunakan kaedah pensampelan semula	0.4332	0.7470	0.4332	0.3374
Menggunakan kaedah SMOTE (<i>Oversampling</i>)	0.4333	0.7471	0.4333	0.3376

Jadual 6 Keputusan Bayes Naif

Metrik Penilaian	Ketepatan	Kejituan	Dapatan semula	F1-skor
Tidak menggunakan kaedah pensampelan semula	0.7322	0.7632	0.7437	0.7354
Menggunakan kaedah SMOTE (<i>Oversampling</i>)	0.7325	0.7632	0.7437	0.7354

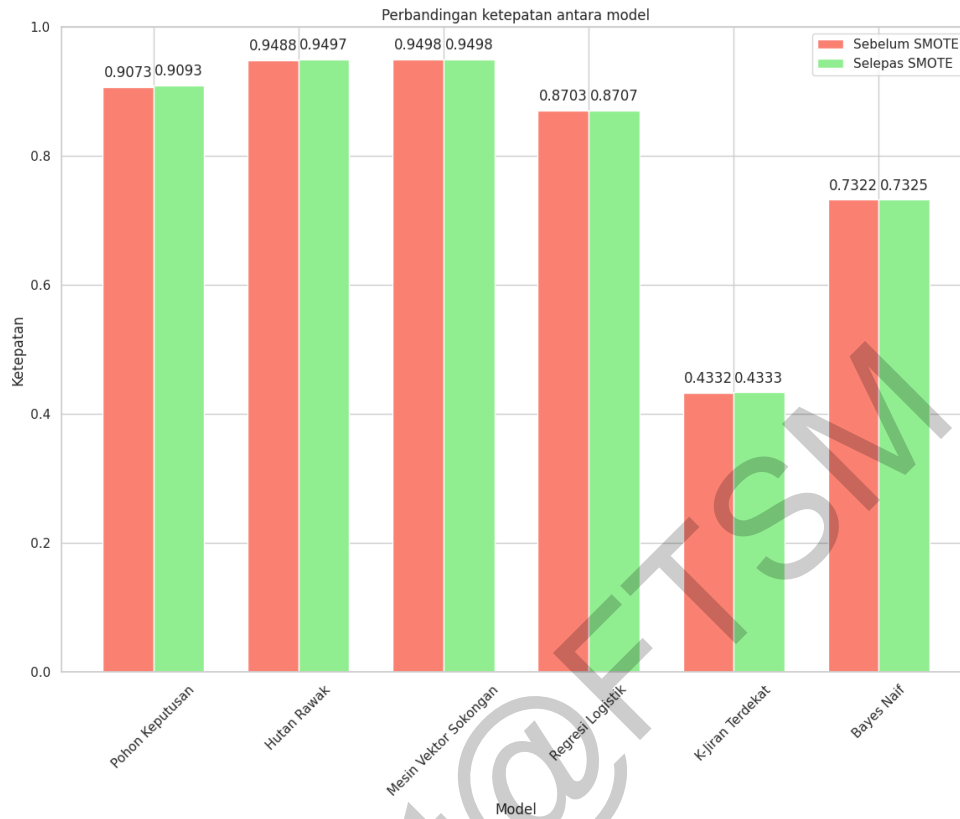
Berdasarkan penilaian pemodelan yang telah dijalankan seperti jadual di atas, saya dapati bahawa model yang menggunakan kaedah pensampelan semula iaitu SMOTE (*Oversampling*) telah memperoleh keputusan akhir yang lebih baik berbanding dengan model yang tidak menggunakan kaedah pensampelan semula. Pemilihan SMOTE sebagai kaedah pensampelan semula adalah tepat dan diperlukan dalam kajian ini. Oleh itu, penggunaan SMOTE terbukti

bermanfaat dalam kes ini kerana ia membantu mengatasi ketidakseimbangan kelas yang merupakan masalah umum dalam data yang mempunyai lebih banyak contoh dari satu kelas berbanding kelas yang lain. Tanpa menggunakan SMOTE, model mungkin cenderung mengabaikan kelas minoriti dan akan menghasilkan ramalan yang bias dan tidak tepat. Oleh itu, Jadual 7 telah merumuskan hasil akhir keputusan bagi semua model ramalan yang dijalankan dalam kajian ini.

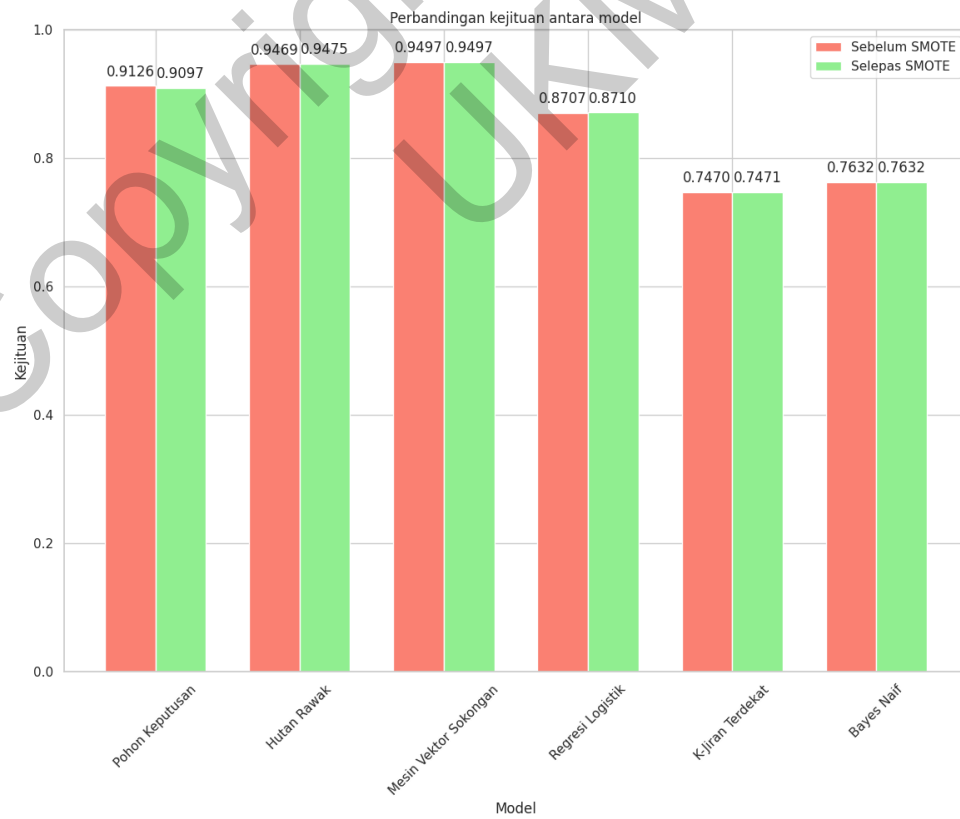
Jadual 7 Keputusan Model Ramalan

Metrik Penilaian	Ketepatan	Kejituan	Dapatan semula	F1-skor
Pohon Keputusan	0.9093	0.9097	0.9072	0.9058
Hutan Rawak	0.9497	0.9475	0.9478	0.9476
Mesin Vektor Sokongan	0.9498	0.9497	0.9498	0.9498
Regresi Logistik	0.8707	0.8710	0.8707	0.8703
K-Jiran Terdekat	0.4333	0.7471	0.4333	0.3376
Bayes Naif	0.7325	0.7632	0.7437	0.7354

Berdasarkan Jadual 7, mesin vektor sokongan menunjukkan prestasi model terbaik dengan ketepatan (0.9498), kejituan (0.9497), dapatan semula (0.9498) dan F1-skor (0.9498). Hutan rawak juga menunjukkan prestasi yang baik, tetapi mencapai nilai yang lebih rendah berbanding mesin vektor sokongan. Berbanding dengan model lain seperti pohon keputusan, regresi logistik, k-jiran terdekat dan bayes naif, mesin vektor sokongan konsisten menunjukkan nilai yang lebih tinggi, terutamanya dalam kejituan dan dapatan semula yang penting dalam konteks kelas yang tidak seimbang. Secara keseluruhan, Rajah 2, Rajah 3, Rajah 4 dan Rajah 5 di bawah menunjukkan graf bar perbandingan prestasi model ramalan sama ada tidak menggunakan kaedah pensampelan semula atau menggunakan kaedah SMOTE (*Oversampling*).



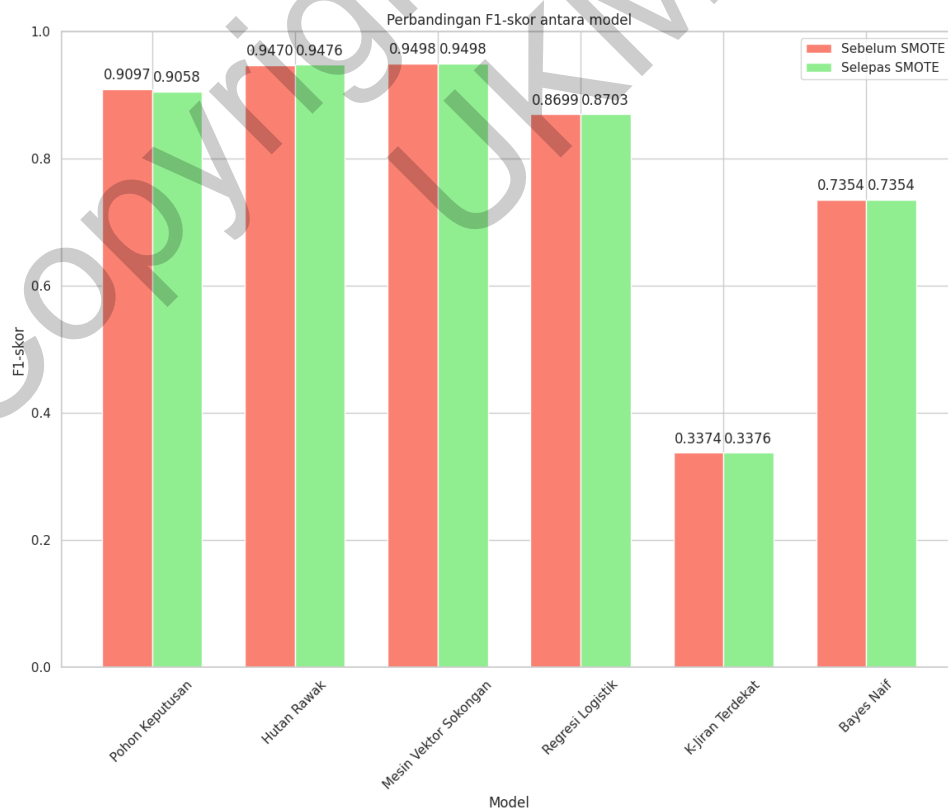
Rajah 2 Perbandingan ketepatan antara model



Rajah 3 Perbandingan kejituan antara model



Rajah 4 Perbandingan dapatan semula antara model



Rajah 5 Perbandingan F1-skor antara model

Streamlit adalah alat yang sangat berguna untuk membina web interaktif bagi analisis data. Ia adalah alat yang sesuai dengan kajian ini iaitu Analisis Sentimen Ulasan Produk Kecantikan dan Penjagaan Kulit. Dengan menggunakan *Streamlit*, papan pemuka (*dashboard*) yang membolehkan pengguna memuat naik ulasan, menganalisis sentimen dan memvisualisasikan hasil analisis dengan mudah dapat dibina. Oleh itu, bagi projek ini, model mesin vektor sokongan telah dipilih sebagai model yang dapat meramal ramalan yang tepat kerana ia menghasilkan keputusan akhir yang tertinggi berbanding model lain. Rajah 6 menunjukkan paparan “*Home*” bagi projek ini iaitu Analisis Sentimen Ulasan Produk Kecantikan dan Penjagaan Kulit. Selain itu, Rajah 7 menunjukkan paparan “*Product Review Analyzer*”. Di bahagian ini, pengguna boleh mengisi nama produk dan ulasan produk untuk menilai ulasan tersebut akan dikelaskan sama ada sebagai positif, negatif atau neutral. Pengguna boleh mengisi ulasan produk tersebut dengan had 300 patah perkataan. Kemudian keputusan akan divisualisasikan kepada pengguna untuk memudahkan kefahaman pengguna dalam pemilihan produk. Seterusnya, Rajah 8 telah menunjukkan paparan “*Product Recommendation*” iaitu senarai 10 produk terbaik yang mempunyai ulasan positif yang tertinggi. Pengguna juga akan lebih yakin jika terdapat majoriti ulasan positif dalam sesuatu produk.



Rajah 6 Paparan “*Home*”

Rajah 7 Paparan “Product Review Analyzer”

	product_name	brand_name	total_positive	to
0	The Moisturizing Soft Cream Moisturizer	La Mer	194	
1	Trinity + Eye and Lip Enhancer Attachment Bundle	NuFACE	152	
2	Algae + Moringa Universal Hydrating Face Oil	MARA	150	
3	Mini Algae + Moringa Universal Face Oil	MARA	150	
4	(Glow)Setting 100% Mineral Powder SPF 35	Supergoop!	143	
5	The Silk Serum Wrinkle-Smoothing Retinol Alternative	Tatcha	140	
6	Absolue Soft Cream Revitalizing & Brightening Moisturizer	Lancôme	136	
7	Face Stone Nourishing Solid Refillable Facial Oil with Tamanu + Blue Tansy	Kate McLeod	133	
8	Clean Pro Rose Quartz	PMD	132	
9	Equilibrium Resurfacing Retinoid Treatment	Hourglass	131	

Rajah 8 Paparan “Product Recommendation”

Seterusnya, Rajah 9 menunjukkan bahawa ulasan produk tersebut telah dikategorikan sebagai positif.

The screenshot displays the 'Product Review Analyzer' interface. At the top, there are navigation links for 'Home', 'Product Review Analyzer', and 'Product Recommendation'. The main input section includes:

- Enter Product Name:** Lancôme
- Enter Product Review:** Excellent product, my skin is softer. Lightweight product and the smell is amazing. Highly recommended!
- Original Text:** Excellent product, my skin is softer. Lightweight product and the smell is amazing. Highly recommended!
- Cleaned Text:** excellent product skin softer lightweight product smell amaze highly recommend

An 'Analyze' button is located below the cleaned text. The 'Results' section shows:

- Sentiment: Positive 😊**
- A table with the following data:

	metric	value
0	compound	0.8716
- A bar chart showing a single blue bar for the 'compound' metric with a value of approximately 0.87. The y-axis is labeled 'value' and ranges from 0.0 to 0.8. The x-axis is labeled 'metric' and has a category 'compound'.

Rajah 9 Keputusan ulasan produk positif

Rajah 10 menunjukkan bahawa ulasan produk tersebut telah dikategorikan sebagai negatif.

The screenshot displays the 'Product Review Analyzer' interface. At the top, there are navigation links for 'Home', 'Product Review Analyzer', and 'Product Recommendation'. The main input section includes:

- Enter Product Name:** A text box containing 'tarte'.
- Enter Product Review:** A text area containing the review: 'Awful!!!! Stinks, sticky and goes bad (oxidation) fast. Worst thing I've ever purchased from Sephora'.
- Original Text:** A text box showing the original review text.
- Cleaned Text:** A text box showing the cleaned review text: 'awful stink sticky go bad oxidation fast worst thing ive ever purchase sephora'.
- Analyze** button.

The **Results** section shows:

- Sentiment: Negative** with a sad face emoji.
- A table with the following data:

	metric	value
0	compound	-0.9296

- A bar chart showing the sentiment score for the 'compound' metric, with a value of -0.9296. The y-axis is labeled 'value' and ranges from -1.0 to 0.0. The x-axis is labeled 'metric' and has a category 'compound'.

Rajah 10 Keputusan ulasan produk negatif

Rajah 11 menunjukkan bahawa ulasan produk tersebut telah dikategorikan sebagai neutral.

The screenshot displays the 'Product Review Analyzer' interface. It features a navigation bar with 'Home', 'Product Review Analyzer', and 'Product Recommendation'. The main form includes input fields for 'Enter Product Name:' (filled with 'clinique'), 'Enter Product Review:' (filled with 'makes skin look healthier with daily use'), and 'Original Text:' (filled with 'makes skin look healthier with daily use'). Below these is a 'Cleaned Text:' field (filled with 'make skin look healthier daily use') and an 'Analyze' button.

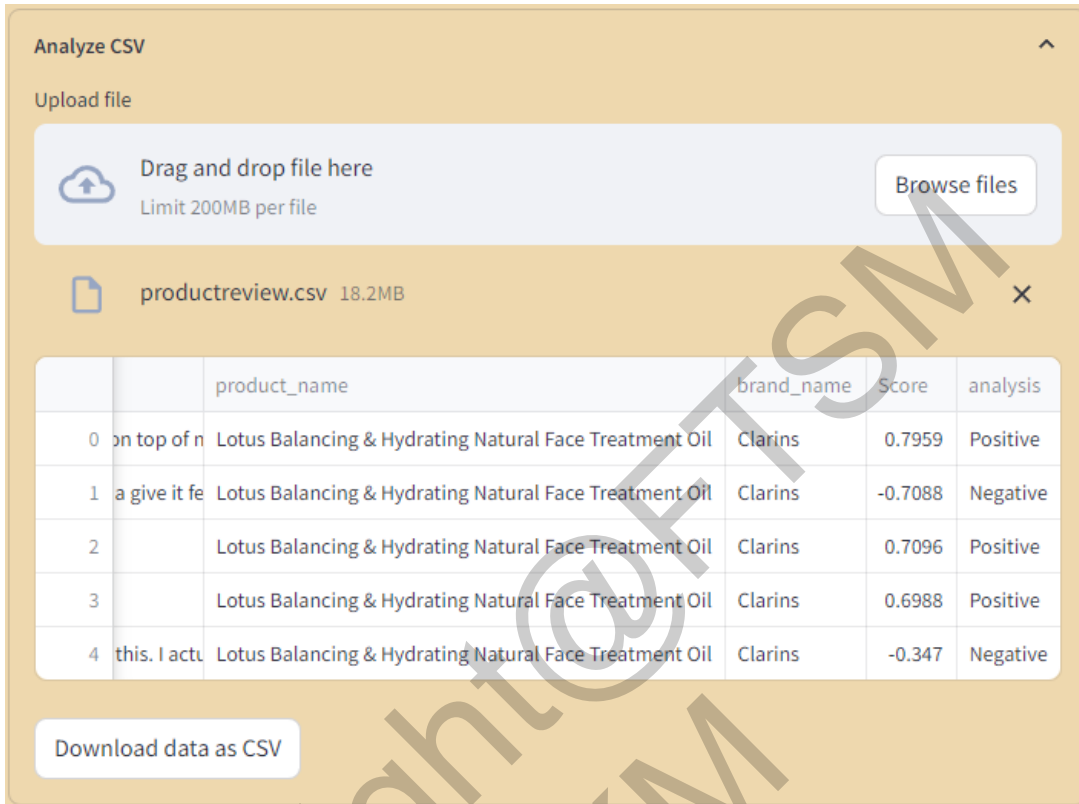
The 'Results' section shows the sentiment analysis: 'Sentiment: Neutral 😐'. A table displays the following data:

	metric	value
0	compound	0

Next to the table is a bar chart with a legend for 'metric' (blue square) and 'compound' (blue square). The y-axis is labeled 'value' with a tick at 0. The x-axis has labels for 'compound' and 'metric'. The chart shows a single bar for 'compound' with a value of 0.

Rajah 11 Keputusan ulasan produk neutral

Akhir sekali, Rajah 12 memaparkan keputusan analisis set data CSV iaitu dengan cara memuat naik fail CSV untuk menganalisis semua ulasan produk yang terkandung dalam set data tersebut.



Analyze CSV

Upload file

Drag and drop file here
Limit 200MB per file

Browse files

productreview.csv 18.2MB

		product_name	brand_name	Score	analysis
0	on top of n	Lotus Balancing & Hydrating Natural Face Treatment Oil	Clarins	0.7959	Positive
1	a give it fe	Lotus Balancing & Hydrating Natural Face Treatment Oil	Clarins	-0.7088	Negative
2		Lotus Balancing & Hydrating Natural Face Treatment Oil	Clarins	0.7096	Positive
3		Lotus Balancing & Hydrating Natural Face Treatment Oil	Clarins	0.6988	Positive
4	this. I actu	Lotus Balancing & Hydrating Natural Face Treatment Oil	Clarins	-0.347	Negative

Download data as CSV

Rajah 12 Keputusan analisis set data CSV

Cadangan Penambahbaikan

Selepas menjalankan kajian yang menyeluruh, Terdapat beberapa cadangan penambahbaikan yang dapat meningkatkan model ramalan pada masa hadapan. Antaranya, meningkatkan kualiti set data merupakan salah satu cadangan penambahbaikan kerana set data yang berkualiti tinggi dapat membantu keupayaan model untuk menganalisis hasil akhir. Hal ini, teknik pensampelan semula perlu dijalankan agar dapat menyelesaikan masalah data yang tidak stabil dan meningkatkan kualiti model. Cadangan penambahbaikan yang seterusnya adalah menambah kapasiti komputer seperti meningkatkan RAM agar proses implementasi dapat dijalankan dengan lancar tanpa sebarang ralat pengkodan berlaku.

KESIMPULAN

Kesimpulannya, dengan memaparkan hasil output yang terbaik, ia bermaksud kajian ini berjaya dalam menganalisis sentimen berdasarkan ulasan produk kecantikan dan penjagaan kulit. Oleh itu, penilaian prestasi model merupakan bahagian paling penting dan langkah yang

tidak boleh diabaikan kerana bahagian ini akan mengenal pasti model yang akan membuat ramalan yang terbaik. Setelah menjalankan beberapa perbandingan antara model, saya telah mengenal pasti bahawa mesin vektor sokongan merupakan model yang mempunyai prestasi terbaik berbanding dengan beberapa model yang lain dalam menganalisis ulasan produk. Selain itu, segala masalah yang terjadi ketika dalam kajian ini terutamanya dalam pengaturcaraan dapat diatasi dengan jayanya. Kajian ini juga berjaya mencapai objektif yang telah ditetapkan dengan mengumpulkan dan menganalisis ulasan pengguna terhadap produk kecantikan dan penjagaan kulit. Walaupun menghadapi beberapa kekangan, kajian ini memberikan pandangan yang berharga mengenai sentimen pengguna dan faktor-faktor yang mempengaruhi kepuasan mereka. Cadangan penambahbaikan yang dikemukakan diharapkan dapat membantu dalam menjalankan kajian yang lebih baik di masa hadapan, serta meningkatkan kualiti produk dan strategi pemasaran dalam industri

Kekuatan Kajian

Beberapa kekuatan telah menyumbang kepada kejayaan kajian ini. Pertama, kajian ini menggunakan data sebenar daripada ulasan produk kecantikan dan penjagaan kulit yang diambil daripada laman web Sephora.com. Ini memastikan hasilnya lebih relevan dan realistik, dan ia meningkatkan kebolehpercayaan. Kedua, sentimen ulasan produk telah dinilai menggunakan pelbagai model pembelajaran mesin iaitu model mesin vektor sokongan (SVM) telah menunjukkan prestasi terbaik. Ini menjamin pendekatan analisis yang komprehensif dan hasil yang optimum. Ketiga, aplikasi web Streamlit mempunyai reka bentuk antara muka pengguna yang hebat yang memudahkan pengguna menggunakan alat analisis sentimen dan menjadikan maklumat lebih mudah difahami.

Tambahan pula, proses pra-pemrosesan data yang teliti termasuk pembersihan data dan penormalan teks telah memastikan bahawa data yang digunakan adalah berkualiti tinggi dan tidak terganggu. Selain itu, kajian ini berjaya mengkategorikan sentimen ulasan pengguna kepada kategori positif, negatif dan neutral. Ini memberikan gambaran yang jelas tentang penerimaan pengguna terhadap produk kecantikan dan penjagaan kulit serta faktor yang mempengaruhi sentimen mereka. Dengan kekuatan-kekuatan ini, kajian ini memberikan sumbangan yang berharga kepada pemahaman tentang sentimen pengguna serta menawarkan pandangan yang dapat digunakan untuk penambahbaikan produk dan strategi pemasaran.

Kelemahan Kajian

Sepanjang menjalankan kajian ini, terdapat beberapa kekangan yang telah dihadapi sebelum projek ini berjaya dilaksanakan. Antara kekangan tersebut adalah dari segi set data. Ini kerana set data yang digunakan adalah tidak seimbang, iaitu ulasan produk bagi nilai positif sangat banyak berbanding dengan negatif dan neutral sehingga menyebabkan keputusan ramalan tidak dapat diramal dengan tepat. Hal ini telah memberi kesan kepada kebolehsanaan model bagi melakukan ramalan yang terbaik. Selain itu, kekangan yang dihadapi adalah kapasiti komputer. Proses menjalankan model pembelajaran mesin memerlukan kapasiti yang mencukupi agar tidak mengganggu hasil akhir keputusan. Jadi, dengan meningkatkan kapasiti komputer atau mengurangkan nilai set data dapat meningkatkan prestasi model.

PENGHARGAAN

Pertama sekali saya ingin mengucapkan syukur Alhamdulillah dan setinggi-tinggi kesyukuran dipanjatkan kehadiran ilahi kerana dengan izin-Nya saya dapat menyiapkan laporan yang bertajuk 'Analisis Sentimen Ulasan Produk Kecantikan dan Penjagaan kulit' ini.

Setinggi-tinggi penghargaan saya ucapkan kepada penyelia saya yang amat dihormati, Prof. Madya Dr. Zalinda binti Othman atas segala tunjuk ajar dan bantuan yang diberikan kepada saya secara langsung atau tidak langsung sepanjang tempoh kajian ini dilaksanakan. Beliau telah membimbing dan memberi nasihat yang amat berharga dalam menyiapkan projek ini. Segala jasa, ilmu dan tunjuk ajar beliau akan saya kenang sepanjang hayat.

Saya juga mengambil peluang ini untuk memberikan penghargaan dan ucapan terima kasih yang tidak terhingga kepada ibu bapa, ahli keluarga dan rakan-rakan atas bantuan dan sokongan yang telah diberikan kepada saya. Bantuan dan sokongan tersebut telah mendorong saya untuk lebih bersemangat dalam menyiapkan kajian ini. Jutaan terima kasih juga diucapkan kepada semua pensyarah Fakulti Teknologi dan Sains Maklumat (FTSM) yang telah memberikan ilmu pengetahuan kepada saya sepanjang pengajian saya di Universiti Kebangsaan Malaysia (UKM).

Akhir sekali, tidak lupa juga ucapkan ribuan terima kasih kepada semua yang terlibat secara langsung atau tidak langsung dalam menjayakan kajian ini. Saya juga ingin memohon ribuan kemaafan sekiranya terdapat kesalahan sepanjang saya menjalankan projek tahun akhir ini.

Sekian, terima kasih.

RUJUKAN

- Astuti, T. & Astuti, Y. 2022. Analisis Sentimen Review Produk Skincare Dengan Naïve Bayes Classifier Berbasis Particle Swarm Optimization (PSO). *Jurnal Media Informatika Budidarma*, 6(4), 1806. <https://doi.org/10.30865/mib.v6i4.4119>
- Devaraj, K. 2023. Top difference between training data and testing data. Testsigma Blog. <https://testsigma.com/blog/difference-between-training-data-and-testing-data/>
- Fadly, Marlina, D., Kuniawan, T.B., Zakaria, M.Z. & Abdullah, S. F. 2022. Sentiment Analysis on Natural Skincare Products. *Journal of Data Science*, 2022(12), 1–17. <http://eprints.intimal.edu.my/id/eprint/1667>
- Majumder, P. 2023. Creating a Movie Reviews Classifier Using TF-IDF in Python. <https://www.analyticsvidhya.com/blog/2021/09/creating-a-movie-reviews-classifier-using-tf-idf-in-python/>
- Putri, N.F., Faraby, S.A. & Dwifabri, M. 2021. Analisis Sentimen pada Produk Kecantikan dari Ulasan Female Daily Menggunakan Information Gain dan Support Vector Machine (SVM) Classifier. *EProceedings of Engineering*, 8(5).
- Sari, D.N., Sari, D.N., Adelia, F., Rosdiana, F., Butar, B.B. & Hariyanto, M. 2020. Analisa Sentimen Terhadap Review Produk Kecantikan Menggunakan Metode Naïve Bayes Classifier. (*Jurnal Informatika*), 4(3), 109. <https://doi.org/10.31000/jika.v4i3.3086>
- Satpathy, S. 2020. SMOTE - A Common Technique to Overcome Class Imbalance Problem. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2020/10/overcoming-class-imbalance-using-smote-techniques/>
- Tolentino, G. 2022. Skin Care Routine Order: A Step By Step Guide. <https://www.dermstore.com/blog/in-what-order-do-i-apply-my-skin-care-products-infographic/>
- Wardani, A.P.P., Adiwijaya, A. & Purbolaksono, M.D. 2022. Sentiment Analysis on Beauty Product Review Using Modified Balanced Random Forest Method and Chi-Square. *Journal of Information System Research (JOSH)*, 4(1), 1–7. <https://doi.org/10.47065/josh.v4i1.2047>

Nurul Atikah binti Mansor (A189076)
 Prof. Madya Dr. Zalinda binti Othman
 Fakulti Teknologi & Sains Maklumat
 Universiti Kebangsaan Malaysia