

**MENGELOMPOKAN, MENGANALISIS SENTIMEN DAN
MERINGKASKAN BERITA KEWANGAN MENGGUNAKAN
PEMBELAJARAN MESIN, PEMBELAJARAN MENDALAM SERTA
KEKOMPUTERAN BAHASA TABII**

SAMIRA NAIR VASU DAVAN

PROF. MADYA TS. DR. RAVIE CHANDREN A/L MUNIYANDI

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

ABSTRAK

Saintis data kewangan menghadapi cabaran apabila menggunakan model seperti BERT untuk menganalisis kandungan khusus industri, seperti perbankan, kerana terpaksa mengendalikan bahasa kewangan yang bernuansa. Sekatan ini secara langsung mempengaruhi ketepatan yang mungkin dijangkakan oleh saintis data kewangan daripada model komputeran bahasa Tabii (NLP) mereka, terutamanya apabila digunakan untuk menganalisis artikel berita yang mempunyai istilah kewangan. Selain itu, dengan banyaknya data berita, analisis manual bagi set data yang begitu luas adalah tidak praktikal kerana memerlukan banyak tenaga buruh serta peruntukan sumber yang tidak praktikal. Walaupun terdapat sesetengah tapak web mempunyai sistem klasifikasi untuk berita, namun keberkesanannya adalah terhad. Untuk menangani masalah ini, kajian ini mencadangkan aplikasi web yang secara sistematik mengekstrak berita kewangan yang relevan daripada data tidak berstruktur menggunakan pengikis web yang memfokuskan pada kata kunci kewangan utama dan mengkategorikan berita ke dalam kategori seperti pasaran saham, perniagaan, hal ehwal antarabangsa, politik dan ekonomi menggunakan pengelompokan K-means dengan Peruntukan Dirichlet Terpendam (PDT). Pendekatan ini memanfaatkan teknik pembelajaran mesin dan pembelajaran mendalam iaitu BERT dengan regresi logistik untuk analisis sentimen dan Facebook Bart untuk ringkasan. BERT mengkategorikan artikel sebagai positif atau negatif, manakala Facebook Bart menghasilkan ringkasan berita yang ringkas berdasarkan sentimen yang diekstrak. Mengikuti metodologi "Air Terjun", projek ini berjalan melalui peringkat yang ditentukan: pengekstrakan berita, pengkategorian, integrasi analisis sentimen, ringkasan dan penciptaan aplikasi bahagian hadapan. Model "Air Terjun" meminimumkan risiko dengan memastikan setiap fasa diselesaikan dengan teliti sebelum melangkah ke fasa seterusnya. Selain itu, seni bina pelayan pelanggan dilaksanakan untuk meningkatkan kecekapan dan kebolehskalaan aplikasi. Produk akhir yang lengkap ialah tapak web mesra pengguna yang menawarkan berita kewangan yang dikategorikan, cerapan sentimen dan ringkasan. Penyelesaian ini mendemokrasi pengetahuan kewangan, menjadikan maklumat penting tersedia untuk khalayak yang lebih luas dan memperkasakan pengguna untuk membuat keputusan kewangan yang termaklum. Projek

ini bertujuan untuk menyediakan sistem klasifikasi yang boleh dipercayai dan berkesan untuk berita kewangan, mengatasi batasan sistem sedia ada dan meningkatkan ketepatan dan kebolegunaan analisis data kewangan dengan ketara.

Kata kunci: Pengelompokan K-Means, Analisis Sentimen, BERT, Facebook Bart

PENGENALAN

Di Malaysia, idea literasi kewangan semakin popular pada awal 2000-an bermula dengan inisiatif Bank Negara Malaysia, iaitu bank sentral Malaysia, dalam mewujudkan Program Pengayaan Bakat Sektor Kewangan (FSTEP). Literasi kewangan melibatkan hubungan antara amalan kewangan yang wajar dan pengetahuan kewangan, menunjukkan bahawa mereka yang lebih berpengetahuan cenderung mengikuti amalan yang disyorkan oleh pakar kewangan. Kursus dalam FSTEP bertujuan meningkatkan literasi kewangan dan memberi akses kepada maklumat dan kebolehan penting. Melalui kempen ini, banyak pertubuhan kerajaan dan bukan kerajaan telah melancarkan pelbagai projek dan program untuk menggalakkan literasi kewangan. Walau bagaimanapun, menurut OECD, hanya 59.7% rakyat Malaysia yang mempunyai literasi kewangan, meninggalkan 40.3% masyarakat yang kurang bernasib baik dalam menilai peluang kewangan.

Pengetahuan seorang penganalisis kewangan adalah penilai utama dalam bidang kewangan. Penganalisis kewangan, berbanding orang awam, adalah bijak dan cukup berpengetahuan untuk memahami berita tentang kewangan, termasuk ekonomi, pasaran saham, perniagaan dan politik. Kemahiran ini membolehkan mereka menghasilkan rancangan kewangan yang bijak yang bertujuan untuk mempercepatkan penjana kekayaan mereka. Selain itu, mereka dapat menentukan sama ada berita itu mempunyai kesan yang baik atau tidak baik terhadap persekitaran kewangan. Orang awam tidak boleh dengan mudahnya menduplikasi keupayaan ini. Orang ramai mungkin mendapati sukar untuk memperoleh manfaat sepenuhnya daripada pengetahuan kewangan yang tersedia jika mereka tidak mempunyai akses kepada penganalisis kewangan. Di sinilah keberkesanan teknologi kekomputeran bahasa tabii (NLP) benar-benar menunjukkan peranannya. Keupayaan untuk menganalisis data kewangan tidak berstruktur, seperti yang terdapat dalam berita, siaran media sosial dan laporan kewangan, telah menjadikan pemprosesan bahasa semula jadi (NLP) sebagai bidang penyelidikan yang penting dalam bidang kewangan. Namun, kekangan besar muncul seiring dengan perkembangan data teks kewangan baru-baru ini. Teks ini dihasilkan dalam jumlah yang sangat besar setiap hari sehingga hampir sukar bagi satu entiti untuk menganalisisnya secara manual bagi mendapatkan konklusi yang berguna. Ini sangat relevan, khususnya dalam lingkungan industri FinTech yang sedang berkembang pesat, di mana pentingnya untuk selalu kekal terkini dengan berita kewangan. Melalui perkembangan teknologi pembelajaran mesin, model perlombongan teks dapat diimplementasikan. Model ini memproses dan menganalisis sejumlah besar teks kewangan secara automatik sambil menggunakan pendekatan statistik untuk mengenal pasti data dan trend penting. Perkembangan ini menyerlahkan sinergi antara analisis profesional dan inovasi teknologi dalam membolehkan orang ramai membuat keputusan kewangan yang bijak.

Model moden seperti BERT (Devlin et al., 2019), yang dikembangkan oleh Google dan telah dilatih terlebih dahulu pada korpus teks besar yang tidak berlabel yang diperoleh dari Wikipedia, telah menunjukkan keberkesanan yang signifikan dalam memahami dan memproses data dari pelbagai domain. Model ini telah menunjukkan prestasi cemerlang dalam cabaran yang melibatkan pelbagai subjek. Walau bagaimanapun, kecekapan mereka berkurangan apabila menggunakan bahasa dan semantik yang khusus untuk industri perbankan. Sekatan ini secara langsung mempengaruhi ketepatan yang mungkin dijangkakan oleh saintis data kewangan daripada model Kekomputeran Bahasa Tabii (NLP) mereka, terutamanya apabila digunakan untuk menganalisis artikel berita yang mempunyai istilah kewangan atau secara tidak langsung boleh dikaitkan dengan kewangan. Walaupun model ini mempunyai ketepatan yang tinggi dalam memahami bahasa secara umum, model ini sukar untuk memahami bahasa kompleks dan khusus yang digunakan dalam situasi kewangan. Ketidakupayaan model untuk memperoleh cerapan fundamental dengan efektifnya daripada berita berkaitan kewangan disebabkan oleh defisit dalam pemahaman khusus tentang kewangan. Selain itu, dengan banyaknya data berita, analisis manual bagi set data yang begitu luas adalah tidak praktikal kerana memerlukan banyak tenaga buruh serta peruntukan sumber yang tidak praktikal.

Walaupun terdapat sesetengah tapak web mempunyai sistem klasifikasi untuk berita, namun keberkesananannya adalah terhad. Sistem klasifikasi yang benar-benar boleh dipercayai dan berkesan untuk berita kewangan adalah kurang dalam persekitaran semasa. Mengikuti peristiwa kewangan penting, seperti ekonomi, pergerakan pasaran saham dan perubahan dalam industri perbankan, terus menjadi tugas yang sukar tanpa cara untuk membahagikan item berita dengan cepat dan betul kepada pelbagai kategori. Membuat pertimbangan segera dan berpendidikan terhalang oleh kekurangan klasifikasi ini, yang melambatkan pengambilan cepat maklumat yang berkaitan dengan pelbagai bidang kewangan. Tambahan pula, kesukaran adalah jauh lebih besar bagi mereka yang mempunyai latar belakang kewangan yang lemah. Individu dari latar belakang ini mungkin hanya tahu cara membaca artikel berita yang dibahagikan kepada kategori tertentu tetapi mereka tidak tahu cara menggunakan maklumat tersebut untuk faedah kewangan mereka. Untuk menentukan sama ada sesuatu berita menghantar sikap positif, negatif atau neutral, metodologi yang kompleks diperlukan kerana sifat penyelidikan sentimen yang rumit dalam kewangan. Ia memerlukan pengetahuan dan akses kepada teknik analisis yang canggih untuk menentukan bagaimana emosi boleh mempengaruhi pasaran kewangan dan aset. Tanpa maklumat ini, orang ramai mungkin merasa amat sukar untuk mengurus kerumitan berita kewangan dan membuat keputusan yang selaras dengan objektif dan keadaan kewangan mereka.

Matlamat utama kajian ini adalah untuk membangunkan aplikasi sendiri yang mengekstrak berita kewangan yang relevan daripada data berita terkumpul tidak berstruktur. Item berita diekstrak menggunakan pengikis web dengan penekanan pada kata kunci penting dalam industri kewangan. Melalui pemodelan topik menggunakan teknik pembelajaran mesin seperti K-means clustering, maklumat ini dikategorikan ke dalam kategori seperti ekonomi, pergerakan pasaran saham, politik, hal ehwal antarabangsa, dan perkembangan perniagaan. K-means clustering digunakan kerana data berita yang tidak dilabelkan membawa kepada

pelaksanaan pembelajaran tanpa selia untuk mengelaskan artikel dengan cekap ke dalam kumpulan tersendiri. Untuk analisis sentimen, alat VADER digunakan untuk melabelkan sentimen secara automatik, diikuti dengan semakan manual untuk memastikan ketepatan. Model analisis sentimen berdasarkan BERT dengan regresi logistik dilatih dan diuji menggunakan set data terbahagi, dan hasilnya disimpan kembali ke pangkalan data. Akhir sekali, model Facebook BART digunakan untuk menjana ringkasan ringkas artikel berita dan frasa utama, yang turut disimpan dalam pangkalan data untuk kemudahan pengguna. Bahagian front-end aplikasi ini akan dibangunkan dengan HTML, CSS, dan Javascript.

METODOLOGI KAJIAN

Projek ini akan menggunakan proses “Waterfall”. Keputusan ini adalah disebabkan oleh ciri projek iaitu setiap peringkat mesti disiapkan sebelum meneruskan pembangunan ciri seterusnya. Memandangkan keperluannya ditakrifkan dengan baik, model “Waterfall” sejajar dengan strukturnya dengan lancar. Selain itu, model “Waterfall” amat sesuai untuk projek ini kerana pendekatannya yang linear dan berurutan. (Thesing et al., 2021). Dalam konteks ini, projek ini dikategorikan kepada fasa yang berbeza, bermula dengan komponen pengekstrakan berita penting. Fasa awal ini diikuti dengan langkah-langkah seterusnya, termasuk pengkategorian berita, integrasi analisis sentimen, pembangunan kamus istilah kewangan, dan akhirnya penghasilan ‘front-end’ aplikasi. Dengan mematuhi model “Waterfall”, projek ini meminimumkan risiko yang wujud, kerana setiap fasa disempurnakan dengan teliti sebelum melangkah ke fasa seterusnya. Perhatian yang teliti terhadap perincian ini bukan sahaja memastikan tahap ketepatan yang tinggi tetapi juga menyumbang kepada kebolehpercayaan keseluruhan aplikasi.

Fasa keperluan

Dalam fasa ini, keperluan analisis sentimen menggunakan NLP pada berita kewangan dinilai. Penyelidikan meluas dilakukan untuk mengesahkan perkaitan sistem dan untuk menentukan alat-alat penting yang diperlukan untuk pembinaannya. Spesifikasi fungsi sistem iaitu ciri pengekstrakan berita, ciri pengelompokkan berita, ciri analisis sentimen dan ringkasan berita kewangan digariskan dengan teliti.

Fasa reka bentuk

Dalam fasa ini, keperluan diubah kepada pelan terperinci. Seni bina sistem direka bentuk terlebih dahulu di mana interaksi antara setiap komponen dipetakan. Selain itu, algoritma reka bentuk setiap ciri akan dilakukan secara terperinci. Kemudian, reka bentuk pangkalan data dipertimbangkan untuk memastikan data disimpan dalam cara berstruktur yang memudahkan pencarian maklumat. Bingkai wayar dan mockup juga akan dilakarkan yang akan memberikan gambaran visual antara muka pengguna untuk menggambarkan produk akhir.

Fasa pelaksanaan

Ini adalah fasa di mana pengkodan berlaku. Mekanisme untuk setiap ciri akan dikodkan dan bahagian antara muka juga akan dilakukan. Penyepaduan semua ciri akan menjadi aspek

penting dalam fasa ini, memastikan aplikasi ini berfungsi secara harmoni sebagai sistem bersatu.

Fasa pengujian

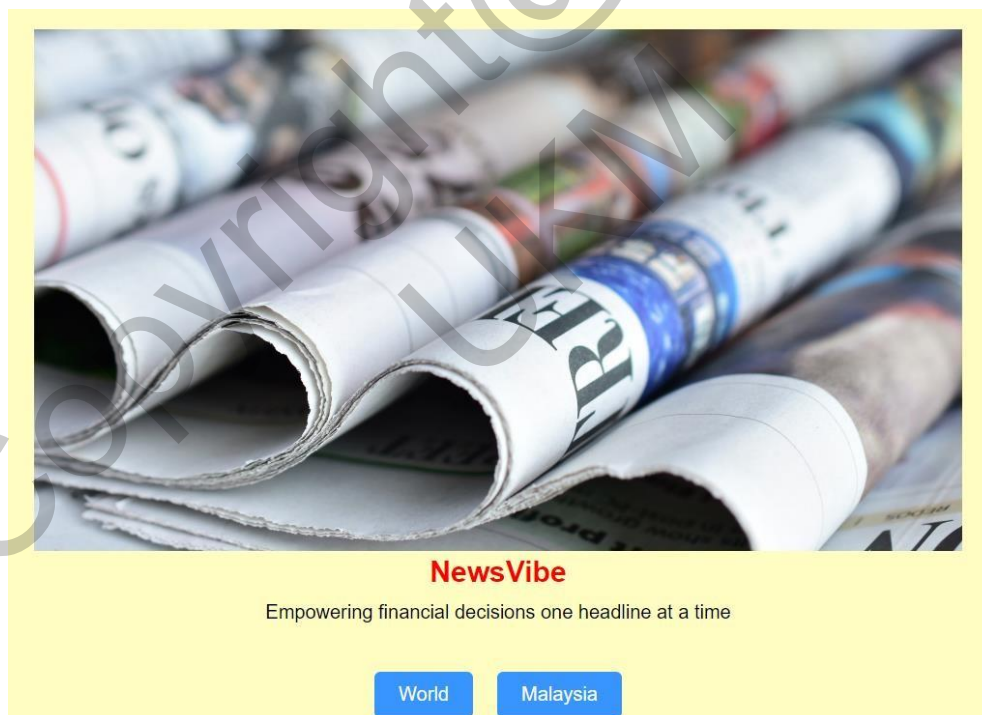
Fasa ini adalah penting dalam memastikan keteguhan dan kebolehpercayaan sistem. Tiga jenis ujian dijalankan, bermula dengan ujian untuk meneliti kualiti pengelompokan K_Means menggunakan Skor Siluet. Kemudiannya, ujian untuk menilai kualiti analisis sentimen BERT. Ujian kebolehgunaan sistem menggunakan skala kebolehgunaan sistem membolehkan pengguna sebenar berinteraksi dengan aplikasi bagi memberikan maklum balas yang tidak ternilai untuk diperhalusi.

- **Kualiti Pengelompokan dengan Skor Siluet** : Skor Siluet dipilih untuk menilai keberkesanan algoritma pengelompokan dalam menyusun artikel berita ke dalam kategori yang bermakna. Metrik ini mengukur sejauh mana kluster ditakrifkan dengan baik dan berbeza, penting untuk memastikan artikel dikumpulkan secara logik berdasarkan persamaan topik. Dengan menggunakan skor Siluet, penilaian kuantitatif telah dijalankan ke atas prestasi pengelompokan untuk mengoptimalkannya dan meningkatkan ketepatan pengkategorian.
- **Pengesahan Analisis Sentimen (Kejituan, Skor capaian semula, Pengukuran_F1, Ketepatan)** : Untuk analisis sentimen, metrik seperti Kejituan, Skor capaian semula, Pengukuran-F1 dan Ketepatan telah dipilih untuk mengesahkan prestasi model BERT. Kejituan mengukur ketepatan ramalan sentimen positif dan negatif, Skor capaian semula menilai sejauh mana sentimen ditangkap secara keseluruhan, Pengukuran-F1 seimbang Kejituan dan Skor capaian semula, dan Ketepatan menyediakan ukuran keseluruhan ramalan yang betul. Metrik ini penting untuk memastikan keputusan analisis sentimen boleh dipercayai dan sejajar dengan jangkaan pengguna dalam mentafsir sentimen artikel dengan tepat.
- **Kebolehgunaan Bahagian Hadapan dengan Skala Kebolehgunaan Sistem: Tinjauan** ini menggunakan alat standard industri yang dibangunkan oleh John Brooke pada tahun 1986, dan telah digunakan secara meluas untuk menilai kebolehgunaan antara muka bahagian hadapan. Ujian dimulakan dengan melibatkan sampel 15 pengguna yang berinteraksi dengan laman web NewsVibe. Setiap pengguna diminta untuk meneroka ciri dan fungsi tapak web. Berikutan interaksi mereka, pengguna diminta untuk mengisi soal selidik skala kebolehgunaan sistem melalui Borang Google. Pendekatan ini memberikan cerapan kuantitatif tentang pengalaman pengguna, menilai faktor seperti kemudahan penggunaan, kejelasan navigasi dan kepuasan keseluruhan. Dengan menggunakan skala ini, maklum balas berstruktur dikumpulkan daripada pengguna, isu kebolehgunaan dikenal pasti dan penambahbaikan diberi keutamaan untuk meningkatkan penglibatan dan kepuasan pengguna dengan aplikasi.

KEPUTUSAN DAN PERBINCANGAN

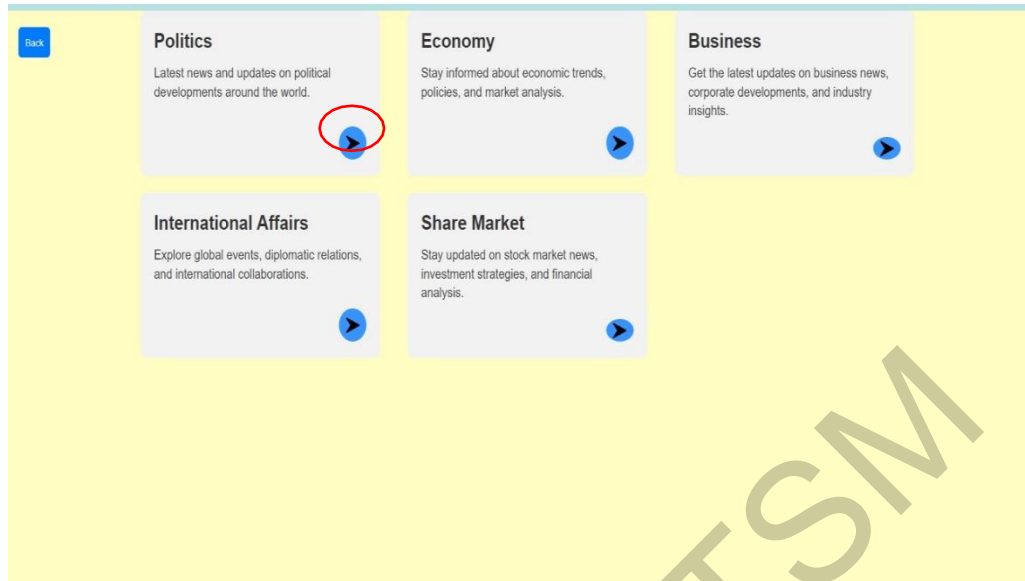
NewsVibe, sebuah aplikasi web untuk menganalisis sentimen berita, telah berjaya dibangunkan dan semua dokumentasinya telah dilengkapkan. Semasa proses pembangunan, sistem ini pertama-tama mengekstrak berita dari platform-platform terkenal dan menyimpannya dalam pangkalan data. Kemudian, proses pra-pemrosesan dilakukan untuk membersihkan data. Berita kemudiannya diklasifikasikan ke dalam pelbagai sub-kategori menggunakan pengkelompokan K-means dengan Peruntukan Dirichlet Terpendam (PDT). Analisis sentimen Vader dilakukan untuk mendapatkan data standard emas yang akan digunakan dalam pemrosesan BERT. Dengan dataset tersebut, BERT dilatih menggunakan regresi logistik untuk melakukan analisis sentimen. Selanjutnya, ringkasan menggunakan Facebook BART dilakukan pada dataset untuk mendapatkan ringkasan pendek artikel berita. Antara muka pengguna aplikasi ini menampilkan semua ini dalam laman webnya di mana pengguna dapat memilih kategori, membaca artikel berita, serta melihat analisis sentimen dan ringkasan artikel. Berikut menunjukkan antara muka dari laman web NewsVibe.

Apabila memasuki laman pertama dalam NewsVibe, pengguna akan melihat dua butang dalam warna biru iaitu butang “World” dan “Malaysia. Pengguna boleh memilih kategori berita yang ingin mereka meneroka.



Rajah 1 Antara Muka Pemilihan Kategori

Selepas pengguna memilih kategori, sistem akan memaparkan sub-kategori berita iaitu “Politics”, “Economy”, “Business”, “International Affairs” dan “Share Market”. Pengguna perlu memilih daripada sub-kategori ini untuk mendapatkan berita yang mereka ingin membaca.



Rajah 2 Antara Muka Pemilihan Sub-Kategori

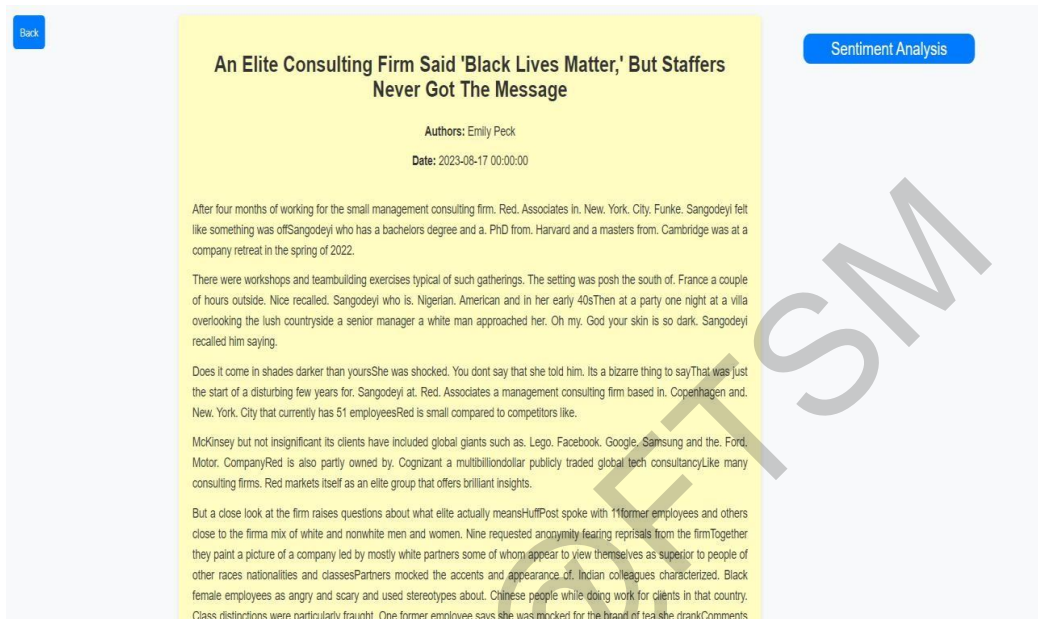
NewsVibe ini akan memaparkan senarai berita dalam sub-kategori tersebut selepas pengguna membuat pilihan mereka seperti yang ditunjukkan dalam Rajah 3. Pengguna boleh memilih artikel berita yang mereka ingin membaca. Pengguna juga boleh membuat tapisan mengikut kriteria tarikh. Selepas penapisan, sistem akan memberikan senarai artikel yang telah dipilih kepada pengguna. Jika tarikh yang dipilih oleh pengguna tidak sah, sistem akan mengembalikan mesej ralat dan meminta pengguna untuk menukar tarikh yang dipilih. Proses ini akan berulang sehingga kriteria tapisan yang sah diberikan oleh pengguna.



Rajah 3 Antara Muka Paparan Senarai Berita

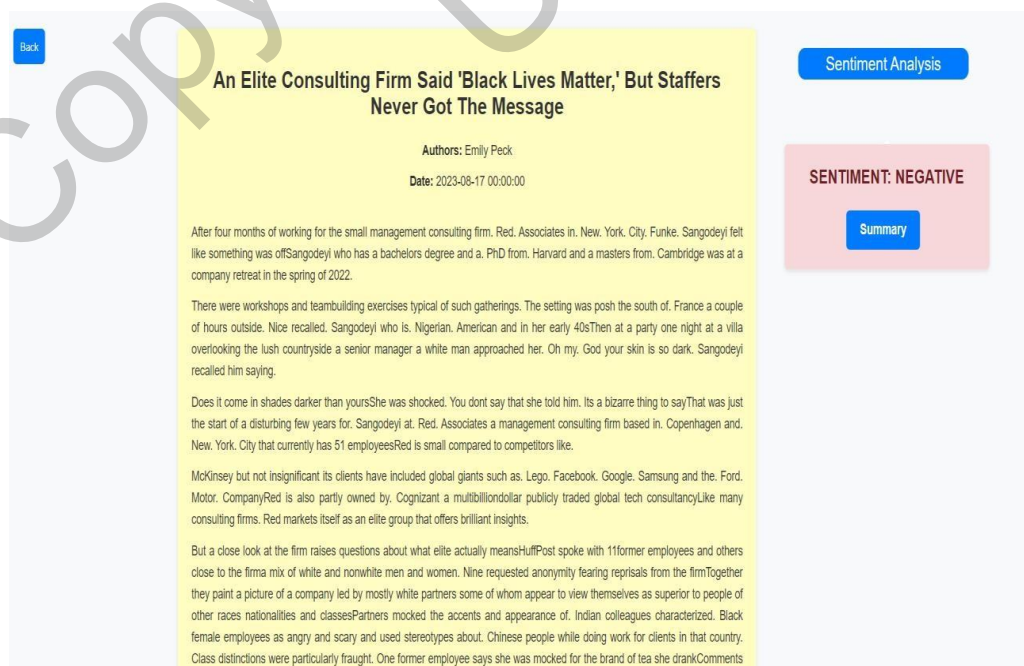
Selepas memilih artikel, pengguna dibawa ke halaman paparan artikel di mana ia direka untuk bacaan tertumpu seperti yang ditunjukkan dalam Rajah 4. Tajuk utama ditunjukkan dengan jelas dalam huruf tebal untuk menarik perhatian, diikuti oleh badan artikel dalam fon Helvetica, yang menggalakkan kebolehbacaan dan kejelasan. Untuk mengenal pasti analisis

sentimen artikel, butang “Sentiment Analysis” dengan kod warna 6439FF diletakkan di bahagian bawah imej untuk membolehkan pengguna mengakses cerapan sentimen dengan segera.



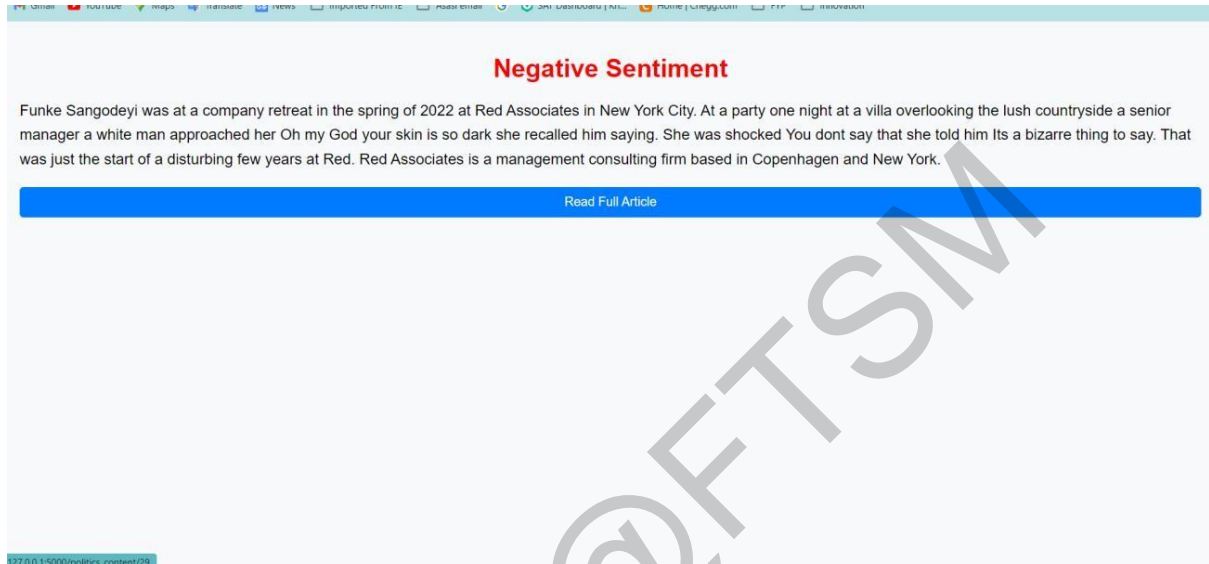
Rajah 4 Paparan Artikel Berita

Apabila pengguna menekan butang analisis sentimen, kotak pop up yang memaparkan hasil analisis sentimen—positif atau negatif—akan muncul. Untuk memudahkan pemahaman dan pembezaan, sentimen positif ditunjukkan dengan teks hijau dan sentimen negatif dengan merah. Tambahan pula, kotak pop up ini mengandungi pilihan "Summary" yang membolehkan penonton meneliti versi ringkas analisis sentimen. Rajah 5 menunjukkan rupa kotak pop-up.



Rajah 5 Paparan Analisis Sentimen Artikel

Dalam paparan ringkasan berita seperti dalam Rajah 6, sentimen dipaparkan sebagai tajuk dan kandungan dipaparkan di bawah. Terdapat juga butang yang dipanggil "Baca Artikel Penuh" yang membolehkan pengguna kembali ke kandungan penuh artikel.



Rajah 6 Paparan Ringkasan Berita

Pengujian Sistem

a. Pengelompokan

Jadual 1 Skor Siluet

Kategori	Skor Siluet
Dunia	0.7148
Malaysia	0.9717

Markah Siluet yang diperoleh untuk kedua-dua kategori 'Dunia' dan 'Malaysia' dalam analisis pengelompokan menggunakan algoritma Peruntukan Dirichlet Terpendam (PDT) memberikan pandangan yang lebih mendalam tentang keberkesanan penyusunan artikel berita dalam kategori yang berbeza dalam aplikasi NewsVibe. Kategori 'Dunia' mencapai skor Siluet 0.7148, menunjukkan kelompok yang ditakrifkan dengan munasabah untuk topik berita global merentas subkategori iaitu Politik, Perniagaan, Hal Ehwal Antarabangsa, Ekonomi dan Pasaran Saham. Skor ini menunjukkan bahawa algoritma PDT berjaya mengelompokkan artikel berita antarabangsa berdasarkan persamaan semantik mereka, yang seterusnya meningkatkan organisasi dan kebolehcapaian kandungan berita global yang pelbagai untuk pengguna.

Sebaliknya, kategori 'Malaysia' memperoleh skor Siluet yang lebih tinggi iaitu 0.9717, mencerminkan kluster yang sangat kohesif khusus untuk subjek berita Malaysia dalam subkategori seperti politik tempatan, perkembangan ekonomi, hal ehwal antarabangsa, arah aliran pasaran saham dan aktiviti perniagaan. Skor yang lebih tinggi ini menunjukkan bahawa

algoritma PDT lebih tepat mengkategorikan artikel berita Malaysia ke dalam subkategori yang berbeza dengan persamaan dalaman yang lebih besar berbanding artikel dalam kelompok lain. Kekukuhan pengkategorian ini boleh dikaitkan dengan beberapa faktor unik kandungan berita Malaysia.

Satu faktor penyumbang kepada skor Siluet yang lebih tinggi dalam kategori 'Malaysia' adalah kekhususan relatif dan jumlah artikel berita Malaysia yang lebih rendah berbanding kategori 'Dunia' yang lebih luas. Topik berita Malaysia cenderung mempersembahkan julat kepelbagaian tematik yang lebih terbatas berbanding topik berita global, menjadikannya lebih mudah bagi algoritma untuk membezakan dan mengkategorikan artikel berdasarkan ciri kandungan yang dikongsi dalam subkategori yang ditentukan ini. Skop yang difokuskan ini meningkatkan kesepaduan dan keistimewaan kelompok lalu memastikan pengguna boleh menavigasi dan mengakses jenis kandungan berita Malaysia yang khusus berdasarkan minat dan keperluan maklumat mereka. Tambahan pula, artikel berita Malaysia sering mempamerkan ciri linguistik dan kontekstual yang berbeza yang memudahkan pengelompokan yang lebih tepat. Algoritma boleh memanfaatkan ciri khusus ini, seperti kata kunci, rujukan geografi, dan nuansa budaya yang unik kepada Malaysia, untuk membentuk kelompok padu yang sejajar rapat dengan jangkaan pengguna dan keperluan mendapatkan maklumat.

Secara ringkasnya, skor Siluet 0.7148 untuk kategori 'Dunia' dan 0.9717 untuk kategori 'Malaysia' menyerlahkan keupayaan algoritma PDT untuk mengkategorikan artikel berita dengan berkesan dalam NewsVibe merentas pelbagai sub-kategori. Walaupun kategori 'Dunia' mendapat manfaat daripada keupayaan algoritma untuk mengumpulkan pelbagai topik berita global, skor yang lebih tinggi dalam kategori 'Malaysia' menggariskan ketepatannya dalam mengelompokkan artikel berita Malaysia dalam sub-kategori tertentu. Keputusan ini menunjukkan kebolehsuaian algoritma dalam mengendalikan kandungan berita yang luas dan khusus, memastikan organisasi dan kebolehaksesan dipertingkatkan untuk pengguna yang mencari kemas kini global dan cerapan setempat.

b. Analisis Sentimen (BERT)

Jadual 2 Metric Prestasi Analisis Sentimen

Kategori	Ketepatan (Accuracy)	Kejituan (Precision)	Skor capaian semula (Recall)	Pengukuran- F1 (F1-Score)
Dunia	0.82	0.81	0.80	0.80
Malaysia	0.88	0.87	0.88	0.87

Keputusan analisis sentimen menggunakan BERT dalam aplikasi NewsVibe bukan sahaja mempamerkan ketepatannya yang tinggi tetapi juga menyerlahkan keberkesannya dari segi Kejituan, Skor capaian semula dan pengukuran-F1 metrik merentas kategori 'Dunia' dan 'Malaysia'. Dalam kategori 'Dunia', BERT mencapai Ketepatan 0.82, menunjukkan ketepatan keseluruhannya dalam ramalan sentimen, di samping Kejituan 0.81, Skor capaian semula 0.80, dan Pengukuran-F1 0.80. Metrik ini menggambarkan keupayaan BERT untuk

mengklasifikasikan sentimen positif dan negatif dengan tepat dalam artikel berita global sambil mengekalkan prestasi seimbang merentas metrik Kejituan dan Skor capaian semula.

Begitu juga, dalam kategori 'Malaysia', BERT menunjukkan prestasi yang lebih baik dengan Ketepatan 0.88, di samping Kejituan 0.87, Skor capaian semula 0.88, dan Pengukuran-F1 yang mengagumkan sebanyak 0.87. Metrik ini menggariskan kepandaian BERT dalam mengenal pasti sentimen positif dengan tepat dalam artikel berita Malaysia, mencapai kadar Kejituan dan Skor capaian yang tinggi, dan dengan itu menyumbang kepada Pengukuran-F1 yang unggul. Prestasi kukuh BERT merentas metrik ini boleh dikaitkan dengan keupayaan pemprosesan bahasa semula jadinya yang maju, yang membolehkannya memahami dan mentafsir nuansa halus dalam sentimen yang dinyatakan dalam artikel berita. Terutama, BERT mempamerkan kecenderungan untuk mengklasifikasikan sentimen positif dengan lebih berkesan daripada sentimen negatif. Trend ini dibuktikan oleh skor Ketepatan yang lebih tinggi untuk sentimen positif dalam kedua-dua kategori 'Dunia' dan 'Malaysia', menunjukkan bahawa BERT boleh mengenal pasti dan mengkategorikan artikel berita dengan lebih tepat yang menyatakan nada emosi positif seperti kejayaan, kemajuan atau keyakinan.

Membandingkan keputusan ini dengan ulasan literatur dan kajian terdahulu mengenai analisis sentimen menggunakan BERT, kajian ini mendedahkan hasil prestasi yang menggalakkan, terutamanya dalam konteks berita kewangan. Kajian terdahulu sering menumpukan pada analisis sentimen berita umum, manakala aplikasi ini menyasarkan pengkategorian berita kewangan secara khusus, yang memberikan cabaran dan kerumitan unik. Sebagai contoh, dalam kajian yang mengkaji analisis sentimen mengenai berita komoditi, BERT mencapai Ketepatan 0.87. Keputusan ini mempamerkan kejayaan perbandingan aplikasi NewsVibe dalam mengklasifikasikan sentimen dengan tepat dalam domain berita kewangan, di mana nuansa dan istilah khusus memainkan peranan penting dalam tafsiran sentimen. Begitu juga, dalam penyelidikan lain yang memfokuskan pada analisis sentimen berita Turki, pelaksanaan BERT mencapai Ketepatan 0.80. Dapatan ini menekankan lagi prestasi aplikasi ini yang lebih kukuh dalam pengkategorian berita berkaitan kewangan berbanding kajian analisis sentimen berita terdahulu, yang selalunya tidak menyelidiki selok-belok bahasa dan konteks kewangan. Dengan menyasarkan berita kewangan secara khusus, kajian ini menangani jurang dalam literatur sedia ada, menunjukkan keberkesanan BERT dalam mengendalikan domain khusus di mana analisis sentimen yang tepat adalah penting untuk membuat keputusan yang berwawasan dan penglibatan pengguna.

c. Skala Kebolegunaan Sistem

Ujian antara muka sistem dimulakan dengan melibatkan sampel 15 pengguna yang berinteraksi dengan laman web NewsVibe. Setiap pengguna diminta untuk meneroka ciri dan fungsi tapak web. Berikutan interaksi mereka, pengguna diminta untuk mengisi soal selidik skala kebolegunaan sistem melalui Borang Google. Langkah ini menangkap persepsi dan pengalaman mereka mengenai kebolegunaan bahagian hadapan. Semasa pengumpulan data, respons daripada Borang Google telah disusun dan dianalisis. Skor daripada tinjauan telah dianalisis untuk mengenal pasti kekuatan dan bidang untuk penambahbaikan dalam kebolegunaan bahagian hadapan.

Jadual 3 menunjukkan kekerapan skor yang diterima daripada setiap item aspek dalam Skala Kebolegunaan Sistem. Hasil tinjauan Skala Kebolegunaan Sistem menunjukkan kepuasan pengguna yang tinggi terhadap antara muka NewsVibe, yang didapati intuitif dan mudah digunakan.

Jadual 3 Skala Kebolegunaan Sistem

No	Item	Kekerapan skor
1	Saya rasa saya ingin menggunakan laman web ini dengan kerap.	5
2	Saya mendapati laman web ini tidak rumit sangat	5
3	Saya fikir laman web ini mudah digunakan.	5
4	Saya rasa saya memerlukan sokongan orang teknikal untuk menggunakan laman web ini.	1
5	Saya dapati pelbagai fungsi dalam laman web ini disepadukan dengan baik.	5
6	Saya fikir terdapat terlalu banyak ketidakkonsistenan dalam laman web ini.	1
7	Saya berpendapat bahawa kebanyakan orang akan belajar menggunakan laman web ini dengan cepat.	5
8	Saya mendapati bahawa laman web ini sangat menyusahkan untuk digunakan.	1
9	Saya berasa sangat yakin menggunakan laman web ini.	5
10	Saya perlu belajar banyak perkara sebelum saya boleh menggunakan laman web ini.	1
11	Sejauh manakah anda akan mengesyorkan tapak web ini kepada orang lain.	5

Berdasarkan jawapan responden dan analisis yang dibuat, dapat disimpulkan bahawa kebolegunaan laman web NewsVibe ini adalah pada skala positif. Majoriti soalan positif menerima skor 5 dalam Skala Likert manakala soalan negatif menerima skor 1. Dapat juga disimpulkan bahawa objektif penilaian projek ini tercapai.

Cadangan Penambahbaikan

Bagi menangani had semasa dan mempertingkatkan lagi sistem NewsVibe, beberapa penambahbaikan masa hadapan dicadangkan. Melaksanakan keupayaan pemprosesan data masa nyata yang berterusan akan memastikan artikel berita dianalisis dan diringkaskan sebaik sahaja ia diterbitkan seurus memberikan pengguna maklumat terkini yang tersedia tanpa berlengah-lengah. Peningkatan ini akan meningkatkan dengan ketara perkaitan dan ketepatan masa kandungan berita yang dihantar kepada pengguna, menjadikan platform lebih dinamik dan responsif kepada acara masa nyata. Selain itu, menambah baik penyepaduan tapak web dan reka bentuk antara muka pengguna akan mewujudkan pengalaman pengguna yang lebih lancar dan menyeronokkan. Ini termasuk mengoptimalkan navigasi, reka letak dan responsif merentas peranti yang berbeza. Dengan memperhalusi antara muka tapak web, pengguna akan mendapati lebih mudah untuk mengakses dan berinteraksi dengan kandungan berita, yang membawa kepada peningkatan penglibatan dan kepuasan.

Di samping itu, dengan memasukkan set data yang lebih besar untuk latihan dan pengesahan, ketepatan model pembelajaran mesin boleh dipertingkatkan dengan ketara, menghasilkan pengkategorian dan analisis sentimen yang lebih tepat. Menggunakan set data yang lebih luas akan membantu dalam menangkap pelbagai nuansa yang lebih luas dalam berita kewangan, yang membawa kepada analisis yang lebih tepat dan boleh dipercayai. Tambahan pula, penyepaduan cadangan kandungan yang diperibadikan akan menyesuaikan berita dan analisis mengikut keutamaan dan minat pengguna individu lalu meningkatkan penglibatan dan kepuasan pengguna. Dengan menganalisis tingkah laku dan pilihan pengguna, sistem boleh menyampaikan kandungan yang lebih relevan dan menarik kepada setiap pengguna, dengan itu meningkatkan pengalaman keseluruhan mereka dengan platform.

KESIMPULAN

Pembangunan dan pelaksanaan NewsVibe telah menunjukkan kemajuan yang ketara dalam meningkatkan kebolehcapaian dan analisis berita kewangan melalui pembelajaran mesin lanjutan dan teknik pembelajaran mendalam. Walaupun terdapat cabaran awal dalam pengekstrakan data, had pemprosesan, dan penyepaduan algoritma, sistem kini dilengkapi dengan keupayaan teguh untuk pengkategorian berita, analisis sentimen, dan ringkasan ringkas. Kekuatannya terletak pada infrastruktur bahagian belakang yang komprehensif, antara muka bahagian hadapan yang intuitif, dan keupayaan untuk menyampaikan cerapan tepat pada masanya kepada pengguna. Melangkah ke hadapan, penambahbaikan berterusan seperti pemprosesan data masa nyata, penyepaduan tapak web yang dipertingkatkan, peningkatan ketepatan data melalui set data yang lebih besar, dan pengesyoran kandungan yang diperibadikan adalah penting untuk memperbaiki fungsi NewsVibe dan memenuhi keperluan pengguna yang berkembang serta kemajuan teknologi dalam landskap berita digital.

Kekuatan Sistem

Sistem NewsVibe mempamerkan beberapa kekuatan yang menonjol dalam mengkategorikan dan menganalisis berita kewangan. Penggunaan algoritma pengkelompokan dan analisis sentimen lanjutan seperti Peruntukan Dirichlet Terpendam (PDT) dan BERT menyediakan pengkategorian yang tepat dan berwawasan serta tafsiran sentimen artikel berita. Selain itu, penyepaduan model BART untuk ringkasan memastikan artikel yang kompleks dikurangkan kepada ringkasan ringkas, meningkatkan kefahaman pengguna. Kekuatan lain adalah infrastruktur bahagian belakang yang komprehensif untuk pemprosesan dan penyimpanan data yang cekap, serta antara muka hadapan yang mesra pengguna dengan maklum balas positif.

Kelemahan Sistem

Walau bagaimanapun, sistem ini mempunyai kekangan, terutama dalam pemprosesan data masa nyata yang berterusan yang mempengaruhi kemas kini artikel berita dan kesesuaian maklumat yang diberikan kepada pengguna. Selain itu, terdapat ruang untuk penambahbaikan

dalam penyepaduan dan reka bentuk antara muka pengguna untuk meningkatkan pengalaman pengguna secara keseluruhan.

PENGHARGAAN

Dengan rasa rendah diri dan rasa terima kasih, penulis kajian ingin merakamkan setinggi-tinggi penghargaan kepada semua pihak yang telah menyumbang dalam menyiapkan tesis ini. Terlebih dahulu, segala puji dan syukur ke hadirat Tuhan Yang Maha Esa, yang telah memberikan bimbingan, kekuatan, dan rahmatnya sepanjang perjalanan penyelidikan ini; semoga segala usaha diberkati olehNya. Terima kasih yang tidak terhingga diucapkan kepada penyelia, Prof. Madya Dr. Ravie Chandren a/l Muniyandi, atas bimbingan, nasihat, dan sokongan beliau yang tidak ternilai sepanjang proses penyelidikan. Juga, ucapan terima kasih atas sumbangan dan kemudahan yang diberikan oleh Fakulti Teknologi Maklumat dan Sains, yang memainkan peranan penting dalam menjalankan kajian ini. Sokongan infrastruktur dan keupayaan penyelidikan amat dihargai. Di samping itu, penghargaan yang tidak terhingga kepada ibu bapa dan individu lain yang dengan murah hati berkongsi kepakaran dan pandangan mereka dalam domain kewangan. Bimbingan mereka amat diperlukan dalam menilai model, merapatkan jurang yang ketara dalam pengetahuan yang diperlukan untuk projek ini, yang memfokuskan pada domain kewangan. Sumbangan mereka tidak ternilai, memastikan ketepatan dan kaitan penemuan penyelidikan.

RUJUKAN

- Abiodun, M. I., Absalom, E. E., Laith, A., Belal, A., Jia, H. 2023. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*. 622
- Abdussalam, M. F. 2022. BERT implementation on news sentiment analysis and analysis benefits on branding. *Jurnal Media Informatika Budidarma*, 6(4), 2064-2073. <https://doi.org/10.30865/mib.v6i4.4579>
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. 2019. Bert: Pre-training of Deep Bidirectional Trans-formers for Language Understanding..
- Jurgelaitis, M. L. čeponienė dan R. Butkienė.2022. Solidity Code Generation From UML State Machines in Model-Driven Smart Contract Development. *IEEE Access*.vol. 10.
- Kang, T. 2023. News Data Analysis System Based on UML and Computer Aided Technology. 2nd International Conference on Artificial Intelligence and Autonomous Robot Systems (AIARS), Bristol, United Kingdom.
- Kevin, K.J.F dan David, W. M. 2023. Addressing UX Practitioners' Challenges in Designing ML Applications: an Interactive Machine Learning Approach. In
- Lo, A.W., and Singh. M. (2023). From ELIZA to ChatGPT: The evolution of NLP and financial applications. *Journal of Portfolio Management*.
- Othman, H. and Yaakub, M.R. 2022. Implementing BERT with Ktrain library for sentiment analysis. *Journal of Visual Language and Computing*.
- Sun, C., Qiu, X., Xu, Y., & Huang, X. 2019. How to Fine-Tune BERT for Text Classification?. *Chinese Computational Linguistics*. CCL 2019. Lecture Notes in Computer Science, 11856, Springer, Cham, Switzerland. doi:10.1007/978-3-030-32381-3_16.

- Talaat, A.S. 2023. Sentiment analysis classification system using hybrid BERT models. *J Big Data* 10, 110
- Thesing, T., Feldmann, C., & Burchardt, M. 2021. Agile Versus Waterfall Project Management: Decision Model for Selecting the Appropriate Approach to a Project. *Procedia Computer Science*.
- Wan, X., Yang, J., Marinov, S., Calliess, J. P., Zohren, S., & Dong, X. 2021. Sentiment correlation in financial news networks and associated market movements. *Scientific Reports*
- Yadav, A., Jha, C.K, Sharan, A. and Vaish, V.. 2020. Sentiment analysis of financial news using unsupervised approach. *Procedia Computer Science* 167.
- Yadav, H., Patel, N., Jani, D. 2023. Fine-Tuning BART for Abstractive Reviews Summarization. In: Shukla, A., Murthy, B.K., Hasteer, N., Van Belle, JP. (eds) *Computational Intelligence. Lecture Notes in Electrical Engineering*, vol 968. Springer, Singapore.

Samira Nair Vasu Davan (A189440)

Prof. Madya Ts. Dr. Ravie Chandren A/L Muniyandi

Ketua Program PascaSiswazah

Fakulti Teknologi & Sains Maklumat

Universiti Kebangsaan Malaysia