CRIMESENSE: KLASIFIKASI JENAYAH BERASASKAN TEKNIK PEMBELAJARAN MESIN

NURUL ATHIRAH BINTI MOHAMMAD

PROF. MADYA TS. DR NORSAMSIAH SANI

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

ABSTRAK

CrimeSense: Klasifikasi Jenayah Berasaskan Teknik Pembelajaran Mesin dibangunkan dengan tujuan untuk meningkatkan keselamatan pengguna melalui penerapan kecerdasan buatan (AI) dan teknik pembelajaran mesin. Sistem ini menganalisis pola jenayah berdasarkan data sejarah seperti lokasi dan waktu kejadian, seterusnya meramalkan kawasan berisiko tinggi. Maklumat ini membantu pengguna untuk memahami taburan jenayah dengan lebih jelas serta mengenal pasti kawasan yang memerlukan perhatian. Masalah utama yang cuba diatasi adalah peningkatan kes jenayah dalam negara, yang menimbulkan kebimbangan terhadap keselamatan awam. Dalam era Revolusi Perindustrian Keempat, banyak kejadian jenayah tidak lagi dilaporkan secara rasmi kepada pihak berkuasa. Sebaliknya, maklumat sering tersebar secara tidak formal di media sosial kerana ia lebih mudah diakses dan tidak memerlukan proses yang rumit. Sistem pemantauan tradisional pula masih bergantung kepada laporan manual yang lambat dan kurang efisien, mengakibatkan kesukaran dalam mengenal pasti pola kejadian secara menyeluruh. Justeru, sistem ini dibangunkan sebagai alternatif moden yang menggunakan pendekatan berasaskan data bagi mengenal pasti kawasan berisiko dan membantu dalam pemahaman trend jenayah. Data yang diperoleh daripada laman web relevan akan melalui fasa pembersihan, pengekstrakan ciri, serta proses pengelasan menggunakan model seperti Support Vector Machine (SVM), Random Forest, dan Decision Tree. Penilaian prestasi model dilakukan menggunakan metrik seperti ketepatan dan recall. Paparan heatmap interaktif turut disediakan bagi membolehkan pengguna membuat penilaian visual terhadap kawasan panas jenayah. Privasi pengguna dan etika pengendalian data juga dijaga dengan teliti sepanjang proses pembangunan sistem ini. Secara keseluruhannya, sistem ini diharapkan dapat membantu mencipta persekitaran yang lebih selamat melalui pendekatan analisis data yang efektif, tidak invasif, dan mesra pengguna.

Kata Kunci: Jenayah, Pembelajaran Mesin Ramalan Jenayah.

PENGENALAN

Kecerdasan buatan (AI) semakin menjadi sebahagian daripada kehidupan seharian kita, dengan aplikasi yang meluas dalam pelbagai bidang, termasuk perniagaan, penjagaan kesihatan, dan keselamatan. Teknologi ini berkembang pesat, menawarkan penyelesaian yang lebih pintar dan lebih cekap untuk menangani cabaran kompleks yang dihadapi oleh masyarakat. Dari analitik ramalan hingga pembelajaran mesin, AI membolehkan kita memahami corak tersembunyi dalam data dan membuat keputusan yang lebih baik berdasarkan maklumat yang tepat.

Menyedari potensi besar kecerdasan buatan (AI) dalam bidang keselamatan, pembangunan Sistem Klasifikasi Jenayah Berasaskan Teknik Pembelajaran Mesin diwujudkan untuk menganalisis data sejarah jenayah dan mengenal pasti corak kejadian yang berlaku di sesebuah kawasan. Sistem ini direka untuk mengklasifikasikan jenis jenayah berdasarkan ciri-ciri tertentu serta memvisualisasikan kawasan berisiko tinggi melalui peta panas (heatmap). Segala analisis dijalankan menggunakan data sedia ada secara anonim, tanpa penglibatan data peribadi pengguna, selaras dengan prinsip privasi dan etika pengendalian data.

Sistem ini menggunakan teknik AI seperti ramalan dan pembelajaran mesin, termasuk pengelompokan, pengesanan anomali dan pemprosesan bahasa semula jadi untuk menganalisis data sejarah jenayah dan laporan masa nyata. Melalui analisis ini, sistem dapat mengenal pasti corak jenayah yang terjadi dan meramalkan kawasan yang berisiko tinggi. Tambahan pula, papan pemuka visualisasi data memberikan pengguna gambaran yang jelas tentang tren keselamatan yang berlaku serta membantu mereka membuat keputusan yang lebih tepat berdasarkan data yang tersedia.

Oleh kerana maklumat jenayah yang tidak telus serta komunikasi yang terhad antara pihak berkuasa dan masyarakat, ia membawa kepada kegagalan dalam menyampaikan maklumat penting kepada orang awam serta menyebabkan ketidakpercayaan dan melemahkan kerjasama antara masyarakat dan pihak berkuasa. Masalah maklumat asimetri yang dihadapi oleh KNP hanya boleh diselesaikan jika KNP membangunkan statistik jenayah yang lebih telus dan meningkatkan saluran komunikasi untuk menyampaikan maklumat jenayah penting kepada orang awam (Lee, & Kim,. 2020). Dengan maklumat jenayah yang penting ini, masyarakat dijangka dapat membuat keputusan yang lebih baik mengenai isu keselamatan mereka. Dengan pendekatan yang proaktif ini, sistem ramalan ini membuktikan bagaimana kecerdasan buatan

boleh digunakan untuk membangunkan penyelesaian keselamatan komuniti yang lebih cekap dan inovatif, sambil mengekalkan perlindungan terhadap privasi pengguna.

KAJIAN LITERATUR

Model klasifikasi jenayah berasaskan pembelajaran mesin bertujuan untuk mengenal pasti serta meramalkan risiko kejadian jenayah di kawasan awam. Pelbagai teknik analisis data digunakan dalam kajian lepas, termasuk mengenal pasti corak risiko berdasarkan lokasi dan masa. Walaupun memberi manfaat besar, cabaran dari segi ketepatan dan kebolehgunaan model dalam masyarakat masih wujud. Oleh itu, kajian lanjut amat penting untuk meningkatkan keberkesanan sistem ini.

Dalam teknik pembelajaran terawasi, model seperti Decision Tree dan Random Forest digunakan untuk mengenal pasti faktor berkaitan kawasan atau masa berisiko tinggi. Logistic Regression pula membantu mengira kebarangkalian jenis jenayah tertentu (Chen & Guestrin, 2019). Untuk data bersiri, model seperti RNN dan LSTM berkesan dalam mengenal pasti pola masa, manakala K-Means dan DBSCAN pula digunakan untuk mengesan kawasan jenayah berisiko tinggi (Sivaranjani, 2022). Ramalan daripada model ini sering disepadukan ke dalam papan pemuka masa nyata, membolehkan pihak berkuasa dan orang awam bertindak lebih awal (Nguyen, 2023).

Dalam kajian oleh Boukabous dan Azizi (2022), model BERT digunakan untuk analisis sentimen daripada 70,000 tweet berkaitan jenayah. Ia mencatat ketepatan tinggi iaitu 94.91%, mengatasi model seperti SVM dan Random Forest. Seterusnya, Zamri dan Tahir (2021) pula menggunakan model CNN seperti VGG19 dan ResNet untuk mengenal pasti jenayah ragut melalui imej CCTV. VGG19 menunjukkan ketepatan tertinggi sekitar 81%.

Dao dan Sappa (2024) menggunakan gabungan LSTM-GRU untuk meramalkan kadar jenayah di AS berdasarkan data sosio-ekonomi. Model berprestasi tinggi di negeri berisiko, namun kurang tepat di negeri berisiko rendah. Manakala, kajian oleh Shohan dan Akash (2022) di Bangladesh menggunakan pelbagai model termasuk Random Forest dan XGBoost pada 6574 data jenayah. Setelah menggunakan teknik SMOTE, model XGBoost mencatat ketepatan 59%.

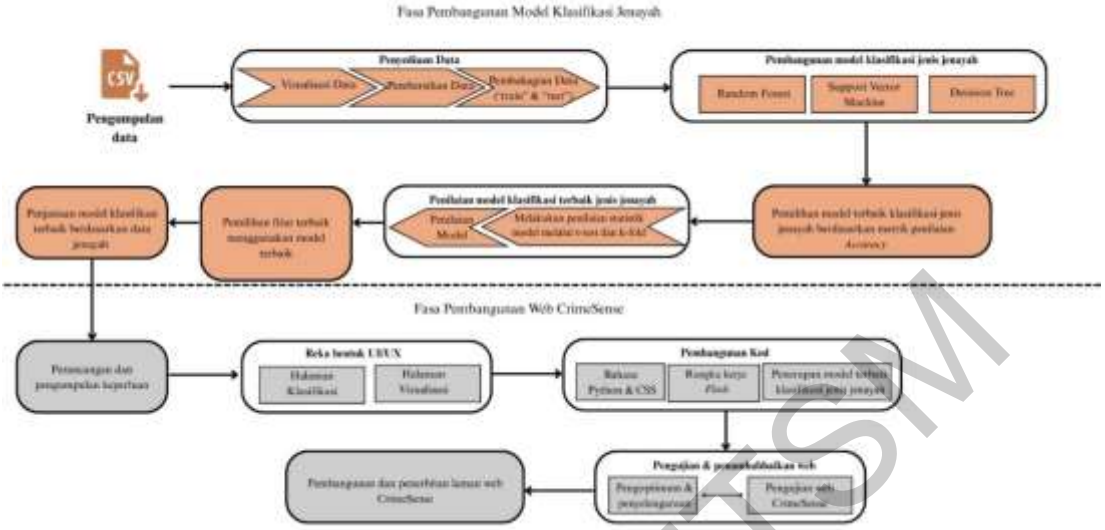
Safat (2021) menggunakan model LSTM, ARIMA, dan XGBoost untuk meramal jenayah di Chicago dan Los Angeles. XGBoost menunjukkan ketepatan tertinggi, manakala ARIMA lebih sesuai untuk ramalan tren jangka panjang. Disamping itu, kajian oleh Abdalrdha et al. (2024) menggunakan kamus pintar dan analisis sentimen pada tweet dalam bahasa Arab. Model Decision Tree dan Random Forest mencapai ketepatan hingga 97%.

Ikwen et al. (2024) pula menggunakan rangkaian neural MLP berdasarkan data dari Cross River State. Model mencatat ketepatan 74% dan F1-score 0.79, membuktikan potensinya dalam meramal kawasan berisiko. Selain itu, Sevillano dan Domínguez-Morales (2023) menggunakan model seperti ANN, SVM, dan Random Forest untuk meramal sama ada mangsa keganasan rumah tangga akan menarik diri dari proses undang-undang. Model ANN mencatat ketepatan tertinggi iaitu 91.16%. Tambahan lagi, kajian oleh Veena (2022) membandingkan SVM, KNN, GMM, dan K-Means untuk mengesan jenayah siber. Model GMM menunjukkan ketepatan tertinggi iaitu 96.56%, manakala SVM mencatat 89%.

Akhir sekali, Prabhu dan Aniruddha (2023) menggunakan lima model termasuk XGBoost Regressor untuk meramal kadar jenayah berdasarkan data dari NCRB. Model XGBoost Regressor mencatat ketepatan tertinggi iaitu 93.20%, menunjukkan keberkesanannya dalam mengenal pasti kawasan berisiko tinggi dan membantu agihan sumber keselamatan.

METODOLOGI

Metodologi CRIPS-DM telah dijadikan asas rujukan dalam merancang dan melaksanakan pembangunan projek ini secara menyeluruh.



Rajah 1 Carta Alir Pembangunan Model dan Sistem

Rajah 1 menunjukkan carta alir bagi pembangunan model klasifikasi jenayah dan sistem web yang diintegrasikan bersama model pembelajaran mesin terlatih. Proses pembangunan model dijalankan menggunakan bahasa pengaturcaraan Python, dengan Jupyter Notebook dan Visual Studio Code sebagai platform. Selepas pemilihan model terbaik daripada tiga model yang diuji, model tersebut diintegrasikan ke dalam sistem web menggunakan Flask sebagai rangka kerja backend, manakala HTML dan CSS digunakan untuk membina antaramuka pengguna. Setelah proses integrasi selesai, sistem diuji untuk memastikan ia memenuhi keperluan dan objektif kajian

Fasa Penyediaan Data

Set data yang digunakan dalam kajian ini diperoleh daripada laman web Kaggle, iaitu Los Angeles Crime Dataset (2020–Present) yang dimuat naik oleh Nathaniel Lybrand dan mengandungi lebih 955,000 rekod insiden jenayah sejak tahun 2020. Data ini merangkumi pelbagai maklumat seperti tarikh dan masa kejadian, lokasi, maklumat mangsa, jenis premis, status kes, dan deskripsi jenayah. Sebelum digunakan dalam model klasifikasi, data ini telah melalui proses pembersihan dan penyediaan seperti membuang nilai tidak lengkap, duplikasi, serta menyeragamkan format teks dan tarikh. Nilai hilang ditangani dengan kaedah yang sesuai mengikut jenis data, manakala satu ciri baharu iaitu “Crime Type” turut dibina berdasarkan pengelompokan jenis jenayah

yang sedia ada. Proses ini penting untuk memastikan kualiti data adalah baik dan bersesuaian sebagai input kepada model pembelajaran mesin yang akan dibangunkan.

Date Rptd	DATE OCC	TIME OCC	AREA	AREA NAME	Rpt Dist No	Part 1-2	Crm Cd	Crm Cd Desc	...	Status	Status Desc	Crm Cd 1	Crm Cd 2
03/01/2020 12:00:00 AM	03/01/2020 12:00:00 AM	2130	7	Wilshire	784	1	510	VEHICLE - STOLEN	...	AA	Adult Arrest	510.0	998.0
02/09/2020 12:00:00 AM	02/08/2020 12:00:00 AM	1800	1	Central	182	1	330	BURGLARY FROM VEHICLE	...	IC	Invest Cont	330.0	998.0
11/11/2020 12:00:00 AM	11/04/2020 12:00:00 AM	1700	3	Southwest	356	1	480	BIKE - STOLEN	...	IC	Invest Cont	480.0	NaN
05/10/2023 12:00:00 AM	03/10/2020 12:00:00 AM	2037	9	Van Nuys	964	1	343	SHOPLIFTING-GRAND THEFT ($950.01 & OVER)	...	IC	Invest Cont	343.0	NaN
08/18/2022 12:00:00 AM	08/17/2020 12:00:00 AM	1200	6	Hollywood	666	2	354	THEFT OF IDENTITY	...	IC	Invest Cont	354.0	NaN

Jadual 2 Menunjukkan 5 baris pertama set data

Fasa Pembangunan Model

Fasa pembangunan model dalam projek ini merangkumi pembinaan dan penilaian tiga algoritma pembelajaran mesin utama: Random Forest (RF), Support Vector Machine (SVM), dan Decision Tree (DT). Model-model ini dilatih menggunakan dataset jenayah yang telah dipraolah, terdiri daripada 50,000 sampel yang mengandungi atribut seperti umur mangsa, jantina mangsa, masa kejadian, lokasi kejadian (nama kawasan), jenis premis, jenis senjata, serta deskripsi jenayah yang telah diklasifikasikan ke dalam kategori tertentu. Proses pembangunan melibatkan beberapa peringkat termasuk pemilihan ciri yang relevan, pembahagian data kepada set latihan dan set ujian, serta penalaan hiperparameter (hyperparameter tuning) untuk mengoptimumkan prestasi setiap model. Fasa ini menggunakan bahasa pengaturcaraan Python dan dijalankan di platform Visual Studio Code serta Google Colaboratory untuk memastikan kelancaran dan keberkesanan proses pembangunan model. Ketiga-tiga model mempunyai ciri tersendiri yang menyumbang kepada ketepatan klasifikasi. RF menggunakan *ensemble* pelbagai pokok keputusan untuk mengurangkan variasi dan *overfitting*. SVM membina *hyperplane* dengan margin maksimum antara kelas dan berupaya menangani data berdimensi tinggi. DT pula menawarkan visualisasi aliran keputusan yang jelas dan mudah ditafsir. Dalam fasa ini, parameter seperti bilangan *estimators* (RF), jenis kernel (SVM), dan kedalaman pokok (DT) disesuaikan bagi meningkatkan prestasi model. Berdasarkan metrik penilaian seperti ketepatan, skor F1, dan matriks kekeliruan, model

terbaik akan dikenalpasti dan dipilih untuk integrasi dalam fasa pembangunan sistem seterusnya.

Fasa Pengujian Model

Dalam fasa pengujian model, prestasi tiga model pembelajaran mesin Random Forest, Support Vector Machine (SVM), dan Decision Tree akan dinilai menggunakan metrik ketepatan (Accuracy) dan skor F1 (F1-Score). Ketepatan ini mengukur sejauh mana model meramalkan kelas dengan betul, manakala skor F1 menilai keseimbangan antara *precision* dan *recall*, penting untuk mengatasi isu ketidakseimbangan data. Selain itu, ujian *T-Test* dan *p-value* dijalankan ke atas dua model terbaik berdasarkan skor F1 untuk menentukan perbezaan prestasi yang signifikan secara statistik. Analisis ini membolehkan pemilihan model akhir yang paling stabil dan sesuai untuk digunakan dalam sistem klasifikasi jenayah.

Fasa Pembangunan Sistem

Pembangunan sistem CrimeSense melibatkan dua komponen utama iaitu front-end dan back-end. Front-end bertanggungjawab untuk membina antara muka pengguna yang mesra dan mudah digunakan, membolehkan pengguna memasukkan data serta menerima maklum balas secara interaktif. Rekabentuk front-end memberi penekanan kepada kesederhanaan dan kecekapan bagi memastikan proses input data dan paparan ramalan berjalan lancar tanpa kekeliruan. Sementara itu, back-end pula mengendalikan logik aplikasi, pengurusan data, serta integrasi model pembelajaran mesin yang digunakan untuk meramal jenis jenayah berdasarkan input pengguna. Dalam pembangunan back-end, rangka kerja Flask dipilih kerana keupayaannya untuk menyokong aplikasi web yang ringan dan mudah diselenggara, selain membenarkan integrasi lancar dengan model yang dibangunkan menggunakan bahasa Python. Pendekatan pembangunan ini memastikan sistem CrimeSense dapat beroperasi secara efisien dan responsif, sekaligus memberikan pengalaman pengguna yang baik serta ketepatan dalam hasil ramalan yang disediakan. Keseluruhan proses pembinaan sistem ini bertujuan untuk menghasilkan aplikasi web yang praktikal dan berfungsi dengan stabil dalam membantu pengguna memahami dan meramal pola jenayah di persekitaran mereka.

Fasa Pengujian dan Penambahbaikan

Setelah pembangunan sistem diselesaikan, proses pengujian dijalankan bagi menilai tahap keberkesanan serta mengenal pasti sebarang kelemahan atau isu dalam fungsi

sistem. Hasil daripada pengujian ini digunakan sebagai asas untuk melakukan penyesuaian dan penambahbaikan, agar sistem dapat beroperasi dengan lebih cekap dan memenuhi keperluan pengguna secara optimum.

KEPUTUSAN DAN PERBINCANGAN

CrimeSense, ialah sebuah sistem klasifikasi jenayah yang telah berjaya dibangunkan dengan lengkap berserta dokumentasinya. Projek ini melibatkan dua bentuk pengujian utama. Pertama ialah pengujian terhadap model pembelajaran mesin seperti SVM, Random Forest dan Decision Tree bagi menilai ketepatan klasifikasi. Seterusnya, ialah pengujian tahap kepuasan pengguna yang dijalankan untuk menilai prestasi dan keberkesanan laman web CrimeSense daripada perspektif pengguna.

Pengujian Model Pembelajaran Mesin CrimeSense

Pengujian model merupakan langkah penting dalam menilai keupayaan dan keberkesanan model pembelajaran mesin yang dibangunkan. Penilaian ini dijalankan dengan membandingkan ketepatan (accuracy) dan skor F1 bagi set latihan dan ujian bagi ketiga-tiga model iaitu Random Forest, SVM dan Decision Tree.

Jadual 1 Perbandingan Ketepatan Antara Model

Model	**Train Accuracy**	**Test Accuracy**	**Train F1**	**Test F1**
Random Forest	1.0000	0.9966	1.0000	0.9964
SVM	0.3950	0.3949	0.33259	0.3194
Decision Tree	0.9998	0.9992	0.9998	0.9991

Berdasarkan Jadual 1, model Random Forest dan Decision Tree menunjukkan prestasi yang sangat baik dengan nilai ketepatan dan F1-score melebihi 99% bagi set latihan dan ujian, sekali gus menandakan keupayaan model dalam mengenal pasti corak jenayah tanpa berlaku overfitting. Sebaliknya, model SVM menunjukkan prestasi yang jauh lebih rendah dengan nilai ketepatan sekitar 39% dan F1-score hanya 31.9%, mencadangkan bahawa model ini kurang sesuai tanpa penalaan parameter atau pemilihan ciri yang lebih spesifik.

Secara keseluruhan, Random Forest dan Decision Tree merupakan model yang paling berkesan bagi sistem CrimeSense kerana berjaya mengekalkan ketepatan tinggi secara konsisten. Namun, model Decision Tree telah dipilih sebagai model akhir kerana ia menawarkan ketelusan dalam penjelasan keputusan *(explainability*), mudah untuk ditafsir oleh pengguna bukan teknikal, dan lebih ringan dari segi masa latihan dan sumber pemprosesan berbanding *Random Forest* yang melibatkan gabungan beratus-ratus pokok keputusan. Keupayaan Decision Tree untuk memberikan prestasi hampir setara dengan Random Forest tetapi dengan struktur yang lebih mudah dan mesra pengguna menjadikannya pilihan paling sesuai untuk diintegrasikan ke dalam sistem CrimeSense.

Bagi memastikan perbezaan prestasi model bukan disebabkan oleh kebetulan semata-mata, ujian statistik t-test dua sampel dilakukan terhadap nilai metrik bagi ketiga-tiga model utama iaitu Random Forest, SVM dan Decision Tree. Jadual 2 menunjukkan nilai *t-statistic* dan *p-value* bagi perbandingan setiap pasangan model.

Jadual 2 Perbandingan T-test Antara Model

Model	T-Statistic	p-Value
Random Forest vs SVM	31.7490	0.000000
Random Forest vs Decision Tree	-2.6817	0.018341
SVM vs Decision Tree	-31.7796	0.000000

Berdasarkan Jadual 2, semua nilai p-value adalah kurang daripada 0.05, menunjukkan perbezaan prestasi antara model adalah signifikan secara statistik. Sebagai contoh, perbandingan antara Random Forest dan SVM menunjukkan nilai t-statistic yang sangat tinggi (31.7490) dengan p-value 0.000000, menandakan Random Forest mempunyai prestasi jauh lebih baik berbanding SVM. Selain itu, perbandingan antara Random Forest dan Decision Tree juga menunjukkan perbezaan signifikan (p-value = 0.018341), namun nilai t-statistic yang negatif menandakan Decision Tree memberikan hasil yang sedikit lebih baik dalam konteks metrik yang diuji. Walaupun Random Forest dan Decision Tree kedua-duanya menunjukkan prestasi yang baik secara umum, hasil t-test dan pertimbangan terhadap kebolehfahaman serta keperluan model yang mudah dijelaskan kepada pengguna bukan teknikal menjadikan Decision Tree dipilih sebagai model akhir bagi sistem ini. Keupayaannya memberikan hasil yang hampir

setara dengan Random Forest, disamping visualisasi yang mudah dan struktur yang boleh dijelaskan dengan telus, menjadikannya pilihan yang lebih sesuai untuk sistem pengklasifikasian jenayah kampus seperti CrimeSense.

Pengujian Kepuasan Pengguna

Ujian kepuasan pengguna dijalankan untuk menilai sejauh mana pengguna berpuas hati terhadap sesuatu sistem atau perkhidmatan berdasarkan pengalaman mereka. Maklum balas biasanya diperoleh melalui soal selidik, temubual atau pemerhatian, yang merangkumi aspek seperti kemudahan penggunaan, fungsi dan reka bentuk. Tujuan utama ujian ini adalah untuk mengenal pasti bahagian yang perlu ditambah baik agar sistem lebih mesra pengguna dan memenuhi keperluan sebenar.

Jadual 3 Tafsiran Skala Skor Min

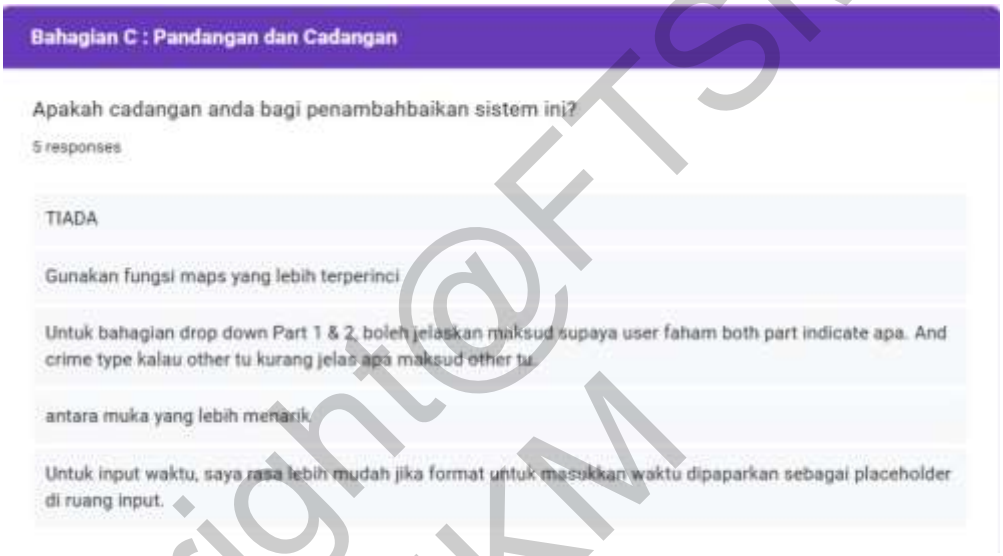
Skor Min	Tafsiran
1.00 – 2.32	Rendah
2.33 – 3.65	Sederhana
3.66 – 5.00	Tinggi

Jadual 3 memperincikan interpretasi bagi skala skor min dalam ujian yang dijalankan. Skor dalam julat 1.00 hingga 2.32 dikategorikan sebagai tahap rendah, manakala julat 2.33 hingga 3.65 mewakili tahap sederhana, dan skor antara 3.66 hingga 5.00 menunjukkan tahap yang tinggi.

Jadual 4: Skor min pengujian kepuasan pengguna

No	Item	Min
1	Adakah antara muka sistem ini mudah difahami dan digunakan?	4.00
2	Adakkah input yang perlu dimasukkan oleh pengguna adalah jelas dan mudah?	4.00
3	Adakah maklumat hasil klasifikasi yang dipaparkan mudah difahami?	3.00
4	Sejauh mana anda berpuas hati dengan sistem ini secara keseuruhan?	3.00
	Min keseluruhan	3.50

Jadual 4 memperincikan skor min bagi pengujian kepuasan pengguna terhadap laman web CrimeSense. Secara keseluruhan, setiap item soal selidik menunjukkan tahap kepuasan yang baik berdasarkan skor min yang berada pada julat sederhana tinggi. Dua soalan pertama mencatatkan skor tertinggi iaitu 4, menggambarkan penerimaan positif pengguna terhadap fungsi atau paparan tertentu dalam sistem. Sementara itu, soalan ketiga dan keempat masing-masing memperoleh skor 3, yang masih menunjukkan tahap kepuasan yang memuaskan namun berpotensi untuk penambahbaikan pada aspek yang dinilai



Rajah 3: Cadangan Penambahbaikan

Berdasarkan Rajah 3 iaitu maklum balas dan cadangan daripada responden, beberapa aspek dalam sistem CrimeSense perlu ditambah baik bagi meningkatkan pengalaman pengguna. Antara cadangan yang dikemukakan termasuklah penambahbaikan fungsi peta agar lebih terperinci, penjelasan yang lebih jelas terhadap bahagian 'Part 1 & 2' serta kategori 'Other' bagi jenis jenayah agar pengguna lebih faham maksud setiap pilihan. Selain itu, antara muka sistem disarankan agar direka bentuk dengan lebih menarik. Manakala bagi input masa, pengguna mencadangkan agar format masa dipaparkan sebagai placeholder untuk memudahkan pengisian.

KESIMPULAN

Secara keseluruhannya, CrimeSense telah berjaya dibangunkan sebagai sebuah sistem klasifikasi jenayah yang menggunakan pembelajaran mesin untuk membantu orang awam memahami jenis-jenis jenayah berdasarkan data sedia ada. Walaupun terdapat

beberapa kekangan semasa pembangunan seperti kekangan peranti dan format input, sistem ini tetap dapat disiapkan dengan mengambil kira maklum balas pengguna. Diharapkan CrimeSense dapat menjadi alat sokongan yang bermanfaat dalam meningkatkan kesedaran keselamatan dan membantu pengguna membuat keputusan yang lebih bijak berasaskan maklumat.

Kekuatan

Antara kekuatan utama projek ini ialah kejayaannya membangunkan aplikasi klasifikasi jenayah berasaskan web yang boleh digunakan secara terus oleh pengguna tanpa memerlukan proses log masuk atau pendaftaran, sekaligus memastikan ciri penggunaan secara anonim. Dengan pemilihan model Decision Tree, sistem ini mampu menghasilkan keputusan klasifikasi dengan ketepatan tinggi (99.92%) serta latihan yang pantas, menjadikannya sesuai untuk diintegrasikan ke dalam antaramuka web Flask dengan prestasi yang konsisten. Selain itu, pemilihan hanya 9 ciri penting secara manual membantu menjadikan borang input lebih ringkas, mudah diisi, dan mudah difahami oleh pengguna biasa.

Kekangan

Terdapat beberapa kekangan yang dihadapi sepanjang pembangunan. Antaranya ialah had masa pemprosesan apabila melibatkan jumlah data yang besar, terutamanya ketika menjalankan latihan model dan penalaan hiperparameter. Kekangan teknikal seperti Timeout Error juga mempengaruhi keupayaan untuk menyesuaikan model secara optimum dalam persekitaran pembangunan yang terhad. Selain itu, meskipun sistem ini direka dengan kesederhanaan dan kefungsian sebagai keutamaan, masih terdapat ruang untuk penambahbaikan dari segi rekabentuk antara muka pengguna serta fungsi tambahan seperti statistik laporan atau peta panas (heatmap) lokasi jenayah. Akhir sekali, sistem ini masih berdasarkan data statik (CSV), dan integrasi dengan pangkalan data masa nyata boleh dipertimbangkan pada masa akan datang untuk meningkatkan keboleh suaian sistem terhadap situasi sebenar.

Cadangan Masa Depan

Bagi memastikan CrimeSense terus relevan dan berskala dalam pembangunan sistem klasifikasi jenayah, beberapa penambahbaikan strategik disarankan berdasarkan pemerhatian sepanjang pelaksanaan projek ini:

i. Ujian Validasi Luaran Melalui Dataset dari Lokasi Berbeza dengan melaksanakan kajian pengesahan silang menggunakan data dari kawasan yang berlainan boleh meningkatkan kebolehpercayaan model dan memastikan bahawa sistem berfungsi secara konsisten merentasi pelbagai konteks persekitaran dan geografi setempat.

ii. Integrasi dengan Sistem Pemantauan Keselamatan Masa Nyata dengan ini, sistem boleh ditambah baik dengan menghubungkannya kepada sistem pemantauan CCTV atau pangkalan data polis bantuan secara masa nyata bagi meningkatkan keupayaan pengesanan awal dan pemberitahuan automatik kepada pihak berkuasa.

iii. Antaramuka Interaktif dan Peta Jenayah Dinamik Meningkatkan pengalaman pengguna dengan menambah visualisasi interaktif seperti peta panas lokasi jenayah dan carta statistik dinamik dapat memperkayakan kefahaman pengguna dan memudahkan tindakan pencegahan berdasarkan data yang boleh ditindak

PENGHARGAAN

Alhamdulillah, saya bersyukur ke hadrat Ilahi atas kurniaan-Nya yang memberi saya kekuatan untuk menyiapkan projek ini sebagai sebahagian daripada syarat Ijazah Sarjana Muda Sains Komputer dengan Kepujian. Saya amat berterima kasih kerana dengan kesabaran dan tekad, segala cabaran yang dilalui dapat diatasi, dan pencapaian ini adaah hasil rahmat-Nya.

Saya ingin merakamkan ucapan terima kasih yang tidak terhingga kepada pensyarah-pensyarah, khususnya penyelia saya , Prof. Madya Ts. Dr. Nor Samsiah Sani, di atas bimbingan dan masa yang dicurahkan untuk membantu saya menyempurnakan projek ini. Ilmu dan pengalaman belliau amat memberi manfaat dalam menjayakan projek ini.

Terima kasih juga kepada Fakulti Teknologi dan Sains Maklumat atas sumber dan fasiliti yang disediakan. Akhir sekali, saya ingin mengucapkan penghargaan kepada Ibu bapa ,keluarga dan rakan-rakan yang telah memberi sokongan padu sepanjang perjalanan menyiapkan projek ini.

RUJUKAN

Abdalrdha., Al-Bakry, M., & Farhan. (2024). Arabic crime tweet filtering and prediction using machine learning. Iraqi Journal for Computers and Informatics, 50(1), 73–85. https://doi.org/10.25195/ijci.v50i1.479

Abu, & Yusoff. (2022). Faktor-faktor penyumbang 'fear of crime' dalam kalangan warga emas di daerah Yan, Kedah. e-BANGI, 19(6), 15–25.

Administrator. (2024, February 8). Data analytics in crime prediction and prevention. iResearchNet. Criminal Justice. https://criminal-justice.iresearchnet.com/criminal-justice-process/impact-of-technology/data-analytics-in-crime-prediction-and-prevention/

Ali, (2022, August 22). Supervised machine learning. DataCamp. https://www.datacamp.com/blog/supervised-machine-learning

Boukabous, & Azizi. (2022). Crime prediction using a hybrid sentiment analysis approach based on the bidirectional encoder representations from transformers. Indonesian Journal of Electrical Engineering and Computer Science, 25(2), 1131. https://doi.org/10.11591/ijeecs.v25.i2.pp1131-1139

Lee, & Kim (2020). A problem of crime control policy in South Korea: A challenge of asymmetric information. Journal of Scientific Research and Reports, 26(3), 80–85. https://doi.org/10.9734/jsrr/2020/v26i330239

Putra, K. (2022, February 14). Support Vector Machine Algorithm. School of Information Systems. https://sis.binus.ac.id/2022/02/14/support-vector-machine-algorithm/

Shilpa. (2023, September 20). Architectural design in software engineering. ArtOfTesting. https://artoftesting.com/architectural-design-in-software-engineering

What is Hyperplane? Role in Machine Learning. (2024, May 27). Deepchecks. https://www.deepchecks.com/glossary/hyperplane/

Zulfa, Y., Sani, N., & Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia. (2024). Sherlocked: Pengiklanan Diperibadikan Melalui Analisis Pemakaian Berdasarkan Teknik Pembelajaran Mendalam. In *Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia*.

Nurul Athirah Binti Mohammad (A193199)
Prof. Madya Ts. Dr. Norsamsiah Sani
Fakulti Teknologi & Sains Maklumat
Universiti Kebangsaan Malaysia