

# **APLIKASI WEB MERAMALKAN PENYENARAIAAN PALSU BERDASARKAN TEKNIK PEMBELAJARAN MESIN**

NUR ATIQA BINTI SUHAIMI  
TS. DR NOR SAMSAH BINTI SANI

## **ABSTRAK**

Internet telah menggantikan media tradisional dan menjadi salah satu platform utama untuk menyebarkan maklumat sama ada ia benar atau palsu. Penipuan dan penyelewengan sering berlaku dalam pasaran hartanah Malaysia kerana ketiadaan undang-undang yang ketat. Ini kerana terdapat agensi yang tidak bertanggungjawab menerbitkan penyenaiaian hartanah palsu atau tidak tersedia pada harga yang murah dan berpatutan sebagai cara mereka menarik pelanggan baharu bagi mengaut keuntungan sendiri. Untuk mengatasi masalah ini dalam sektor hartanah, aplikasi web semakin pantas akan dibangunkan dengan menggunakan teknik pembelajaran mesin untuk meramal penyenaiaian palsu dalam hartanah yang tersebar di internet. Dalam kajian ini, data dari dua laman web hartanah tempatan dikumpul bagi meramal penyenaiaian palsu dalam hartanah. Seterusnya, metodologi yang digunakan bagi membangunkan aplikasi web ini adalah model CRISP-DM yang terdiri daripada 6 fasa. Fasa pertama iaitu fasa pemahaman perniagaan dimana objektif kajian ditentukan. Fasa kedua iaitu pemahaman data di mana data dari dua laman web hartanah tempatan dikumpul bagi meramal penyenaiaian palsu dalam hartanah. Set data ini merangkumi klasifikasi penyenaiaian hartanah dan 12 ciri hartanah dikumpulkan dengan beberapa data nilai yang hilang. Bagi fasa ketiga fasa penyediaan data, proses pra-pemprosesan seperti mengisi data yang hilang dan atribut teks diklasifikasikan sebagai angka dijalankan. Fasa pemodelan merupakan fasa keempat di mana tiga model yang berbeza diterapkan dan dibandingkan iaitu mesin sokongan vektor (SVM), hutan rawak dan pokok keputusan. Ketiga-tiga model ini diuji dan dinilai bagi mendapatkan keputusan yang terbaik. Analisis prestasi perbandingan antara ketiga-tiga model tersebut dalam fasa lima iaitu fasa penilaian menunjukkan bahawa hutan rawak merupakan model terbaik dalam meramalkan penyenaiaian palsu dalam kajian ini kerana berjaya mencapai ketepatan yang tertinggi iaitu 99.58%. Pendekatan ini dapat diterapkan sebagai petunjuk bagi meramal penyenaiaian palsu dalam hartanah. Seterusnya bagi fasa keenam iaitu fasa penggunaan, aplikasi web dibangunkan dengan menintegrasikan model pembelajaran mesin yang terbaik iaitu model hutan rawak. Menggunakan teknik pembelajaran mesin ini, sistem boleh mengenal pasti penyenaiaian palsu dalam hartanah.

## **1 PENGENALAN**

Di era moden yang didukung oleh teknologi canggih menyebabkan dunia berubah dengan cepat dengan wujudnya beberapa kelebihan dunia digital tetapi ianya juga mempunyai keburukannya. Antara keburukannya ialah penyebaran berita palsu dalam internet. Berita palsu ini tersebar untuk menjatuhkan reputasi seseorang atau organisasi. Salah satu teknologi canggih yang kian berleluasa ialah penggunaan teknologi kecerdasan buatan (AI). Hal ini dapat dilihat dari impak teknologi AI terhadap kehidupan seharian manusia. Bidang kecerdasan buatan (AI) sedang mengalami tempoh kemajuan yang sengit disebabkan oleh penyatuan beberapa teknologi utama. Teknologi ini boleh belajar, meramal, menganalisis dan membuat kesimpulan dengan cepat daripada manusia. Ia dibangunkan untuk menyelesaikan masalah dari pelbagai bidang, pengimejan dan pengecaman pertuturan. Ia menyelesaikan masalah dengan menggunakan pelbagai teknik seperti pembelajaran mesin dan pembelajaran mendalam

berdasarkan ketepatan yang diberikan oleh teknik tersebut. Mesin akan diprogramkan untuk mencuba beribu-ribu tugas. Semasa proses ini, mesin akan memahami dan mempelajari konsep yang diperlukan untuk menyelesaikan tugas. Selain itu, ia hanya memahami bahasa manusia bagi membuat keputusan tanpa membuat kesilapan. Teknologi ini akan mengumpul data dan seterusnya menganalisis bagi membuat keputusan.

Pembelajaran mesin adalah sebahagian daripada kecerdasan buatan yang membantu dalam membuat sistem yang boleh belajar dan melakukan tindakan yang berbeza (Donepu, 2019). Pembelajaran mesin digunakan di sektor yang berbeza untuk melakukan tugas yang berbeza. Kebanyakan algoritma pembelajaran mesin digunakan untuk tujuan ramalan atau mengesan sesuatu yang tersembunyi. Penggunaan pembelajaran mesin adalah terbukti membantu dalam mengesan atau meramal penipuan. Dalam bidang hartanah, teknologi ini digunakan bagi meramal kebarangkalian penyenaian palsu. Algoritma pembelajaran mesin akan mengesan berita palsu secara automatik sebaik sahaja mereka telah dilatih. Pembelajaran mesin digunakan dalam pelbagai tugas pengkomputeran di mana mereka bentuk dan pengaturcaraan jelas algoritma dengan prestasi yang baik adalah sukar atau tidak boleh dilaksanakan. Pembelajaran mesin berkait rapat dengan (dan sering bertindih dengan) statistik pengiraan, yang turut memfokuskan kepada membuat ramalan melalui penggunaan komputer. Dalam bidang analisis data, pembelajaran mesin ialah kaedah yang digunakan untuk merangka model dan algoritma kompleks bagi melakukan ramalan.

Para penyelidik menggunakan pelbagai algoritma yang berbeza untuk mengesan berita palsu. Penyelidik (Wang, 2017) berpendapat bahawa pengesanan berita palsu adalah satu cabaran yang besar. Mereka telah menggunakan pembelajaran mesin untuk mengesan berita palsu. Ramai saintis percaya bahawa isu berita palsu ini boleh ditangani dengan mengaplikasikan teknik pembelajaran mesin dan kecerdasan buatan (AI). Hal ini kerana, algoritma pembelajaran mesin mula berfungsi dengan baik bagi menangani pelbagai masalah seperti pengesanan suara, pengecaman imej malah ia juga digunakan untuk mengesan berita palsu. Dalam penulisan (Rada Mihalcea), penulis telah membangunkan dua sistem untuk pengesanan penipuan dengan menggunakan mesin vektor sokongan dan Naïve Bayes. Keduanya memberikan ketepatan yang baik. Seterusnya, penyelidik (Zhou et al., 2019) mendapati bahawa berita palsu semakin meningkat mengikut peredaran masa. Hal ini merupakan salah satu sebabnya keperluan untuk mengesan berita palsu dari terus tersebar. Bagi mengatasi masalah ini, algoritma pembelajaran mesin telah dilatih bagi memenuhi tujuan ini.

Pembangunan hartanah pada dasarnya adalah proses membangunkan bangunan atau tanah kepada nilai penggunaan yang lebih tinggi. Membeli hartanah tidak semudah yang disangka kerana terdapatnya penipuan merata-rata yang sering digunakan oleh agensi haram bagi mengaut keuntungan. Agensi yang tidak bertanggungjawab ini mempunyai insentif yang sangat kuat untuk meletakkan penyenaian hartanah palsu sebagai umpan mudah untuk mereka mendapatkan pelanggan. Sebagai pembeli yang bijak, setiap kali melihat tawaran hartanah yang hebat untuk penjualan atau sewa di dalam portal dengan harga yang tidak masuk akal jangan terlalu mempercayainya kerana itu mungkin suatu umpan. Hal ini kerana, dengan meletakkan penyenaian palsu ini dapat meningkatkan peluang agensi tersebut untuk berhubung dengan mana-mana pelanggan yang berharga iaitu mereka yang baru ingin mencari harta. Sebaik sahaja agensi ini mendapatkan nombor telefon pelanggan berharga mereka, mereka akan cuba membuat panggilan untuk menarik pelanggan tersebut dengan mencadangkan hartanah yang lebih mahal dan mewah. Kebiasaanya, satu hartanah tunggal boleh didapati dari pelbagai agensi dan hanya agensi tertentu yang mempunyai pelanggan sahaja yang berjaya menutup perjanjian tersebut. Ini adalah salah satu sebabnya, mereka meletakkan banyak penyenaian palsu di internet. Dalam projek ini, pembelajaran mesin digunakan untuk meramal penyenaian palsu hartanah berdasarkan pelbagai ciri hartanah.

Mengikut penyelidikan sedia ada, kebanyakan penyelidik berpendapat bahawa pelbagai algoritma pembelajaran mesin boleh digunakan. Untuk mendapatkan ramalan penyenaian palsu yang tepat, hendaklah dilatih dan diuji untuk mencapai ketepatan yang baik. Dalam membangunkan model ramalan penyenaian palsu dalam hartanah ini, 3 algoritma yang berbeza digunakan iaitu mesin vektor sokongan (SVM), pokok keputusan dan hutan rawak. Ketiga-tiga algoritma ini akan dilatih bagi mengenalpasti algoritma yang berjaya mencapai ketepatan tertinggi.

## **2 PENYATAAN MASALAH**

Secara umumnya, pasaran hartanah tidak dikawal oleh undang-undang yang ketat. Hal ini membawa kepada jalan mudah untuk pihak agensi yang tidak bertanggungjawab ini menerbitkan penyenaian palsu atau tidak tersedia kebanyakannya pada harga di bawah purata di pasaran sebagai kaedah bagi memperoleh pelanggan yang baru mencari harta hidup baru. Pelanggan yang tidak menyedari bahawa mereka telah ditipu ini sangat percaya kepada agensi

tersebut kerana terhasut dengan harga murah yang ditawarkan sedangkan mereka tidak pernah melihat tempat yang diinginkan bahkan mereka juga mungkin ditipu dengan skim penipuan untuk pembayaran pendahuluan sebelum perjanjian. Selain itu, sebagai manusia mungkin wujud di mana sangat sukar untuk memutuskan sama ada penyenaian itu benar atau palsu kerana adanya agensi yang tidak bertanggungjawab. Hal ini membuatkan pelanggan sukar untuk mencari harta yang boleh dipercayai. Di Malaysia, pembeli tidak dapat mengenalpasti samaada hartanah yang diinginkan palsu atau nyata. Mereka hanya mengetahui penyenaian itu palsu secara manual. Oleh itu terdapat keperluan teknik pembelajaran mesin untuk diaplikasikan dalam membina aplikasi web meramal penyenaian palsu dalam hartanah. Dengan mengaplikasikan teknik pembelajaran mesin, dapat mengatasi masalah penyenaian palsu ini dengan mengenalpasti kesahihan penyenaian tersebut secara automatik. Hal ini kerana teknik pembelajaran mesin dapat meramal dengan tepat sekiranya menggunakan algoritma yang memberikan ketepatan terbaik. Setakat ini, belum lagi ada sistem di Malaysia yang dibangunkan untuk meramal hartanah sama ada palsu atau tidak dari segi ciri-ciri hartanah menggunakan algoritma pembelajaran mesin. Selain itu, pelanggan yang ingin mencari hartanah dengan harga yang berpatutan tanpa ditipu, aplikasi web ini dapat membantu dengan meramal hartanah tersebut sama ada palsu atau tidak berdasarkan ciri- ciri hartanah tersebut.

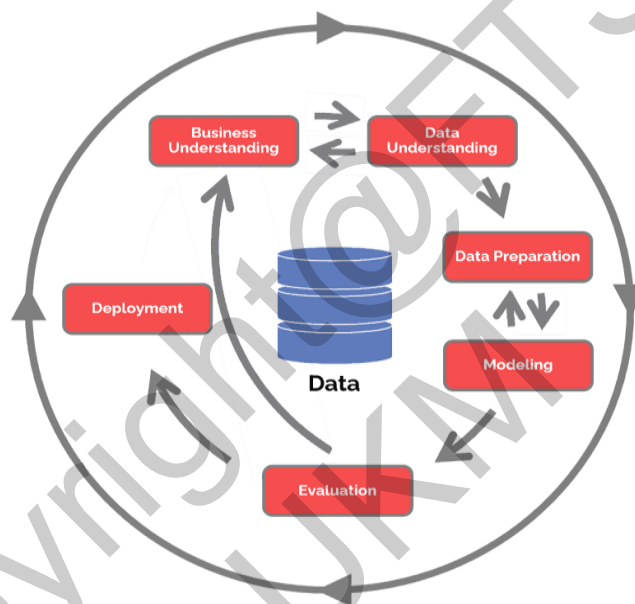
### **3 OBJEKTIF KAJIAN**

Objektif kajian projek ini adalah untuk membangunkan aplikasi web yang menggunakan teknik pembelajaran mesin bagi mengesan penyenaian palsu hartanah. Objektif bagi kajian ini adalah seperti berikut :

- a. Untuk membangunkan model pembelajaran mesin bagi mengesan penyenaian palsu hartanah.
- b. Untuk mengenalpasti algoritma pembelajaran mesin yang terbaik bagi mengesan penyenaian palsu hartanah. Tiga jenis algoritma akan diuji dan model dengan prestasi terbaik akan dipilih.
- c. Untuk membangunkan satu aplikasi web yang dapat mengesan penyenaian palsu hartanah bagi mengurangkan risiko pembeli hartanah daripada ditipu dengan menyediakan ruang untuk pengguna memasukkan ciri-ciri yang ditetapkan.

#### 4 METODOLOGI KAJIAN

Metodologi yang akan digunakan bagi membangunkan projek ini adalah model CRISP- DM. Model CRISP-DM bermaksud proses rentas industri untuk perlombongan data dan ianya digunakan secara meluas di global. Model ini bertujuan untuk membuat projek perlombongan data besar lebih dipercayai, lebih terurus dan lebih cepat termasuklah menilai kemajuan projek berbanding objektif asal. Selain itu dengan menggunakan model ini sebarang perisian boleh digunakan bagi analisis. Model ini mempunyai 6 fasa iaitu pemahaman perniagaan, pemahaman data, penyediaan data, pemodelan, penilaian dan penggunaan.



Rajah 1 Fasa CRISP-DM

##### Fasa 1: Pemahaman Perniagaan

Dalam fasa ini, objektif projek ialah untuk mengesan penyenaian palsu hartanah. Selain itu, mengekstrak data untuk mendapatkan pemahaman data yang menyebabkan penipuan dalam hartanah.

##### Fasa 2: Pemahaman data

Data awal dikumpulkan dan mengenal pasti atribut yang diperlukan untuk mengesan penyenaian palsu.

### Fasa 3: Penyediaan data

Menyediakan data untuk diimport ke dalam aplikasi web ramalan penyenaaran palsu. Fasa di mana pembersihan data dilakukan, mengelaskan atribut sebagai tidak relevan, kategori dan berangka untuk digunakan semasa latihan dan ujian dalam algoritma yang berbeza.

### Fasa 4: Pemodelan

Tiga algoritma berbeza (mesin vektor sokongan, pokok keputusan dan hutan rawak) digunakan dalam projek ini untuk latihan, pengesanan dan ujian bagi mengesan penyenaaran palsu.

### Fasa 5: Penilaian

Dalam fasa ini, pemeriksaan dijalankan untuk memastikan telah menggunakan alat terbaik untuk perlombongan data dan mengesahkan bahawa data tersebut seperti apa yang digambarkan dan difahami dalam fasa Pemahaman Perniagaan. Seterusnya, keputusan akhir adalah pemilihan algoritma terbaik berdasarkan ketepatan yang dicapai.

### Fasa 6: Penggunaan

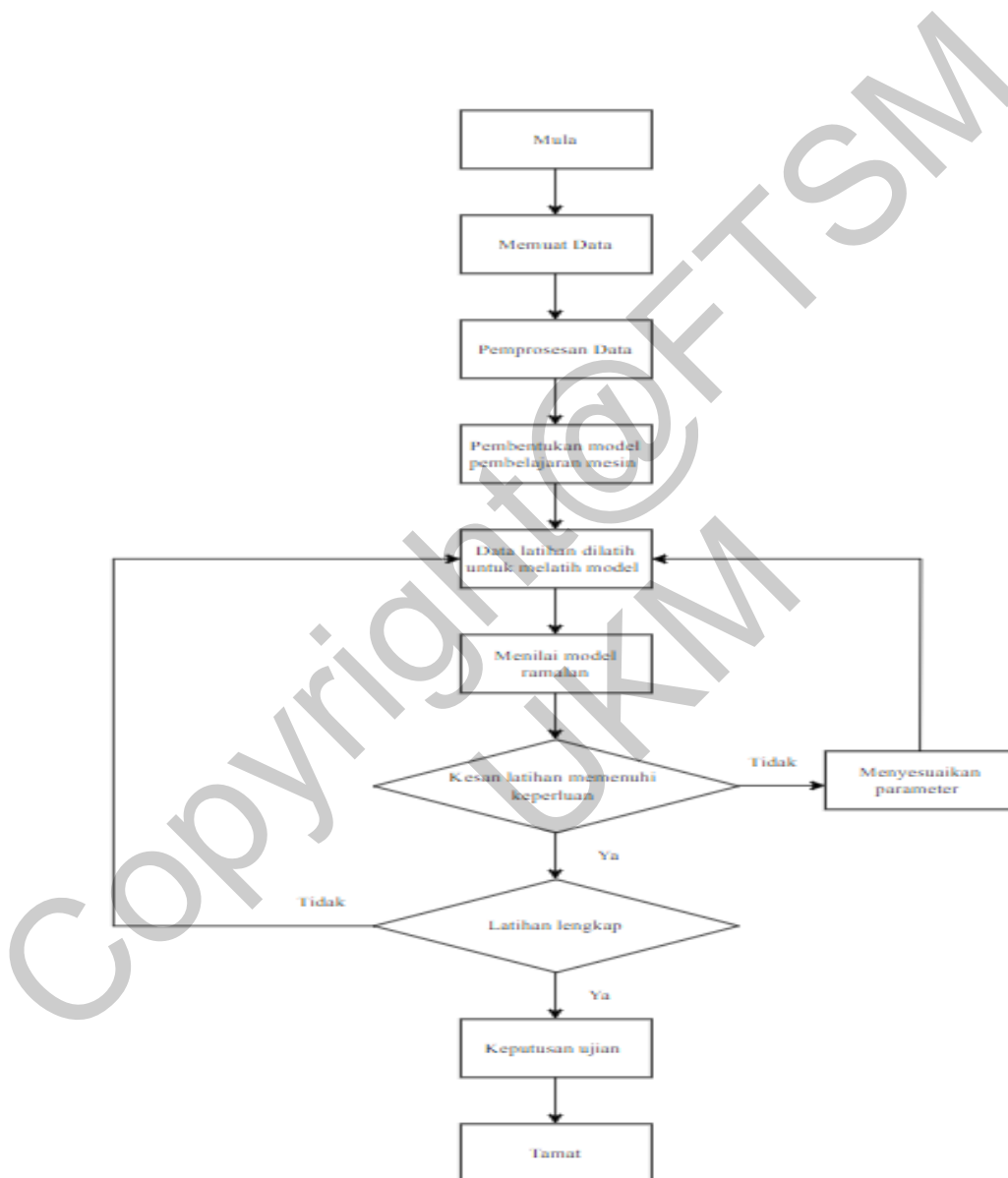
Dalam fasa ini, pelaksanaan dibuat apabila telah bersedia dengan reka bentuk. Penyenaaran dalam hartanah dihasilkan secara berterusan dan hal ini juga akan membuatkan penyenaaran palsu baru muncul jadi projek ini harus sentiasa dikemas kini dengan menambah set data yang baru dan melatih semula algoritma untuk mencapai ketepatan yang maksimum.

## 5 HASIL KAJIAN

Klasifikasi bagi kelas 1 iaitu palsu adalah berdasarkan penerangan setiap set data tersebut. Antara kata kunci bagi penerangan set data yang dilabel sebagai kelas 1 adalah:

- For illustration only
- Illusion only
- Purpose only
- Photo reference

- Tenant's privacy
- Not actual unit
- Photos shown may or may not be actual unit
- Images photos displayed are artist impression
- For display



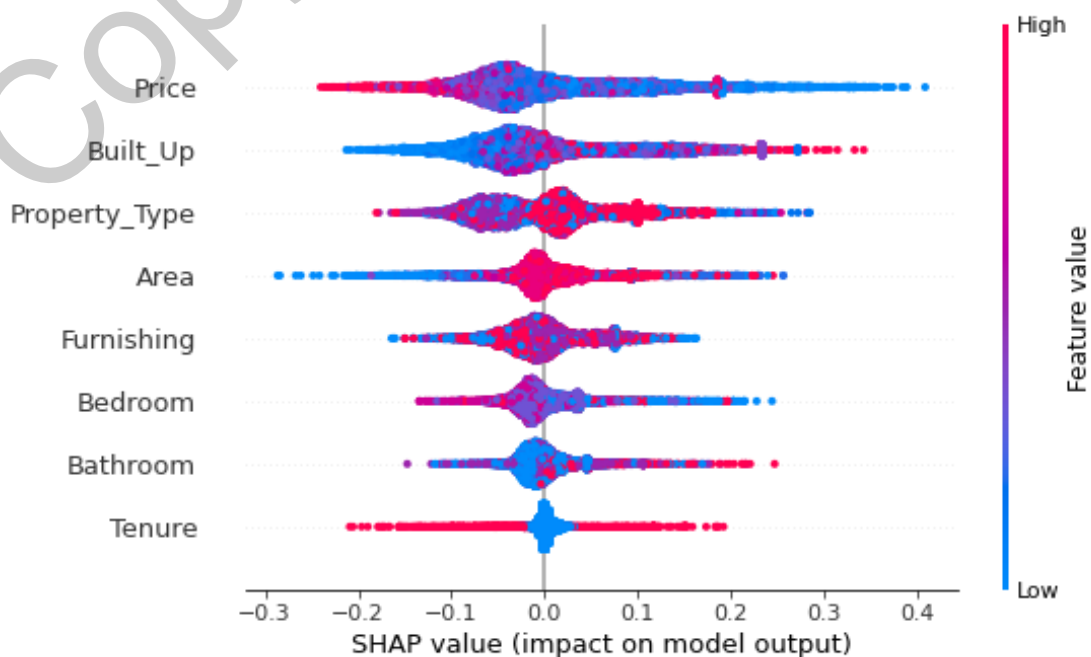
Rajah 2 Aliran dalam membangunkan model pembelajaran mesin

Rajah 2 menunjukkan aliran dalam membangunkan setiap model pembelajaran mesin. Selepas selesai proses pemprosesan data, ialah proses pembentukan model pembelajaran mesin di mana 3 algoritma pembelajaran mesin digunakan iaitu mesin sokongan vektor (SVM), hutan rawak dan pokok keputusan. Ketiga-tiga model ini dinilai untuk menguji prestasi yang berjaya dicapai oleh setiap model dengan data ujian. Seterusnya model dengan prestasi yang terbaik akan dipilih untuk membangunkan aplikasi web Meramalkan Penyenaian Palsu dalam Hartanah (FPLP).

Set Data	Hutan Rawak	SVM	Pokok Keputusan
Set data latihan	99.29 %	79.42%	99.45%
Set data ujian	99.58%	79.38%	73.61%

Jadual 3 Ketepatan bagi setiap model

Jadual 3 menunjukkan, ketepatan bagi setiap model. Model hutan rawak mempunyai ketepatan yang tertinggi bagi kedua-dua set data latihan dan ujian iaitu sebanyak 99.29% dan 99.58%. Manakala model SVM mempunyai ketepatan paling rendah bagi kedua-dua set data latihan dan ujian iaitu sebanyak 79.42% dan 79.38%. Oleh itu, model pembelajaran mesin yang mencapai prestasi terbaik adalah hutan rawak dan ianya akan dipilih sebagai model terbaik bagi membangunkan aplikasi web.



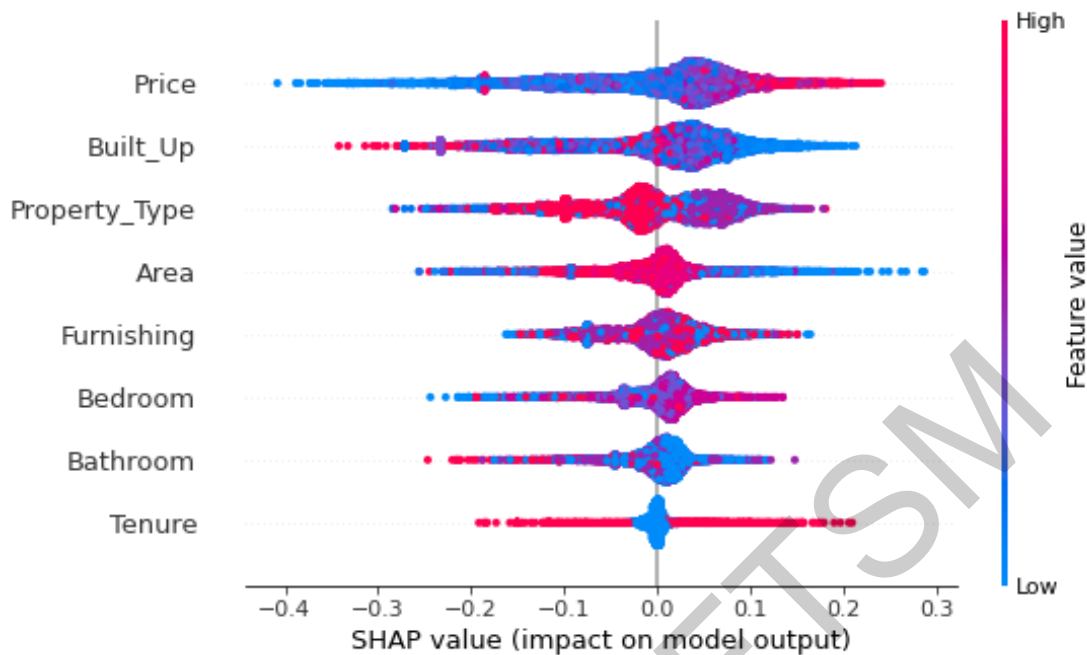


## Rajah 4 Summary plot bagi kelas 1

Seterusnya, analisis SHAP digunakan bagi menerangkan ramalan kejadian dengan mengira sumbangan setiap atribut kepada ramalan. Plot nilai SHAP boleh menunjukkan hubungan positif dan negatif peramal dengan pembolehubah sasaran. Setiap titik bagi kedua-dua rajah di atas mempunyai tiga ciri iaitu :

- Lokasi menegak menunjukkan atribut yang digambarkan seperti *Price*.
- Warna menunjukkan sama ada ciri itu tinggi atau rendah bagi setiap baris set data itu.
- Lokasi mendatar menunjukkan sama ada kesan nilai tersebut menyebabkan ramalan yang lebih tinggi atau lebih rendah.

Rajah 4 menunjukkan summary plot bagi kelas 1. *Price* yang tinggi mempunyai impak negatif terhadap ramalan manakala *Price* yang rendah mempunyai impak positif untuk model meramal sebagai kelas 1. Bagi *Built Up* yang tinggi mempunyai impak positif terhadap ramalan manakala *Built Up* yang rendah memberi impak negatif terhadap ramalan. *Property Type* yang tinggi mempunyai impak positif terhadap ramalan manakala *Property Type* yang rendah mempunyai impak negatif terhadap ramalan. *Area* yang tinggi mempunyai impak positif terhadap ramalan manakala *Area* yang rendah mempunyai impak negatif terhadap ramalan. *Furnishing* yang tinggi mempunyai impak positif terhadap ramalan manakala *Furnishing* yang rendah mempunyai impak negatif terhadap ramalan. *Bedroom* yang rendah mempunyai impak positif terhadap ramalan manakala *Bedroom* yang tinggi mempunyai impak negatif terhadap ramalan. *Bathroom* yang tinggi mempunyai impak positif terhadap ramalan manakala *Bathroom* yang rendah mempunyai impak negatif terhadap ramalan dan *Tenure* yang tinggi mempunyai impak positif terhadap ramalan manakala *Tenure* yang rendah mempunyai impak negatif terhadap ramalan.



Rajah 5 Summary plot bagi kelas 0

Rajah 5 menunjukkan summary plot bagi kelas 0. Price yang tinggi mempunyai impak positif terhadap ramalan manakala Price yang rendah mempunyai impak negatif untuk model meramal sebagai kelas 0. Built Up yang rendah memberi impak positif terhadap ramalan manakala Built Up yang tinggi memberi impak negatif terhadap ramalan. Property Type yang rendah memberi impak positif terhadap ramalan manakala Property Type yang tinggi memberi impak negatif terhadap ramalan. Area yang rendah memberi impak positif terhadap ramalan manakala Area yang tinggi memberi impak negatif terhadap ramalan. Furnishing yang tinggi memberi impak positif terhadap ramalan manakala Furnishing yang rendah memberi impak negatif terhadap ramalan. Bedroom yang tinggi memberi impak positif terhadap ramalan manakala Bedroom yang rendah memberi impak negatif terhadap ramalan. Bathroom yang rendah memberi impak positif terhadap ramalan manakala Bathroom yang tinggi memberi impak negatif terhadap ramalan. Tenure yang tinggi memberi impak positif terhadap ramalan manakala Tenure yang rendah memberi impak negatif terhadap ramalan.

Semasa proses menganalisis data dan pembangunan model pembelajaran mesin, bahasa pengaturcaraan yang digunakan ialah bahasa Python. Manakala bahasa yang digunakan semasa membangunkan web aplikasi ialah Python, HTML, CSS dan JavaScript. Perisian yang digunakan ialah Visual Studio Code.



Rajah 6 Antara muka hadapan laman utama web aplikasi

Rajah 6 menunjukkan muka hadapan aplikasi web di mana pengguna hendaklah menekan butang Start Predict bagi mengisi butiran di dalam borang untuk meramal penyenaian.

The image shows a 'PREDICTION' form within the application. The form is titled 'PREDICTION' in blue text. Below the title, it says 'Enter the following values to predict the fake listing in property'. The form contains several input fields and radio buttons: 'Price:' followed by a text input field; 'Tenure:' with two radio button options, 'Freehold' and 'Leasehold'; 'Property Type:' followed by a dropdown menu with 'Flat' selected; 'Built Up:' followed by a text input field; and 'Furnishing:' with two radio button options, 'Fully Furnished' and 'Semi Furnished'. A large, semi-transparent watermark 'Copyright @ UKM' is visible across the form.

Property Type: Flat

Built Up:

Furnishing:

Fully Furnished

Semi Furnished

Unfurnished

Bedroom:

Bathroom:

Area: Kuantan

Predict

Rajah 7 Antara muka borang ramalan

Rajah 7 menunjukkan antara muka borang ramalan di mana pengguna diminta untuk memasukkan data mengikut atribut yang telah dinyatakan. Selepas selesai memasukkan data. Seterusnya dengan menekan butang Predict, pengguna dapat melihat keputusan ramalan sama ada hartanah yang dimasukkan palsu atau benar berdasarkan teknik pembelajaran mesin.

**PREDICTION**

Enter the following values to predict the fake listing in property

Price: 205000

Built Up: 550000

Furnishing:

Fully Furnished

Semi Furnished

Unfurnished

**Saved data**

Price: 227000

Tenure: 111

Property Type: 11111

Built Up: 550000

Rajah 8 Borang dengan rekaan hanya nilai sahaja

Rajah 8 menunjukkan ruang untuk memasukkan nilai bagi atribut Price, Built Up, Bedroom dan Bathroom. Ruang ini hanya boleh memasukkan nilai angka sahaja.

Freehold  
 Leasehold  
 Property Type:   
 Built Up:   
 Furnishing:  
 Fully Furnished  
 Semi Furnished  
 Unfurnished  
 Bedroom:   
 Bathroom:   
 Area:

Rajah 9 Borang dengan rekaan drop down box

Rajah 9 menunjukkan ruang nilai nominal bagi atribut *Property Type* dan *Area* direka dengan format *drop-down* untuk memudahkan pemilihan nilai. Ruang ini hanya boleh memilih nilai sedia ada sahaja.

Tenure:  
 Freehold  
 Leasehold  
 Property Type:   
 Built Up:   
 Furnishing:  
 Fully Furnished  
 Semi Furnished  
 Unfurnished  
 Bedroom:   
 Bathroom:

Rajah 10 Borang dengan rekaan radio button

Rajah 10 menunjukkan ruang nilai nominal bagi atribut *Furnishing* dan *Tenure* direka dengan format *radio button* untuk memudahkan pemilihan nilai. Ruang ini hanya boleh memilih nilai sedia ada sahaja.

The image shows a web form titled "PREDICTION". The main heading is "Enter the following values to predict the fake listing in property". Below this, there are three input fields: "Price" (a text box with a warning message "Please fill out this field."), "Tenure" (radio buttons for "Freehold" and "Leasehold"), and "Property Type" (a dropdown menu with "Flat" selected).

Rajah 11 Mesej amaran kekurangan data

Rajah 11 menunjukkan amaran kekurangan kemasukan nilai akan diberitahu kepada pengguna selepas menekan butang *predict*.

## 6 KESIMPULAN

Secara keseluruhannya, Web Aplikasi Meramalkan Penyenaiaan Palsu Berdasarkan Teknik Pembelajaran Mesin (FPLP) berjaya dibangunkan walaupun terdapat sedikit masalah semasa model meramal penyenaiaan. Hal ini kerana, masalah ketidakseimbangan data oleh kerana label kelas *true* lebih banyak berbanding kelas *fake*. Web aplikasi ini diharapkan dapat membantu pengguna yang ingin mengetahui status penyenaiaan hartanah yang diingini. Walaupun terdapat beberapa kekurangan, diharapkan web aplikasi ini dapat dijadikan titik kajian untuk kajian pada masa hadapan seterusnya dapat melakukan penambahbaikan bagi kekurangan yang terdapat.

## 7 RUJUKAN

- Sridharan, M. (2018, September 25). Think insights - CRISP-DM: A framework for data mining & analysis. Retrieved November 3, 2021, from Thinkinsights.net website: <https://thinkinsights.net/digital/crisp-dm/>
- Donepudi, P. K., & UST-Global, Inc. (2019). Automation and machine learning in transforming the financial industry. *Asian Business Review*, 9(3), 129–138.
- Wang, W. Y. (2017). “liar, liar pants on fire”: A new benchmark dataset for fake news detection. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Stroudsburg, PA, USA: Association for Computational Linguistics.
- Mihalcea, R., & Strapparava, C. (n.d.). The lie detector: Explorations in the automatic recognition of deceptive language. Retrieved November 3, 2021, from Aclanthology.org website: <https://aclanthology.org/P09-2078.pdf>
- Zhou, X., Zafarani, R., Shu, K., & Liu, H. (2019). Fake news: Fundamental theories, detection strategies and challenges. Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining - WSDM '19. New York, New York, USA: ACM Press.
- Hotz, N. (2018, September 10). *What is CRISP DM?* Data Science Process Alliance. <https://www.datascience-pm.com/crisp-dm-2/>
- Crisp DM methodology*. (2020, June 17). Smart Vision Europe. <https://www.sv-europe.com/crisp-dm-methodology/>
- Understanding machine learning with SHAP analysis. (2021, February 17). Acerta. <https://acerta.ai/blog/understanding-machine-learning-with-shap-analysis/>
- Molnar, C. (2022, March 29). 9.6 SHAP (SHapley Additive exPlanations). Github.Io. <https://christophm.github.io/interpretable-ml-book/shap.html>

Advanced uses of SHAP values. (2021, November 9). Kaggle.Com; Kaggle.  
<https://www.kaggle.com/code/dansbecker/advanced-uses-of-shap-values/tutorial>

Copyright@FTSM  
UKM

Nur Atiqah Binti Suhaimi (A175100)  
Ts. Dr Nor Samsiah Binti Sani  
Fakulti Teknologi & Sains Maklumat,  
Universiti Kebangsaan Malaysia