

# MODEL PREDIKTIF BAGI HARGA PASARAN RUMAH MENGUNAKAN PEMBELAJARAN MESIN

ALIF AKMAL BIN NOR HISHAM

RUZZAKIAH BINTI JENAL

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

## ABSTRAK

Harga pasaran rumah kini menjadi suatu perdebatan dalam kalangan pembeli berikutan kenaikan yang mendadak nilai hartanah ketika ini. Harga rumah atau nilai rumah di Malaysia turun naik dengan ketara dalam dekad yang lalu. Keraguan seperti ini menimbulkan kesan yang jelas kepada mereka yang ingin melabur atau membeli rumah di Malaysia. Kaedah tradisional bagi penilaian nilai hartanah selalunya bergantung pada penilaian manual, yang boleh memakan masa, subjektif dan terdedah kepada kesilapan manusia. Selain itu, ia mungkin tidak menggambarkan spektrum penuh bagi pembolehubah yang mempengaruhi harga hartanah. Hal ini membawa kepada ketidakcekapan dan ketidaktepatan dalam proses penilaian yang berpotensi menyebabkan nilai hartanah menjadi di bawah harga atau terlalu mahal berbanding nilai yang sepatutnya. Terdapat keperluan yang mendesak untuk penyelesaian melalui penyelesaian berasaskan data untuk menangani kelemahan ini dan menawarkan anggaran nilai hartanah yang tepat. Oleh itu, model Prediktif bagi Harga Pasaran Rumah menggunakan Pembelajaran Mesin adalah sebuah model yang dibangunkan bagi menganalisis pelbagai ciri rumah dan persekitaran kediaman tersebut untuk membuat ramalan yang lebih jitu. Tujuan utama model adalah untuk menyediakan anggaran harga yang tepat dan berguna untuk pembeli, penjual, pelabur dan untuk kepentingan lain dalam pasaran hartanah. Objektif model adalah untuk membangunkan algoritma atau model pengiraan yang boleh meramalkan harga pasaran rumah berdasarkan pelbagai atribut atau ciri. Kombinasi alat seperti TensorFlow dan Jupyter Notebooks digunakan untuk membangunkan model supaya menjadi lebih baik dan untuk mencapai objektif dengan membawa pendekatan sistematik yang merangkumi pengumpulan data, prepemprosesan, pemilihan ciri, pemilihan dan latihan model, penilaian dan penggunaan. Algoritma pembelajaran mesin yang berpotensi termasuk regresi linear dan pepohon keputusan untuk membina model ramalan. Penyelesaian yang direka bentuk untuk mengendalikan pelbagai jenis hartanah kediaman dan boleh disesuaikan dengan keadaan pasaran yang berbeza. Pembangunan Model Prediktif bagi Harga Pasaran Rumah menggunakan Pembelajaran mesin memacu kemajuan yang ketara dalam bidang penilaian hartanah. Dengan memanfaatkan kuasa yang dipacu oleh data, proses penilaian hartanah dapat diselaraskan dan juga dapat menyediakan anggaran yang tepat dan boleh dipercayai. Inovasi ini memberi manfaat kepada pelbagai pihak berkepentingan dalam pasaran hartanah, daripada pemilik rumah individu kepada pelabur dan professional industri yang akhirnya akan menyumbang kepada keputusan yang telus dan urus niaga yang cekap.

Kata kunci: Harga rumah, ramalan, Pembelajaran Mesin

## PENGENALAN

Sektor hartanah di Malaysia pada waktu ini mengalami kepesatan yang mengagumkan dalam dekad terakhir yang mana begitu banyak pembangunan yang berlaku lebih-lebih lagi di kawasan bandar. Pasaran harta kediaman di Malaysia mengalami peningkatan harga yang drastik sejak lima belas tahun yang lalu (Husain et al. 2011). Kepesatan yang berlaku antaranya disebabkan oleh faktor keadaan ekonomi semasa dan juga kepesatan sesetengah lokasi yang menyediakan banyak infrastruktur, aksesibiliti dan fasiliti awam. Sebagai contoh kawasan-kawasan di sekitar ibu kota, Kuala Lumpur dan bandar-bandar utama seperti Petaling Jaya, Shah Alam dan Johor Bahru cenderung mempunyai harga rumah yang lebih tinggi berbanding dengan kawasan luar bandar atau kawasan-kawasan kecil. Selain daripada itu, faktor-faktor seperti ketika pasaran hartanah dan faktor ekonomi global juga boleh memberi kesan kepada harga rumah di Malaysia (Loanstreet n.d.).

Penting untuk diingat bahawa harga rumah adalah subjektif dan boleh berubah dari masa ke masa. Oleh itu, prospek pembeli dan pelabur hartanah perlu mempertimbangkan pelbagai faktor sebelum membuat keputusan untuk membeli atau melabur. Memiliki rumah untuk perlindungan dan keselesaan adalah penting bagi kebanyakan orang, dan sering dilihat sebagai tanda kejayaan peribadi. Kaedah penilaian hartanah biasanya bergantung pada proses manual yang memakan masa, subjektif, dan terdedah kepada kesilapan manusia.

Untuk mengatasi kelemahan-kelemahan ini, banyak pakar hartanah dan syarikat penilai kini beralih kepada teknik penilaian yang lebih berstruktur dan berpandukan data, seperti analisis harga jualan terkini, model penilaian hartanah automatik, dan pendekatan komputer untuk membantu membuat penilaian yang lebih tepat dan objektif. Dalam beberapa tahun kebelakangan ini, perkembangan pesat penggunaan data besar telah menjadikan pembelajaran mesin sebagai pendekatan ramalan yang penting, kerana ia mampu meramalkan harga rumah dengan lebih tepat berdasarkan atributnya.

Projek yang dibangunkan menggunakan set data bersepadu dan algoritma yang pelbagai untuk mendapatkan hasil yang tepat. Atribut yang mempengaruhi hasil kajian diperoleh daripada beberapa sumber, seperti Jabatan Statistik Malaysia (DOSM). Data-data ini akan melalui proses pemprosesan untuk dimanipulasi, dibersihkan, dan dianalisis, bagi mengekstrak maklumat yang bermakna sebelum melalui pembelajaran mesin. Proses ini memfokuskan pada pembangunan algoritma dan model yang membolehkan komputer belajar dan menghasilkan ramalan serta keputusan.

## METODOLOGI KAJIAN

Analisis dan spesifikasi keperluan adalah langkah awal yang penting dalam pembangunan sistem atau penyelesaian teknologi maklumat atau perniagaan. Proses ini melibatkan memahami apa yang diinginkan oleh pengguna dan organisasi. Bab ini juga membincangkan analisis dan spesifikasi keperluan termasuk keperluan pengguna, keperluan fungsian sistem,

keperluan bukan fungsian sistem, serta perkakasan dan perisian yang digunakan. Bab ini juga membahas model ramalan harga rumah menggunakan pembelajaran mesin. Keperluan pengguna merujuk kepada apa yang diharapkan oleh sistem atau aplikasi perisian dari perspektif pengguna akhir. Memahami dan mendokumentasikan keperluan pengguna adalah penting kerana ia membentuk asas untuk reka bentuk, pelaksanaan, dan pengujian perisian.

Keperluan sistem merujuk kepada spesifikasi perkakasan, perisian, dan komponen rangkaian yang diperlukan untuk sistem beroperasi dengan baik. Keperluan fungsian sistem fokus pada tugas yang perlu dilakukan oleh sistem, manakala keperluan bukan fungsian menerangkan aspek prestasi dan kualiti sistem (Gorbachenko, 2024). Spesifikasi keperluan perkakasan membincangkan sistem operasi, pemrosesan, dan aspek perkakasan lain. Spesifikasi keperluan perisian menerangkan fungsi perisian dan bagaimana ia dijangka berfungsi. Ia juga menunjukkan perisian dan model sistem melalui gambar rajah konteks dan rajah kes guna. Bab ini juga menerangkan metodologi pembangunan perisian menggunakan reka bentuk CRISP-DM, yang sesuai untuk proses data sains.

### **Fasa Pemahaman Perniagaan**

Langkah pertama dalam proses pembangunan projek data sains. Tujuannya adalah untuk memahami apa yang perniagaan ingin capai dan bagaimana projek data sains dapat membantu mencapai matlamat tersebut. Langkah ini melibatkan beberapa aktiviti utama antaranya pemahaman matlamat projek, objektif dan keperluan daripada perspektif perniagaan, kemudian menukar pengetahuan ini kepada definisi masalah perlombongan data dan pelan awal dengan set tugas khusus dan hasil yang diinginkan untuk mencapai objektif peringkat projek. Dalam konteks pasaran hartanah, fasa pemahaman perniagaan melibatkan kajian tentang bagaimana trend pasaran hartanah berkembang dan bagaimana maklumat tersebut boleh digunakan untuk membuat keputusan perniagaan yang lebih baik. Dengan maklumat ini, saya dapat merancang langkah-langkah untuk mengumpul dan menganalisis data hartanah, membangun model pembelajaran mesin, dan menguji model tersebut untuk memastikan ketepatan ramalan.

### **Fasa Pemahaman Data**

Fasa ini memfokuskan kepada mengumpul dan meneroka data yang berkaitan dengan projek, terutama data dari jabatan yang mengawal selia hartanah. Proses mengumpul set data yang tersedia dan memeriksa secara menyeluruh untuk memahami ciri-cirinya. Ini termasuk melihat kualiti data, seperti nilai yang hilang dan nilai terpencil (outlier). Menangani isu-isu ini penting untuk memastikan analisis dan pemodelan yang tepat. Selepas itu, proses mendokumentasikan penemuan tentang struktur dan kandungan data. Ini membantu dalam memahami bagaimana data tersebut boleh digunakan untuk pembangunan model ramalan. Akhir sekali, pemilihan data yang paling relevan dari set data yang dikumpul untuk digunakan dalam pembangunan model. Ini memastikan hanya data yang berguna dan berkualiti tinggi digunakan dalam langkah-langkah seterusnya.

### **Fasa Penyediaan Data**

Langkah penting dalam projek untuk memastikan data berkualiti tinggi untuk analisis. Pemilihan bahagian data yang berkaitan dari set data yang diperolehi. Ini memastikan hanya data yang relevan digunakan untuk analisis selanjutnya. Kemudiannya, proses pembersihan data untuk meningkatkan kualitinya. Ini melibatkan menangani isu seperti nilai yang hilang, nilai terencil (outlier), dan ketidakkonsistenan. Tujuannya adalah untuk memastikan data bersih dan sesuai untuk analisis. Selanjutnya, proses pembinaan data yang diperlukan. Ini termasuk menghasilkan atribut baru dari atribut yang sedia ada atau mengubah nilai atribut yang ada. Atribut baru ini membantu dalam analisis yang lebih mendalam. Akhir sekali, proses pengintegrasian dan memformat data. Ini melibatkan menggabungkan data dari pelbagai sumber dan memastikan format data adalah konsisten. Proses ini menghasilkan set data baru yang siap untuk analisis dan pemodelan.

### **Fasa Pemodelan**

Fasa pemodelan melibatkan pemilihan teknik pemodelan yang sesuai untuk membuat ramalan dalam pembelajaran mesin. Tiga teknik utama yang sering digunakan adalah Regresi Linear yang memodelkan hubungan antara dua pembolehubah dengan mencari garis lurus yang terbaik yang menggambarkan hubungan ini, Regresi Hutan Rawak (Random Forest Regression) yang menggunakan banyak pokok keputusan untuk membuat ramalan yang lebih tepat melalui setiap pokok membuat ramalan, dan keputusan akhir adalah purata dari semua pokok yang akan menjadikan teknik ini baik untuk set data yang kompleks dan mengurangkan risiko membuat kesilapan besar (overfitting) dan yang terakhir adalah XGBoost yang menggabungkan banyak model kecil (pokok keputusan) secara berurutan, di mana setiap model cuba membetulkan kesilapan model sebelumnya. Teknik ini sangat baik dalam menangani data yang kompleks dan mengelakkan overfitting melalui pelbagai kaedah penyusunan semula dan pengoptimuman. XGBoost memerlukan penalaan parameter yang teliti untuk mencapai prestasi terbaik..

### **Fasa Penilaian**

Pada peringkat ini, keberkesanan sejauh mana model ramalan menjadi penentu. Prosedur penilaian direka bentuk bagi mengkaji kecekapan dan ketepatan model dengan penekanan utama adalah kepada kejayaan model dalam memenuhi objektif projek. Fasa penilaian juga melibatkan penilaian sebarang hasil perlombongan data lain yang dihasilkan. Keputusan perlombongan data melibatkan model yang semestinya berkaitan dengan objektif projek asal dan semua penemuan lain yang tidak semestinya berkaitan dengan objektif projek asal, tetapi mungkin juga dapat mendedahkan cabaran, maklumat atau petunjuk tambahan untuk hala tuju masa hadapan.

### **Fasa Penggunaan**

Dalam peringkat penggunaan, model ramalan yang dinilai digabungkan ke dalam web HouseForecast. Terdapat beberapa fasa dalam peringkat ini, di mana pelan penggunaan dibangunkan dengan komprehensif untuk penggunaan model ke dalam persekitaran operasi. Selepas itu, pemantauan dan penyelenggaraan model selepas pembangunan dilaksanakan dengan melibatkan penyediaan mekanisma untuk mengenal pasti prestasi model dalam

senario dunia sebenar juga mengesan isu dan melaksanakan kemas kini atau penambahbaikan tepat pada masanya. Hasil daripada penilaian ke atas model, pendokumentasian laporan akhir meringkaskan keseluruhan projek. Laporan mungkin termasuk penemuan utama, cerapan dan pengesyoran yang diperolehi daripada model. Semakan retrospektif dijalankan selepas itu ke atas keseluruhan projek. Semakan membantu dalam pembelajaran daripada hasil pengalaman projek semasa dan memperlajusi proses perlombongan data untuk usaha masa hadapan.

## KEPUTUSAN DAN PERBINCANGAN

Pembangunan model prediktif harga rumah menggunakan pembelajaran mesin berjaya dibangunkan dan juga berintegrasi ke dalam aplikasi HouseForecast. Sepanjang proses pembangunan model, bahasa pengaturcaraan yang digunakan adalah bahasa pengaturcaraan *Python* yang digunakan di dalam *Jupyter Notebook*. Pembangunan web aplikasi HouseForecat pula menggunakan bahasa pengaturcaraan HTML menggunakan perisian Sublime.

Melalui proses pemodelan, proses perbandingan dapat dijalankan bagi ketiga-tiga model yang dibangunkan iaitu model Regresi Linear, model Regresi Hutan Rawak dan juga Regresi *XGBoost*. Proses tersebut akan membandingkan metrik prestasi Min Ralat Punca Kuasa Dua (*Root Mean Square Error*), Min Ralat Mutlak (*Root Absolute Error*), dan  $R^2$  (*R-Square*). Hasil keputusan perbandingan berikut akan menentukan model yang terbaik yang akan diserap ke dalam aplikasi HouseForecast. Proses ini juga akan memecahkan kepada nisbah pemisahan bagi set latihan dan set ujian yang berbeza. Nisbah 0.1, 0.2, 0.3 dan 0.4 mewakili perkadaran set data yang diperuntukkan kepada set ujian semasa proses pemisahan data.

Jadual 1 RMSE Bagi Setiap Model

Nisbah	Regresi Linear		Hutan Rawak		XGBoost	
	Set Validasi	Set Ujian	Set Validasi	Set Ujian	Set Validasi	Set Ujian
(90:10)	0.5996	0.5961	0.4698	0.4946	0.3413	0.3831
(80:20)	0.5775	0.6086	0.4645	0.4890	0.3410	0.3857
(70:30)	0.5632	0.4422	0.4658	0.4983	0.3410	0.3927
(60:40)	0.5801	0.4610	0.4590	0.4895	0.3342	0.3959

Jadual 2 MAE Bagi Setiap Model

Nisbah	Regresi Linear		Hutan Rawak		XGBoost	
	Set Validasi	Set Ujian	Set Validasi	Set Ujian	Set Validasi	Set Ujian
(90:10)	0.4697	0.4676	0.3596	0.3765	0.2563	0.2928
(80:20)	0.4523	0.4804	0.3556	0.3666	0.2558	0.2903
(70:30)	0.4422	0.4751	0.3576	0.3729	0.2562	0.2948
(60:40)	0.4610	0.4687	0.3504	0.3652	0.2523	0.2934

Jadual 3 R<sup>2</sup> Bagi Setiap Model

Nisbah	Regresi Linear		Hutan Rawak		XGBoost	
	Set Validasi	Set Ujian	Set Validasi	Set Ujian	Set Validasi	Set Ujian
(90:10)	0.6346	0.6275	0.7657	0.7436	0.8763	0.8461
(80:20)	0.6712	0.6253	0.7710	0.7581	0.8766	0.8495
(70:30)	0.6850	0.6270	0.7697	0.7475	0.8766	0.8431
(60:40)	0.6429	0.6320	0.7764	0.7557	0.8815	0.8402

Hasil daripada proses latihan tersebut telah mengeluarkan keputusan yang diukur dengan menggunakan metrik prestasi seperti di atas. Bagi Jadual 1 dan Jadual 2, nilai RMSE dan MAE yang rendah menunjukkan model yang baik manakala bagi Jadual 3 pula nilai R<sup>2</sup> yang tinggi akan menunjukkan model yang lebih baik. Daripada ketiga-tiga metrik tersebut dapat dilihat model *XGBoost* menunjukkan keputusan yang lebih konsisten dan jitu bagi mengeluarkan keputusan ramalan yang baik.

Daripada keputusan berikut *XGBoost* akan melalui proses penalaan hiperparameter melalui teknik *GridSearchCV* bagi mengoptimumkan lagi keputusan ramalan *XGBoost*. Kaedah tersebut akan melibatkan pencarian melalui subset khusus hiperparameter untuk mencari gabungan yang menghasilkan prestasi terbaik seperti yang diukur oleh skor yang ditetapkan. Gabungan parameter yang terbaik ditentukan melalui *GridSearchCV* yang memperlihatkan model yang terpilih dilatih menggunakan set data gabungan daripada set latihan dan pengesahan untuk mengoptimumkan latihan kepada jumlah data yang tersedia untuk pembelajaran. '*learning\_rate*'= 0.1, '*max\_depth*'= 4, '*n\_estimators*'= 300 sepertimana yang terlihat pada Jadual 4.

Jadual 4 Parameter Yang Terpilih Daripada *GridSearchCV*

Hiperparameter	Nilai
<i>n_estimators</i>	300
<i>learning_rate</i>	0.1
<i>max_depth</i>	4

Bagi membandingkan dua model, *T-Test* ialah ujian statistik yang digunakan untuk membandingkan min dua kumpulan dan menentukan sama ada mereka berbeza secara signifikan antara satu sama lain. Ia membantu dalam menguji hipotesis tentang cara populasi. *T-Test* menghasilkan statistik ujian yang dipanggil nilai-t, yang mengukur saiz perbezaan berbanding dengan variasi dalam data sampel. *P-value* pula ialah kebarangkalian yang mengukur bukti terhadap hipotesis nol. Ia memberitahu betapa berkemungkinan untuk memperoleh keputusan sekurang-kurangnya sama ekstrem seperti yang diperhatikan, dengan mengandaikan hipotesis nol adalah benar.

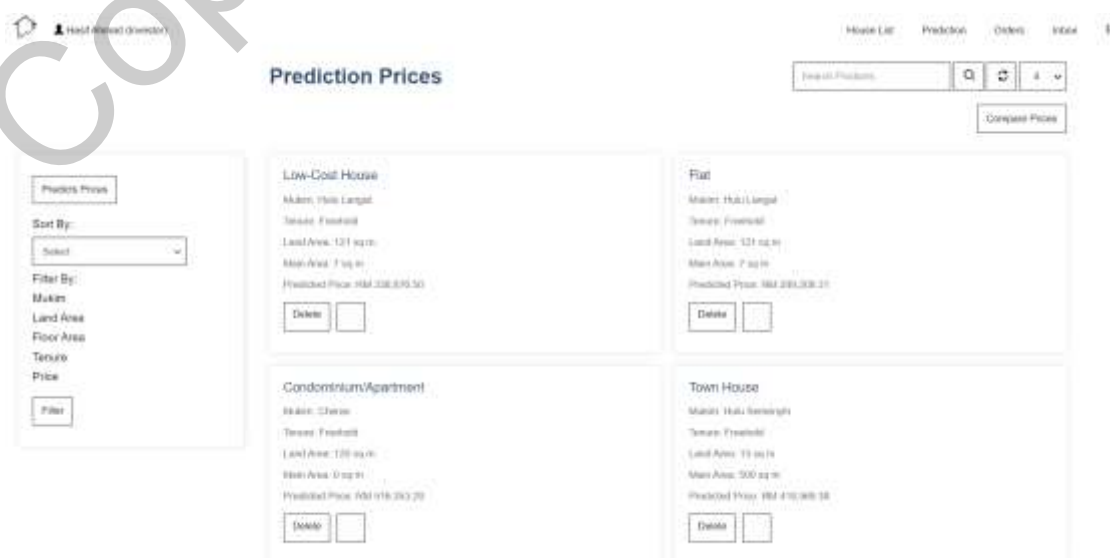
Jadual 5 Nilai Ujian *T-Test* Dan *P-value*.

Metrik Prestasi	Nilai
<i>T-Test</i>	2.5647628339435915
<i>P-value</i>	0.010693252842586987

Jadual 5 menunjukkan nilai *T-Test* dan *P-value* yang didapati hasil daripada ujian yang dilakukan di antara model *XGBoost* tanpa penalaan hiperparameter dan juga model *XGBoost* selepas penalaan hiperparameter. Memandangkan nilai p berada di bawah ambang biasa 0.05, anda akan menolak hipotesis nol. Keputusan ini menunjukkan bahawa terdapat bukti signifikan secara statistik untuk menyatakan bahawa terdapat perbezaan antara cara kedua-dua model yang diuji.

Proses menyepadukan model pembelajaran mesin ke dalam pembangunan web melibatkan beberapa langkah penting. Selepas membina model pembelajaran mesin yang terbaik dengan menggunakan Jupyter Notebook, web sebagai tapak disediakan untuk menyampaikan ramalan model mengenai data baru dalam masa nyata. Proses seterusnya adalah pembangunan melalui API untuk memudahkan interaksi antara model dan web bagi membolehkan pengguna menyerahkan data dan menerima ramalan secara dinamik. Proses pembangunan bahagian hadapan memastikan antara muka pengguna adalah intuitif dan responsif supaya dapat menawarkan pengalaman yang lancar. Proses ini menghubungkan model dengan tapak front-end melalui API *Flask*. Selepas API *Flask* berjaya berjalan, model pembelajaran mesin dapat digunakan di dalam pelayar web.

Proses pembangunan antara muka pula dilakukan bagi membolehkan model prediktif dapat dilakukan secara praktikal oleh pengguna. Proses ini akan menitikberatkan pemilihan atribut yang penting bagi membolehkan input yang penting daripada pengguna sahaja akan diproses ke dalam model ramalan.



Rajah 1 Antara Muka Halaman Prediksi HouseForecast

Rajah 1 menunjukkan antara muka halaman prediksi HouseForecast di mana butang 'Predicts Prices' pada kiri halaman akan membawa pengguna ke borang ramalan. Dapat dilihat juga rekod yang tersimpan hasil keputusan ramalan terdahulu yang telah dibuat oleh pengguna yang disimpan di dalam pangkalan data.

The screenshot shows a web form titled "Predicts House Prices" with a close button (x) in the top right corner. The form contains the following fields:

- Property Type:** A dropdown menu with the placeholder text "Please select one from the list".
- Mukim:** A dropdown menu with the placeholder text "Please select one from the list".
- Tenure:** A dropdown menu with the placeholder text "Please select one from the list".
- Land/Parcel Area (sq m):** A text input field with the placeholder text "Please enter a value".
- Main Floor Area (sq m):** A text input field with the placeholder text "Please enter a value".
- Predict & Save:** A button located at the bottom of the form.

Rajah 2 Antara Muka Borang Ramalan dalam HouseForecast.

Rajah 2 menunjukkan antara muka bagi borang ramalan harga rumah di mana pengguna diminta untuk memilih dan memasukkan data mengikut atribut yang dinyatakan. Pengguna yang mengisi kesemua atribut sahaja yang berjaya melakukan ramalan apabila menekan butang Predict & Save. Butang tersebut akan membawa pengguna untuk melihat keputusan ramalan harga rumah yang dikehendaki.

### **Pengujian Prestasi dan Kebolehgunaan**

Maklum balas yang dikumpulkan melalui Google Form yang direka bagi menilai prestasi dan kebolehgunaan model ramalan di dalam HouseForecast. Pengujian dilakukan daripada beberapa aspek seperti kebolehgunaan dan pengalaman pengguna, kefungsiannya, dan ketepatan ramalan. Penilaian membantu dalam usaha meningkatkan dan menambah baik sistem sama ada dari front-end mahupun back-end bagi memenuhi keperluan pengguna pada masa akan datang.

Jadual 2 menunjukkan skor min yang diterima daripada setiap item aspek hubungan Ibu Bapa dan Anak. Item 1, 2, 8 dan 9 adalah yang tertinggi dengan skor min 4.87 di mana item 4 adalah yang paling rendah dengan skor min 4.60. Walaubagaimanapun, min keseluruhan ialah 4.77 yang dianggap Tinggi kerana ia melebihi 3.65 seperti mengikut Jadual 1.

Jadual 2 Skor Min Hubungan Ibu Bapa-Anak



No.	Soalan
1	<i>How easy was it to navigate and use this form?</i>
2	<i>Were the instructions and labels on the form clear and understandable?</i>
3	<i>How satisfied are you with the design and layout of the form?</i>
4	<i>Were the dropdown options (for Property Type, Mukim, and Tenure) comprehensive and relevant?</i>
5	<i>Were the input fields for Land/Parcel Area and Main Floor Area appropriate and easy to fill out?</i>
6	<i>How satisfied are you with the accuracy of the house price predictions?</i>
7	<i>How satisfied are you with the speed at which the predictions were generated?</i>
8	<i>Do you feel the predicted house prices are relevant to the current market conditions?</i>
9	<i>How does the accuracy of this prediction tool compare to other house price prediction tools you have used?</i>
10	<i>For which purposes do you find the house price predictions most useful?</i>
11	<i>How would you rate your overall experience using this house price prediction form?</i>

Berdasarkan hasil penilaian terhadap projek HouseForecast, terdapat beberapa kesimpulan penting yang boleh diambil mengenai prestasi dan kebolegunaan sistem ini. Pertama sekali, kemudahan navigasi dan penggunaan borang ramalan mendapat maklum balas positif dari majoriti pengguna. Sebanyak 95% daripada responden menilai borang tersebut sebagai mudah atau sangat mudah untuk digunakan, menunjukkan bahawa borang ini direka dengan baik dan mesra pengguna.

Arahan dan label pada borang juga mendapat penilaian tinggi. Semua responden mendapati arahan dan label adalah jelas dan mudah difahami. Ini menunjukkan bahawa maklumat pada borang disampaikan dengan berkesan, membantu pengguna dalam mengisi borang dengan betul tanpa kekeliruan.

Reka bentuk dan susun atur borang diterima baik oleh pengguna, dengan 95% responden menyatakan kepuasan yang tinggi terhadap aspek ini. Ini menunjukkan bahawa borang tersebut bukan sahaja menarik dari segi visual tetapi juga disusun dengan logik dan mudah diikuti. Pilihan luncur turun dalam borang dianggap komprehensif dan relevan oleh semua responden, yang menunjukkan bahawa borang ini memenuhi keperluan pengguna dari segi kegunaan.

Medan input untuk Luas Tanah dan Luas Lantai Utama dianggap sesuai dan mudah diisi oleh 90% responden. Walaupun terdapat sedikit ketidakpastian (10% menjawab "Mungkin"), ini menunjukkan bahawa medan input umumnya diterima baik dan hanya memerlukan sedikit penambahbaikan untuk mencapai kepuasan sepenuhnya.

Kebanyakan pengguna berpuas hati dengan ketepatan ramalan harga rumah, dengan 90% menyatakan kepuasan yang tinggi. Ini mencerminkan prestasi ramalan yang baik dari model yang digunakan dalam projek ini. Selain itu, kelajuan ramalan juga diterima dengan baik, dengan 75% responden berpuas hati atau sangat berpuas hati, menunjukkan bahawa

sistem berfungsi dengan efisien.

Ramalan harga rumah dianggap relevan dengan keadaan pasaran semasa oleh semua responden, dengan 100% menyatakan persetujuan. Ini menunjukkan bahawa model ramalan selaras dengan dinamik pasaran semasa, memberikan ramalan yang boleh dipercayai. Alat ramalan harga rumah ini juga dianggap lebih baik daripada alat lain oleh majoriti responden, dengan 85% menilai alat ini sebagai lebih baik atau jauh lebih baik, menunjukkan keunggulan projek ini berbanding pesaing.

Dari segi tujuan penggunaan, ramalan harga rumah paling berguna untuk membeli dan menjual hartanah, diikuti oleh pelaburan dan penilaian untuk tujuan cukai. Ini menunjukkan pelbagai kegunaan alat tersebut dalam pasaran hartanah. Pengalaman keseluruhan pengguna sangat positif, dengan 100% responden menyatakan pengalaman yang baik atau cemerlang, menunjukkan bahawa alat ini sangat dihargai oleh pengguna.

Secara keseluruhannya, penilaian menunjukkan bahawa projek HouseForecast telah berjaya memenuhi keperluan pengguna dari segi kebolegunaan, ketepatan, dan prestasi. Terdapat tahap kepuasan yang tinggi dalam semua aspek yang dinilai, menunjukkan bahawa projek ini direka dan dilaksanakan dengan baik. Beberapa kawasan kecil untuk penambahbaikan mungkin wujud, tetapi secara keseluruhan, alat ini mendapat maklum balas yang sangat positif.

### **Cadangan Penambahbaikan**

Cadangan penambahbaikan adalah penting untuk memastikan HouseForecast kekal relevan dalam jangka masa panjang. Salah satu penambahbaikan utama yang dipertimbangkan ialah pengumpulan data yang lebih bervariasi dari segi atribut serta lokasi. Dengan menambah pelbagai atribut dan memperluas lokasi pengumpulan data, HouseForecast boleh menjadi lebih inklusif dan berguna kepada pengguna di pelbagai tempat. Ini bukan sahaja meningkatkan ketepatan ramalan, tetapi juga menjadikan alat ini lebih relevan kepada pengguna yang berbeza.

Penambahan visualisasi bagi keputusan ramalan yang meliputi garis masa yang berbeza juga merupakan penambahbaikan yang penting. Dengan menyediakan visualisasi yang padat dan berinformasi, pengguna dapat memahami tren dan perubahan dalam data ramalan dengan lebih baik. Visualisasi yang baik boleh membantu pengguna membuat keputusan yang lebih tepat dan berinformasi, sekaligus meningkatkan pengalaman pengguna secara keseluruhan.

Selain itu, menambah baik ketepatan keputusan ramalan melalui pengujian model pada set data yang lebih banyak dan berbeza adalah satu lagi cadangan penting. Dengan menguji model pada pelbagai set data, HouseForecast dapat menjadi lebih fleksibel dan mampu menangani kepelbagaian data dengan lebih baik. Ini akan meningkatkan ketepatan

dan kebolehpercayaan ramalan, menjadikan alat ini lebih berguna dan dipercayai oleh pengguna. Secara keseluruhan, penambahbaikan ini akan menjadikan HouseForecast lebih baik dan mesra pengguna, memastikan ia kekal relevan dan berguna dalam jangka masa panjang.

## KESIMPULAN

Pembangunan model ramalan berintegrasi web HouseForecast telah disiapkan dalam tempoh masa yang ditetapkan. Pembangunan ini menepati objektif kajian di mana berjaya membangunkan model prediktif bagi harga pasaran rumah serta menilai prestasi akhir bagi keputusan ramalan. Sistem ini dibangun bagi memberi manfaat kepada pengguna yang terlibat dalam urusan pembelian dan penjualan rumah. Keperluan sistem dan reka bentuk sistem juga telah dijelaskan dan dipenuhi. Bab ini akan menghuraikan lebih lanjut mengenai kekangan dan cadangan penambahbaikan model ramalan pada masa hadapan.

### **Kekuatan Sistem**

Kekuatan model ramalan yang berintegrasi dengan web HouseForecast boleh dihurai daripada dua aspek iaitu melalui prestasi model ramalan dan juga kebolehgunaan bahagian ramalan di dalam web. Melalui pembangunan model ramalan yang dilatih menggunakan 3 model berbeza juga ditambah dengan menambah baik model sedia ada melalui penalaan hiperparameter telah membantu mengeluarkan keputusan ramalan yang lebih jitu dan juga mengurangkan peratusan ralat dengan seminimanya. Hal ini memberikan prestasi terbaik dari segi keputusan ramalan untuk disampaikan kepada pengguna. Selain itu, keputusan ramalan yang dijana akan disimpan di dalam pangkalan yang menjadikan web HouseForecast adalah tempat lebih baik daripada sesetengah laman web yang lain. Kebolehgunaan model ramalan juga menjadi tunjang kepada projek ini memandangkan laman antara muka bagi borang ramalan di dalam HouseForecast adalah mesra pengguna. Pengecilan terhadap setiap ruang input yang perlu diisi pada borang ramalan menjadikan ia lebih ringkas dan mudah bagi pengguna.

### **Kelemahan Sistem**

Namun dalam masa yang sama terdapat kekangan dan halangan bagi projek ini, misalnya kekurangan sumber data transaksi yang terkini mengurangkan sedikit terhadap pengeluaran keputusan ramalan yang lebih baik. Ketiadaan beberapa pembolehubah penting yang sepatutnya ada di dalam set data telah menghadkan wawasan bagi projek ini. Di samping itu, sumber data yang sepatutnya dalam jangka masa yang panjang juga tidak berjaya didapatkan sepenuhnya. Hal ini menyebabkan beberapa keperluan dan kefungsi projek perlu disemak semula memandangkan bentuk data yang dirancang sebelum ini mampu menghasilkan keputusan ramalan yang lebih bersifat trend ataupun grafik jangka masa.

## PENGHARGAAN

Penulis kajian ini ingin mengucapkan setinggi-tinggi penghargaan kepada penyelia saya iaitu Dr.Ruzakkiah Binti Jenal di atas segala bantuan dan bimbingan yang dicurahkan bagi menyiapkan projek ini.

Penulis kajian ini juga ingin mengucapkan terima kasih kepada semua pihak yang membantu secara langsung mahupun tidak langsung dalam menyempurnakan projek ini. Segala bantuan yang telah dihulurkan amatlah dihargai kerana tanpa bantuan mereka, projek ini tidak dapat dilaksanakan dengan baik. Semoga tuhan merahmati dan memberikan balasan yang terbaik.

## RUJUKAN

- Team Loanstreet. 2021. 4 Faktor yang Mempengaruhi Harga Hartanah. <https://loanstreet.com.my/ms/pusat-pembelajaran/faktor-mempengaruhi-harga-hartanah> [25 January 2024].
- Bernamea. 2023. Pasaran perumahan terus mantap pada tempoh separuh kedua 2022 - BNM | Astro Awani. AstroAwani <https://www.astroawani.com/berita-malaysia/pasaran-perumahan-terus-mantap-pada-tempoh-separuh-kedua-2022-bnm-413090> [25 January 2024].
- Hotz, N. 2023. What is CRISP DM? — Data Science Process Alliance. <https://www.datascience-pm.com/crisp-dm-2/> [25 January 2024].
- What Is XGBoost and How Does It Improve Machine Learning? 2024. Analytics Vidhya <https://www.analyticsvidhya.com/blog/2018/09/an-end-to-end-guide-to-understand-the-math-behind-XGBoost/> [25 January 2024].
- Beheshti. 2023. Random Forest Regression. A basic explanation and use case in 7... Towards Data Science. Towards Data Science <https://towardsdatascience.com/random-forest-regression-5f605132d19d> [25 January 2024].
- Sruthi, E R. 2024. Building a Random Forest Model: A Step-by-Step Guide. Analytics Vidhya <https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/> [25 January 2024].
- Zainuddin. 2022. Permintaan rumah di Malaysia semakin meningkat. BeritaHarian <https://www.bharian.com.my/bisnes/hartanah/2022/11/1032959/permintaan-rumah-di-malaysia-semakin-meningkat> [25 January 2024].

*Alif Akmal Bin Nor Hisham (A187266)*

*Dr. Ruzakkiah Binti Jenal*

Fakulti Teknologi & Sains Maklumat

Universiti Kebangsaan Malaysia