

# **SISTEM RAMALAN HARGA SEWA PANGSAPURI DAN KONDOMINIUM MENGGUNAKAN MODEL PEMBELAJARAN MESIN (RENTIMATE)**

Loo Yong Li  
Nor Samsiah Sani

Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia

## **ABSTRAK**

Didorong oleh perkembangan transformasi digital yang semakin giat, harta tanah dijual atau disewakan dengan menggunakan platform dalam talian telah menjadi fenomena sosial di Malaysia. Kedua-dua pihak penawaran dan permintaan dalam perniagaan harta tanah berhadapan dengan gangguan pasaran yang disebabkan oleh persaingan pasaran dan ketelusan maklumat yang belum pernah terjadi sebelum ini. Dalam projek ini, data kondominium dan pangsapuri dikumpul dari laman web perumahan tempatan untuk menerapkan pemodelan ramalan dalam ramalan harga sewa harta tanah. Empat model yang berbeza diterapkan dan dibandingkan, iaitu Regresi Linear, Regresi Lasso, Regresi hutan rawak dan regresi xgboost. Set data yang merangkumi harga sewa dan 13 ciri pangsapuri dan kondominium dengan beberapa nilai yang hilang di 13 daerah berbeza di Selangor, Malaysia telah dikumpulkan. Hasilnya menunjukkan bahawa di antara algoritma yang digunakan, model regresi xgboost yang menggunakan 9 atribut yang mempunyai korelasi yang tinggi dengan atribut sasaran mencatatkan ralat kuasa dua punca minimum yang paling rendah iaitu (RMSE) 0.012. Pendekatan ini dapat diterapkan untuk menilai nilai pasar harta tanah sewa, dan hasil ramalan dapat digunakan sebagai petunjuk dari berbagai fenomena bandar dan memberikan rujukan praktikal untuk pemilik rumah dan penyewa. Aplikasi web dibangunkan dengan mengintegrasikan model pembelajaran mesin yang terbaik daripada ujian sebagai model ramalan berdasarkan maklumat yang dimasukkan.

## **1 PENGENALAN**

Fakta menunjukkan bahawa hartanah merupakan salah satu aset utama yang menyumbang kepada perkembangan ekonomi Malaysia. Hartanah juga mempengaruhi kestabilan sosial oleh itu

merupakan unsur yang sering kali digunakan sebagai rujukan kepada keadaan ekonomi semasa. Menurut laporan pasaran hartanah Malaysia daripada Jabatan Penilaian dan Perkhidmatan Harta Malaysia (NAPIC) pada suku pertama tahun 2020, pemulihan sektor hartanah bergantung kepada faktor domestik dan faktor luaran seperti kestabilan politik, keadaan minyak global dan perkembangan berkaitan pandemik Covid-19. NAPIC mendedahkan bahawa pangsapuri perkhidmatan Malaysia yang berstatus tergantung menambah 3.3% iaitu 31,661 unit bernilai RM20.03 bilion pada suku pertama 2020, pada suku yang sama pada tahun sebelum ini, bilangan residen tergantung adalah 30,664 unit bernilai RM18.82 bilion. NAPIC juga menyatakan bahawa Index Harga Perumahan Malaysia (MHPI) terus berkembang secara sederhana. Bilangan pangsapuri tergantung bertambah sebanyak 26.5% iaitu 21,683 unit bernilai RM18.64 bilion. Dalam satu laporan terkini oleh kumpulan penyelidikan pelaburan daripada Maybank menyatakan ekonomi negara pada masa kini menunjukkan fenomena yang lemah disebabkan penutupan perniagaan dan kadar pemberhentian kerja yang meningkat.

Harga sewa pangsapuri dan kondominium dipengaruhi oleh pelbagai faktor. Dalam penyelidikan yang dijalankan oleh ahli sains komputer, model-model menganalisis data kebanyakannya hanya meramal keputusan berdasarkan modal ramalan yang mudah. Disebabkan situasi ini berlakunya sisihan yang besar dalam perubahan jangka masa yang pendek. Isu-isu yang timbul akan mempengaruhi keputusan meramal, antara isunya termasuk plot grafik berlebihan dan kekurangan atau data pemfaktoran data yang bising. Plot grafik berlebihan dan kekurangan akan menghasilkan kesalahan ramalan pada data ujian manakala Penyelidikan ini akan memperkenalkan pembelajaran mesin berasaskan teknik penggabungan. Data set perumahan yang berjumlah 10,713 yang mengandungi 13 ciri-ciri telah dikumpul dari laman web hartanah tempatan untuk penyelidikan ini. Data set yang dikumpul akan dibahagikan mengikut ratio untuk kegunaan pembelajaran, pengesahan dan percubaan.

## **2 PENYATAAN MASALAH**

Kegiatan penyewaan hartanah di Malaysia didorong oleh pelbagai faktor seperti penyewaan untuk pekerjaan atau urusan sementara di sesuatu tempat. Selain itu, harga jual hartanah yang semakin meningkat sehingga tidak mampu ditanggung oleh kebanyakan orang telah menukar pilihan orang ramai untuk memilih menyewa rumah untuk menetap sebagai alternatif utama terutamanya di bandar-bandar besar seperti Kuala Lumpur dan Klang Valley. Pihak pembekal dan pihak penuntut

dalam situasi ekonomi negara yang tidak stabil ini, amat berwaspada dalam melakukan aktiviti jual beli. Kini dengan peningkatan keperluan dalam pasaran penyewaan ini, penyewa mempunyai pilihan yang lebih banyak dari segi lokasi, kualiti dan nilai hartanah yang ingin disewa. Pemilik hartanah perlu menjaga kelengkapan fasiliti untuk menarik perhatian para penyewa. Trend mempamerkan hartanah untuk disewa atas talian telah mendorong penambahan bilangan hartanah yang dipaparkan untuk disewa di pelantar-pelantar hartanah atas talian. Oleh itu, untuk meningkatkan persaingan hartanah dalam pasaran yang kian berkembang besar ini, ciri-ciri hartanah yang dipaparkan menyumbang kepada peningkatan nilai sewa sesebuah hartanah. Antara ciri-ciri asas yang akan dinyatakan termasuk bilangan katil dan bilik air, lokasi, harga sewa, kemudahan yang dibekalkan dan pelbagai informasi mengenai fasiliti yang dibekalkan. Setakat ini, belum lagi ada sistem di Malaysia yang dibangunkan untuk meramal nilai pangsapuri dan kondominium dari segi ciri-ciri hartanah menggunakan algoritma pembelajaran mesin. Walaupun nilai sesebuah hartanah juga dipengaruhi oleh faktor luaran seperti bilangan kedai makan dan kemudahan awam di sekelilingnya, namun begitu dalam projek ini akan memberikan fokus kepada ramalan nilai sesebuah hartanah berdasarkan faktor dalaman terutamanya fasiliti yang dibekalkan. Keputusan ramalan mempunyai nilai rujukan bukan sahaja kepada para penyewa, tetapi juga kepada pemilik hartanah dan juga pihak pemaju. Untuk penyewa yang ingin mencari hartanah dengan harga berpatutan, model ini dapat membantu untuk menapis pilihan hartanah yang sesuai berdasarkan ciri-ciri pilihan mereka. Kepada pihak pemilik hartanah dan pemaju pula, modal ini mempunyai kegunaan seperti mengenal pasti nilai hartanah yang dimiliki, sebagai rujukan pasaran hartanah di sesuatu kawasan dan sebagai panduan untuk meningkatkan nilai hartanah.

### **3 OBJEKTIF KAJIAN**

Objektif kajian projek ini adalah untuk membangunkan web aplikasi berintegrasi model pembelajaran mesin yang mampu meramal harga sewa sesebuah pangsapuri atau kondominium berdasarkan ciri-ciri yang dikumpul.

Web aplikasi ini dibangunkan dengan langkah seperti :

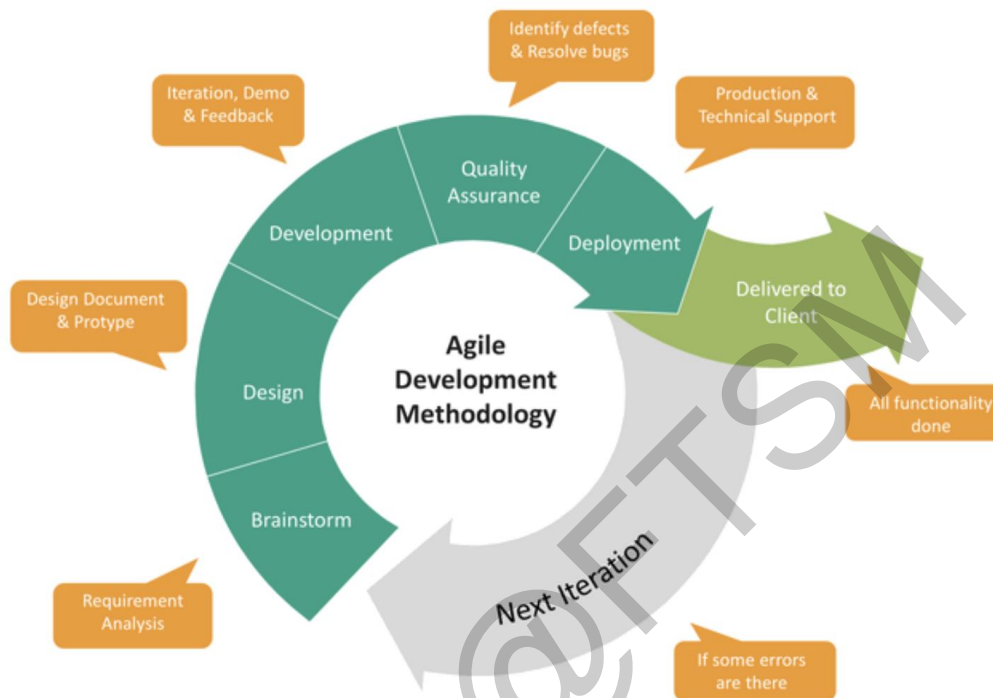
- a) Mengumpul data dari laman web hartanah menggunakan teknik mengikis laman web (*web scrapping*).

b) Membangunkan model pembelajaran mesin dengan data yang dikumpul untuk menjalankan ramalan harga sewa hartanah. Empat jenis algoritma akan diuji dan model dengan prestasi terbaik akan dipilih.

c) Membina satu aplikasi web berintegrasi model pembelajaran mesin yang dipilih. Web aplikasi ini menyediakan ruang untuk memasukkan ciri-ciri dari pengguna untuk menjana keputusan harga sewa Ramlan hartanah.

#### **4 METODOLOGI**

Dalam projek ini, kaedah Kitar Hayat Pembangunan Sistem (SDLC) yang akan digunakan adalah model tangkas (*Agile model*) untuk membina web aplikasi ramalan harga sewa ini. Model tangkas merupakan proses pembangunan perisian yang terdiri daripada beberapa peringkat. Model ini menitik berat kepada kelengkapan komponen fungsi sistem yang akan dibangunkan dalam masa yang singkat. Pendekatan ini tidak mementingkan penetapan tugas dan jadual masa, semua matlamat dan masa dipadatkan dinamakan dengan istilah “pecut”. Setiap pecutan mempunyai satu jangka masa dengan senarai matlamat yang mampu disiapkan dalam masa yang ditetapkan. Setelah setiap pecutan habis, keputusan dan prestasi sistem akan dinilai untuk mendapat maklumat pengubahsuai untuk pecutan seterusnya. Rajah menunjukkan fasa-fasa metodologi bagi modal tangkas.



*Rajah 1 Fasa-fasa metodologi model tangkas*

### **i. FASA PERCAMBAHAN FIKIRAN (*BRAINSTORM*)**

Dalam Fasa ini, untuk memastikan sistem yang bakal dibangunkan memenuhi keperluan sasaran pengguna, pelbagai artikel, jurnal, buletin dan laporan kajian mengenai domain ini perlu dikumpulkan untuk memahami situasi semasa. Pengetahuan mengenai domain kajian adalah sangat penting untuk membangunkan sistem yang mempunyai nilai perniagaan yang diakui oleh pengguna sistem ini. Selain itu, dalam peringkat ini juga perlu memikirkan kerangka jangkaan fungsian sistem dari segi kegunaan dan perbincangan keseluruhan sistem secara abstrak.

### **ii. FASA REKA BENTUK (*DESIGN*)**

Dalam fasa ini, model dan prototaip yang mewakili sistem akan dibangunkan. Sistem yang direka akan berdasarkan reka bentuk yang meningkatkan kebolegunaan sistem oleh pengguna. Dalam fasa ini, reka bentuk untuk setiap keadaan akan direka dengan berdasarkan kesesuaian pengguna untuk memasukkan maklumat. Reka bentuk sistem haruslah mudah faham oleh pelbagai pengguna dan tidak melibatkan operasi sistem yang rumit untuk mendapatkan keputusan.

**iii. FASA PEMBANGUAN (*DEVELOPMENT*)**

Dalam fasa ini, dengan semua sumber dan keperluan yang dikumpul, prototaip sistem akan dibangunkan demi melakukan demo terhadap komponen fungsi. Dalam kajian ini, prestasi pelbagai modal pembelajaran mesin akan dinilai dari segi ketepatan ramalan harga sewa berbanding dengan harga sewa sebenar. Keselesaan menggunakan sistem dari segi pengguna akan dibuat ulasan.

**iv. FASA JAMINAN QUALITI (*QUALITY ASSURANCE*)**

Fasa ini akan mengenal pasti kekurangan yang menyebabkan prestasi yang tidak memuaskan dan kesilapan dalam algoritma yang digunakan. Masalah-masalah yang tertimbul akan direkodkan untuk diubahsuai pada pecutan yang seterusnya.

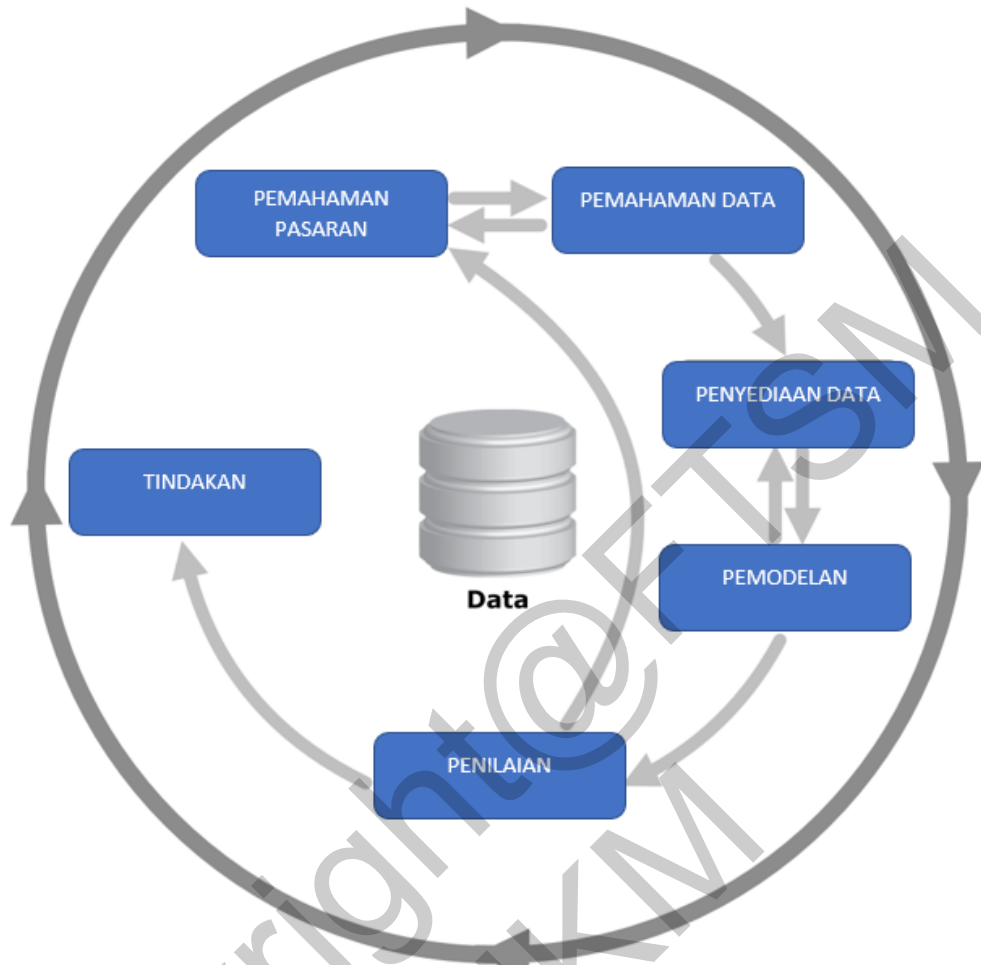
**v. FASA PENYEBARAN (*DEPLOYMENT*)**

Menurut catatan prestasi sistem pada fasa yang sebelumnya, fasa ini akan menentukan sama ada sistem yang dibangunkan sudah bersedia untuk disebar ke server untuk kegunaan awam. Sistem juga akan diperlengkap dari segi sokongan teknikal dan pengeluaran. Sistem yang dibangunkan hanya akan disebar sekiranya semua komponen fungsi mencapai prestasi yang memuaskan.

**iv. ULANGAN SETERUNYA (*NEXT ITERATION*)**

Sekiranya berlaku kesalahan dan tidak kepuasan terhadap prestasi sistem, ulangan yang seterusnya akan dijalankan bagi mengubahsuai dan menambah baik sistem berdasarkan ulasan daripada aliran yang sebelum ini.

Corak reka bentuk yang digunakan dalam projek ini metodologi *CRISP-DM*. Corak ini sesuai digunakan untuk menjalankan projek pembinaan web aplikasi yang berintegrasi model pembelajaran mesin. Corak ini mengandungi 6 fasa untuk membina projek secara teratur dan fasa-fasa ini boleh berkisar berdasarkan keperluan pembangun. Antara fasa yang terkandung termasuk fasa pemahaman perniagaan (*business understanding*), fasa pemahaman data (*data understanding*), fasa penyediaan data (*data preparation*), fasa pemodelan (*modeling*), fasa penilaian (*evaluation*) dan fasa tindakan(*deployment*).



*Rajah 2 Methodologi pembangunan model (CRISP-DM)*

### **Pemahaman Perniagaan**

Pada fasa pemahaman perniagaan, pembangun akan memahami konteks kepada tujuan projek supaya pembangun memperoleh tanggapan yang jelas tentang data yang berkaitan dengan model perniagaan. Nilai harga sewa hartanah berubah-ubah dari semasa ke semasa. Maklumat-maklumat yang memberikan kesan kepada perubahan ini tidak boleh diputuskan dengan pengetahuan am dan pendapat subjektif. Untuk mempelajari sebab perubahan nilai harga sewa hartanah dalam aspek komputeran, kajian-kajian lepas dan laporan pasaran yang mengenai tajuk ini telah dikumpul dan dikaji. Projek ini menggunakan teknik pembelajaran mesin untuk menjalankan fungsi ramalan, oleh itu kajian mengenai kaedah untuk mengumpul data dan

menjalankan latihan mesin perlu dilakukan dengan tatacara yang lebih berkomputeran dalam projek ini. Hasil fasa ini ialah pembangun dapat memahami secara menyeluruh mengenai konteks projek yang dilakukan dan mengumpul data merujuk kepada ciri-ciri berguna yang tertentu dari laman web hartanah dunia sebenar. Penggunaan data dari dunia sebenar boleh meningkatkan nilai keputusan ramalan model pembelajaran mesin yang dibangunkan untuk membuat ramalan harga sewa hartanah.

### Pemahaman Data

Pada fasa pemahaman data pula bertujuan untuk memahami jenis dan bentuk yang perlu diperolehi. Dalam fasa kualiti data akan dianalisis dari pelbagai aspek seperti kelengkapan, nilai, pertaburan dan pematuhan tadbir urus data. Dalam fasa ini juga perlu membincangkan nilai yang paling berguna untuk dikumpulkan untuk kegunaan pembinaan model pembelajaran mesin dengan merujuk kepada kajian lepas yang merangkumi bidang yang berkaitan. Pembangun sistem dapat mengenal pasti jenis data yang perlu dikumpul dan mengubahsuaikan format data kepada keadaan yang sesuai untuk diproses oleh algoritma yang bakal digunakan untuk menjalankan pembelajaran mesin. Dengan menggunakan data yang diproses, keputusan ramalan harga sewa dapat dijanakan.

Sumber	Data	Keterangan
Laman web hartanah tempatan	Harga sewa	Harga sewa setiap bulan
	Harga persegi	Harga setiap unit meter persegi
	Ruang	Kebesaran hartanah dalam meter persegi
	Bilik tidur	Bilangan bilik tidur
	Bilik air	Bilak bilik air
	Jenis	Jenis hartanah, kondominium atau pangsapuri
	Tingkat	Kedudukan tingkat hartanah
	Perabot	Keadaan perabot dalam hartanah tersebut



	Tahun bina	Tahun hartanah dibina
	Ciri-ciri	Kemudahan-kemudahan yang disediakan oleh hartanah
	Lokasi	Kedudukan lokasi hartanah yang dikumpul

Jadual 1 Ringkasan pengenalan sumber data kasar

Sumber data yang dikumpul dalam menjalankan projek ini adalah daripada pelbagai laman web tempatan Malaysia. Data yang diperoleh merupakan maklumat pada *December* 2020. 14000 data mentah dengan 11 atribut telah dikumpul secara rawak dan tiada data peribadi yang dikumpul untuk memastikan integriti projek ini.

## Penyediaan Data

Fasa yang seterusnya ialah fasa penyediaan data yang akan menjalankan proses *ETL* (*Extract, Transform, Load*). Proses ini menjadikan data yang dikumpul ke dalam keadaan yang berguna untuk melaksanakan algoritma pembelajaran mesin. Pembersihan data bertujuan untuk mengisi data yang hilang, menggantikan data yang berulang atau membuang data yang tidak berkualiti. Data yang hilang bermaksud set data tersebut adalah tidak lengkap.

address	price	bed	bath	sqft	psf	developer	buildyear	floorlevel	furnish	tenure	type	feature	facilities
-Jalan Kibang Besar, Kampung Batang Timur 5, Klebu	1,500	3	2	1,400	1.07	N/A	N/A	Middle Fl Fully Furr	Freehold	Condomin	{}	{feature: ["TV"], facilities: ["Air-Conditioning"], ["Audio System"], ["Balcony"], ["Bathub"], ["Bed"], ["Cable TV"], ["Chinacapsule"], ["Dishwasher"], ["Gymnasium room"], ["Playground"], ["24 hours security"]]}	
-Jalan Melaka Raya 23, Taman Hillir Kota 1, Bandar Hilir	1,800	3	2	1,000	0.85	N/A	N/A	Middle Fl Partially Leasehold	Apartment	{}	{feature: ["TV"], facilities: ["Business Centre"], ["Facilities-Car Park"], ["Facilities-Children's Playground"], ["Facilities-Dining range"], ["Facilities-Eateries"], ["Facilities-Playground"], ["Facilities-Bar"], ["Facilities-Dining Room Furniture"], ["Facilities-Fridge"], ["Facilities-Kitchen Utensils"], ["Facilities-Living Room Furniture"], ["Facilities-Television"]]}		
Melaka Tengah, Melaka	900	1	2	900	0.92	N/A	N/A	Middle Fl Partially N/A	Apartment	{}	{feature: ["TV"], facilities: ["Covered car park"], ["Facilities-24 hours security"]]}		
Melaka Tengah, Melaka	1,300	4	2	1,300	1	Hubdaya S	2006	N/A	Partially Freehold	Apartment	{feature: ["TV"], facilities: ["Covered car park"], ["Facilities-Driving range"], ["Facilities-Eateries"], ["Facilities-Lift lobby"], ["Facilities-Mini-Mart"], ["Facilities-Restaurant"], ["Facilities-Sauna"], ["Facilities-Swimming pool"], ["Facilities-Tennis court"]]}		
Melaka Tengah, Melaka	1,900	3	2	1,088	1.47	N/A	N/A	High Floor Fully Furr N/A	N/A	Apartment	{feature: ["TV"], facilities: ["Barbecue Area"], ["Facilities-Cafe"], ["Facilities-Car Park"], ["Facilities-Children's Playground"], ["Facilities-Game room"], ["Facilities-Gymnasium room"], ["Facilities-Sauna"], ["Facilities-Swimming pool"]]}		
casalaga, Bandar Hillir, Melaka	3,800	4	2	1,440	2.64	N/A	N/A	High Floor Fully Furr N/A	Freehold	Condomin	{feature: ["TV"], facilities: ["Playground"], ["Facilities-Sauna"], ["Facilities-24 hours security"], ["Facilities-Squash court"], ["Facilities-Swimming pool"], ["Facilities-Tennis court"]]}		
Casa Lago apartment, Melaka Tengah, Melaka	1,300	2	2	858	1.52	N/A	N/A	High Floor Fully Furr Leasehold	Apartment	{feature: ["TV"], facilities: ["Perimeter Fencing"], ["Facilities-24 hours security"]]}			
Bukit लगुंग, Melaka Tengah, Melaka	1,300	2	2	1,100	1	N/A	N/A	Freehold	Condomin	{feature: ["TV"], facilities: ["Covered car park"], ["Facilities-Gymnasium room"], ["Facilities-Playground"], ["Facilities-24 hours security"], ["Facilities-Swimming pool"]]}			
Jalan Tun Perak, Melaka Tengah, Melaka	2,200	3	2	1,088	2.02	N/A	N/A	Middle Fl Fully Furr N/A	Freehold	Condomin	{feature: ["TV"], facilities: ["Barbecue Area"], ["Facilities-Car Park"], ["Facilities-Covered car park"], ["Facilities-Children's Playground"], ["Facilities-Clubhouse"], ["Facilities-Furniture"], ["Facilities-Sauna"], ["Facilities-Swimming pool"]]}		
Bukit लगुंग, Melaka Tengah, Melaka	2,200	1	2	1,689	1.74	N/A	N/A	Middle Fl Fully Furr N/A	Freehold	Condomin	{feature: ["TV"], facilities: ["Covered car park"], ["Facilities-Gymnasium room"], ["Facilities-Jogging track"], ["Facilities-Playground"], ["Facilities-24 hours security"], ["Facilities-Swimming pool"], ["Facilities-Tennis court"]]}		
Melaka Tengah, Melaka	1,200	2	2	1,645	1.86	N/A	N/A	High Floor Fully Furr N/A	Freehold	Condomin	{feature: ["TV"], facilities: ["Covered car park"], ["Facilities-Clubhouse"], ["Facilities-Playground"], ["Facilities-24 hours security"], ["Facilities-Swimming pool"], ["Facilities-Tennis court"]]}		
Bukit Beruang, Melaka	1,500	2	2	1,033	1.45	N/A	N/A	Middle Fl Fully Furr Freehold	Freehold	Condomin	{feature: ["TV"], facilities: ["Covered car park"], ["Facilities-Sauna"], ["Facilities-24 hours security"], ["Facilities-Squash court"], ["Facilities-Swimming pool"], ["Facilities-Tennis court"]]}		
Melaka Raya, Melaka	1,500	2	2	870	1.72	N/A	N/A	High Floor Fully Furr N/A	Freehold	Condomin	{feature: ["TV"], facilities: ["Barbecue Area"], ["Facilities-Gymnasium room"], ["Facilities-Playground"], ["Facilities-24 hours security"], ["Facilities-Swimming pool"]]}		
Jalan Murchi Abdullah, Batu Berendam, Melaka	1,850	3	2	980	1.89	N/A	N/A	High Floor Fully Furr N/A	Freehold	Condomin	{feature: ["TV"], facilities: ["Covered car park"], ["Facilities-Playground"], ["Facilities-24 hours security"]]}		
Melaka Tengah, Melaka	1,600	4	2	1,300	1.23	N/A	N/A	Middle Fl Fully Furr Freehold	Apartment	{feature: ["TV"], facilities: ["Covered car park"], ["Facilities-Playground"], ["Facilities-24 hours security"]]}			
Melaka Tengah, Melaka	3,800	3	2	1,200	3.17	N/A	N/A	N/A	N/A	Condomin	{}		
A'Famosa Resort Alor Gajah, Alor Gajah, Melaka	450	1	N/A	821	0.55	N/A	N/A	N/A	Leasehold	Condomin	{feature: ["TV"], facilities: ["Covered car park"], ["Facilities-Gymnasium room"], ["Facilities-Jogging track"], ["Facilities-Mini-Mart"], ["Facilities-Playground"], ["Facilities-24 hours security"], ["Facilities-Swimming pool"]]}		
Parkland no-006, Melaka Tengah, Melaka	1,800	3	2	1,088	1.65	N/A	N/A	High Floor Fully Furr N/A	Freehold	Condomin	{}		
Jalan Persiaran Bunga Raya, Melaka Tengah, Melaka	3,200	3	2	1,300	2.46	N/A	N/A	High Floor Fully Furr Freehold	Apartment	{feature: ["TV"], facilities: ["Barbecue Area"], ["Facilities-Covered car park"], ["Facilities-Clubhouse"], ["Facilities-Gymnasium room"], ["Facilities-Jacuzzi"], ["Facilities-Playground"], ["Facilities-Sauna"], ["Facilities-Swimming pool"]]}			
Bukit Beruang, Melaka	1,000	3	2	1,100	0.91	N/A	N/A	Low Floor Fully Furr N/A	Apartment	{}			
Kibang, Melaka	1,100	3	2	980	1.12	N/A	N/A	Middle Fl Fully Furr Freehold	Condomin	{feature: ["TV"], facilities: ["Covered car park"], ["Facilities-Gymnasium room"], ["Facilities-24 hours security"], ["Facilities-Swimming pool"]]}			
-, Taman Melaka Raya Taman Melaka Raya, Melaka	1,400	3	2	1,200	1.17	N/A	N/A	Middle Fl Fully Furr N/A	Apartment	{feature: ["TV"], facilities: ["Car Park"]]}			
Bukit Beruang, Melaka	850	3	2	980	0.87	N/A	N/A	High Floor Fully Furr N/A	Apartment	{feature: ["TV"], facilities: ["Covered car park"], ["Facilities-Gymnasium room"], ["Facilities-Playground"], ["Facilities-24 hours security"], ["Facilities-Swimming pool"]]}			
Melaka Tengah, Melaka	1,200	2	2	1,645	1.86	N/A	N/A	Middle Fl Fully Furr N/A	Freehold	Condomin	{feature: ["TV"], facilities: ["Playground"], ["Facilities-24 hours security"], ["Facilities-Swimming pool"]]}		
Bukit baru jaya, Melaka Tengah, Melaka	1,500	4	2	1,230	1.22	N/A	N/A	Middle Fl Fully Furr N/A	Condomin	{feature: ["TV"], facilities: ["24 hours security"]]}			
Melaka Tengah, Melaka	1,400	1	2	1,689	2.03	N/A	N/A	Middle Fl Fully Furr N/A	Condomin	{feature: ["TV"], facilities: ["Basement hall"], ["Facilities-Basketball court"], ["Facilities-Barbecue Area"], ["Facilities-Covered car park"], ["Facilities-Gymnasium room"], ["Facilities-Sauna"], ["Facilities-Swimming pool"]]}			
- Jalan Kota Laksmanna 3/3, Taman Maharyaya, Melaka	900	3	2	850	0.71	N/A	N/A	High Floor Fully Furr N/A	Partially Leasehold	Apartment	{feature: ["TV"], facilities: ["Open car park"]]}		
Melaka Raya, Melaka	1,100	1	2	1,435	2.53	N/A	N/A	High Floor Fully Furr N/A	Condomin	{feature: ["TV"], facilities: ["Gymnasium room"], ["Facilities-Swimming pool"]]}			
Jalan Tun Perak, Melaka Tengah, Melaka	1,800	3	2	1,088	1.65	N/A	N/A	Middle Fl Fully Furr Freehold	Condomin	{feature: ["TV"], facilities: ["Basketball court"], ["Facilities-Basement car park"], ["Facilities-Cafe"], ["Facilities-Covered car park"], ["Facilities-Driving range"], ["Facilities-Perimeter Fencing"], ["Facilities-Sauna"], ["Facilities-Swimming pool"]]}			
Melaka Tengah, Melaka	2,200	3	2	1,152	1.91	N/A	N/A	High Floor Fully Furr N/A	Condomin	{}			
Ayer Keroh, Melaka	1,100	3	2	850	1.29	N/A	N/A	N/A	Freehold	Condomin	{feature: ["TV"], facilities: ["Barbecue Area"], ["Facilities-Car Park"], ["Facilities-Covered car park"], ["Facilities-Gymnasium room"], ["Facilities-Library"], ["Facilities-Playground"], ["Facilities-Sauna"], ["Facilities-Swimming pool"]]}		
Kibang, Melaka	1,500	3	2	1,000	1.5	N/A	N/A	Ground Fl Fully Furr N/A	Condomin	{feature: ["TV"], facilities: ["Basement hall"], ["Facilities-Covered car park"], ["Facilities-24 hours security"], ["Facilities-Swimming pool"]]}			
Intan court view pasir, Ujong Pasir, Melaka	900	3	2	1,008	0.84	N/A	N/A	Middle Fl Fully Furr N/A	Condomin	{feature: ["TV"], facilities: ["Covered car park"], ["Facilities-Gymnasium room"], ["Facilities-Playground"], ["Facilities-24 hours security"], ["Facilities-Swimming pool"]]}			

Rajah 3 Data mentah yang dikumpul dari laman web

Rajah di atas menunjukkan salah satu set data mentah yang dikumpulkan daripada laman web hartanah tempatan. Terdapat pelbagai keadaan yang perlu ditangani (*noise*) dalam set data ini.

Contohnya, ciri-ciri *feature* dan *facilities* mengandungi maklumat mengenai kemudahan yang dibekalkan oleh hartanah tertentu. Keadaan data dalam ciri-ciri ini adalah dalam bentuk yang tidak mampu digunakan oleh algoritma pembelajaran mesin. Dengan menggunakan *pandas*, data yang diperlukan dikenal pasti dan kolom baru yang bersifat *boolean* diwujudkan, seperti kemunculan penghawa dingin, kolam renang dan tempat letak kereta.

bed	bath	sqft	psf	developer	buildyear	floorlevel	furnishing	tenure	type	feature	facilities	location	aircond	parking	pool	balcony	bedroom	area	gym	rent
3	2	1058	1.23	SYF Development		Low Floor	Partially F	Freehold	Condomin	["feature"	["facilities"	Kajang	1	0	1	1	3	1058	1	1300
3	2	1045	0.91				Fully Furni	Freehold	Apartment	["feature"	["facilities"	Kajang	0	0	0	0	3	1045	0	950
3	2	1124	1.33	Cheras Ho	2014	Low Floor	Fully Furni	Freehold	Apartment	["feature"	["facilities"	Kajang	1	0	1	1	3	1124	1	1500
Studio		450	2.22	Binastra Li	2017			Freehold	Condomin	["feature"	["facilities"	Kajang	0	1	1	0	1	450	1	1000
4	2	1425	1.26	TLS Group	2018		Fully Furni	Freehold	Condomin	["feature"	["facilities"	Kajang	0	0	1	0	4	1425	1	1800
3	2	1110	0.81	Permata C	2015	Middle Flo	Unfurnish	Freehold	Condomin	["feature"	["facilities"	Kajang	0	0	1	1	3	1110	1	900
3	2	920	1.47				Partially Furnished		Condomin	["feature"	["facilities"	Kajang	1	0	1	1	3	920	0	1350
3	2	1249	1				Partially Furnished		Condomin	["feature"	["facilities"	Kajang	0	0	0	0	3	1249	0	1250
3	2	1056	0.95			Middle Flo	Unfurnish	Freehold	Condomin	["feature"	["facilities"	Kajang	0	0	0	0	3	1056	0	1000
3	2	1021	0.98	Cheras Ho	2014	Middle Flo	Partially F	Freehold	Condomin	["feature"	["facilities"	Kajang	0	0	1	0	3	1021	1	1000
3	2	900	1.22	SHL Conso	2006		Fully Furni	Freehold	Condomin	["feature"	["facilities"	Kajang	0	1	1	0	3	900	1	1100
3	2	1080	1.76				Fully Furni	Freehold	Condomin	["feature"	["facilities"	Kajang	0	0	0	0	3	1080	1	1900
Studio		1450	2.11	Binastra Li	2017		Fully Furni	Freehold	Condomin	["feature"	["facilities"	Kajang	1	1	1	1	1	450	1	950
3	2	1345	1.41				Fully Furni	Freehold	Condomin	["feature"	["facilities"	Kajang	1	0	1	0	3	1345	1	1900
4	2	1292	1.93	Sime Darb	2017	Middle Flo	Fully Furni	Freehold	Apartment	["feature"	["facilities"	Kajang	0	0	1	0	4	1292	1	2500
3	2	1200	1	TLS Group	2013		Partially F	Freehold	Condomin	["feature"	["facilities"	Kajang	0	0	1	0	3	1200	1	1200
3	2	872	1.61	Sime Darb	2017	Middle Flo	Unfurnish	Freehold	Apartment	["feature"	["facilities"	Kajang	0	0	1	0	3	872	1	1400
3	3	1023	1.17	Cheras Ho	2019			Freehold	Condomin	["feature"	["facilities"	Kajang	0	0	1	0	3	1023	1	1200
3	2	1080	1.48				Fully Furni	Freehold	Condomin	["feature"	["facilities"	Kajang	1	0	0	1	3	1080	1	1600
3	2	951	1.16	Mitraland	2018		Partially F	Freehold	Condomin	["feature"	["facilities"	Kajang	0	0	1	0	3	951	1	1100
3	2	1080	1.34				Fully Furni	Freehold	Condomin	["feature"	["facilities"	Kajang	0	0	0	0	3	1080	1	1450
3	2	1080	1.3				Partially F	Freehold	Condomin	["feature"	["facilities"	Kajang	1	0	0	1	3	1080	1	1400
3	2	1300	1.46	TLS Group	2015	Middle Flo	Fully Furni	Freehold	Condomin	["feature"	["facilities"	Kajang	1	0	1	1	3	1300	1	1900
3	2	1058	0.95	SYF Development		Middle Flo	Unfurnish	Freehold	Condomin	["feature"	["facilities"	Kajang	1	0	1	1	3	1058	1	1000
3	2	1024	1.07	Cheras Ho	2014	Middle Flo	Partially F	Freehold	Apartment	["feature"	["facilities"	Kajang	0	0	1	0	3	1024	1	1100
4	2	1700	1.12	TLS Group	2019	Middle Flo	Fully Furni	Freehold	Condomin	["feature"	["facilities"	Kajang	1	1	1	1	4	1700	1	1900

Rajah 4 Data yang telah diproses untuk penggunaan pembelajaran mesin

Pengeluaran data adalah proses untuk membuang atribut atau data daripada set data setelah mengkaji kepentingan sama ada akan memberi impak atau tidak kepada hasil kajian. Pengeluaran data terbahagi kepada dua iaitu pengurangan jumlah data dan pengurangan jumlah ciri-ciri. Data yang hilang boleh digantikan dengan data purata untuk ciri-ciri yang dirujuk atau terus mengeluarkan baris data yang tidak lengkap daripada set data. Seperti kolom *developer* yang banyak muncul kehilangan data akan dikeluarkan daripada set data kerana atribut tersebut tidak memberikan impak yang besar kepada keputusan ramalan. Ciri-ciri bernilai dipilih adalah merujuk kepada kajian-kajian lepas kepada topik yang sama mengenai ciri-ciri yang memberi impak kepada keputusan ramalan. Selepas melakukan transformasi terhadap data mentah, 10,713 data telah dikekalkan dengan 14 atribut.

BIL	Ciri-ciri	Bentuk	Keterangan
1.	Daerah	Nominal	Kedudukan hartanah

2.	Harga sewa	Ratio	Harga sewa dalam Ringgit Malaysia
3.	Ruang	Ratio	Keluasan hartanah dalam meter persegi
4.	Bilik tidur	Ratio	Bilangan bilik yang boleh dijadikan bilik tidur
5.	Bilik air	Ratio	Bilangan bilik air
6.	Tingkat	Ordinal	Kedudukan hartanah dari segi tingkat
7.	Perabot	Nominal	Kelengkapan perabot yang disediakan
8.	Harga meter persegi	Ratio	Harga meter persegi dalam Ringgit Malaysia
9.	Jenis hartanah	Nominal	Jenis hartanah sama ada ialah kondominium atau pangsapuri
10.	Penghawa dingin	Nominal	Menyediakan kemudahan penghawa dingin atau tidak
11.	Balkoni	Nominal	Mempunyai balkoni atau tidak
12.	Tempat letak kereta	Nominal	Menyediakan tempat letak kereta atau tidak
13.	Kolam renang	Nominal	Mempunyai kolam renang atau tidak
14.	Gym	Nominal	Mempunyai kemudahan latihan atau tidak

*Jadual 2 Senarai atribut yang dikekalkan*

Proses transformasi data yang dilakukan dalam projek ini ialah pembinaan atribut (*attribute construction*). Selepas menjalankan teknik *one hot encoding*, data yang bersifat *ratio* pada asalnya iaitu daerah, dan jenis hartanah akan ditransformasikan kepada data nominal.

## **Pemodelan**

Fasa pemodelan merupakan teras kepada projek ini. Setelah memproses set data kepada bentuk yang sesuai untuk melakukan pembinaan model pembelajaran mesin, pelbagai model dibina dan diuji keputusan. Model dengan keputusan terbaik digunakan untuk melaksanakan fungsi ramalan dalam web aplikasi. Antara teknik yang dipilih untuk menjalankan ujian pembinaan model sistem dalam projek ini ialah Regresi Linear, regresi Lasso, regresi hutan rawak dan regresi xgboost. Setiap algoritma akan dihasilkan dengan model latihan dan diuji dengan data ujian. Model dengan keputusan pemodelan terbaik akan dipilih untuk pembangunan web aplikasi.

## Penilaian

Pada fasa penilaian, pembangun akan menilai sama ada keputusan yang dijana adalah tepat dan betul. Jikalau keputusan yang diperoleh kurang memuaskan, corak seni bina ini menjalankan corak untuk mengkaji semula daripada fasa yang pertama untuk memahami punca kekurangan kepuasan terhadap keputusan terkini. Dalam projek ini, data yang dikumpul akan diasingkan kepada set latihan dan set ujian untuk mengesahkan ketepatan model dalam kenyataan semasa fasa pembinaan model.

Dengan menguji model terlatih dan menandakan set data dengan ramalan, adalah mungkin untuk mengira ralat ramalan model purata model terlatih dan seberapa baik ramalan tersebut sesuai dengan nilai (label) yang betul secara keseluruhan. Ralat mutlak min (MAE) adalah perbezaan mutlak min antara ramalan label dan label sebenar. MAE menunjukkan ralat rata-rata untuk model terlatih. Kesalahan jenis ini tidak menunjukkan seberapa besar kesalahan tersebut. Maksud mutlak kesalahan dikira sebagai :

$$MAE = \frac{\sum | \text{label ramalan} - \text{label sebenar} |}{n}$$

Di mana  $n$  adalah bilangan entri data yang akan diramalkan.

Ralat kuasa dua rata-rata (RMSE) seperti MAE juga merupakan ukuran ralat antara label ramalan dan label sebenar. Perbezaan antara dua cara ini mengukur ralat adalah bahawa RMSE mengambil kira variasi kesalahan. Ini bermaksud bahawa jika terdapat variasi besar dari kesalahan

dari ramalan ke ramalan, maka RMSE semakin besar. Ini mempunyai kesan bahawa RMSE sama dengan atau (seperti dalam kebanyakan kes) lebih besar daripada MAE. Ralat kuasa dua bermaksud boleh dikira sebagai:

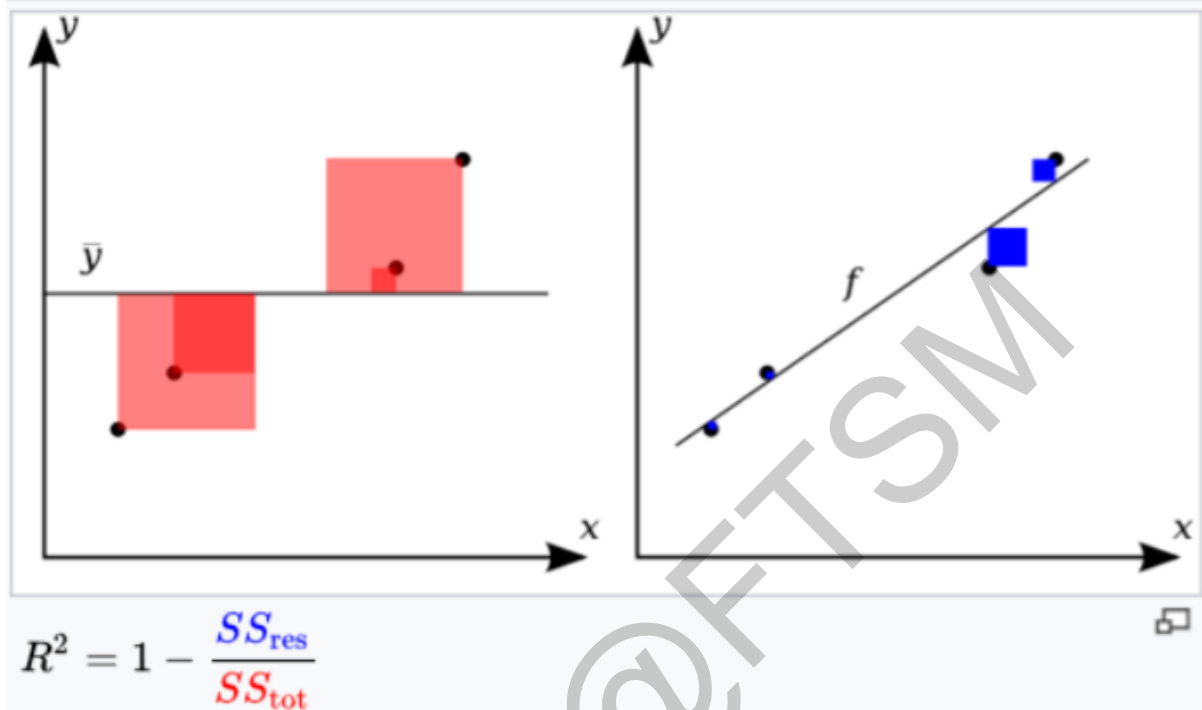
$$RMSE = \sqrt{\frac{\sum(\text{label ramalan} - \text{label sebenar})^2}{n}}$$

Di mana  $n$  adalah bilangan entri data yang akan diramalkan.

Pekali penentuan atau r-kuasa dua boleh digunakan sebagai ukuran untuk menerangkan bagaimana model berfungsi semasa diperkenalkan ke set data ujian. Medan mentakrifkan kuasa dua sebagai ukuran yang memberikan maklumat mengenai hubungan antara atribut. Begitu juga Kit Kemahiran Akademik, Universiti Newcastle menggambarannya sebagai korelasi antararamalan buat dan nilai sebenar. Pekali penentuan diukur antara 0-1 di mana 1 bermaksud model terlatih mempunyai kesesuaian yang sempurna, dan 0 bermaksud terlatih model adalah rawak sepenuhnya. R-kuasa dua dikira sebagai:

$$R^2 = 1 - \frac{\sum(\text{label sebenar} - \text{label ramalan})^2}{\sum(\text{label sebenar} - \text{purata}(\text{label sebenar}))^2}$$

Dalam statistik, pekali penentuan, dilambangkan  $R^2$  adalah bahagian varians dalam pemboleh ubah bersandar yang dapat diramalkan dari pemboleh ubah tidak bersandar. Semasa menilai kebaikan-fit dari simulasi ( $Y_{pred}$ ) berbanding nilai yang diukur ( $Y_{obs}$ ), tidak sesuai untuk mendasarkan ini pada  $R^2$  dari regresi linear (iaitu,  $Y_{obs} = m \times Y_{pred} + b$ ). The  $R^2$  mengukur tahap sebarang korelasi linier antara  $Y_{obs}$  dan while, sementara untuk penilaian kebaikan hanya satu korelasi linear khusus yang harus dipertimbangkan:  $Y_{obs} = 1 \times Y_{pred} + 0$  (iaitu, garis 1: 1)



Semakin baik regresi linier (di sebelah kanan) sesuai dengan data dibandingkan dengan purata yang lebih sederhana (pada grafik kiri), semakin dekat nilai  $R^2$  dengan 1. Kawasan residu kuasa dua biru sehubungan dengan regresi linear. Luas kotak merah mewakili sisa-sisa squared berkenaan dengan nilai purata.

Satu set data mempunyai  $n$  nilai yang ditandai  $y_1, \dots, y_n$  (secara kolektif dikenali sebagai  $y_i$  atau sebagai vektor  $y = [y_1, \dots, y_n]^T$ ), masing-masing dikaitkan dengan nilai yang dipasang (atau dimodelkan, atau diramalkan)  $\hat{y}_1, \dots, \hat{y}_n$ .

Tentukan baki sebagai  $e_i = y_i - \hat{y}_i$ . Sekiranya  $y$  adalah min bagi data yang diperhatikan, kita ada

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Jadi, kebolehubahan set data dapat diukur dengan menggunakan rumus tiga jumlah kotak:

- a. Jumlah kuasa dua (sebanding dengan varians data):

$$SS_{tot} = \sum_i (\hat{y}_i - \bar{y})^2$$

b. Jumlah regresi kuadrat, juga disebut jumlah kuadrat yang dijelaskan:

$$SS_{reg} = \sum_i (\hat{y}_i - \bar{y})^2$$

c. Jumlah kuadrat sisa, juga disebut jumlah petak baki:

$$SS_{res} = \sum_i (y_i - \hat{y}_i)^2 = \sum_i e_i^2$$

Definisi yang paling umum bagi pekali ( $R$  kuasa dua) penentuan adalah:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

$R^2$  adalah statistik yang akan memberikan beberapa maklumat mengenai kebaikan kesesuaian model. Andaikan  $R^2 = 0.49$ . Ini menunjukkan bahawa 49% kebolehubahan pemboleh ubah bersandar telah diambil kira untuk, dan baki 51% kebolehubahan masih belum dijelaskan. Dalam regresi,  $R^2$  pekali penentuan adalah ukuran statistik sejauh mana ramalan regresi kira titik data sebenar. Seorang  $R^2$  dari 1 menunjukkan bahawa ramalan regresi sangat sesuai data.

## Tindakan

Langkah yang terakhir merupakan fasa tindakan. Dalam projek ini, model yang dibina akan disebarkan dalam bentuk aplikasi web. Web aplikasi ini menyediakan ruang untuk membolehkan pengguna untuk mengisi data dan mendapatkan keputusan ramalan daripada sistem. Langkah ini merupakan fasa terakhir yang tidak terkandung dalam kitaran seni bina ini. Untuk projek ini, web aplikasi telah dibina dengan menggunakan HTML dan dihias dengan CSS. Selepas itu, web aplikasi akan dimuat naik ke server melalui *github* dan *heroku*.

## 5 HASIL KAJIAN

Hasil kajian projek ini adalah mampu melakukan pembinaan model pembelajaran mesin dengan menggunakan pelbagai kombinasi atribut iaitu semua (semua atribut), atribut kuat (nilai korelasi

0.3-0.5) dan sederhana (nilai korelasi 0.2-0.29). Semua model pembelajaran mesin telah dinilai keputusannya dengan menggunakan penilai metrik R kuasa dua dan rmse. Keputusannya regresi xgboost mencatatkan nilai yang terbaik untuk R kuasa dua (0.999) dan rmse (0.012) oleh itu dipilih sebagai model untuk membangunkan web aplikasi. Web aplikasi telah direka dengan HTML dan dihias dengan CSS, dengan menggunakan Flask, web aplikasi berintegrasi model pembelajaran mesin telah berjaya dibangunkan.

### Penilaian Model

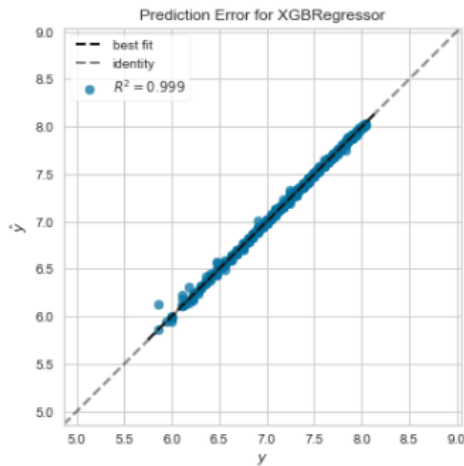
#### R kuasa dua

R kuasa dua adalah pengukuran statistik yang mewakili bahagian varians bagi pemboleh ubah bersandar yang dijelaskan oleh pemboleh ubah bebas. Ketepatan setiap mesin algoritma pembelajaran dapat dikemukakan dengan R kuasa dua. Dari jadual di bawah, hutan rawak dan *XgBoost* mencatatkan nilai yang sangat cemerlang apabila semua atribut digunakan untuk pembangunan model. Ini bermaksud keputusan ramalan bagi dua model ini adalah sangat mendekati nilai sebenar data pengujian.

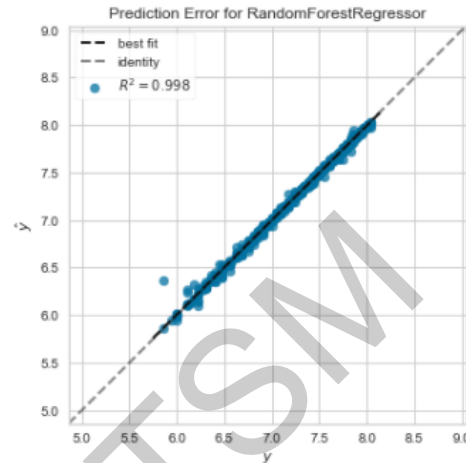
<i>Algoritma</i>	<i>Semua</i>	<i>Kuat</i>	<i>Sederhana</i>
<i>Regresi Linear</i>	<b>0.612</b>	<b>0.446</b>	<b>0.178</b>
<i>Regresi Lasso</i>	<b>0.612</b>	<b>0.446</b>	<b>0.177</b>
<i>Regresi Hutan Rawak</i>	<b>0.998</b>	<b>0.656</b>	<b>0.236</b>
<i>Regresi XGBoost</i>	<b>0.999</b>	<b>0.660</b>	<b>0.236</b>

*Jadual 3 Nilai R kuasa 2 setiap model*





Rajah 5 Kesalahan ramalan XgBoost



Rajah 6 Kesalahan ramalan hutan rawak

Rajah di atas menunjukkan rajah R kuasa dua daripada model XgBoost dan hutan rawak yang menggunakan data latihan semua atribut. Kedua-dua model ini mempunyai keputusan R kuasa dua yang terbaik. Seperti yang ditunjukkan, nilai yang diramal amat mendekati keputusan sebenar, hal ini boleh dikenalpasti dengan lokasi titik dengan garis paling sesuai.

### **Root mean square error (RMSE)**

RMSE adalah sisihan piawai bagi kesalahan ramalan. Ia mewakili sisihan piawai sampel perbezaan antara nilai ramalan dan nilai yang diperhatikan (disebut baki pemboleh ubah tidak bersandar). Nilai yang lebih rendah ialah ralat yang menunjukkan kesesuaian algoritma. Bacaan terbaik dicatatkan oleh algoritma hutan rawak dan *XgBoost* apabila menggunakan semua atribut untuk menjalankan latihan model. Kedua-dua model ini mencatatkan nilai 0.017 dan 0.012 apabila diuji dengan set data ujian. Kesimpulannya untuk langkah pembangunan web aplikasi, model *XgBoost* dipilih sebagai model ramalan web aplikasi yang dibangunkan.

<i>Algoritma</i>	<i>Semua</i>	<i>Kuat</i>	<i>Sederhana</i>
<i>Regresi Linear</i>	<b>0.244</b>	<b>0.291</b>	<b>0.355</b>
<i>Regresi Lasso</i>	<b>0.244</b>	<b>0.291</b>	<b>0.354</b>
<i>Regresi Hutan Rawak</i>	<b>0.017</b>	<b>0.229</b>	<b>0.342</b>
<i>Regresi XGBoost</i>	<b>0.012</b>	<b>0.227</b>	<b>0.342</b>

*Jadual 4 Nilai RMSE setiap model*

Akhirnya, saya akan meramal harga sewa menggunakan model *XgBoost* yang dilatih dengan data latihan yang mengandungi semua atribut untuk menguji dengan set data ujian. Untuk melihat perbezaan antara nilai keputusan dan nilai sebenar, semua harga sewa akan disusun ke label bingkai data dan label akan membandingkan harga dari set data ujian. Model juga meramalkan nilai dari indeks yang berbeza. Dari jadual di bawah, kita dapat menilai bahawa, nilai ramalan yang diperoleh sangat hampir dengan nilai sebenar.

<b>Index</b>	<b>Nilai Sebenar</b>	<b>Nilai Ramalan</b>
6693	7.438384	7.43424
6092	6.684612	6.6862698
1111	7.130899	7.134758

*Jadual 5 Pengujian Keputusan ramalan menggunakan model XgBoost*

### **Pembangunan web aplikasi berintegrasi model pembelajaran mesin**

Penerapan model pembelajaran mesin adalah proses membuat model tersedia dalam produksi di mana aplikasi web, perisian perusahaan dan API dapat menggunakan model terlatih dengan memberikan input data baru dan menghasilkan ramalan. Menerusi penilaian yang dibuat model *xgboost* dipilih sebagai model ramalan kepada web aplikasi ini.

Selepas membina model pembelajaran mesin yang terbaik dengan menggunakan *Jupyter Notebook*, seterusnya ialah mula membina aplikasi web yang dapat menghubungkan model dan menghasilkan ramalan mengenai data baru dalam masa nyata. Terdapat dua bahagian aplikasi ini:

- *Front-end* (direka dengan HTML)
- *Back-end* (Disebarkan melalui *Flask* dalam *Python*)

**RENTIMATE**

Enter the following values to predict  
the property renting price

Property Type

Area (100-6000) <input type="text"/>	Psf (0.2-10) <input type="text"/>
Number of bathroom <input type="text"/>	Number of bedroom <input type="text"/>
Furnishing <input type="text"/>	Pool <input type="text"/>
Aircond <input type="text"/>	Gym <input type="text"/>

*Rajah 7 Antara muka web aplikasi*

Gambar rajah di atas menunjukkan rekaan akhir antara muka web aplikasi. Terdapat ruang-ruang yang disediakan untuk memasukkan data hartanah yang ingin diramal. Selepas mengisi semua ruang, butang *predict* boleh ditekan untuk melakukan ramalan. Dengan menekan butang ini, semua pengguna boleh menjana keputusan ramalan harga sewa dengan menggunakan model pembelajaran mesin yang dibina. Manakala butang *Restore Input Values* pula disediakan untuk mengembalikan nilai yang dimasukkan oleh pengguna sebelum ini.

## 6 KESIMPULAN

Sistem ramalan harga sewa pangsapuri dan kondominium ini dijangka dapat memberikan rujukan tambahan kepada mana-mana pihak yang terlibat dalam aktiviti bidang hartanah. Kemudahan ini memainkan peranan penting di negara yang sedang membangun seperti Malaysia untuk mewujudkan satu protokol penilaian yang telus dan adil. Pembangunan sistem ini juga memberikan pemahaman dan inspirasi yang mendalam kepada ahli peniagaan dan penyelidik yang kurang jelas tentang teknik pembelajaran mesin dalam melakukan ramalan. Sistem ini membolehkan sesiapa sahaja untuk menggunakan model pembelajaran mesin untuk melakukan ramalan dengan mudah. Pengumpulan data mengalami kekangan semasa projek ini dilakukan kerana data hartanah merupakan data yang perlu dikumpul dengan berwaspada. Oleh itu pelaksanaan pembinaan model mempunyai kekurangan kelengkapan ciri-ciri data dari segi

pembinaan model yang lebih lengkap. Akibat kekurangan pengalaman dalam pengumpulan dan pengendalian data, pelbagai tutorial dan latihan dijadikan rujukan untuk menyempurnakan projek ini.

## 7 RUJUKAN

- Fan, C., Cui, Z., and Zhong, XHouse Prices Prediction with Machine Learning Algorithms. In Proceedings of the 2018 10th International Conference on Machine Learning and Computing <https://doi.org/10.1145/3195106.3195133>
- Mary Lotz Waterfall vs. Agile: Which is the Right Development Methodology for Your Project? 5 July 2018 <https://www.seguetech.com/waterfall-vs-agile-methodology/>
- LALAINÉ C. DELMENDO , Malaysia's great house price boom is over, 27 Jun 2020 <https://www.globalpropertyguide.com/Asia/Malaysia/Price-History>
- A. Varma, A. Sarma, S. Doshi and R. Nair, "House Price Prediction Using Machine Learning and Neural Networks," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 1936-1939, doi: 10.1109/ICICCT.2018.8473231.
- T.M. Oshiro, P.S. Perez and J.A. Baranauskas, "How many trees in a random forest?", Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7376, pp. 154-168, 2012, ISBN 9783642315367.
- R.J. Shiller, Understanding recent trends in house prices and home ownership, National Bureau of Economic Research, Working Paper 13553, Oct. 2007.
- Yingyu Feng, Kelvyn Jones: Comparing multilevel modelling and artificial neural networks in house price prediction. ICSDM 2015: 108-114.
- Jiang, L.; Phillips, P.C.B.; Yu, J. A New Hedonic Regression for Real Estate Prices Applied to the Singapore Residential Market. Technical Report, Cowles Foundation Discussion Paper No. 1969. 2014. Available online: <https://ssrn.com/abstract=2533017> (accessed on 2 November 2020).
- Jinze Li.: Monthly Housing Rent Forecast Based on LightGBM (Light Gradient Boosting) Model. ISSN: 2307-0692, Volume 7, Issue 6, December, 2018