

**LAPORAN**  
**MEMPREDIKSI PENJUALAN HARGA RUMAH DARI**  
**DATA KING COUNTRY MENGGUNAKAN**  
**ALGORITMA *LINEAR REGRESSION***



**UNIVERSITAS  
TEKNOLOGI  
SUMBAWA**

Dosen Pengampu:

**Herfandi, M. Kom**

Disusun Oleh:

- |                     |                 |
|---------------------|-----------------|
| 1. Mar'I Yustiardin | (20.01.013.009) |
| 2. Nizar Mulyawan   | (20.01.013.011) |
| 3. Sulastri         | (20.01.013.015) |
| 4. Andini Wulandari | (20.01.013.020) |
| 5. Fitrah Ramdhani  | (20.01.013.021) |

**PROGRAM STUDI TEKNIK INFORMATIKA**  
**FAKULTAS REKAYASA SISTEM**  
**UNIVERSITAS TEKNOLOGI SUMBAWA**  
**2022**

## **KATA PENGANTAR**

Segala puji bagi Allah SWT. yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan laporan dengan judul "Memprediksi Harga Rumah Dari Data King Country Menggunakan Algoritma Linear Regression" sebagai pelengkap nilai Ujian Akhir Semester 3 pada mata kuliah Artificial Intelligence.

Diharapkan dalam penulisan laporan ini dapat melengkapi nilai Ujian Akhir Semester 3 ini sehingga dapat lulus pada mata kuliah Artificial Intelligence yang diampuh oleh Bapak Herfandi, M. Kom. dengan nilai terbaik.

Penulis menyadari bahwa masih banyak kesalahan dan kekurangan dalam penulisan laporan ini, maka berbagai saran dan kritik membangun sangat diharapkan bagi penulis untuk diperbaiki. Semoga laporan ini dapat bermanfaat khususnya bagi penulis serta bagi pembaca pada umumnya.

Sumbawa, 13 Januari 2022

Penulis

## ABSTRAK

Perkembangan teknologi zaman sekarang sangatlah pesat sehingga manusia sendiri mulai susah untuk mengikuti perkembangan teknologi tersebut. Semua ini dikarenakan mesin dibuat untuk belajar dari semua data-data yang dimuat oleh memory sehingga membuat mesin perkembangannya lebih pesat dari manusia yang telah membuat yang dinamakan “Machine Learning”. Dikarenakan perkembangan mesin tersebut, semua pekerjaan yang tadinya dilakukan oleh manusia diambil alih oleh mesin dengan berasaskan data-data yang tersimpan di memory. Adapun bidang-bidang yang diambil alih itu tidaklah sedikit, diantaranya adalah, bidang pendidikan, ekonomi, bisnis, bahkan sampai bidang kesehatan. Dengan adanya pendekatan *Forecasting* atau prediksi kita dapat mengendalikan pasar dan mengambil keuntungan dari suatu barang. Selain itu, penggunaan Algoritma Linear Regression ini sangat cocok untuk memprediksi harga suatu barang. Kita dapat mengendalikan harga pasar untuk rumah berdasarkan data dari King Country.

### **Kata Kunci:**

*Machine Learning, Forecasting, algoritma Linear Regression, Rumah*

## DAFTAR ISI

KATA PENGANTAR .....	ii
ABSTRAK .....	iii
DAFTAR ISI.....	iv
BAB I PENDAHULUAN .....	1
1.1. LATAR BELAKANG.....	1
1.2. RUMUSAN MASALAH .....	1
1.3. BATASAN MASALAH .....	1
1.4. TUJUAN PENELITIAN .....	1
1.5. MANFAAT PENELITIAN.....	1
1.6. SISTEMATIKA PENULISAN .....	2
BAB II LANDASAN TEORI .....	3
2.1. TINJAUAN PUSTAKA.....	3
2.2. DASAR TEORI.....	3
2.2.1. ALGORITMA LINEAR REGRESSION .....	3
A.PENGERTIAN REGRESSION.....	3
B.REGRESI LINEAR SEDEHANA.....	3
C.REGRESI LINEAR BERGANDA .....	4
D.KOEFISIEN DETERMINASI.....	4
2.2.2. RAPIDMINER.....	4
A.PENGERTIAN RAPIDMINER .....	4
B.SEJARAH RAPIDMINER .....	5
2.2.3. CROSS-INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM) .....	5
A. BUSINESS UNDERSTANDING .....	5
B. DATA UNDERSTANDING .....	5
C. DATA PREPARATION.....	6
D. MODELING .....	6
E. BAB III METODOLOGI PENELITIAN .....	7
3.1. ALAT DAN BAHAN PENELITIAN .....	7
3.1.1. ANALISA KEBUTUHAN PERANGKAT .....	7
3.1.2. ANALISA KEBUTUHAN DATA .....	7

3.2. WAKTU DAN TEMPAT PENELITIAN .....	8
3.2.1. WAKTU PENELITIAN.....	8
3.2.2. TEMPAT PENELITIAN .....	8
BAB IV PEMBAHASAN.....	9
4.1. TAHAPAN PREDIKSI PENJUALAN HARGA RUMAH.....	9
4.1.1. BUSINESS UNDERSTANDING.....	9
4.1.2. DATA UNDERSTANDING .....	9
4.1.3. DATA PREPARATION .....	11
4.1.4. MODELLING .....	13
BAB V PENUTUP.....	16
5.1. KESIMPULAN .....	16
5.2. SARAN .....	16
DAFTAR PUSTAKA .....	17

## DAFTAR TABEL

Tabel 3.1 10 Data teratas Penjualan Rumah .....	7
---	---

## DAFTAR GAMBAR

Gambar 4.1 Data awal Harga Pejualan Rumah King Country.....	9
Gambar 4.2 Data Penjualan Rumah Setelah ditambah Timeseries.....	10
Gambar 4.3 Memilih Menu Blank Process .....	11
Gambar 4.4 Pilih Data yang akan Dikelola.....	11
Gambar 4.5 Pilih Semua Cell Data .....	12
Gambar 4.6 Penetapan Label pada attribute Price .....	12
Gambar 4.7 Pilih Lokasi Penyimpanan data .....	13
Gambar 4.8 Penampakan data di RapidMiner .....	13
Gambar 4.9 Drag Data dari Repository ke Process.....	14
Gambar 4.10 Drag Algortima Linear Regression ke Process .....	14
Gambar 4.11 Sambungkan Data, Algoritma, dan Result .....	14
Gambar 4.12 Statistic dari Result.....	15

# **BAB I PENDAHULUAN**

## **1.1.LATAR BELAKANG**

Perkembangan teknologi zaman sekarang sangatlah pesat sehingga manusia sendiri mulai susah untuk mengikuti perkembangan teknologi tersebut. Semua ini dikarenakan mesin dibuat untuk belajar dari semua data-data yang dimuat oleh memory sehingga membuat mesin perkembangannya lebih pesat dari manusia yang telah membuat yang dinamakan “Machine Learning”.

Dikarenakan perkembangan mesin tersebut, semua pekerjaan yang tadinya dilakukan oleh manusia diambil alih oleh mesin dengan berdasarkan data-data yang tersimpan di memory. Adapun bidang-bidang yang diambil alih itu tidaklah sedikit, diantaranya adalah, bidang pendidikan, ekonomi, bisnis, bahkan sampai bidang kesehatan.

Dalam bidang bisnis, mesin mengambil alih dalam hal prediksi. Prediksi dalam bisnis sangat menguntungkan, karena dengan memprediksi harga suatu barang, kita dapat mengendalikan pasar dan mendapat keuntungan yang lebih.

Pada laporan ini, penulis akan menjelaskan tentang bagaimana cara memprediksi harga jual rumah dari data king country yang akan merupakan titik tumpu dari prediksi yang akan dilakukan oleh penulis.

## **1.2.RUMUSAN MASALAH**

Berdasarkan latar belakang diatas, penulis merumuskan masalah yang akan dipecahkan yaitu “Bagaimana cara atau metode yang digunakan dalam memprediksi harga rumah?”.

## **1.3.BATASAN MASALAH**

Berdasarkan rumusan masalah, penulis membatasi permasalahan yaitu memprediksi penjualan harga rumah hanya dengan menggunakan algoritma *linear regression*. Algoritma tersebut akan menjadi titik fokus penulis pada laporan ini.

## **1.4.TUJUAN PENELITIAN**

Adapun tujuan dari penelitian ini adalah memenuhi tugas akhir penulis pada matakuliah Artificial Intelligence yang diampuh oleh Bapak Herfandi, M. Kom. Selain itu, dapat menambah wawasan bagi para pembaca yang mudah-mudahan akan bermanfaat disuatu hari nanti.

## **1.5.MANFAAT PENELITIAN**

Adapun manfaat dari penelitian ini terbagi sebagai berikut:

- a) Bagi Pembaca, sebagai referensi atau titik tumpu untuk menulis atau melanjutkan penelitian ini. Sekaligus menambah wawasan.
- b) Bagi Penulis, sebagai pemenuhan tugas akhir dan menambah wawasan.

#### **1.6.SISTEMATIKA PENULISAN**

Penelitian ini menerapkan sistematika penulisan sebagai berikut.

- a) BAB I PENDAHULUAN, pada bab ini penulis menjelaskan tentang latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan dalam penelitian ini.
- b) BAB II LANDASAN TEORI, pada bab II ini penulisan memaparkan beberapa referensi yang menjadi titik fokus pada penelitian ini, yang tentunya akan dibahas pada bab pembahasan.
- c) BAB III METODOLOGI PENELITIAN, tak jauh berbeda dengan bab sebelumnya, namun pada bab ini penulis menjelaskan lebih dalam mengenai metode apa saja yang digunakan pada penelitian ini.
- d) BAB IV PEMBAHASAN, pada bab ini penulis memaparkan data yang akan diterapkan dengan metode yang dijelaskan pada bab sebelumnya.
- e) BAB V PENUTUP, pada bab ini penulis menarik kesimpulan tentang penelitian. Selain itu, bab ini juga berisi saran dan masukan untuk pembaca.



## **BAB II LANDASAN TEORI**

### **2.1.TINJAUAN PUSTAKA**

Adapun referensi dari pembuatan laporan ini adalah dari Adipati Martulandi, yang dimana dalam penelitian beliau juga menggunakan metode yang sama, dan mendapatkan hasil yang baik.

Data yang digunakan oleh beliau adalah penjualan harga rumah dibenua eropa yang datanya akan sangat banyak bila dipaparkan. Dari hasil penelitian beliau, Beliau menggunakan 7 tahap dalam pembentukan data terprediksi yakni, *meng-import* library pandas, numpy dan seaborn pada python; menyiapkan data yang akan digunakan; memntukan karakteristik data *from the surface*; mengecek data yang kiranya ada data yang kosong; melakukan exploratory data analysis; modeling; dan yang terakhir melakukan prediksi harga rumah sesuai dengan keinginan.(Adipati Martulandi, 2019)

### **2.2.DASAR TEORI**

#### **2.2.1. ALGORITMA LINEAR REGRESSION**

##### **A. PENGERTIAN REGRESSION**

Analisis regresi (regression analysis) merupakan suatu teknik untuk membangun persamaan dan menggunakan persamaan tersebut untuk membuat perkiraan (prediction). Dengan demikian, analisis regresi sering disebut sebagai analisis prediksi. Dikatakan prediksi karena nilai prediksi tidak selalu tepat dengan nilai riilnya. Semakin kecil tingkat penyimpangan antara nilai prediksi dengan nilai riilnya, maka semakin tepat persamaan regresi yang bentuk. Hal ini dapat didefinisikan bahwa analisa regresi adalah metode statistika yang digunakan untuk menentukan kemungkinan bentuk hubungan antara variabel-variabel dengan tujuan pokok dalam penggunaan metode untuk meramalkan atau memperkirakan nilai dari suatu variabel lain yang diketahui.

##### **B. REGRESI LINEAR SEDEHANA**

Regresi linier sederhana merupakan suatu proses untuk mendapatkan hubungan matematis dalam bentuk suatu persamaan antara variabel tak bebas tunggal dengan variabel bebas tunggal atau dengan kata lain, regresi linier yang hanya melibatkan satu peubah bebas X yang dihubungkan dengan satu peubah tak bebas Y. Bentuk umum model regresi linier sederhana yaitu:

$$Y = b_0 + b_1X_1 + e_i$$

Di mana :  $Y$  = variabel tak bebas (dependen)

$b_0$  = parameter intersep

$b_1$  = koefisien regresi (slop)

$X_1$  = variabel bebas (independen)

$e_i$  = kesalahan penduga

### C. REGRESI LINEAR BERGANDA

Disamping hubungan linier dua variabel, hubungan linier lebih dari dua variabel dapat juga terjadi. Pada hubungan ini, perubahan satu variabel dipengaruhi oleh lebih dari satu variabel lain. Maka regresi linier berganda adalah analisis regresi yang menjelaskan hubungan antara peubah respon (variable dependent) dengan faktor-faktor yang mempengaruhi lebih dari satu predaktor (variable independent).

### D. KOEFISIEN DETERMINASI

Koefisien determinasi yang dinyatakan dengan  $R^2$  untuk pengujian regresi linier berganda yang mencakup lebih dari dua variabel adalah untuk mengetahui proporsi keragaman total dalam variabel tak bebas ( $Y$ ) yang dapat dijelaskan atau diterangkan oleh variabel-variabel bebas ( $X$ ) yang ada dalam model persamaan regresi linier berganda secara bersama-sama.

## 2.2.2. RAPIDMINER

### A. PENGERTIAN RAPIDMINER

RapidMiner adalah platform perangkat lunak ilmu data yang dikembangkan oleh perusahaan bernama sama dengan yang menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran dalam, penambangan teks, dan analisis prediktif. Hal ini digunakan untuk bisnis dan komersial, juga untuk penelitian, pendidikan, pelatihan, rapid prototyping, dan pengembangan aplikasi serta mendukung semua langkah dalam proses pembelajaran mesin termasuk persiapan data, hasil visualisasi, validasi model, dan optimasi. RapidMiner dikembangkan pada model inti terbuka. Dengan RapidMiner Studio Free Edition, yang terbatas untuk 1 prosesor logika dan 10.000 baris data, tersedia di bawah lisensi AGPL.

## **B. SEJARAH RAPIDMINER**

RapidMiner sebelumnya dikenal sebagai YALE (Yet Another Learning Environment), mulai dikembangkan pada tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer dari Unit Kecerdasan Buatan Universitas Teknik Dortmund. Mulai tahun 2006, perkembangannya didorong oleh Rapid-I, sebuah perusahaan yang didirikan oleh Ingo Mierswa dan Ralf Klinkenberg pada tahun yang sama. Pada tahun 2007, nama perangkat lunak itu berubah dari YALE menjadi RapidMiner. Pada tahun 2013, perusahaan melakukan rebranding dari Rapid-I menjadi RapidMiner.

### **2.2.3. CROSS-INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM)**

Cross-Industry Standard Process for Data Mining atau CRISP-DM adalah salah satu model proses datamining (datamining framework) yang awalnya (1996) dibangun oleh 5 perusahaan yaitu Integral Solutions Ltd (ISL), Teradata, Daimler AG, NCR Corporation dan OHRA. Framework ini kemudian dikembangkan oleh ratusan organisasi dan perusahaan di Eropa untuk dijadikan methodology standard non-proprietary bagi data mining. Adapun tahapan CRISP-DM dibagi menjadi enam yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Berikut ini penjelasan singkat mengenai tahapan CRISP-DM:

#### **A. BUSINESS UNDERSTANDING**

Ini adalah tahap pertama dalam CRISP-DM dan termasuk bagian yang cukup vital. Pada tahap ini membutuhkan pengetahuan dari objek bisnis, bagaimana membangun atau mendapatkan data, dan bagaimana untuk mencocokkan tujuan pemodelan untuk tujuan bisnis sehingga model terbaik dapat dibangun. Kegiatan yang dilakukan antara lain: menentukan tujuan dan persyaratan dengan jelas secara keseluruhan, menerjemahkan tujuan tersebut serta menentukan pembatasan dalam perumusan masalah data mining, dan selanjutnya mempersiapkan strategi awal untuk mencapai tujuan tersebut.

#### **B. DATA UNDERSTANDING**

Secara garis besar untuk memeriksa data, sehingga dapat mengidentifikasi masalah dalam data. Tahap ini memberikan fondasi analitik untuk sebuah penelitian dengan membuat ringkasan (summary) dan mengidentifikasi potensi masalah

dalam data. Tahap ini juga harus dilakukan secara cermat dan tidak terburu-buru, seperti pada visualisasi data, yang terkadang insight-nya sangat sulit didapat jika dihubungkan dengan summary data nya. Jika ada masalah pada tahap ini yang belum terjawab, maka akan mengganggu pada tahap modeling.

### **C. DATA PREPARATION**

Secara garis besar untuk memperbaiki masalah dalam data, kemudian membuat variabel derived. Tahap ini jelas membutuhkan pemikiran yang cukup matang dan usaha yang cukup tinggi untuk memastikan data tepat untuk algoritma yang digunakan. Bukan berarti saat Data Preparation pertama kali dimana masalah-masalah pada data sudah diselesaikan, data sudah dapat digunakan hingga tahap terakhir. Tahap ini merupakan tahap yang sering ditinjau kembali saat menemukan masalah pada saat pembangunan model. Sehingga dilakukan iterasi sampai menemukan hal yang cocok dengan data.

### **D. MODELING**

Secara garis besar untuk membuat model prediktif atau deskriptif. Pada tahap ini dilakukan metode statistika dan Machine Learning untuk penentuan terhadap teknik data mining, alat bantu data mining, dan algoritma data mining yang akan diterapkan. Lalu selanjutnya adalah melakukan penerapan teknik dan algoritma data mining tersebut kepada data dengan bantuan alat bantu. Jika diperlukan penyesuaian data terhadap teknik data mining tertentu, dapat kembali ke tahap data preparation.

Dikarenakan pengetahuan penulis yang masih terbatas, penulis membatasi tahapan CRISP-DM yang diterapkan pada data. Penulis membatasi tahapan hanya sampai dengan tahap modeling.

## BAB III METODOLOGI PENELITIAN

### 3.1. ALAT DAN BAHAN PENELITIAN

#### 3.1.1. ANALISA KEBUTUHAN PERANGKAT

Adapun kebutuhan perangkat dari penelitian ini diantaranya sebagai berikut.

##### A. Perangkat Keras,

Perangkat keras yang digunakan pada penelitian ini adalah laptop/PC untuk pengolahan data dengan spek yang memadai. Selain itu, Flashdisk juga dibutuhkan sebagai media perpindahan data.

##### B. Perangkat Lunak,

Terdapat 2 perangkat lunak atau software yang dibutuhkan dalam penelitian ini. Diantara adalah RapidMiner dan Visual Studio Code.

Penggunaan Rapidminer disini adalah sebagai perangkat pengolah data serta pada rapidminer ini nanti kita dapat menemukan algoritma linear regression didalamnya.

Sedangkan Visual Studio Code atau yang kerap disingkat VSCode berguna sebagai perangkat lunak yang digunakan pada tahap modeling. Dengan menggunakan bahasa pemrograman Python.

#### 3.1.2. ANALISA KEBUTUHAN DATA

Berikut ini 10 data teratas yang akan digunakan sebagai data training.

id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition	grade	sqft_above	sqft_basement	yr_built	yr_renovated	zipcode	lat	long	sqft_living15	sqft_lot15
7129300520	24 Oktober 2014	221900	3	1	1180	5650	1	0	0	3	7	1180	0	1955	0	98178	47.5112	-122.257	1340	5650
6414100192	12 September 2014	538000	3	2.25	2570	7242	2	0	0	3	7	2170	400	1951	1991	98125	47.721	-122.319	1690	7639
5631500400	25 Februari 2015	180000	2	1	770	10000	1	0	0	3	6	770	0	1933	0	98028	47.7379	-122.233	2720	8062
2487200875	09 Desember 2014	604000	4	3	1960	5000	1	0	0	5	7	1050	910	1965	0	98136	47.5208	-122.393	1360	5000
1954400510	18 Februari 2015	510000	3	2	1680	8080	1	0	0	3	8	1680	0	1987	0	98074	47.6168	-122.045	1800	7503
7237550310	12 Mei 2015	1.23E+06	4	4.5	5420	101930	1	0	0	3	11	3890	1530	2001	0	98053	47.6561	-122.005	4760	101990
1321400060	27 Juni 2014	257500	3	2.25	1715	6819	2	0	0	3	7	1715	0	1995	0	98003	47.3097	-122.327	2238	6819
2008000270	15 Januari 2015	291850	3	1.5	1060	9711	1	0	0	3	7	1060	0	1963	0	98198	47.4095	-122.315	1650	9711
2414600126	15 April 2015	229500	3	1	1780	7470	1	0	0	3	7	1050	730	1960	0	98146	47.5123	-122.337	1780	8113
3793500160	12 Maret 2015	323000	3	2.5	1890	6560	2	0	0	3	7	1890	0	2003	0	98038	47.3684	-122.031	2390	7570

**Tabel 3.1 10 Data teratas Penjualan Rumah**

Sumber : [www.kingcountry.com](http://www.kingcountry.com)

Berdasarkan data diatas penulis menemukan 21 attributes dalam data teratas penjualan rumah, dengan attribute *prize* sebagai *label* dan attribute *date* sebagai *timeseries*. Sedangkan attribute lain adalah attribute yang membentuk attribute label. Pada data ini semuanya terisi dan dan tidak ada cell yang tak terisi.

### **3.2.WAKTU DAN TEMPAT PENELITIAN**

#### **3.2.1. WAKTU PENELITIAN**

Penelitian ini dimulai pada tanggal 30 Desember 2021 atau sepekan lalu untuk memenuhi tugas akhir semester 3 pada matakuliah Artificial Intelligence.

#### **3.2.2. TEMPAT PENELITIAN**

Penelitian ini dilakukan masing-masing secara daring dengan tetap bekerja sama antar anggota kelompok.

## BAB IV PEMBAHASAN

### 4.1. TAHAPAN PREDIKSI PENJUALAN HARGA RUMAH

Berikut ini tahapan dalam memprediksi harga Rumah dari label *price* menggunakan aplikasi RapidMiner.

#### 4.1.1. BUSINESS UNDERSTANDING

Adapun latar belakang pembuatan laporan ini adalah untuk mengendalikan dengan cara memprediksi pasar dari harga jual rumah menggunakan algoritma *linear regression*.

#### 4.1.2. DATA UNDERSTANDING

Data yang akan digunakan ini bersumber dari King Country yang dimana awalnya data ini attribute timeseries-nya bertipe data polinomial dan bukan bertipe data date. Seperti yang diketahui pada pendekatan Forecasting membutuhkan attribute timeseries dengan type data date yaitu yang menerangkan tentang waktu. Adapapun bentuk awal dari data sebagai berikut.

#	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	yr_built	waterfront	view	condition	grade	sqft_above	sqft_basement	yr_renovated	zipcode	lat	long	sqft_living15	sqft_lot15	
1123939523	201413177000000	221900	3	2	2180	3690	2	0	0	3	7	1080	0	2090	0	98176	47.5122	-122.237	1340	3690
843430182	201412097000000	558000	3	2.25	2573	7242	2	0	0	3	7	1370	490	1951	2091	98125	47.721	-122.519	1690	7638
9831506480	201402257000000	180000	2	1	770	10000	1	0	0	3	6	770	0	1928	0	98028	47.7376	-122.233	2726	8062
348720875	201412097000000	824000	4	3	3363	5000	1	0	0	5	7	3258	910	1985	0	98136	47.5288	-122.585	1380	5000
2954405532	201502187000000	510000	3	2	3880	9380	1	0	0	3	8	1680	0	1987	0	98074	47.8188	-122.045	1800	7563
7237918330	201405127000000	1.231-08	4	4.5	5430	203830	2	0	0	3	11	3890	1030	2001	0	98053	47.8381	-122.005	4760	203930
1111830060	201406377000000	257900	3	2.25	1711	6819	2	0	0	3	7	1711	0	1990	0	98001	47.3897	-122.117	2238	6819
3088025270	201302117000000	381800	3	1.5	2043	8711	1	0	0	3	7	1362	0	1983	0	98188	47.4684	-122.335	1480	9711
3619802128	201303197000000	229800	3	1	1742	7878	1	0	0	3	7	1362	0	1986	0	98188	47.5129	-122.337	1780	8117
4793936180	201303197000000	518800	4	2.5	3890	8360	1	0	0	3	7	3890	0	2001	0	98009	47.3884	-122.111	2090	7170
1788806230	201304077000000	862900	5	2.5	3980	8798	1	0	0	3	8	3890	1790	1989	0	98037	47.8037	-122.149	2210	8920
8212903280	201305377000000	488000	3	2	3180	8000	1	0	0	4	7	860	0	1982	0	98131	47.89	-122.282	1930	8000
134315138	201405387000000	517000	3	1	3439	19901	1.5	0	0	4	7	1439	0	1927	0	98028	47.7558	-122.229	1780	12097
8284850070	201413077000000	400000	3	1.75	3570	9980	1	0	0	4	7	1570	0	1977	0	98074	47.8127	-122.045	1370	10258
117500570	201305127000000	530000	5	2	3830	4850	1.5	0	0	3	7	1810	0	1990	0	98137	47.87	-122.394	1300	4050
8287830013	201301347000000	830000	4	3	2950	5500	2	0	3	3	8	1880	970	1976	0	98126	47.5714	-122.571	2540	4000
1875500960	201407177000000	395000	3	2	3880	14940	1	0	0	3	7	1880	0	1994	0	98029	47.7277	-121.983	1800	14938
6865301467	201405387000000	485000	4	1	3800	4300	1.5	0	0	4	7	1800	0	1918	0	98111	47.8648	-122.243	1810	4300
58002367	201412037000000	180000	2	1	1302	3650	1	0	0	4	7	1302	0	1921	0	98001	47.3388	-122.31	1680	3093
7982330602	201104370000000	230000	3	2	1200	6778	1	0	0	4	7	1200	0	1989	0	98001	47.3348	-122.336	1380	8003
4600938876	201305347000000	381600	4	1.75	3623	6983	1	0	0	4	7	860	0	1987	0	98111	47.7025	-122.381	1400	6983
2524649179	201408287000000	2.046-08	3	1.75	3050	68667	1	0	0	4	8	2850	720	1984	0	98040	47.8108	-122.231	6110	28584
7137970849	201307077000000	281600	3	2.5	3271	6109	2	0	0	3	8	1270	0	1991	0	98092	47.9186	-121.889	2310	7089
8891402200	201308197000000	252700	2	1.5	3970	8943	1	0	0	3	7	1870	0	1989	0	98001	47.8181	-122.186	1220	8186
3834796200	201411277000000	325000	3	2.25	3450	4500	1	0	0	4	8	2450	0	1989	0	98001	47.3789	-122.172	2070	4893
1285006200	201411077000000	233000	3	1	3723	4897	1.5	0	0	3	6	1713	0	1941	0	98001	47.3648	-122.238	1070	4795
1794936383	201406287000000	937000	5	1.75	3490	3991	2	0	0	3	8	1790	790	1915	0	98128	47.8386	-122.36	1780	3573

Gambar 4.1 Data awal Harga Pejualan Rumah King Country

Dikarenakan timeseries adalah attribute yang penting dalam pendekatan Forecasting. Jadi, penulis menulis manual timeseries tersebut dengan berlandaskan data polinomial tersebut. Berikut penampakan data setelah diperbaiki.

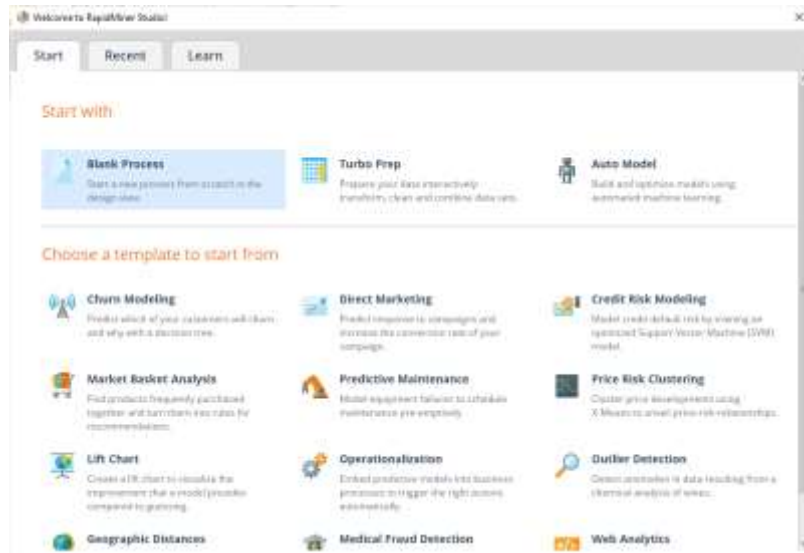
id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floor	waterfront	view	condition	grade	sqft_above	sqft_basement	yr_built	yr_renovated	zipcode	lat	long	sqft_living15	sqft_lot15
7129005020	24 Oktober 2014	221900	3	1	1180	5650	1	0	0	3	7	1180	0	1955	0	98178	47.5112	-122.257	1340	5650
6414100192	12 September 2014	538000	3	2.25	2570	7242	2	0	0	3	7	2170	400	1951	1991	98125	47.721	-122.319	1690	7639
5631500400	25 Februari 2015	180000	2	1	770	10000	1	0	0	3	6	770	0	1933	0	98028	47.7379	-122.233	1720	8862
2487100875	09 Desember 2014	604000	4	3	1860	5000	1	0	0	5	7	1050	910	1965	0	98136	47.5308	-122.393	1360	5000
1254400510	18 Februari 2015	510000	3	2	1680	8280	1	0	0	3	8	1680	0	1987	0	98074	47.6168	-122.045	1800	7503
7237503010	12 Mei 2015	1.23E+06	4	4.5	5420	101930	1	0	0	3	11	3890	1530	2001	0	98053	47.6561	-122.005	4760	101930
1321400060	27 Juni 2014	257500	3	2.25	1715	6819	2	0	0	3	7	1715	0	1995	0	98003	47.3097	-122.327	2238	6819
2008000270	15 Januari 2015	291850	3	1.5	1060	9711	1	0	0	3	7	1060	0	1963	0	98198	47.4095	-122.315	1650	9711
2414600126	15 April 2015	229500	3	1	1780	7470	1	0	0	3	7	1050	730	1960	0	98146	47.5123	-122.337	1780	8113
3795500160	12 Maret 2015	323000	3	2.5	1890	6560	2	0	0	3	7	1890	0	2003	0	98038	47.2894	-122.031	2390	7570
1788005020	03 April 2015	662500	3	2.5	3560	9796	1	0	0	3	8	1960	1700	1965	0	98007	47.6007	-122.145	2210	8925
9212900260	27 Mei 2014	468000	2	1	1160	6000	1	0	0	4	7	850	300	1942	0	98115	47.69	-122.252	1330	6000
1141015116	28 Mei 2014	310000	3	1	1430	19901	1.5	0	0	4	7	1430	0	1927	0	98028	47.7558	-122.229	1780	12697
6054650070	07 Oktober 2014	400000	3	1.75	1370	9680	1	0	0	4	7	1370	0	1977	0	98074	47.6127	-122.045	1370	10208
1175000570	12 Maret 2015	530000	5	2	1810	4850	1.5	0	0	3	7	1810	0	1900	0	98107	47.67	-122.394	1360	4850
1297300055	12 Januari 2015	650000	4	3	2950	5000	2	0	3	3	9	1980	970	1979	0	98126	47.5714	-122.575	2140	4000
1875500060	31 Juli 2015	395000	3	2	1890	14040	2	0	0	3	7	1890	0	1994	0	98019	47.7277	-121.962	1890	14018
6865200140	29 Mei 2015	485000	4	1	1600	4300	1.5	0	0	4	7	1600	0	1916	0	98103	47.6648	-122.343	1610	4300
16000397	05 Desember 2014	189000	2	1	1200	9850	1	0	0	4	7	1200	0	1921	0	98002	47.3089	-122.21	1060	5095
7983100060	24 April 2015	230000	3	1	1250	9774	1	0	0	4	7	1250	0	1969	0	98003	47.3343	-122.306	1280	8850
6300500875	14 Mei 2014	385000	4	1.75	1630	4980	1	0	0	4	7	1630	760	1947	0	98133	47.7025	-122.341	1400	4980
2524049179	26 Agustus 2014	2.00E+06	3	2.75	3050	44867	1	0	4	3	9	2330	720	1968	0	98040	47.5316	-122.233	4110	20336
7137970340	03 Juli 2014	285000	5	2.5	2270	6300	2	0	0	3	8	2270	0	1995	0	98092	47.3266	-122.169	2240	7005
8091400200	16 Mei 2014	252700	2	1.5	1070	9613	1	0	0	3	7	1070	0	1985	0	98030	47.3533	-122.166	1220	8186
3814700000	20 November 2014	329000	3	2.25	2450	6500	2	0	0	4	8	2450	0	1985	0	98030	47.3739	-122.172	2200	6865
1200200000	03 November 2014	233000	3	2	1710	4697	1.5	0	0	5	6	1710	0	1941	0	98002	47.3048	-122.218	1030	4705
1794500383	26 Juni 2014	937000	3	1.75	2450	2691	2	0	0	3	8	1750	700	1915	0	98119	47.6386	-122.36	1760	3573
3303700376	01 Desember 2014	667000	3	1	1400	1581	1.5	0	0	5	8	1400	0	1909	0	98112	47.6221	-122.314	1860	3861
5101402488	24 Juni 2014	438000	3	1.75	1520	6380	1	0	0	3	7	790	730	1948	0	98115	47.695	-122.304	1520	6235
7137100393	02 Maret 2015	719000	4	2.5	2570	7173	2	0	0	3	8	2570	0	2005	0	98052	47.7073	-122.11	2630	6026
8562750320	10 November 2014	580500	3	2.5	2320	3980	2	0	0	3	8	2320	0	2003	0	98027	47.5391	-122.07	2580	3980
2426039314	01 Desember 2014	280000	2	1.5	1130	1265	3	0	0	3	7	1190	0	2005	0	98133	47.7274	-122.357	1390	1756
461000390	24 Juni 2014	687500	4	1.75	2330	5000	1.5	0	0	4	7	1510	820	1929	0	98117	47.6823	-122.368	1460	5000
7589100193	10 November 2014	535000	3	1	1090	3000	1.5	0	0	4	8	1090	0	1929	0	98117	47.6889	-122.375	1570	5080
7955080270	03 Desember 2014	322500	4	2.75	2060	6659	1	0	0	3	7	1280	780	1981	0	98058	47.4276	-122.157	2020	8720
8547205180	13 Juni 2014	696000	3	2.5	2300	3080	1.5	0	0	3	8	1510	790	1930	2002	98115	47.6827	-122.31	1590	2624
9453500030	28 Mei 2014	550000	4	1	1660	3488	2	0	0	1	5	930	730	1933	0	98052	47.6621	-122.132	2160	11467
2768000400	30 Desember 2014	540000	4	2	2360	6000	2	0	0	4	8	2360	0	1904	0	98107	47.6702	-122.362	1730	4700
7895500070	13 Februari 2015	240000	4	1	1220	8075	1	0	0	2	7	890	330	1969	0	98001	47.3341	-122.282	1290	7800
2078500320	20 Juni 2014	605000	4	2.5	2620	7553	2	0	0	3	8	2620	0	1996	0	98056	47.5301	-122.18	2620	11884
5547700270	15 Juli 2014	625000	4	2.5	2570	5520	2	0	0	3	9	2570	0	2000	0	98074	47.6145	-122.027	2470	5669
7766200013	11 Agustus 2014	775000	4	2.25	4220	24186	1	0	0	3	8	2600	1620	1984	0	98166	47.445	-122.547	2410	30617
7203220400	07 Juli 2014	861999	5	2.75	3595	5619	2	0	0	3	9	3595	0	2014	0	98053	47.6848	-122.016	3625	5619
9270020160	28 Oktober 2014	885000	3	1	1570	2280	2	0	0	3	7	1570	0	1922	0	98119	47.6413	-122.364	1380	2640
1432701230	29 Juli 2014	309000	3	1	1280	9656	1	0	0	4	6	920	360	1959	0	98058	47.4485	-122.175	1340	8808
8035350320	18 Juli 2014	488000	3	2.5	3160	13603	2	0	0	3	8	3160	0	2003	0	98019	47.7443	-121.977	3050	9232
8945200830	25 Maret 2015	210490	3	1	990	8528	1	0	0	3	6	990	0	1966	0	98023	47.3066	-122.171	1228	8840
4178300310	16 Juli 2014	785000	4	2.5	2290	13416	2	0	0	4	9	2290	0	1981	0	98007	47.6194	-122.351	2680	13685
9215400105	28 April 2015	450000	3	1.75	1250	5963	1	0	0	4	7	1250	0	1953	0	98115	47.6796	-122.301	970	1001
822029084	11 Maret 2015	1.35E+06	3	2.5	2753	65005	1	1	2	13	9	2165	580	1953	0	98070	47.4041	-122.451	2680	72513
5245600105	16 September 2014	228000	3	1	1190	9199	1	0	0	3	7	1190	0	1955	0	98148	47.4258	-122.322	1190	9364
7231300125	17 Februari 2015	345000	5	2.5	3150	9134	1	0	0	4	8	1640	1510	1966	0	98056	47.4934	-122.189	1990	9133
7518505990	31 Desember 2014	600000	3	1.75	1410	4080	1	0	0	4	7	1000	410	1950	0	98117	47.6808	-122.384	1410	4080
3626039271	05 Februari 2015	585000	2	1.75	1980	8550	1	0	0	4	7	990	990	1981	0	98117	47.6989	-122.369	1480	8738
4217401195	03 Maret 2015	920000	5	2.25	2730	6000	1.5	0	0	3	8	2130	600	1927	0	98105	47.6571	-122.281	2730	6000
9822700295	12 Mei 2014	885000	4	2.5	2830	5000	2	0	0	3	9	2830	0	1995	0	98105	47.6597	-122.29	1950	5000
9478500640	19 Agustus 2014	292500	4	2.5	2250	4485	2	0	0	3	7	2250	0	2008	0	98042	47.3663	-122.114	2250	4500
2799800710	07 April 2015	301000	3	2.5	2420	4750	2	0	0	3	8	2420	0	2003	0	98042	47.3663	-122.122	2690	4750
7922800400	27 Agustus 2014	951000	5	3.25	3250	14342	2	0	4	4	8	3250	0	1968	0	98008	47.588	-122.116	2960	11044
8079040320	23 Februari 2015	430000	4	3	1850	9976	2	0	0	3	8	1850	0	1991	0	98059	47.5059	-122.149	2270	8542
1516000055	10 Desember 2014	650000	3	2.25	2150	12235	1	0	3	4	8	1590	560	1950	0	98166	47.4336	-122.339	2570	18900
9558000045	28 Agustus 2014	289900	3	1.75	1260	8400	1	0	0	3	7	1260	0	1954	0	98148	47.4366	-122.335	1290	8750
5072410070	21 Oktober 2014	605000	3	1.75	2510	8690	2	0	0	5	8	2510	0	1973	0	98166	47.4428	-122.344	2500	9500
9528102996	07 Desember 2014	549000	3	1.75	1540	1044	3	0	0	3	8	1540	0	2014	0	98115	47.6765	-122.32	1580	3090
1189001180	03 Juni 2014	425000	3	2.25	1660	6000	1	0	0	3	7	1110								



### 4.1.3. DATA PREPARATION

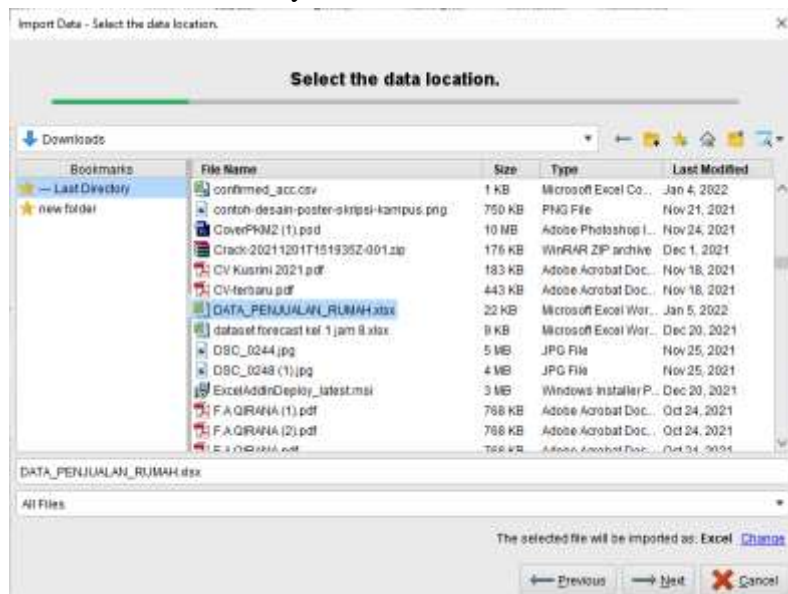
Pada tahap ini, data yang tadinya sudah ditambah attribute timeseries akan dimasukkan kedalam aplikasi RapidMiner. Berikut langkah-langkah memasukkan data ke RapidMiner

1. Buka Aplikasi RapidMiner pada PC atau Laptop.
2. Pilih Blank Process untuk memulai.



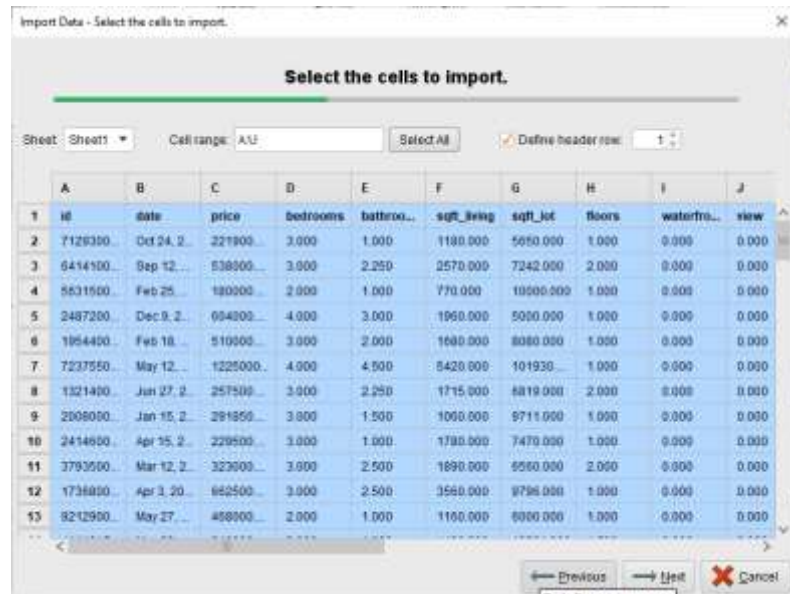
*Gambar 4.3 Memilih Menu Blank Process*

3. Setelah itu pada menu import data, pilih data yang akan dimasukkan, Contohnya DATA\_PENJUALAN\_RUMAH



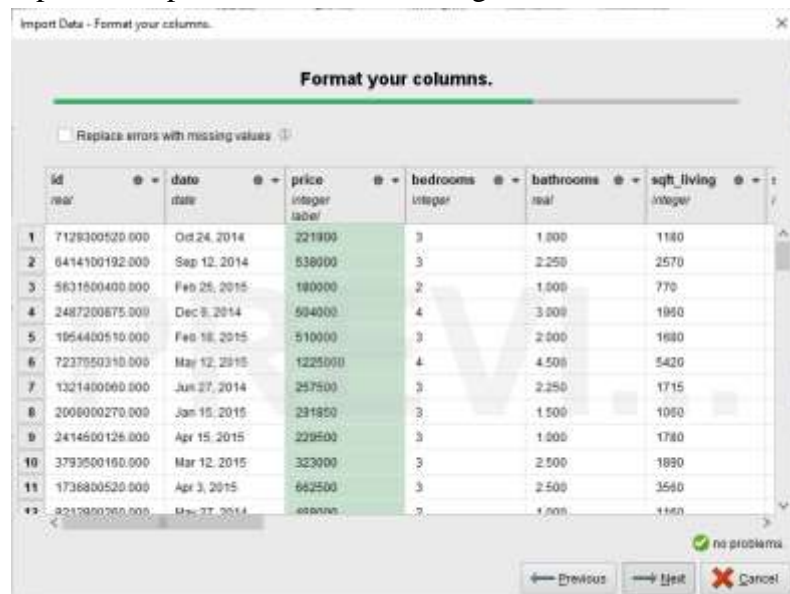
*Gambar 4.4 Pilih Data yang akan Dikelola*

- Setelah memilih data akan muncul tampilan seperti berikut. Pastikan semua cell data telah di select atau dipilih. Lalu pilih Next



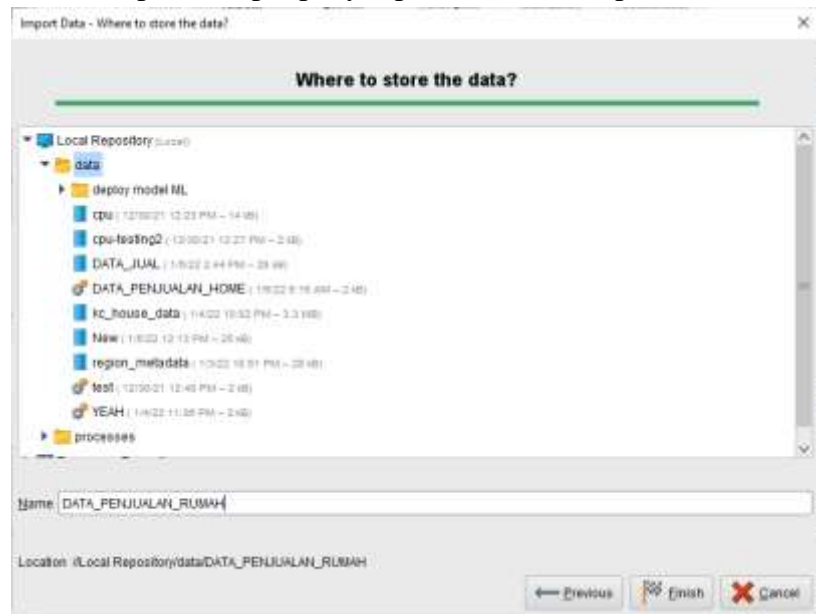
**Gambar 4.5 Pilih Semua Cell Data**

- Pilih type data dari setiap attribute dan sesuaikan seperti memilih label. Pada data ini labelnya adalah *price*. Dan jangan lupa menetapkan attribute *date* sebagai timeseries.



**Gambar 4.6 Penetapan Label pada attribute Price**

6. Terakhir, pilih tempat penyimpanan data. Lalu pilih Finish.



**Gambar 4.7 Pilih Lokasi Penyimpanan data**

#### 4.1.4. MODELLING

Pada tahap modeling ini, kita akan mengatur model dari data nantinya. Ditahap ini juga kita akan memasukkan algoritma linear regression yang dimana menjadi titik pusat dari laporan ini. Berikut langkah-langkah modeling pada RapidMiner:

1. Langkah pertama, siapkan data seperti yang kita lakukan sebelumnya. Berikut ini penampakan data saat di RapidMiner.



**Gambar 4.8 Penampakan data di RapidMiner**

- Masuk ke Menu Design, sekarang saatnya perancangan design. Sekarang drag DATA\_PENJUALAN\_RUMAH dari Repository ke Process.



**Gambar 4.9 Drag Data dari Repository ke Process**

- Masukkan juga Algoritma Linear Regression, dengan cara ke menu Operator dan search “linear regression” lalu drag juga Algoritma.



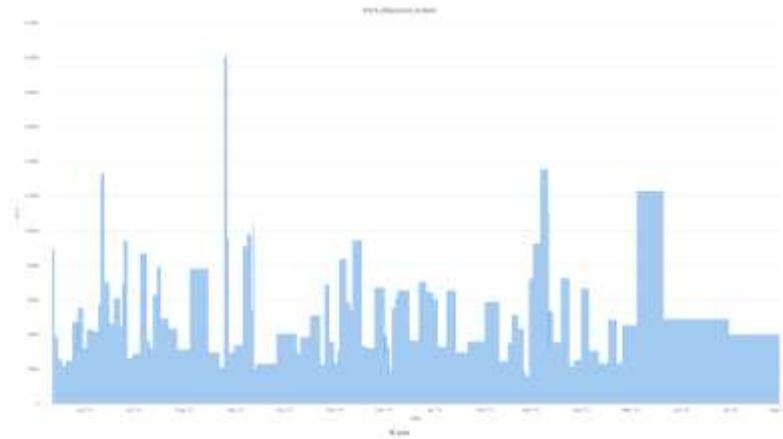
**Gambar 4.10 Drag Algoritma Linear Regression ke Process**

- Sambungkan Output Data\_penjualan\_rumah pada Training Algoritma Linear Regression. Sedangkan modeling pada result. Dan tekan Run.



**Gambar 4.11 Sambungkan Data, Algoritma, dan Result**

Berikut ini Statistic dari Result yang kita kelola:



***Gambar 4.12 Statistic dari Result***

Dari gambar 4.12 dapat dinyatakan bahwa pada kisaran bulan agustus sampai september 2014 harga rumah meningkat sampai menyentuh angka 2,000k. Sedangkan pada akhir bulan februari terjadi penurunan hingga menyentuh angka kisaran 170k.

## **BAB V PENUTUP**

### **5.1.KESIMPULAN**

Berdasarkan hasil dari penelitian ini, penulis dapat menarik kesimpulan bahwa dengan adanya Machine Learning ini akan mempermudah manusia dalam memprediksi harga suatu barang hanya dengan mempelajari dari data-data sebelumnya.

Forecasting adalah teknik untuk memprediksi nilai yang akan ada dimasa selanjutnya(masa depan). Tentunya akan sangat menguntungkan apabila bias menguasai teknik forecasting ini.

### **5.2.SARAN**

Adapun saran dari penulis adalah dalam penelitian ini masih memiliki banyak kekurangan yaitu terlalu terpaku pada algoritma linear regression, dan tidak membandingkannya dengan algoritma yang lain.

## DAFTAR PUSTAKA

Martulandi A., “*Memprediksi Harga Rumah dengan Machine Learning: Multiple Linear Regression*”(2019)

Indrawati T., “*PENGGUNAAN METODE LINEAR REGRESSION UNTUK PREDIKSI PENJUALAN SMARTPHONE*” (2018)