

ANALISIS PERBANDINGAN METODE *ARTIFICIAL NEURAL NETWORK*, REGRESI LINEAR BERGANDA DAN REGRESI POLINOMIAL UNTUK MEMPREDIKSI HARGA JUAL RUMAH

SKRIPSI



ST. HESTIANA KADIR

H13116504

PROGRAM STUDI ILMU KOMPUTER DEPARTEMEN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

OKTOBER 2020



ANALISIS PERBANDINGAN METODE *ARTIFICIAL NEURAL NETWORK*, REGRESI LINEAR BERGANDA DAN REGRESI POLINOMIAL UNTUK MEMPREDIKSI HARGA JUAL RUMAH

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Ilmu Komputer Departemen Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

ST. HESTIANA KADIR

H13116504

PROGRAM STUDI ILMU KOMPUTER DEPARTEMEN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

OKTOBER 2020



LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

ANALISIS PERBANDINGAN METODE *ARTIFICIAL NEURAL NETWORK*, REGRESI LINEAR BERGANDA DAN REGRESI POLINOMIAL UNTUK MEMPREDIKSI HARGA JUAL RUMAH

adalah benar hasil karya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 6 Oktober 2020




ST. HESTIANA KADIR

NIM. H13116504



ANALISIS PERBANDINGAN METODE *ARTIFICIAL NEURAL NETWORK*, REGRESI LINEAR BERGANDA DAN REGRESI POLINOMIAL UNTUK MEMPREDIKSI HARGA JUAL RUMAH

Disetujui oleh:



Pembimbing Utama

Pembimbing Pertama

Dr. Hendra, S.Si., M.Si.
NIP. 19760102 200212 1 001

Supri Bin Hj Amir, S.Si., M.Eng.
NIP. 19880504 201903 1 012

Pada 6 Oktober 2020



HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : St. Hestiana Kadir
NIM : H13116504
Program Studi : Ilmu Komputer
Judul Skripsi : Analisis Perbandingan Metode *Artificial Neural Network*,
Regresi Linear Berganda dan Regresi Polinomial Untuk
Memprediksi Harga Jual Rumah

Telah berhasil mempertahankan di hadapan dewan penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

Tanda Tangan

1. Ketua : Dr. Henda, S.Si., M.Si. (.....)
2. Sekretaris : Supri Bin Hj. Amir, S.Si., M.Eng. (.....)
3. Anggota : Dr. Muhammad Hasbi, M.Sc. (.....)
4. Anggota : Edy Saputra, S.Si., M.Si. (.....)

Ditetapkan di : Makassar

Tanggal : 6 Oktober 2020



KATA PENGANTAR

Segala puji bagi Allah *Subhanahu Wa ta'ala*, Tuhan alam semesta yang telah memberikan nikmat kesempatan, kesehatan dan kemampuan sehingga penulisan skripsi ini bisa selesai. Shalawat serta salam senantiasa tercurah kepada *Rasulullah* Muhammad *Shallallahu Alaihi Wasallam*, yang merupakan teladan dalam menjalankan kehidupan di dunia.

Alhamdulillah, skripsi dengan Judul “Analisis Perbandingan Metode *Artificial Neural Network*, Regresi Linear Berganda dan Regresi Polinomial Untuk Memprediksi Harga Jual Rumah” yang disusun sebagai salah satu syarat akademik untuk meraih gelar Sarjana Sains pada Program Studi Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin ini dapat diselesaikan. Tentunya, dalam penulisan skripsi ini, penulis mampu menyelesaikan tepat waktu berkat bantuan dan dukungan dari berbagai pihak. Oleh karena itu, ucapan terima kasih dan apresiasi yang tak terhingga kepada kedua orang tua penulis, Ayahanda **H. Kadir** dan Ibunda **Hj. Upe** yang tak kenal lelah dalam memanjatkan doa serta memberikan nasihat dan motivasi kepada penulis. Tugas akhir ini hanya setitik kebahagiaan kecil yang bisa penulis persembahkan. Tidak lupa pula terima kasih kepada saudara kandung saya, **Siti Hartina Kadir**, **Sri Surahmi**, dan **Nurul Mutia H. Kadir**, kakak ipar saya **H. Danil**, dan keponakan saya **Hijriyah Danil** dan **Khadijah Danil** yang senantiasa memberikan dukungan dan semangat.

Terima kasih juga penulis ucapkan kepada:

1. Rektor Universitas Hasanuddin, Ibu **Prof. Dr. Dwia Aries Tina Pulubuhu** beserta jajarannya.
2. Dekan Fakultas Matematika dan Ilmu Pengetahuna Alam (FMIPA), **Dr.Eng. Amiruddin** beserta jajarannya.
3. Bapak **Dr. Nurdin, S.Si., M.Si.**, sebagai Ketua Departemen Matematika FMIPA Unhas. Penulis juga berterima kasih atas dedikasi dosen-dosen pengajar, serta Departemen atas ilmu dan bantuan yang bermanfaat.



4. Bapak **Dr. Henda, S.Si., M.Si.**, sebagai dosen pembimbing utama sekaligus ketua tim penguji atas semua ilmu yang telah diberikan selama proses perkuliahan dan senantiasa memotivasi penulis dalam penulisan skripsi ini.
5. Bapak **Supri Bin Hj. Amir, S.Si., M.Eng.**, sebagai dosen pembimbing pertama sekaligus sekteraris tim penguji atas ilmu yang diberikan selama proses perkuliahan dan bimbingan, serta segala bentuk bantuan yang telah diberikan dalam penyusunan skripsi ini.
6. Bapak Alm **Dr. Diaraya, M.Ak.**, sebagai anggota tim penguji atas segala ilmu yang telah beliau berikan selama proses perkuliahan serta berbagai masukan dan kritik yang membangun dalam proses penyusunan skripsi ini.
7. Bapak **Dr. Muhammad Hasbi, M.Sc.**, sebagai anggota tim penguji atas segala ilmu yang telah diberikan selama proses perkuliahan serta berbagai masukan dan kritik yang membangun dalam proses penyusunan skripsi ini.
8. Bapak **Edy Saputra, S.Si., M.Si.**, sebagai anggota tim penguji atas segala ilmu yang telah diberikan selama proses perkuliahan serta berbagai masukan dan kritik yang membangun dalam proses penyusunan skripsi ini.
9. Saudaraku **Zinedine Kahlil Gibran Zidane, S.Si.**, salah satu orang yang sangat berperan penting dalam proses pengerjaan skripsi penulis. Dengan ikhlas dan kerendahan hatinya membantu serta mengajarkan banyak hal dan selalu setia mendengarkan keluh kesah penulis selama mengerjakan skripsi ini sampai selesai.
10. Saudara-saudara ku **Rio Mukhtarom, S.Si., Widya Nauli Amalia Puteri, S.Si., Tasnia Akil, Nurmayulina, Suci Rahmadana Anwar, Berlian Adriani Putri, Ainun Mardiyah Istiqamah, S.Si, Nisrina Syadza Dewanty, Rizka Syahfitri, Marselia Ghanyyu Wahdini, Nirwana Sari Hamka, dan ILMU KOMPUTER 2016**, atas kebersamaan, kepedulian, suka-duka, canda tawa yang telah kita lewati selama ini. Semoga persahabatan kita yang telah terjalin tidak pernah usai.
11. Saudara-Saudara ku **PDAM. Andi Dyan Rezki Devi, Fathatul Amma**



ir, Nanda Mutiara Zani, S.T., Nurul Rizqah Al-Fathzi, S.Tr.Gz., Vita , Faradilla Ramadani, Nur Zakiran Ridwan, S.Si., Nur Alfriani P. da, S.Tr.Kes., atas kebersamaan, Kepedulian, suka-duka, canda tawa

yang telah kita lewati selama ini. Semoga persahabatan kita yang telah terjalin tidak pernah usai.

12. Keluarga besar **KKN Tematik Luwu Utara** yang secara ikhlas dan tulus mengabdikan kepada masyarakat.
13. Keluarga besar **A16ORITMA 2016** atas segala bentuk dukungan dan bantuan selama proses perkuliahan. Semoga kesuksesan selalu kita dapatkan dalam setiap langkah-langkah kita.
14. Seluruh pihak yang tidak dapat disebutkan satu per satu atas segala bentuk kontribusi, partisipasi, serta motivasi yang diberikan kepada penulis selama ini. Semoga apa yang kita berikan, dilipatgandakan oleh Allah Subhana Wata'ala.

Penulis menyadari bahwa masih banyak kekurangan dalam tugas akhir ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga karya ilmiah ini memberikan manfaat untuk pembaca.

Makassar, 6 Oktober 2020

St. Hestiana Kadir



PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : St. Hestiana Kadir
NIM : H13116504
Programa Studi : Ilmu Komputer
Departemen : Matematika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Prediktor Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas tugas akhir saya yang berjudul:

“Analisis Perbandingan Metode *Artificial Neural Network*, Regresi Linear Berganda dan Regresi Polinomial Untuk Memprediksi Harga Jual Rumah”

berserta perangkat yang ada (jika diperlukan). Terkait dengan hal diatas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian surat pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada 6 Oktober 2020

Yang menyatakan



(St. Hestiana Kadir)



ABSTRAK

Tempat tinggal merupakan kebutuhan yang sangat mendasar bagi kehidupan manusia. Selain untuk tempat bernaung, beberapa diantara mereka menggunakan tempat tinggal sebagai investasi pada masa yang akan datang. Dalam menentukan harga jual unit rumah pihak pengembang memiliki peranan yang penting. Sebelum menetapkan harga jual unit rumah, pihak pengembang perlu menghitung dengan cermat agar semua unit rumah laku terjual. Maka dari itu diperlukan beberapa cara untuk menentukan harga jual rumah, yakni dengan mempelajari dataset penjualan rumah “*King Country, Washington, AS*” yang ada di *Kaggle* dari bulan Mei 2014 sampai Mei 2015 dengan Algoritma *Artificial Neural Network* dengan metode *backpropagation*, Regresi Linear Berganda dengan metode *Ordinary Least Squared* (OLS) dan Regresi Polinomial dengan metode *Ordinary Least Squared* (OLS). Dari ketiga pemodelan tersebut didapatkan hasil nilai evaluasi untuk *Artificial Neural Network* dengan *R-Squared* 79.3% dengan *Root Mean Squared Error* 106424.29, Regresi Linear Berganda dengan *R-Squared* 64.5% dengan *Root Mean Squared Error* 139560.83 dan Regresi Polinomial dengan *R-Squared* 70.3% dengan *Root Mean Squared Error* 127741.25. Dari ketiga pemodelan tersebut maka model *Artificial Neural Network* merupakan pemodelan terbaik karena memiliki nilai *R-Squared* yang lebih baik dan nilai *Root Mean Squared Error* yang lebih rendah.



ABSTRACT

A residence is a fundamental need in our life. Some of us use these residences, in this case houses, as an investment in the future. In determining the selling price of the house, the landlord has an important role. Before determining the selling price of the house, the landlord should calculate carefully so that all housing units are sold. Therefore, several ways are needed to determine the selling price of a house, namely by studying the house sale dataset "King Country, Washington, USA" in Kaggle from May 2014 to May 2015 with Artificial Neural Network modeling with the backpropagation method, Multiple Linear Regression with Ordinary Least Squared (OLS) and Polynomial Regression with Ordinary Least Squared (OLS). From the three models, the evaluation results for Artificial Neural Network with R-Squared 79.3% with Root Mean Squared Error 106424.29, Multiple Linear Regression with R-Squared 64.5% with Root Mean Squared Error 139560.83 and Polynomial Regression with R-Squared 70.3% with Root Mean Squared Error 127741.25. Of the three models, the Artificial Neural Network model is the best model because it has the highest R-Squared value and the lowest Root Mean Squared Error value.



DAFTAR ISI

LEMBAR PERNYATAAN KEOTENTIKAN.....	iii
PERSETUJUAN PEMBIMBING.....	iv
HALAMAN PENGESAHAN.....	v
KATA PENGANTAR	vi
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	ix
ABSTRAK	x
ABSTRACT.....	xi
DAFTAR ISI.....	xii
DAFTAR GAMBAR	xv
DAFTAR TABEL.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Harga Jual Rumah	5
2.2 <i>Artificial Neural Network</i>	6
2.3 Fungsi aktivasi.....	7
2.4 <i>Multilayer Perceptron</i>	8
<i>Feed Forward</i>	9
<i>Backpropagation</i>	11
Regresi Linear Berganda	13



2.6	Regresi Polinomial	15
2.7	Koefisien Determinasi	16
2.8	<i>Root Mean Squared Error (RMSE)</i>	16
2.9	<i>K-Fold Cross Validation</i>	17
BAB III METODOLOGI PENELITIAN.....		19
3.1	Alur Penelitian.....	19
3.2	Waktu dan tempat.....	20
3.3	Tahapan Penelitian	20
3.3.1	Tahapan Pra Penelitian (persiapan).....	20
3.3.2	Penelitian.....	20
3.4	Deskripsi Data	21
BAB IV HASIL DAN PEMBAHASAN		22
4.1	Deskripsi Data	22
4.2	<i>Preprocessing</i>	24
4.2.1	Distribusi Data	25
4.2.2	Presentasi Data	25
4.2.3	Menentukan Umur Rumah.....	26
4.2.4	Uji F dan Uji T	27
4.2.5	Eliminasi Variabel.....	28
4.2.6	<i>Min-Max Scaler</i>	30
4.3	<i>Training Model</i>	30
4.3.1	Proses <i>Training Artificial Neural Network</i>	31
4.3.2	Proses <i>Training Regresi Linear Berganda</i>	32
4.3.3	Proses <i>Training Regresi Polinomial</i>	33
	Evaluasi Model.....	36
	Memprediksi Harga Jual Rumah.....	38



BAB V KESIMPULAN DAN SARAN.....	40
5.1 Kesimpulan.....	40
5.2 Saran.....	41
Lampiran	44



DAFTAR GAMBAR

Gambar 2.1 <i>Artificial Neural Network</i>	7
Gambar 2.2 Grafik fungsi ReLU.....	8
Gambar 2.3 Arsitektur <i>Multilayer Perceptron</i>	9
Gambar 2.4 Arsitektur <i>Backpropagation</i>	11
Gambar 2.5 Ilustrasi <i>4-Fold Cross Validation</i>	17
Gambar 3.1 Diagram Alur Penelitian.....	19
Gambar 4.1 Distribusi Data	25
Gambar 4.2 Umur Rumah dan Tahun Renovasi Rumah.....	27
Gambar 4.3 Fitur Ranking.....	29
Gambar 4.4 Proses <i>Training R-Squared Artificial Neural Network</i>	31
Gambar 4.5 Proses <i>Training Loss Artificial Neural Network</i>	32
Gambar 4.6 Grafik Data <i>Training R-Squared</i>	35
Gambar 4.7 Grafik Data <i>Training RMSE</i>	35
Gambar 4.8 Grafik Data <i>Testing R-Squared</i>	37
Gambar 4.9 Grafik Data <i>Testing RMSE</i>	38



DAFTAR TABEL

Tabel 4.1 Variabel Data	22
Tabel 4.2 Tampilan 5 Data Pertama.....	23
Tabel 4.3 Presentasi Data.....	26
Tabel 4.4 Uji T atau Uji Parsial	28
Tabel 4.5 <i>Min-Max Scaler</i>	30
Tabel 4.6 Hasil <i>Training Artificial Neural Network</i>	32
Tabel 4.7 Hasil <i>Training Regresi Linear Berganda</i>	33
Tabel 4.8 Hasil <i>Training Regresi Polinomial</i>	33
Tabel 4.9 Hasil Akhir Perhitungan Data <i>Training R-Squared</i>	34
Tabel 4.10 Hasil Akhir Perhitungan Data <i>Training Root Mean Squared Error</i> ...	35
Tabel 4.11 Nilai <i>R-Squared</i> pada hasil <i>testing</i>	36
Tabel 4.12 Nilai <i>Root Mean Square Error</i> pada hasil <i>Testing</i>	36
Tabel 4.13 Hasil Akhir Data <i>Testing R-Squared</i>	37
Tabel 4.14 Hasil Akhir Data <i>Testing Root Mean Squared Error</i>	37
Tabel 4.15 Dataset Baru.....	38
Tabel 4.16 Hasil Prediksi Harga Jual Rumah	39



BAB I

PENDAHULUAN

1.1 Latar Belakang

Tempat tinggal merupakan kebutuhan yang sangat mendasar bagi kehidupan manusia. Tanpa tempat tinggal manusia tidak dapat hidup dengan layak, berkreasi, bekerja, berkumpul, membesarkan anak-anak, beristirahat, dan berlindung dari cuaca (panas, hujan, angin). Tempat tinggal bukan hanya dijadikan sebagai tempat untuk bernaung, tetapi banyak dikalangan masyarakat menengah keatas menjadikan tempat tinggal ataupun perumahan sebagai investasi yang cukup dengan pengembalian berupa penghasilan.

Membeli rumah bukanlah urusan yang mudah dan simpel. Konsumen akan berupaya melakukan pencarian informasi tambahan baik secara internal maupun eksternal hingga pada akhirnya sebuah keputusan pembelian dibuat. Konsumen akan mempertimbangkan tentang harga rumah maupun terkait cara pembayarannya baik secara tunai atau kredit dan besarnya diskon atau hadiah yang ditawarkan pengembang, sebab konsumen akan selalu membandingkan harga rumah – rumah yang ditawarkan menurut kebutuhan dan kemampuannya (Primananda, 2010). Penetapan harga oleh pengembang akan berpengaruh terhadap perilaku pembelian konsumen, harga yang dapat dijangkau oleh konsumen akan cenderung membuat konsumen melakukan pembelian terhadap produk tersebut (Tedjakusuma & Muryani, 2001)

(Assauri & Sofyan, 2007) menyatakan bahwa, strategi pemasaran adalah serangkaian tujuan dan sasaran, kebijakan dan aturan yang memberi arah kepada usaha pemasaran perusahaan dari waktu ke waktu pada masing-masing tingkatan dan acuan serta alokasinya, terutama sebagai tanggapan perusahaan dalam menghadapi lingkungan dan keadaan persaingan yang selalu berubah. Dalam rangka inilah maka setiap perusahaan selalu menetapkan dan menerapkan marketing mix strategy yang terdiri dari 4P yaitu *Product, Price, Place, Promotion* dalam menjalankan setiap kegiatan usahanya. Strategi ini berkaitan dengan bagaimana perusahaan menyajikan penawaran produk disertai strategi yang lain berupa strategi harga, serta strategi penetapan lokasi, pada segmen tertentu yang merupakan sasaran pasarnya. *Marketing mix strategy* yang



tepat, dapat memberikan peranan yang penting terhadap keberhasilan suatu perusahaan untuk dapat tetap melangsungkan usahanya. Dengan penerapan strategi pemesanan yang tepat ini diharapkan dapat membuat konsumen untuk membuat suatu keputusan pembelian. Selain itu, strategi penetapan harga juga sangat penting untuk dilakukan.

Dalam penentuan harga jual unit rumah pihak pengembang memiliki peranan yang penting. Sebelum menetapkan harga jual unit rumah, pihak pengembang perlu menghitung dengan cermat agar semua unit rumah laku terjual. Selain itu, harga jual unit rumah yang ditetapkan dapat diterima oleh pengembang melebihi biaya-biaya yang dikeluarkan dan harga tersebut juga dapat diterima oleh konsumen. Oleh sebab itu, harga jual adalah faktor utama yang wajib diperhatikan pihak pengembang karena harga jual rumah merupakan acuan pertimbangan pemilihan rumah oleh pembeli, disamping berbagai fasilitas yang ditawarkan.

Dalam penelitian sebelumnya berdasarkan dataset yang terkait. *Department of Applied Economics and Statistics, University of Delaware, Newark, DE 19716, U.S.A* telah meneliti data tersebut dengan menggunakan perbandingan algoritma sebanyak 7. Yaitu *Linear Regression* dengan *R-squared* 86,7%, *Random Forest* dengan *R-squared* 88,8%, *Gradient Boosting* dengan *R-squared* 90,7%, *SVM Polynomial* dengan *R-squared* 89,2%, *SVM Radial* dengan *R-squared* 89,3%, *Neural Network* dengan *R-squared* 88,4%, dan *Ensemble Model* dengan *R-squared* 90,9% (Yunjiao Cai, 2007).

Dalam penelitian ini, prediksi untuk mempelajari harga jual rumah menggunakan data set yang ada di Kaggle dengan nama “*King Country, Washington, AS*”. Untuk mempelajari harga jual rumah tersebut penulis melakukan perbandingan algoritma sebanyak 3, yakni: *Artificial Neural Network* (ANN) dengan metode *Backpropagation*, Regresi Linear Berganda dengan metode OLS (*Ordinary Least Square*) dan Regresi Polinomial dengan metode OLS (*Ordinary Least Square*). Perbandingan tersebut nantinya akan dilihat bahwa model yang lebih bagus dapat dilihat dari nilai *R-squared* (R^2) dan nilai *Root Mean Squared Error*

dari ketiga metode tersebut.

rumusan Masalah

apapun rumusan masalah dalam penelitian ini yaitu:



1. Bagaimana model *Artificial Neural Network*, Regresi Linear Berganda dan Regresi Polinomial untuk memprediksi harga jual rumah?
2. Bagaimana perbandingan nilai *R-squared* (R^2) yang diperoleh antara *Artificial Neural Network*, Regresi Linear Berganda dan Regresi Polinomial untuk memprediksi harga jual rumah?
3. Bagaimana perbandingan nilai *Root Mean Squared Error* (RMSE) yang diperoleh antara *Artificial Neural Network*, Regresi Linear Berganda dan Regresi Polinomial untuk memprediksi harga jual rumah?

1.3 Batasan Masalah

Berikut ini merupakan beberapa batasan dalam penelitian ini.

1. Data set harga jual rumah yang digunakan adalah *King Country, Washington, AS*.
2. Metode *Artificial Neural Network* yang digunakan adalah *Backpropagation*.
3. Metode Regresi linear berganda yang digunakan adalah OLS (*Ordinary Least Square*).
4. Metode Regresi polinomial yang digunakan adalah OLS (*Ordinary Least Square*).

1.4 Tujuan Penelitian

Tujuan dari penelitian ini yaitu:

1. Menganalisis model *Artificial Neural Network*, Regresi Linear Berganda dan Regresi Polinomial untuk memprediksi harga jual rumah.
2. Menganalisis perbandingan nilai *R-squared* (R^2) yang diperoleh antara *Artificial Neural Network*, Regresi Linear Berganda dan Regresi Polinomial untuk memprediksi harga jual rumah.
3. Menganalisis perbandingan nilai *Root Mean Squared Error* (RMSE) yang diperoleh antara *Artificial Neural Network*, Regresi Linear Berganda dan Regresi Polinomial untuk memprediksi harga jual rumah.

1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat bermanfaat:

sebagai rujukan untuk mengetahui harga jual rumah berdasarkan metode yang akan.



2. Menjadi sumber informasi mengenai performa *Artificial Neural Network*, Regresi Linear Berganda dan Regresi Polinomial.
3. Menjadi sumber informasi penggunaan *Artificial Neural Network* dengan menggunakan metode *Backpropagation*, Regresi Linear Berganda menggunakan metode OLS (*Ordinary Least Square*), dan Regresi polinomial menggunakan metode OLS (*Ordinary Least Square*).



BAB II

TINJAUAN PUSTAKA

2.1 Harga Jual Rumah

Suatu perusahaan harus menentukan penetapan harga untuk pertama kalinya ketika perusahaan tersebut mengembangkan produk baru maupun ketika perusahaan memperkenalkan produknya pada saluran distribusi yang baru. Makin jelas tujuan suatu perusahaan, makin mudah menetapkan harganya. Setiap harga akan menghasilkan tingkat permintaan yang berbeda, hal itu akan memberi pengaruh yang berbeda terhadap tujuan pemasaran suatu perusahaan. Dalam kasus biasa, permintaan harga berbanding terbalik. Makin tinggi harganya, makin rendah permintaan, makin rendah harganya maka permintaan akan cenderung tinggi. Perusahaan ingin menetapkan harga yang dapat menutup biaya produksi, distribusi, dan penjualan produk, termasuk tingkat pengembalian investasi yang memadai atas usaha dan resiko yang dilakukannya. Dalam rentang kemungkinan-kemungkinan harga yang ditentukan permintaan pasar dan biaya perusahaan, perusahaan tersebut harus memperhitungkan biaya, harga dan kemungkinan reaksi pesaing (Kotler, 2009).

Tetapi suatu perusahaan harus memiliki data agar dapat memudahkan dalam menentukan harga jual rumah tersebut. Pada dataset penjual rumah yang akan digunakan adalah data dari bulan Mei 2014 sampai Mei 2015. Adapun variabel yang terkait dari dataset penjualan rumah tersebut, antara lain:

- *Id*: ID unik untuk setiap rumah yang terjual.
- *Date*: Tanggal penjualan rumah.
- *Price*: Harga setiap rumah dijual.
- *Bedrooms*: Jumlah kamar tidur.
- *Bathrooms*: Jumlah kamar mandi.
- *Sqft_living*: Tampilan persegi dari ruang tamu interior rumah.
- *Sqft_lot*: Tampilan luas ruang tanah.
- *Levels*: Jumlah lantai.
- *Waterfront*: Apakah apartemen menghadap ke tepi laut atau tidak.
- *View*: Indeks dari 0 hingga 4 seberapa bagus tampilan properti itu.



- *Condition*: Indeks dari 1 hingga 5 pada kondisi rumah.
- *Grade*: Indeks dari 1 hingga 13, di mana 1-3 gagal dalam konstruksi dan desain bangunan, 7 memiliki tingkat konstruksi dan desain rata-rata, dan 11-13 memiliki tingkat kualitas konstruksi dan desain yang tinggi.
- *Sqft_above*: Tampilan luas ruang perumahan interior yang berada di atas permukaan tanah.
- *Sqft_basement*: Tampilan persegi ruang perumahan interior yang berada di bawah permukaan tanah.
- *Yr_built*: Tahun rumah awalnya dibangun.
- *Yr_renovated*: Tahun renovasi terakhir rumah.
- *Zipcode*: Area pos apa yang ada di rumah.
- *Lat*: Garis Lintang.
- *Long*: Garis Bujur.
- *Sqft_living15*: Tampilan luas ruang tinggal perumahan interior untuk 15 tetangga terdekat.
- *Sqft_lot15*: Tampilan luas tanah dari 15 tetangga terdekat.

Dari beberapa variabel tersebut, akan digunakan sesuai dengan kebutuhan nantinya dalam menentukan harga jual rumah berdasarkan pertimbangan yang ada dengan menggunakan ke 3 metode tersebut untuk diolah nantinya.

2.2 *Artificial Neural Network*

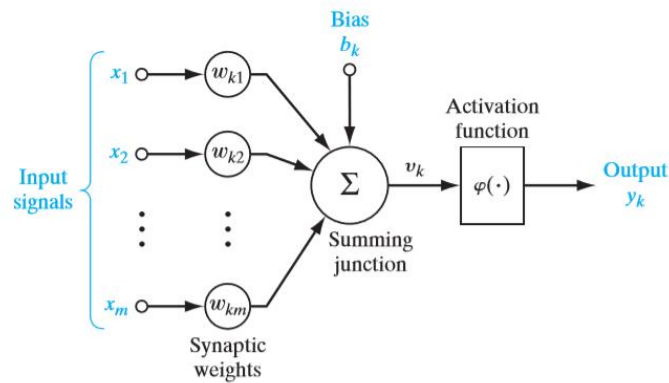
Menurut (Fausett, 1994) *Artificial Neural Network* (ANN) adalah proses sistem informasi yang mempunyai performa karakteristik yang jelas dalam kesamaannya dengan jaringan saraf dalam bidang biologi. ANN sudah dikembangkan sebagai generalisasi model matematika dari pengertian manusia atau saraf biologi, berdasarkan asumsi bahwa:

1. Proses informasi yang terjadi di banyak elemen sederhana yang disebut neuron.
2. Sinyal dikirimkan melalui hubungan koneksi di antara neuron – neuron.
3. Setiap hubungan koneksi mempunyai bobot yang saling berhubungan, di mana jaringan saraf yang khas memperbanyak sinyal yang ditransmisi.

↳ neuron mengaplikasikan sebuah fungsi aktivasi ke jaringan inputnya (hasil dari bobot sinyal input) untuk menentukan sinyal outputnya.



Neuron adalah unit pengolah informasi yang merupakan dasar dari proses sebuah Jaringan Saraf. Adapun model Neuron pada Jaringan Syaraf (Haykin & Simon, 2009):



Gambar 2.1 Artificial Neural Network

Keterangan:

Input: x_1, x_2, \dots, x_m adalah sinyal yang masuk ke sel saraf.

Bobot (*weight*): $w_{k1}, w_{k2}, \dots, w_{km}$ adalah faktor bobot yang berhubung dengan masing-masing node.

Summing function: Sebuah penambah untuk menjumlahkan sinyal-sinyal input. Ditimbang dari kekuatan sinaptik masing-masing neuron.

Fungsi aktivasi: fungsi yang mengubah bentuk linear menjadi nonlinear.

2.3 Fungsi aktivasi

Fungsi aktivasi adalah untuk memberikan kemampuan network agar dapat melakukan tugas non-linear. Tanpa fungsi aktivasi, *neural network* hanyalah kombinasi operasi linear yang hanya dapat melakukan tugas-tugas yang linear pula. Padahal kebanyakan kasus nyata di lapangan merupakan kasus non-linear (Umam, 2018)

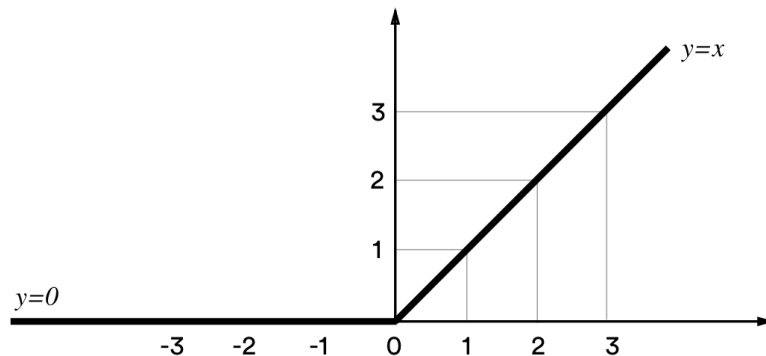
Fungsi aktivasi *Rectified Linear Unit* (ReLU)

salah satu fungsi aktivasi yang sering digunakan pada Neural Network (Chen, 2018). Bentuk fungsi ReLU:

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$



Pada fungsi aktivasi ReLU, semua nilai x negatif akan dipetakan ke 0, seperti pada gambar 2.2:



Gambar 2.2 Grafik fungsi ReLU

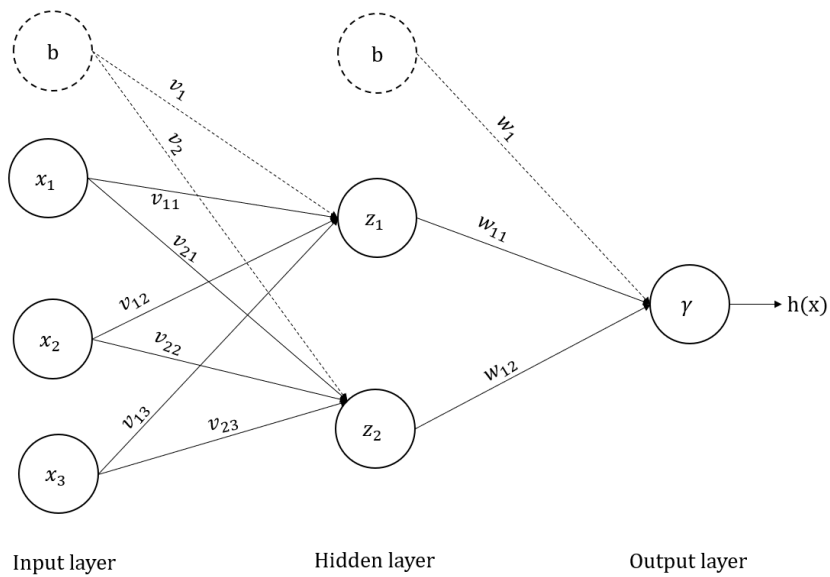
Berdasarkan gambar 2.2, jika nilainya negatif atau di bawah nol, maka hasilnya tetap nol. Tetapi jika nilainya di atas nol, maka hasilnya tetap nilai tersebut.

2.4 *Multilayer Perceptron*

Ada tiga unit yang dapat dibedakan dari *Multilayer Perceptron* yakni: Lapisan input (*input layer*), lapisan tersembunyi (*hidden layer*), lapisan luaran (*output layer*) (WIKIPEDIA, 2020).

1. Lapis masukan (*input layer*) terdiri dari *neuron* yang menerima data masukan dari variabel X . Semua *neuron* pada lapis ini dapat terhubung ke *neuron* pada lapisan tersembunyi atau langsung ke lapisan luaran jika jaringan tidak menggunakan lapisan tersembunyi.
2. Lapisan tersembunyi (*hidden layer*) terdiri dari *neuron* yang menerima data dari lapisan masukan.
3. Lapisan luaran (*output layer*) terdiri dari *neuron* yang menerima data dari lapisan tersembunyi atau langsung dari lapisan masukan yang nilai luarannya melambangkan hasil kalkulasi dari X menjadi nilai Y .





Gambar 2.3 Arsitektur *Multilayer Perceptron*

Keterangan:

x : Input vektor x_1, x_2, x_3

v : Bobot input $v_{11}, v_{12}, v_{13}, v_{21}, v_{22}, v_{23}$

b : Bias yang ada di input layer dan hidden layer

Z : Hidden vektor z_1, z_2

W : Bobot hidden w_{11}, w_{12}

Y : Nilai output hasil perhitungan dari input oleh suatu fungsi aktivasi.

2.4.1 *Feed Forward*

Feed Forward adalah jaringan yang direpresentasikan sebagai satu set neuron yang terhubung bersama, di mana informasi mengalir hanya dalam arah maju, dari input, hidden layer dan ke output. (Dreyfus & Gerard, 2002)

Berikut cara Kerja dari *Feed Forward*

1. Input layer merupakan sebuah vektor

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Nilai matriks bobot yang ada di V akan di transpose.

$$v = \begin{bmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \\ v_{13} & v_{23} \end{bmatrix}$$



$$\text{Jadi } v^T = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \end{bmatrix}$$

Setelah itu input dan bobot dijumlahkan:

$$z_{in_j} = \sum_{i=1}^n x_i v_{ij}$$

Dimana:

$$z_1 = v_{11}x_1 + v_{12}x_2 + v_{13}x_3$$

$$z_2 = v_{21}x_1 + v_{22}x_2 + v_{23}x_3$$

2. Bias yang ada di input layer di tambahkan juga pada akumulasi nilai dari input dan bobot.

$$b = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$z_{in_j} = b + \sum_{i=1}^n x_i v_{ij}$$

Dimana:

$$z_1 = v_{11}x_1 + v_{12}x_2 + v_{13}x_3 + v_1$$

$$z_2 = v_{21}x_1 + v_{22}x_2 + v_{23}x_3 + v_2$$

Setelah itu, gunakan fungsi aktivasi untuk menghitung sinyal outputnya

$$z_j = f(z_{in_j})$$

Pada hidden layer, jumlahkan nilai $[z_1 \quad z_2]$ dengan bobot yang ada di hidden layer dan tambahkan nilai bias yang ada di hidden layer.

$$z_{in_k} = b + \sum_{j=1}^p z_j w_{jk}$$



, $w = \begin{bmatrix} w_{11} \\ w_{12} \end{bmatrix}$. Nilai bobot (w) di transpose, $[w_{11} \quad w_{12}]$ dan nilai bias yang
 iden layer $b = [w_1]$

Dimana:

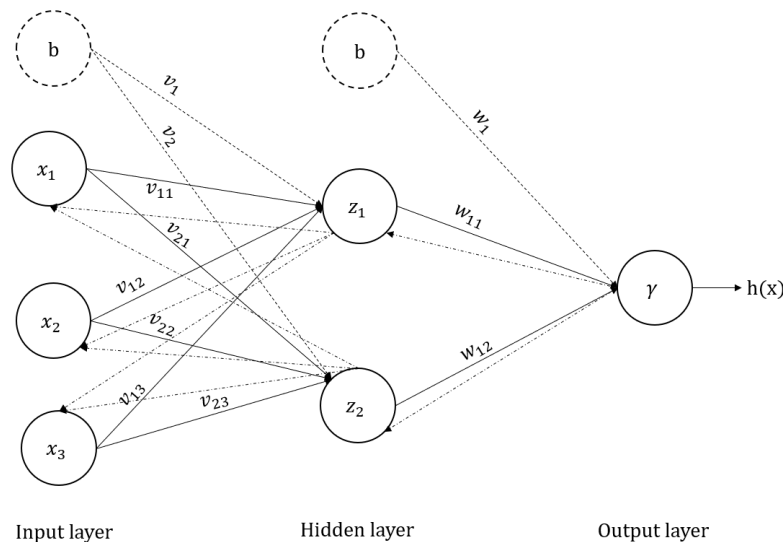
$$y = z_1w_{11} + z_2w_{12} + w_1$$

2.4.2 Backpropagation

Backpropagation merupakan salah satu algoritma pelatihan terarah. Algoritma backpropagation biasanya digunakan oleh *perceptron* dengan banyak lapisan untuk mengubah bobot-bobot yang terhubung dengan neuron-neuron yang ada pada lapisan tersembunyinya. Algoritma Backpropagation menggunakan error output untuk mengubah nilai bobot-bobotnya dalam arah mundur (*backward*). Untuk mendapatkan error tersebut, tahap perambatan maju (*forward propagation*) harus dikerjakan terlebih dahulu. (Fausett, 1994)

Pada dasarnya, pelatihan dengan metode backpropagation terdiri atas tiga langkah, yaitu sebagai berikut:

- Data dimasukkan ke input jaringan (*feedforward*)
- Perhitungan dan propagasi balik dari error yang bersangkutan
- Pembaharuan (adjustment) bobot.



Gambar 2.4 Arsitektur *Backpropagation*

Jika nilai error yang dihasilkan lebih besar dari batas error yang digunakan dalam sistem, maka akan dilakukan koreksi bobot. Koreksi bobot dapat dilakukan menambah atau menurunkan nilai bobot.



Jika sinyal keluaran terlalu besar dari target yang ditentukan maka bobotnya akan dikurangi, sebaliknya jika sinyal keluaran terlalu kecil dari target yang ditentukan

maka bobotnya dinaikkan. Koreksi bobot akan dilakukan sampai selisih target dan sinyal keluaran sekecil mungkin atau sama dengan batas error. Untuk melakukan koreksi bobot akan dilakukan penelusuran ke belakang seperti ditunjukkan dengan tanda panah mundur.

Adapun cara kerja dari *Backpropagation*:

1. Tiap-tiap output menerima target pola yang berhubungan dengan input pembelajaran, hitung informasi errornya.

$$\delta_k = \hat{y} - y$$

Kemudian hitung koreksi bobot (yang akan digunakan untuk memperbaiki nilai bobot w_{jk}) dengan laju pembelajaran α

$$\Delta w_{jk} = \alpha \delta_k Z_j, (k = 1, 2, \dots, m \text{ dan } j = 0, 1, 2, \dots, p)$$

2. Tiap-tiap unit tersembunyi ($Z_j, j = 1, 2, 3 \dots p$) menjumlahkan delta inputnya (dari unit-unit yang ada pada lapisan dibawahnya)

$$\delta_{in_j} = \sum_{k=1}^m \delta_k w_{jk}$$

Hitung informasi errornya, dimana $\theta^{(l)T}$ adalah matriks bobot yang ada di hidden ke layer yang ada di belakangnya dan $\delta^{(l+1)}$ adalah hidden layer ke $i+1$ dari arah belakang kemudian dikalikan dengan fungsi aktivasi.

$$\delta_j = \theta^{(l)T} \delta^{(l+1)} * (z_{in_j})$$

Kemudian hitung koreksi bobot (yang nantinya akan digunakan untuk memperbaiki nilai v_{ij} dengan laju pembelajaran α

$$\Delta v_{ij} = \alpha \delta_j x_i, (j = 1, 2 \dots p; i = 0, 1, 2 \dots n)$$

3. Hitung semua perubahan bobot
 - a. Perubahan bobot garis yang menuju ke unit keluaran

$$w_{jk}(\text{baru}) = w_{jk}(\text{lama}) + \Delta w_{jk}$$



Perubahan bobot garis yang menuju ke unit tersembunyi

$$v_{ij}(\text{baru}) = v_{ij}(\text{lama}) + \Delta v_{ij}$$

Proses iterasi selesai

Keterangan symbol:

δ_k = Informasi tentang kesalahan pada unit Y_k yang disebarkan kembali ke unit tersembunyi

δ_j = Informasi tentang kesalahan dari lapisan output ke unit tersembunyi

α = Laju Pembelajaran (*Learning Rate*)

2.5 Regresi Linear Berganda

Regresi Linier Berganda adalah hubungan secara linear antara dua variabel atau lebih variabel independent (x_1, x_2, \dots, x_n) dengan variabel dependent (Y). (Hasan, 2003) Bentuk umum persamaan regresi linier berganda dapat dituliskan sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \varepsilon$$

Ket:

Y = Variabel dependent

X = Variabel independent

β_0 = Konstanta

β_n = parameter dari model

ε = error pada model

Misalkan Y adalah matriks berukuran $m \times 1$, X adalah matriks berukuran $m \times n$, b adalah matriks berukuran $m \times 1$ dan e adalah matriks berukuran $m \times 1$. Berikut adalah detailnya.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1n} \\ 1 & x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}, b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \text{ dan } \varepsilon = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix}$$

maka kita akan mendapatkan model sebagai berikut:

$$y_1 = b_0 + b_1 x_{11} + b_2 x_{12} + \dots + b_p x_{1p} + e_1$$

$$y_2 = b_0 + b_1 x_{21} + b_2 x_{22} + \dots + b_p x_{2p} + e_2$$

⋮

$$y_m = b_0 + b_1 x_{m1} + b_2 x_{m2} + \dots + b_p x_{mp} + e_m$$



Dalam notasi matriks, dapat dituliskan sebagai berikut:

$$Y = Xb + \varepsilon$$

Untuk meminimalkan kuadrat error nya maka:

$\varepsilon = y - Xb$, merupakan bentuk dari metode kuadrat terkecil, Dimana

$$\varepsilon^2 = (y - Xb)^2 \text{ atau } \varepsilon^T \varepsilon = (y - Xb)^T (y - Xb)$$

$$\varepsilon^T \varepsilon = (y^T - X^T b^T)(y - Xb)$$

$$\varepsilon^T \varepsilon = y^T y - y^T Xb - X^T b^T y + X^T b^T Xb , \text{ atau } \varepsilon^2 = y^T y - 2 X^T b^T y + X^T b^T Xb$$

Untuk meminimalkan kuadrat errornya dengan menghitung turunan pertama terhadap b dan disamakan dengan 0.

$$\frac{\partial(\varepsilon^T \varepsilon)}{\partial b} = 0$$

$$\frac{\partial(\varepsilon^T \varepsilon)}{\partial b} = \frac{\partial(y^T y - 2 X^T b^T y + X^T b^T Xb)}{\partial b}$$

$$0 = -2X^T y + 2X^T Xb$$

$$2X^T y = 2X^T Xb$$

$$X^T y = X^T Xb$$

$$\frac{1}{X^T X} \cdot X^T y = b \text{ atau } b = (X^T X)^{-1} X^T y$$

Maka untuk mendapatkan nilai dari parameter dari model β_n dapat digunakan rumus

$$b = (X^T X)^{-1} X^T y$$

Atau,

$$X^T = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & x_{31} & \cdots & x_{m1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & x_{3n} & \cdots & x_{nm} \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$



asil perkalian matriks dari $(X^T X)$ diperoleh, maka akan di inverskan.

$$X^T = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & x_{31} & \cdots & x_{m1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & x_{3n} & \cdots & x_{nm} \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

Kemudian hasil dari matriks invers yang tadi akan dikalikan dengan hasil matriks $X^T y$. Dari hasil parameter model β_n , sehingga prediksi dari fungsi regresinya sebagai berikut:

$$\hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_n X_n.$$

2.6 Regresi Polinomial

Regresi Polinomial adalah pengembangan dari regresi linear untuk mengatasi data-data non-linear. Adapun rumus untuk mencari koefisien regresinya β , sama dengan yang telah kita jabarkan sebelumnya pada Regresi Linear Berganda. (Umam, 2018).

$$b = (X^T X)^{-1} X^T y$$

Secara formula matematis, yang membedakan antara regresi linear polinomial untuk menghitung nilai koefisien regresi β hanyalah di matriks desain X. Jika di regresi linear dengan dua variabel independent x dengan jumlah datanya adalah m, matriks desainnya adalah:

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}$$

Maka untuk regresi polinomial dengan satu variabel independent x dengan jumlah datanya adalah n dan orde polinomialnya adalah m, matriks desainnya diformulakan sebagai berikut:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix}, b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \varepsilon = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$



regresi polinomial:

$$\hat{y}_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_m x_i^m + \varepsilon_i \quad (i = 1, 2, 3 \dots n)$$

Penjabaran rumus sama dengan regresi linear berganda, Maka untuk mendapatkan nilai dari parameter dari model β dapat digunakan rumus $b = (X^T X)^{-1} X^T y$. Dengan matriks desain di atas, koefisien regresi polinomial dapat kita hitung sama seperti sebelumnya pada regresi linear berganda. Setelah kita mendapatkan nilai koefisien regresinya, kita dapat hitung prediksi output Y dengan $Y = Xb$. Untuk orde polinomial= m , semakin tinggi nilainya, maka model regresi akan kemampuan yang lebih untuk memodelkan data yang kompleks (Umam, 2018)

2.7 Koefisien Determinasi

Koefisien determinasi *R-squared* (R^2) dilakukan untuk melihat adanya hubungan yang sempurna atau tidak, yang ditunjukkan pada apakah perubahan variabel bebas akan diikuti oleh variabel terikat pada proporsi yang sama. Pengujian ini dengan melihat nilai *R-squared* (R^2). Nilai koefisien determinasi adalah antara 0 sampai dengan 1. Selanjutnya nilai *R-squared* (R^2) yang kecil berarti kemampuan variabel-variabel independent dalam menjelaskan variasi variabel dependent amat terbatas. Nilai yang mendekati 1 berarti variabel-variabel independent memberikan hampir semua informasi yang dibutuhkan untuk memprediksi variasi dependent (Ghozali & Imam, 2005).

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Keterangan:

y_i = nilai pada variabel independent.

\hat{y}_i = nilai prediksi.

\bar{y} = rata-rata nilai variabel independent.

2.8 Root Mean Squared Error (RMSE)

Untuk mengukur performa dari model regresi yang di peroleh, nilai error yang ada di model prediksi dan *ground truth* labelnya dapat dijadikan acuan. Dengan nilai *ground truth* label adalah Y atau variable terikat dan nilai prediksi adalah \hat{Y} . (Umam, 2018). Adapun rumus dari RMSE adalah:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}}$$



Keterangan:

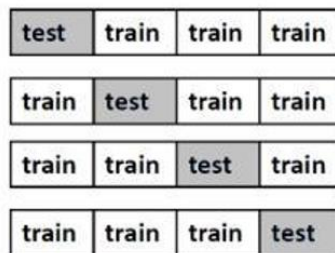
Y = Nilai pada variable indepent atau variable terikat.

\hat{Y} = Nilai Prediksi

n = Jumlah data

2.9 *K-Fold Cross Validation*

Model yang di peroleh dari ketiga metode yang digunakan yakni *Articial Neural Network*, Regresi Linear Berganda dan Regresi Polinomial kemudian dilakukan pengujian menggunakan *K-fold cross validation*. *Cross validation* adalah bentuk sederhana dari teknik statistik. Data yang digunakan dibagi secara acak ke dalam k subset yaitu D_1, D_2, \dots, D_k dengan ukuran yang sama. Dataset akan dibagi menjadi data training dan data testing. Proses *training* dan *testing* dilakukan sebanyak k kali secara berulang-ulang. Pada iterasi ke- i , partisi D_i disajikan sebagai data testing dan partisi sisanya digunakan secara bersamaan dan berurutan sebagai data training. Iterasi kedua, subset D_1, D_2, \dots, D_k akan dites pada D_2 , dan selanjutnya hingga D_k (Witten H, 2001).



Gambar 2.5 Ilustrasi *4-Fold Cross Validation*

Berdasarkan gambar.4 ditunjukkan bahwa nilai fold yang digunakan adalah 4-fold cross validation. Berikut diberikan langkah-langkah pengujian data dengan 4-fold cross validation

- Dataset yang digunakan dibagi menjadi 4 bagian, yaitu $D_1, D_2, D_3,$ dan D_4 . $D_t, t = (1, 2, 3, 4)$ digunakan sebagai data testing dan dataset lainnya sebagai data training.



di mana R -Squared dan *Root Mean Squared Error* dihitung pada setiap iterasi (iterasi-1, iterasi-2, iterasi-3, iterasi-4), kemudian dihitung rata-rata tingkat R -

Squared dan *Root Mean Squared Error* dari seluruh iterasi untuk mendapatkan tingkat *R-Squared* dan *Root Mean Squared Error* data keseluruhan.

