

Penerapan Metode Linear Regression untuk Memprediksi Peserta Training pada PT. NetSolution

Nendi dan Fatonah

Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika
Jl. Radin Inten II Duren Sawit, DKI Jakarta, Jakarta Timur
E-mail : nendijuve@gmail.com, fatonahapril4@gmail.com

Abstrak

Beriring berkembangnya teknologi yang pesat ini, penggunaan teknologi informasi pun sangat di butuhkan banyak perusahaan juga diri sendiri. Untuk itu banyak sekali orang yang memiliki minat untuk melakukan training IT, selain untuk menambah pengetahuan, training IT juga bagus untuk meningkatkan karir. PT Netsolution merupakan salah satu perusahaan yang menyediakan layanan training IT. Dalam bisnisnya pasti mempunyai sekumpulan data peserta training yang sangat banyak, oleh karena itu data harus dapat diolah dengan baik dan rapi. Penelitian ini dilakukan dalam rangka membantu tim Sales & Marketing untuk memprediksi jumlah peserta training baru, sehingga admin dapat menyiapkan jadwal training yang diprediksi akan mengalami kenaikan dalam tiap tahunnya. Dalam kasus ini metode Linear Regression adalah sebuah metode yang bisa membantu mengestimasi atau memprediksi peserta training baru dimasa yang akan datang. Software yang akan digunakan untuk menerapkan metode Linear Regression yaitu RapidMiner dan menghasilkan Performance dengan nilai RMSE sebesar 0.503 yang menunjukkan performa yang bagus dan hasil prediksi yang cukup akurat.

Kata kunci : Linear Regression, RapidMiner, Prediksi, Peserta training.

Pendahuluan

Teknologi informasi dan komunikasi (TIK) semakin berkembang pesat dalam beberapa tahun terakhir dan menjadi bagian penting dalam kegiatan bisnis. Hal tersebut menuntut setiap perusahaan untuk terus berinovasi dan mengembangkan kemampuan karyawan dalam bidang Teknologi informasi dan komunikasi.

Pelatihan atau training adalah suatu investasi penting bagi setiap perusahaan untuk meningkatkan keterampilan dan pengetahuan karyawan mereka dalam bidang Teknologi Informasi dan Komunikasi (TIK). PT. NetSolution menawarkan layanan pelatihan atau training IT sebagai salah satu solusi untuk meningkatkan kompetensi dan keterampilan dalam bidang teknologi informasi bagi individu maupun organisasi yang membutuhkan. Dalam menyediakan layanan training, PT. NetSolution harus memiliki kemampuan untuk memprediksi jumlah peserta training yang akan hadir, agar dapat mengoptimalkan penggunaan sumber daya yang diperlukan, seperti menyusun jadwal yang tepat, menyiapkan ruang kelas yang memadai, menentukan instruktur yang berkualitas, dan menyiapkan materi yang sesuai dengan kebutuhan peserta training.

Peramalan atau forecasting dapat menjadi salah satu metode yang baik untuk diterapkan oleh pe-

rusahaan agar dapat memprediksi permintaan di masa depan, sehingga dapat mengambil keputusan yang lebih baik [1]. Prediksi adalah proses memperkirakan peristiwa masa depan secara sistematis berdasarkan dari informasi masa lalu dan masa kini atau sekarang [2]. Salah satu metode yang sering digunakan untuk memprediksi suatu nilai dengan akurat adalah linear regression [3]. Linear regression merupakan metode untuk mencari hubungan menggunakan pola garis lurus antar variabel dependant (terikat) dengan variable dependent (bebas) [4]. Hubungan antara variabel terkait dengan variabel bebas bergantung pada beberapa bentuk persamaan yaitu mencakup hubungan linear, eksponensial dan berganda [5]. Tujuan dari metode regresi ini yaitu untuk meramalkan nilai variabel terkait dalam hubungan dengan nilai variabel bebas tertentu.[6].

Metode regresi linier terdiri dari 2 jenis yaitu metode regresi linier sederhana dan metode regresi linier berganda [7]. Penelitian ini menggunakan metode regresi linier berganda dengan sepuluh variabel bebas. Regresi linier berganda adalah teknik untuk menentukan hubungan antara variabel yang ingin diprediksi (variabel terkait) dan variabel lain (variabel bebas). Peramalan regresi linier berganda digunakan sebagai teknik untuk mempelajari bagaimana variabel berhubungan dengan variabel lain dalam proses peramalan data. Hal ini

menjadikan metode regresi linier berganda sebagai metode yang tepat untuk menyelesaikan masalah prediksi atau peramalan [8].

Beberapa penelitian telah membuktikan metode ini, seperti penelitian yang dilakukan oleh [9] Memprediksi persediaan ITAS bulan depan menggunakan metode regresi linier. Pada hasil perhitungan MAPE yang telah dilakukan, nilai MAPE terbaik adalah pada bulan Oktober 2019 dengan nilai MAPE sebesar 0.15503 %. Nilai perhitungan akurasi error seluruh data uji peramalan atau prediksi menggunakan MAPE hasilnya kurang dari 10%.

Penelitian yang dilakukan oleh [10] menerapkan metode regresi linier berganda untuk memprediksi stok persediaan barang di bulan berikutnya. Hasil perhitungan menunjukkan bahwa metode regresi linier nilai MAPE sebesar 27,291 dan MAD sebesar 9,916. Penelitian yang dilakukan oleh [11] menggunakan metode regresi linear untuk melakukan prediksi hasil nilai ujian siswa dengan datasheet yang terdiri dari 395 data dan 33 atribut. Hasil evaluasi performance nilai RMSE adalah 0,675. Berdasar hasil tersebut dapat disimpulkan bahwa model yang dihasilkan dapat direkomendasikan untuk digunakan dalam memprediksi nilai ujian siswa.

Penelitian yang dilakukan oleh [12] menerapkan metode linear regression untuk melakukan estimasi jumlah penduduk provinsi sumatera utara. Dataset yang digunakan merupakan tahun 2010-2020 yang diolah menggunakan software RapidMiner. Penelitian yang dilakukan oleh [13] dilakukan prediksi terhadap harga Bitcoin dengan membandingkan empat model algoritma yaitu Linear Regression, Neural Network, Deep Learning, dan k-Nearest Neighbor (k-NN). Tingkat akurasi dari tiap model algoritma akan diuji dengan metode validasi K-Fold Cross Validation dan dievaluasi menggunakan Root Mean Square Error (RMSE). Hasil dengan uji T-Test dalam penelitian ini menyimpulkan bahwa model terbaik untuk memprediksi harga Bitcoin adalah model algoritma Linear Regression dan Neural Network, yaitu dengan hasil RMSE 296.227 +/- 60.125 (micro average: 301.655 +/- 0.000) dan 338.988 +/- 47.837 (micro average: 342.000 +/- 0.000).

Berdasarkan latar belakang tersebut dapat disimpulkan bahwa metode Linear Regression sangat baik dalam melakukan prediksi. Untuk fungsi regresi, terdapat rumus yang dapat digunakan untuk membuat persamaan regresi baru sebagai berikut:

$$y = \alpha + \beta X \quad (1)$$

dengan y adalah variabel dependent (terikat), X adalah variabel independet (bebas), α adalah intercept, dan β adalah koefisien regresi/slop. Selanjutnya Regresi Linear Berganda yang merupakan lanjutan dari regresi linier sederhana, berikut merupakan rumusnya:

$$y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n \quad (2)$$

dengan y adalah variabel dependent (terikat), β_1 adalah intercept dari model, $\beta_2, \beta_3, \dots, \beta_n$ adalah koefisien- koefisien parsial dari variable dependen ke-I, dan X_2, X_3, \dots, X_n adalah variabel-variabel independent ke-I dengan parameternya

Metode regresi linier telah banyak dibahas, untuk penggunaan metode ini ada berbagai aplikasi yang digunakan, salah satunya termasuk aplikasi RapidMiner [14]. RapidMiner adalah aplikasi atau perangkat lunak yang berfungsi sebagai alat pembelajaran dalam ilmu data mining. Platfrom dikembangkan oleh perusahaan yang didedikasikan untuk semua langkah yang melibatkan sejumlah besar data dalam bisnis komersial, penelitian, pendidikan, pelatihan, dan pembelajaran [3].

Implementasi metode linear regression untuk memproses data akan menggunakan tools rapidminer dalam pengolahannya [15]. Terdapat 10 variabel ang ada pada dataset, akan tetapi hanya 6 variabel yang mempengaruhi keluaran prediksi. Dari hasil pengujian menggunakan metode linear regression pada tools Rapidminer akan di evaluai dengan nilai RMSE. RMSE merupakan akar kuadrat rata-rata error dari suatu metode yang digunakan untuk mengevaluasi kesesuaian data sebenarnya dengan model prediksi. Semakin mendekati 0 nilai RMSE, maka semakin akurat prediksinya [16].

Tujuan dari penelitian ini yaitu untuk memprediksi jumlah peserta training baru menggunakan metode linear regression pada PT. NetSolution. Penelitian yang lakukan diharapkan dapat bermanfaat bagi PT. NetSolution dalam memperbaiki kinerja tim Sales & Marketing dan perusahaan dapat mengetahui dengan pasti kebutuhan sumber daya dan mengoptimalkan penggunaannya. Selain itu, hasil penelitian juga dapat memberikan informasi berharga untuk pengambilan keputusan terkait penyediaan layanan training, seperti mengevaluasi dan memperbaiki kurikulum training, meningkatkan kualitas layanan, atau menentukan harga yang lebih kompetitif.

Metode Penelitian

Melalui metode penelitian ini, langkah-langkah proses penelitian akan dijelaskan secara sistematis, sehingga penelitian yang dilakukan lebih terarah dengan baik. Metodologi yang digunakan yaitu metode CRISP-DM (Cross Industry Standard Process for Data Mining). Metode CRISP-DM merupakan salah satu metodologi yang umum digunakan dalam melakukan penelitian dengan data mining. Metode ini populer karena menyediakan pendekatan yang terstruktur dan sistematis untuk mengelola data. CRISP-DM terdiri dari enam tahap yaitu: Business Understanding (pemahaman bisnis), Data Understanding (pemahaman data), Data

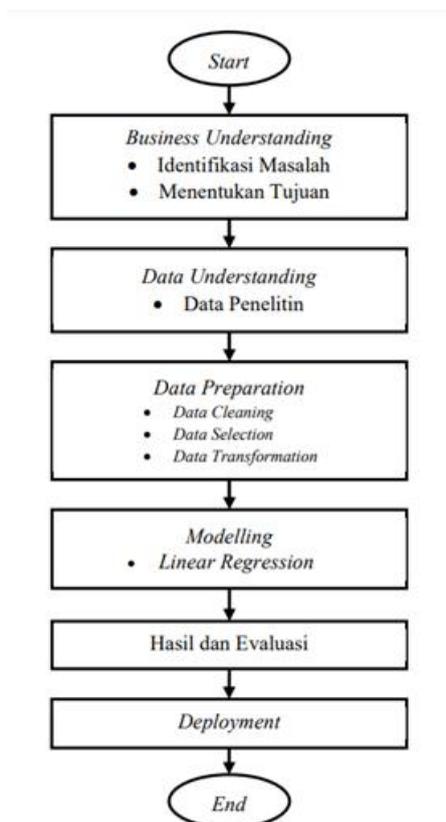
Preparation (persiapan data), Modelling (pemodelan), hasil dan evaluasi, dan Deployment (penyebaran).

Dataset

Data yang digunakan dalam penelitian ini bersifat private, dimana data tersebut merupakan data internal milik perusahaan. Oleh karena itu, peneliti harus memperoleh izin khusus dari perusahaan untuk dapat menggunakan dataset tersebut sebagai bahan penelitian. Data yang digunakan merupakan data peserta training tahun 2018 s/d 2022 pada perusahaan PT. NetSolution. Data ini adalah hasil ekspor dari database perusahaan yang berjenis file xlsx. Total sampel data yang diambil secara acak dalam dataset yaitu sebanyak 60 data dan 10 atribut, diantaranya periode, prakerja, berbayar, gratis, network technician, fiber optic, CCNA, training office, training IT dan member.

Tahapan Penelitian

Pada tahapan penelitian ini, penulis menentukan langkah-langkah apa saja yang dilakukan. Dalam membuat tahapan penelitian dari tahapan awal mulai penelitian sampai akhir penelitian, penulis menggunakan metode CRISP-DM (Cross Industry Standard for Data mining). Dalam tahapan ini akan dilakukan langkah-langkah pengujiannya yaitu seperti pada Gambar 1.



Gambar 1: Tahapan Rancangan Penguujian

Hasil dan Pembahasan

Business Understanding (Pemahaman bisnis)

Tahap ini berfokus pada pemahaman awal mengenai tujuan dan kebutuhan penelitian. Perlu memahami masalah yang ingin dipecahkan, menentukan tujuan, dan mengidentifikasi faktor-faktor kunci untuk memprediksi jumlah peserta training baru. Penulis telah mengidentifikasi permasalahan dalam penelitian ini, diantaranya:

1. Belum adanya solusi untuk memprediksi peserta training baru di PT. NetSolution.
2. Belum diketahui variabel apa saja yang akan digunakan dalam melakukan prediksi.
3. Belum diketahui apakah solusi yang digunakan cukup efektif dalam melakukan prediksi.

Tujuan yang ingin dicapai dari penelitian ini yaitu untuk mendapatkan solusi terbaik dalam memprediksi jumlah peserta training baru di masa depan pada PT. NetSolution.

Data Understanding (Pemahaman data)

Tahap ini berkaitan dengan pengumpulan dan eksplorasi data yang tersedia. Dilakukan proses mengidentifikasi atribut-atribut pada dataset, seperti pemilihan atribut-atribut yang akan digunakan dan yang tidak digunakan dalam permodelan dan pendefinisian atribut yang akan digunakan. Sampel data yang diambil secara acak dalam dataset yaitu sebanyak 60 data dan terdapat 10 atribut, diantaranya: 'Periode' adalah waktu training untuk setiap bulannya, 'Prakerja' adalah jumlah peserta dari kartu prakerja, 'Berbayar' adalah jumlah peserta training yang berbayar, 'Gratis' adalah jumlah peserta training yang tidak bayar atau gratis, 'Network Technician' adalah jumlah peserta training yang mengikuti kelas jaringan komputer dasar, 'Fiber Optic' adalah jumlah peserta training yang mengikuti kelas fiber optik, 'CCNA' adalah jumlah peserta training yang mengikuti kelas ccna, 'Training Office' adalah jumlah peserta training yang mengikuti kelas office, 'Training IT' adalah jumlah peserta training yang mengikuti kelas IT, dan 'Member' adalah seluruh peserta yang mendaftar training.

Data Preparation (Persiapan data)

Pada tahap ini, data dipersiapkan dan diubah agar dapat digunakan dalam pemodelan.

1. Data Cleaning proses mengidentifikasi, memperbaiki, dan menghapus kesalahan, ketidakakuratan, atau inkonsistensi dalam data. Sebelum diolah dengan data mining penulis

melakukan langkah-langkah seperti menghapus data duplikat, mengisi nilai yang hilang, menghapus entri yang tidak valid, dan menormalisasi format data.

2. Data Selection dilakukan proses pemilihan data dari sekumpulan dataset yang ada sebelum masuk ke tahap selanjutnya (lihat Gambar 2). Proses ini dilakukan karena tidak semua data dibutuhkan dalam proses data mining. Hasil dari proses ini, maka diambil dataset sebanyak 60 data dan 10 atribut. Tahap selection ini masih dilanjutkan menggunakan software microsoft excel dengan format xlsx.

Periode	Prakerja	Berbayar	Gratis	Network Technician	Fiber Optic	CCNA	Training Office	Training IT	Member
Dec-22	0	42	0	0	6	0	36	0	42
Nov-22	0	43	0	9	23	3	0	8	43
Oct-22	0	12	0	0	0	12	0	0	12
Sep-22	0	33	0	9	11	13	0	0	33
Aug-22	0	0	0	0	0	0	0	0	0
Jul-22	0	26	0	4	10	1	0	11	26
Jun-22	0	16	0	0	10	6	0	0	16
May-22	0	11	0	6	5	0	0	0	11
Apr-22	0	3	0	3	0	0	0	0	3
Mar-22	0	35	0	4	4	11	0	0	19
Feb-22	0	22	0	8	7	3	2	2	22
Jan-22	0	2	0	0	0	0	2	0	2
Dec-21	0	23	0	7	5	6	0	5	23
Nov-21	12	7	0	0	0	3	2	14	19
Oct-21	98	25	0	0	4	8	10	101	123
Sep-21	19	9	0	0	3	0	0	25	28
Aug-21	0	17	0	2	8	7	0	0	17
Jul-21	0	0	0	0	0	0	0	0	0
Jun-21	0	16	23	6	0	10	0	23	39
May-21	0	0	29	0	0	0	0	29	29

Gambar 2: Data Selection

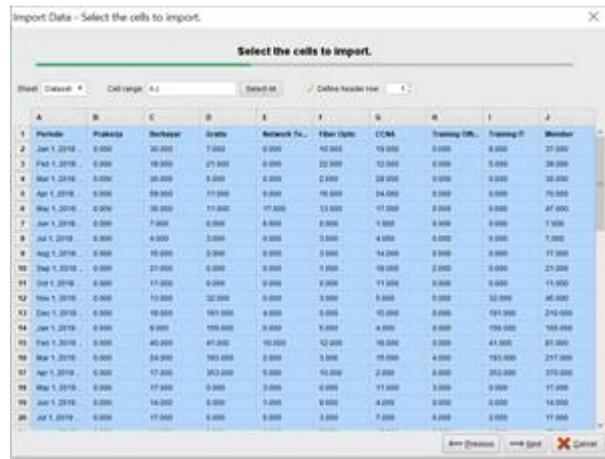
3. Data Transformation

Pada tahap ini, dilakukan proses mengubah format, skala, atau struktur data agar sesuai dengan kebutuhan analisis atau model yang akan digunakan. Atribut Peserta Training dilakukan perubahan nama atribut, diganti dengan Member yang kemudian dijadikan Label yang akan di terapkan pada aplikasi Rapidminer.

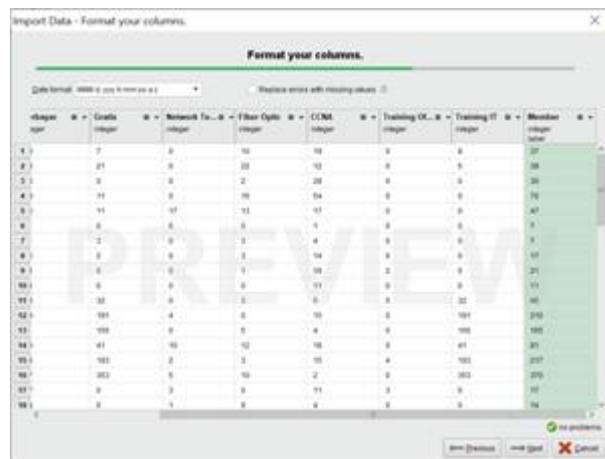
4. Import Data

Hal pertama yang dilakukan yaitu import data, dataset yang telah siap kemudian dimasukkan kedalam Respository Rapidminer (lihat Gambar 3).

5. Menentukan tipe Variabel Atribut (lihat Gambar 4) Periode menggunakan tipe date_time, dan untuk atribut Prakerja, Berbayar, Gratis, Network Technician, Fiber Optic, CCNA, Training Office, Training IT dan Member menggunakan tipe variabel integer karena berisi data angka yang dapat dihitung. Untuk atribut Member akan di jadikan sebagai Label karena atribut tersebut yang akan menghasilkan keluaran prediksi



Gambar 3: Import Data



Gambar 4: Menentukan tipe Variabel

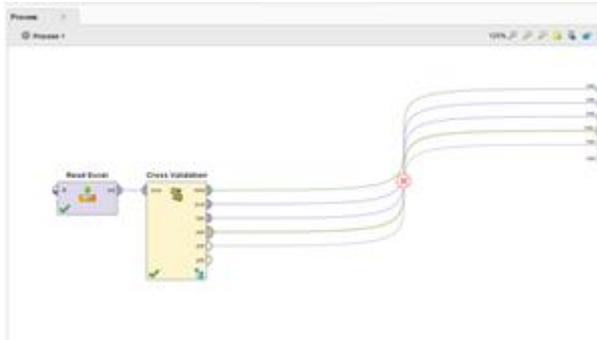
Modelling (Permodelan)

Pada tahap pemodelan akan dilakukan penerapan algoritma Linear Regression menggunakan aplikasi Rapidminer. Mmembangun model Linear Regression dengan menggunakan variabel independen sebagai fitur dan variabel dependen sebagai target. Akan tetapi pada tahap ini tidak dilakukan pembagian data training dan data uji karena akan menggunakan operator Cross Validation dalam tahap pemodelannya. Operator Cross Validation adalah operator yang mengandung dua subproses, yaitu subproses pelatihan untuk melatih model dan subproses pengujian untuk menguji sekaligus mengukur kinerja model. Jelasnya operator ini dapat membagi dataset yang tersedia menjadi data training dan data uji.

Selanjutnya double click pada operator cross validation, maka akan diarahkan ke dalam menu process cross validation (lihat Gambar 5). Dalam proses ini terdapat dua kolom yaitu Training yang akan dimasukan operator Linear Regression (lihat Gambar 6), dan pada kolom Testing masukan Apply Model. Operator Apply Model digunakan untuk menerapkan model yang sebelumnya dilatih

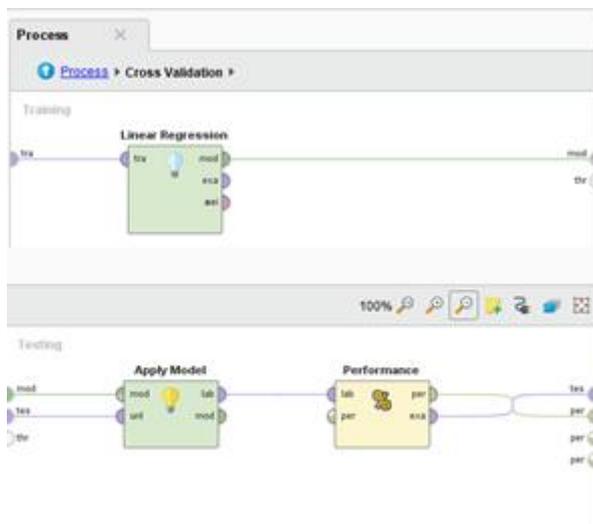
menggunakan data pelatihan ke data yang tidak berlabel (pengujian data).

Proses ini dilakukan dengan cara mengkoneksikan dataset pada Cross Validation guna dalam pembagian data training dan testing.



Gambar 5: Cross Validation

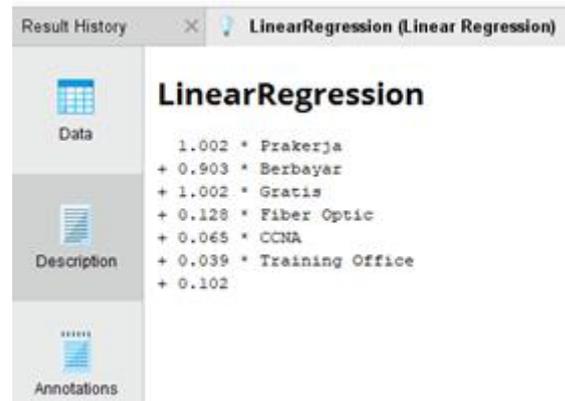
Kemudian masukan operator Performance guna mengevaluasi kinerja model yang secara otomatis memberikan daftar nilai kriteria kinerja untuk tugas tertentu, lalu pastikan kabel terkoneksi dengan benar agar tidak terjadi error.



Gambar 6: Implementasi Metode Linear Regression

Hasil

Setelah dilakukan pengujian menggunakan aplikasi Rapid Miner, maka didapatkan beberapa variabel yang mempengaruhi keluaran prediksi (lihat Gambar 7).



Gambar 7: Variabel yang Mempengaruhi

Dari 10 variabel, terdapat 6 variabel yang mempengaruhi keluaran prediksi. Lebih jelasnya dapat dilihat pada Table 1.

Table 1: Nilai Intercept dan Coefficients

Variabel	Atribut	Coefficients (β)
X1	Prakerja	1,002
X2	Berbayar	0,903
X3	Gratis	1,002
X4	Fiber Optic	0,128
X5	CCNA	0,065
X6	Training Office	0,039
	Intercept (β_1)	0,102

Setelah didapatkan hasil koefisien dari variable-variabel yang mempengaruhi keluaran prediksi, maka nilai tersebut akan diterapkan pada rumus linear regression. Berikut adalah perhitungan prediksi dengan Regresi Linear Berganda yang memiliki rumus:

$$y = \beta_1 + \beta_2X_2 + \beta_3X_3 + \dots + \beta_nX_n \quad (3)$$

Keterangan:
 y = Variabel dependent (terikat)
 β_1 = Intercept dari model
 $\beta_2, \beta_3, \dots, \beta_n$ = Koefisien- koefisien parsial dari variable dependen ke-I
 X_2, X_3, \dots, X_n = Variabel-variabel independent ke-I dengan parameternya

Kemudian nilai Intercept dan Coefficients yang sudah didapatkan dimasukkan ke dalam rumus tersebut. $y = 0.102 + (Prakerjax1.002) + (Berbayarx0.903) + (Gratisx1.002) + (FiberOpticx0.128) + (CCNAX0.065) + (TrainingOfficex0.039)$ (4)

Perhitungan manual:
 1. $y = 0.102 + (0x1.002) + (42x0.903) + (0x1.002) + (6x0.128) + (0x0.065) + (0x0.039) = 38,796$
 2. $y = 0.102 + (0x1.002) + (43x0.903) + (0x1.002) + (23x0.128) + (3x0.065) + (8x0.039) = 42,382$

$$3. y = 0.102 + (0x1.002) + (12x0.903) + (0x1.002) + (0x0.128) + (12x0.065) + (0x0.039) = 11,718$$

Penulis juga membuat perhitungan prediksi secara manual dengan menggunakan microsoft excel (lihat Gambar 8).

Periode	Prakerja	Berbayar	Gratis	Network Technician	Fiber Optic	CCNA	Training Office	Training IT	Member	Prediksi Manual
Des-22	0	42	0	0	6	0	36	0	42	38,796
Nov-22	0	43	0	9	23	3	0	8	41	42,382
Oktober-22	0	12	0	0	0	13	0	0	12	11,718
Sep-22	0	33	0	9	11	13	0	0	33	32,154
Agus-22	0	0	0	0	0	0	0	0	0	0,392
Juli-22	0	26	0	4	10	1	0	11	26	25,254
Juni-22	0	16	0	0	10	6	0	0	16	16,22
May-22	0	11	0	6	5	0	0	0	11	10,675
Apr-22	0	3	0	3	0	0	0	0	3	2,811
Mar-22	0	35	0	4	4	11	0	0	35	32,934
Feb-22	0	22	0	8	7	3	2	2	22	21,117
Jan-22	0	2	0	0	0	0	2	0	2	1,908
Des-21	0	23	0	7	5	6	0	5	23	22,096
Nov-21	12	7	0	0	0	3	2	14	19	18,188
Oktober-21	98	25	0	0	4	8	10	101	123	125,844
Sep-21	19	9	0	0	1	0	0	25	28	28,626
Agus-21	0	17	0	2	8	7	0	0	17	16,532
Juli-21	0	0	0	0	0	0	0	0	0	0,592
Juni-21	0	16	23	6	0	10	0	23	39	39,143
May-21	0	0	29	0	0	0	0	29	29	30,292

Gambar 8: Hasil Prediksi Secara Manual

Hasil pengujian model yang dihasilkan pada Gambar 8 dapat dilihat dengan membandingkan nilai Member dengan nilai Prediction (Member).

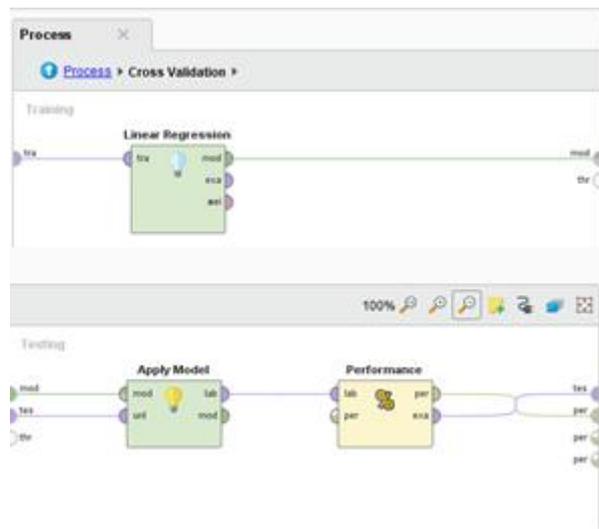
Evaluation

Tahap selanjutnya yaitu pengujian dan evaluasi dengan menggunakan Root Mean Square Error (RMSE) untuk mengukur akurasi nilai error hasil prediksi. Dengan cara memasukan operator performance (regression) pada rapidminer maka memperoleh nilai.

Dapat dilihat dari hasil evaluasi dengan menggunakan RMSE mendapatkan nilai 0.503, itu menunjukkan nilai error yang cukup kecil. Yang artinya menunjukkan bahwa model relatif dapat memprediksi data secara akurat. Karena semakin kecil (mendekati 0) nilai RMSE, maka semakin akurat hasil prediksinya (lihat Gambar 9 dan 10).



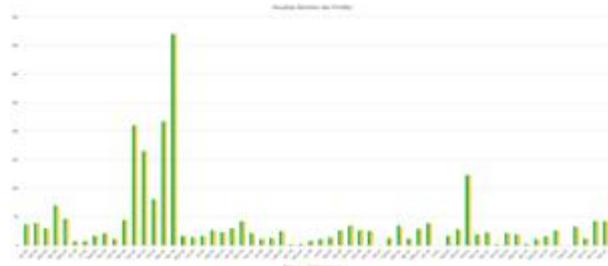
Gambar 9: Hasil Prediksi Dengan RapidMiner



Gambar 10: Nilai RMSE

Deployment

Dapat dilihat pada diagram Clustered Column (lihat Gambar 11) bahwa skema data Member dan Prediksi memiliki kemiripan. Hal ini menunjukkan jika penggunaan metode Linear Regression untuk pelaksanaan prediksi terhadap peserta training baru pada PT. NetSolution berhasil.



Gambar 11: Clustered Column Chart

Dapat dilihat pada Gambar 12, untuk perbandingannya hasil prediksi secara manual dengan perhitungan prediksi yang dihasilkan dengan RapidMiner.

Penutup

Penelitian yang dilakukan telah berhasil membangun model untuk memprediksi peserta training baru dengan metode Linear Regression. Penelitian ini dirancang kedalam tahapan-tahapan metode CRISP-DM (Cross Industry Standard for Data mining) sebagai framework model dalam melakukan pengujian data mining. Dari hasil pengujian yang penulis lakukan, data yang ada diolah sehingga menghasilkan 10 variabel. Kemudian dari 10 variabel yang ada pada dataset, hanya 6 variabel yang mempengaruhi keluaran prediksi. Dari hasil pengujian menggunakan metode Linear Regression serta tools Rapidminer diketahui nilai RMSE sebesar 0.503, ini

menunjukkan performa yang bagus dan hasil prediksi ini sudah tercapai. yang cukup akurat sehingga tujuan dari penelitian

Ro...	Member	prediction(M...	Periode ↓	Prak...	Berbayar	Gratis	Network Te...	Fiber Optic	CCNA	Training Off...	Training IT
36	42	42.000	Dec 1, 2022 ...	0	42	0	0	6	0	36	0
6	43	43.000	Nov 1, 2022 1...	0	43	0	9	23	3	0	8
35	12	12.000	Oct 1, 2022 1...	0	12	0	0	0	12	0	0
30	33	32.034	Sep 1, 2022 ...	0	33	0	9	11	13	0	0
24	0	0.106	Aug 1, 2022 1...	0	0	0	0	0	0	0	0
54	26	26.000	Jul 1, 2022 1...	0	26	0	4	10	1	0	11
18	16	16.226	Jun 1, 2022 1...	0	16	0	0	10	6	0	0
23	11	10.671	May 1, 2022 1...	0	11	0	6	5	0	0	0
60	3	2.764	Apr 1, 2022 1...	0	3	0	3	0	0	0	0
5	19	19.000	Mar 1, 2022 1...	0	35	0	4	4	11	0	0
48	22	20.996	Feb 1, 2022 1...	0	22	0	8	7	3	2	2
59	2	1.968	Jan 1, 2022 1...	0	2	0	0	0	0	2	0
29	23	21.801	Dec 1, 2021 ...	0	23	0	7	5	6	0	5
53	19	19.000	Nov 1, 2021 1...	12	7	0	0	0	3	2	14
4	123	123.000	Oct 1, 2021 1...	98	25	0	0	4	8	10	101
12	28	27.580	Sep 1, 2021 ...	19	9	0	0	3	0	0	25
47	17	16.872	Aug 1, 2021 1...	0	17	0	2	8	7	0	0
22	0	0.106	Jul 1, 2021 1...	0	0	0	0	0	0	0	0
11	39	38.133	Jun 1, 2021 1...	0	16	23	6	0	10	0	23
10	29	29.117	May 1, 2021 1...	0	0	29	0	0	0	0	29
42	12	11.555	Apr 1, 2021 1...	0	12	0	6	4	2	0	0
58	35	32.566	Mar 1, 2021 1...	0	35	0	0	6	8	0	21
21	13	13.049	Feb 1, 2021 1...	0	6	7	0	1	6	0	6
46	0	0.124	Jan 1, 2021 1...	0	0	0	0	0	0	0	0
34	25	25.000	Dec 1, 2020 ...	18	7	0	2	1	2	20	0
41	26	27.434	Nov 1, 2020 1...	23	3	0	0	3	0	23	0
45	34	36.459	Oct 1, 2020 1...	34	0	0	0	0	0	34	0
57	26	25.495	Sep 1, 2020 ...	10	16	0	7	1	8	10	0
17	14	13.915	Aug 1, 2020 1...	4	10	0	2	2	6	4	0
3	11	11.000	Jul 1, 2020 1...	0	11	0	0	8	0	3	0
40	8	8.539	Jun 1, 2020 1...	8	0	0	0	0	0	8	0
16	3	2.989	May 1, 2020 1...	0	3	0	0	0	3	0	0
28	2	2.033	Apr 1, 2020 1...	0	2	0	0	0	2	0	0
56	25	24.241	Mar 1, 2020 1...	0	14	11	1	0	6	6	12
39	13	13.080	Feb 1, 2020 1...	0	13	0	0	6	7	0	0
52	11	11.000	Jan 1, 2020 1...	0	11	0	4	0	7	0	0
9	22	20.731	Dec 1, 2019 ...	0	22	0	11	3	7	1	0
38	42	42.098	Nov 1, 2019 1...	0	26	16	5	13	11	2	11
2	30	30.000	Oct 1, 2019 1...	0	27	3	4	18	5	3	0
20	23	23.032	Sep 1, 2019 ...	0	16	7	6	7	9	0	1
8	27	26.311	Aug 1, 2019 1...	0	24	3	6	5	15	0	1
51	17	17.000	Jul 1, 2019 1...	0	17	0	5	3	7	0	2
27	14	14.148	Jun 1, 2019 1...	0	14	0	1	9	4	0	0
15	17	16.232	May 1, 2019 1...	0	17	0	3	0	11	3	0
50	370	370.000	Apr 1, 2019 1...	0	17	353	5	10	2	0	353
33	217	217.000	Mar 1, 2019 1...	0	24	193	2	3	15	4	193
55	81	79.729	Feb 1, 2019 1...	0	40	41	10	12	18	0	41
26	165	165.459	Jan 1, 2019 1...	0	9	156	0	5	4	0	156
32	210	210.000	Dec 1, 2018 ...	0	19	191	4	0	15	0	191
14	45	44.774	Nov 1, 2018 1...	0	13	32	0	3	5	5	32
31	11	11.000	Oct 1, 2018 1...	0	11	0	0	0	11	0	0
49	21	21.000	Sep 1, 2018 ...	0	21	0	0	1	18	2	0
44	17	16.953	Aug 1, 2018 1...	0	15	2	0	3	14	0	0
7	7	7.350	Jul 1, 2018 1...	0	4	3	0	3	4	0	0
37	7	6.454	Jun 1, 2018 1...	0	7	0	6	0	1	0	0
43	47	46.235	May 1, 2018 1...	0	36	11	17	13	17	0	0
13	70	69.812	Apr 1, 2018 1...	0	59	11	0	16	54	0	0
25	30	29.080	Mar 1, 2018 1...	0	30	0	0	2	28	0	0
1	39	39.000	Feb 1, 2018 1...	0	18	21	0	22	12	0	5
19	37	36.697	Jan 1, 2018 1...	0	30	7	0	10	19	0	8

Gambar 12: Hasil Prediksi Dengan RapidMiner

Beberapa saran yang perlu dipertimbangkan sebagai bentuk dari pengembangan penelitian yang penulis lakukan diantaranya yaitu:

1. Menambahkan atribut yang lebih spesifik dalam menentukan prediksi peserta training baru agar prediksi semakin akurat.
2. Dibuatnya sistem yang terhubung langsung ke mesin, sebab penelitian ini masih berbentuk prototipe yang mana belum digunakan langsung pada perusahaan.
3. Penelitian ini dapat dikembangkan lagi dengan menggabungkan metode-metode lain sehingga dapat menghasilkan hasil prediksi yang lebih variatif

Daftar Pustaka

- [1] N. Darozat, W. Wahyudin dan H. Hamdani, "Penerapan Metode Peramalan Permintaan pada Produk Piece Pivot di PT. Trijaya Teknik Karawang", *J. Serambi Eng.*, vol. 7, no. 2, 2022.
- [2] H. W. Herwanto, T. Widiyaningtyas and P. Indriana, "Application of Linear Regression Algorithm for Predicting Rice Crop Yield", *Natl. J. Electr. Eng. Inf. Technol.*, vol. 8, no. 4, p. 364, 2019.
- [3] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono dan C. Lauw, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression", *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 1, pp. 8–17, 2021.
- [4] M. M. Dewi, L. D. Farida dan M. Nuraminudin, "Regresi Linier untuk Prediksi Konsumsi dan Produksi Daging Unggas (Studi Kasus: Provinsi Jawa Barat)", *J. Inf. Syst. Manag. e-ISSN*, vol. 4, no. 2, pp. 81–85, 2023.
- [5] R. Puspasari, S. Effendi, H. Kurniawan, M. Ayoe dan E. Nasution, "Penentuan Prediksi Hasil Panen Kelapa Sawit Menggunakan Metode Regresi Linier", *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 4, pp. 91–98, 2022.
- [6] I. Wahyuni, N. Nafi'iyah dan Masruroh, "Sistem Peramalan Penjualan Perumahan di Kabupaten Lamongan dengan Menggunakan Metode Regresi Linier Berganda", *Semin. Nas. Sist. Inf.* 2019, no. September, pp. 1969–1973, 2019.
- [7] N. Almumtazah, N. Azizah, Y. L. Putri dan D. C. R. Novitasari, "Prediksi Jumlah Mahasiswa Baru Menggunakan Metode Regresi Linier Sederhana", *J. Ilm. Mat. Dan Terap.*, vol. 18, no. 1, pp. 31–40, 2021.
- [8] A. Fitri Boy, "Implementasi Data Mining Dalam Memprediksi Harga Crude Palm Oil (CPO) Pasar Domestik Menggunakan Algoritma Regresi Linier Berganda (Studi Kasus Dinas Perkebunan Provinsi Sumatera Utara)", *J. Sci. Soc. Res.*, vol. 4307, no. 2, pp. 78–85, 2020.
- [9] N. L. W. Arya Della, R. A. N. Diaz, dan K. D. P. Novianti, "Penerapan Metode Regresi Linier untuk Memprediksi Permohonan ITAS", *J. Eksplora Inform.*, vol. 10, no. 2, pp. 92–100, 2021.
- [10] N. Nafi'iyah, "Prediksi Jumlah Penjualan pada Toko Makmur Jaya Elektronik dengan Regresi Linier", *Res. Comput. Inf. Syst. Technol. Manag.*, vol. 2, no. 2, p. 47, 2019.
- [11] M. Sholeh, E. K. Nurnawati dan U. Lestari, "Penerapan Data Mining dengan Metode Regresi Linear untuk Memprediksi Data Nilai Hasil Ujian Menggunakan RapidMiner", *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 8, no. 1, pp. 10–21, 2023.
- [12] I. Indriani, D. Siregar dan A. P. Windarto, "Penerapan Metode Linear Regression dalam Mengestimasi Jumlah Penduduk", *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 1112, 2022.
- [13] Mikhael, F. Andreas dan U. Enri, "Perbandingan Algoritma Linear Regression, Neural Network, Deep Learning, Dan K-Nearest Neighbor (K-Nn) Untuk Prediksi Harga Bitcoin", *J. Sist. Inf.*, vol. 14, no. 1, pp. 2450–2464, 2022.
- [14] A. Hidayat, M. T. Ziyad dan C. Juliane, "Memprediksi Volume Sampah di Jawa Barat dengan Metode Regresi Linier", *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 1, pp. 590–603, 2023.
- [15]] S. Supriyanto, M. Ilhamsyah dan U. Enri, "Prediksi Harga Minyak Kelapa Sawit Menggunakan Linear Regression Dan Random Forest", *J. Ilm. Wahana Pendidik.*, vol. 8, no. 7, pp. 1–8, 2022.
- [16] Y. Aqsho Ramadhan, A. Faqih dan G. Dwilestari, "Prediksi Penjualan Handphone di Toko X menggunakan Algoritma Regresi Linear", *J. Inform. Terpadu*, vol. 9, no. 1, pp. 40–44, 2023.