**BULLETIN OF INFORMATION TECHNOLOGY (BIT)**
Vol 4, No 2, September 2023, Hal 299 - 305
ISSN 2722-0524 (media online)
DOI 10.47065/ bit.v3i1. 733
https://journal.fkpt.org/index.php/BIT

# Prediksi Jumlah Kasus Klaim Indemnity Dengan Menggunakan Algoritma Regresi Linear Pada Asuransi Mandiri Inhealth

**Qori Yumansya, Ahmad Turmudi Zy, Muhamad Fatchan**
Fakultas Teknik, Program Studi Teknik Informatika, Universitas Pelita Bangsa, Bekasi, Indonesia
Email: [1]qoryum04@gmail.com, [2]turmudi@pelitabangsa.ac.id, [3]fatchan@pelitabangsa.ac.id
Email Penulis Korespondensi: qoryum04@gmail.com

**Abstrak**− Asuransi merupakan salah satu jenis lembaga keuangan yang bertujuan untuk memberikan jaminan kepada nasabah terhadap risiko yang mungkin terjadi pada masa yang akan datang. Pada penelitian ini dengan memanfaatkan beberapa data kasus klaim indemnity pada assuansi inhealth melalui pendekatan metode prediksi dan dapat diterapkan dalam menganalisis data guna melakukan prediksi data assuransi dimasa mendatang berdasar dari tingkat kebutuhan. Proses prediksi algoritma Regresi Linear sederhana dapat diimplementasikan dimana hasilnya juga menunjukkan sebuah wawasan baru bagi kebutuhan prediksi terhadap data klaim .Pengujian dengan menggunakan rapidminer menghasilkan performa yang relevan dengan skenario yang dimodelkan. Model persamaan *Regresi Linear* sederhana setelah dibandingkan hasil perhitungan secara manual dan juga dengan aplikasi *Rapid Miner* secara umum menunjukkan data yang sama. Nilai RMSE juga didapat saat melakukan evaluasi performa model yang diterapkan, dengan nilai RMSE sebesar 0,273 dengan standar deviasi +- 0,0.

**Kata Kunci :** Regrensi Linear, Asuransi Kesehatan, Prediksi, Data Mining, Kesehatan

***Abstract***− Insurance is a type of financial institution that aims to provide guarantees to customers against risks that may occur in the future. In this study, by utilizing some data on indemnity claim cases on inhealth insurance through a prediction method approach and can be applied in analyzing data to make predictions of future insurance data based on the level of need. The prediction process of a simple Linear Regression algorithm can be implemented where the results also provide new insights for the prediction needs of claim data. Tests using rapidminer produce performance that is relevant to the scenario being modeled. The simple Linear Regression equation model after comparing the results of calculations manually and also with the Rapid Miner application generally shows the same data. The RMSE value is also obtained when evaluating the performance of the applied model, with an RMSE value of 0.273 with a standard deviation of +- 0.0.

**Keywords:** Linear Regression, Health Insurance, Prediction, Data Mining, Health

## 1. PENDAHULUAN

Asuransi merupakan salah satu jenis lembaga keuangan yang bertujuan untuk memberikan jaminan kepada nasabah terhadap risiko yang mungkin terjadi pada masa yang akan datang. Salah satu jenis asuransi yang banyak ditawarkan oleh perusahaan asuransi adalah asuransi kesehatan. Asuransi kesehatan memiliki manfaat yang sangat besar bagi nasabah, terutama bagi mereka yang tidak memiliki kemampuan finansial yang cukup untuk menanggung biaya perawatan kesehatan yang tinggi. Dalam dunia asuransi pelayanan menjadi salah satu faktor yang harus diperhatikan dalam bisnis Asuransi Kesehatan. Pelayanan mencakup salah satunya adalah penyelesaian penanganan di rumah sakit. Pelayanan dapat dikatakan baik jika penyelesaian penanganan kasus dapat diselesaikan sesuai dengan (SLA) *Service Level Agreement* yang telah disepakati oleh pihak tertanggung dan Pihak Perusahaan Asuransi.

Asuransi Mandiri *Inhealth* merupakan salah satu perusahaan asuransi kesehatan yang terkemuka di Indonesia. Meskipun Asuransi Mandiri *Inhealth* telah beroperasi selama 30 tahun berjalan dan telah menangani banyak kasus klaim *indemnity*, namun tidak ada yang tahu pasti berapa jumlah kasus klaim *indemnity* yang akan terjadi di masa yang akan datang, terkadang karna jumlah kasus yang membengkak membuat karyawan kualahan untuk memprosesnya. Oleh karena itu, penting untuk memprediksi jumlah kasus klaim indemnity pada Asuransi Mandiri Inhealth agar perusahaan dapat mempersiapkan diri dengan baik untuk menangani kasus-kasus tersebut, karna jumlah kasus yang tidak menentu dengan rata - rata perbulan 5000 sampai dengan 9000 kasus klaim dan di kerjakan oleh 2 orang admin klaim dengan rata – rata target adalah 100 kasus sehingga jumlah kasus membengkak dan membuat karyawan kualahan untuk memprosesnya data kasus klaim.

Dalam mengukur keakuratan dari hasil menggunakan *regresi linier*, terdapat penelitian lain yang mendukungnya. Untuk dapat menyelesaikan permasalahan yang ada maka salah satu cara yang dapat dilakukan untuk memprediksi pengguna layanan kesehatan dengan menggunakan teknik *data mining*. Adapun teknik yang digunakan dalam hal ini adalah Algoritma *Regresi Linier*. Hasil dari penelitian ini adalah jumlah kasus klaim yang memiliki keterkaitan atau hubungan yang diolah dengan teknik *data mining* menggunakan algoritma *regresi linier* dapat membantu pihak manajemen perusahaan dalam menentukan pelaynan kesehatan.

Penelitian sebelumnya mengenai, machine linear untuk analisis regresi linier biaya asuransi kesehatan dengan menggunakan python jupyter notebook[1]. Dataset terdiri dari 1338 dan 7 kolom. Metode penelitian dilakukan dengan memeriksa data dari data yang salah atau dapat mengganggu proses analisis, melakukan analisis pada dataset serta

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 4, No 2, September 2023, Hal 299 - 305
ISSN 2722-0524 (media online)
DOI 10.47065/ bit.v3i1. 733
https://journal.fkpt.org/index.php/BIT

membagi data menjadi data training dan data test. Proses pembagian data adalah 80% digunakan untuk data training dan 20% untuk data test. Semua proses diolah dengan menggunakan pemrograman Python. Library Python yang digunakan numpy, pandas, matplotlib, seaborn, sklearn. Proses analisis dikerjakan dengan Jupyter Notebook. Hasil penelitian menghasilkan model regresi linear ganda y = -12436.85+ 270.35 X1 -188.37 X2+342.77 X3+474.07 X4+24320.10 X5 -385.60 X6 dengan Coefficient of determination 0.7244150380582826 dan MSE 34608265.193358265. Hasil akhir analisis dilakukan perbandingan antara y aktual dengan y prediksi baik dalam bentuk tabel maupun grafik.

Komparasi deteksi kecurangan pada data klaim asuransi pelayanan kesehatan menggunakan metode support vector machine (svm) dan extreme gradient boosting (xgboost)[2]. eh karena itu, penelitian ini membahas bagaimana cara mendeteksi fraud pada pelayanan kesehatan dengan cara machine learning. Machine learning adalah cara peningkatan kemampuan mesin dalam menyelesaikan masalah yang baru. Metode machine learning yang digunakan adalah klasifikasi Support Vector Machine (SVM) dan metode klasifikasi Extreme Gradient Boosting (XGBoost) yang hasilnya dibandingkan untuk melihat model yang lebih baik. Hasil yang didapatkan adalah hasil yang berhasil mendeteksi data fraud pada data pelayanan kesehatan tersebut dengan performa klasifikasi yang baik dalam membantu memberikan referensi pada penyedia layanan dalam mendeteksi *fraud* . *Metode XGBoost* menghasilkan performa klasifikasi yang baik dengan menghasilkan nilai *Balanced Accuracy* dan nilai Recall sebesar 0.9995 dan 0.9994.

Penerapan data mining untuk prediksi perilaku pelanggan menggunakan multiple linear regression[3]. Namun dengan adanya perkembangan membuat masyarakat sangat selektif dalam menantukan gaya hidup mereka. Hal tersebut seolah mengharuskan perusahaan untuk lebih baik lagi dalam mengatur strategi penjualannya, terutama dalam memahami serta memprediksi perilaku palanggan terhadap produsen. Dengan adanya Data Mining diupayakan untuk dapat memprediksi perilaku pelanggan dengan adanya Produk Abaya yang diinginkan baik dari Model, Jenis Kelamin, Berat Badan, SKU, PB, LD, dan Harga. Hasil dari penelitian ini dapat ditentukan menggunakan Algoritma Linier Regresi Berganda dengan nilai akurasi 100 %.

Analisis pengelompokan k-means untuk data bivariat laju kunjungan dan rasio rujukan (studi data pada seluruh fasilitas kesehatan tingkat i di bpjs kesehatan surakarta[4]. Analisis hubungan didasarkan pada data di tiap kelompok hasil pengelompokkan sehingga selanjutnya dapat diketahui bagaimana keterhubungan data laju kunjungan dan ratio rujukan faskes-faskes tingkat I di tiap kelompok. Analisis hubungan dilakukan dengan menggunakan pendekatan copula (bivariat). Ukuran keterhubungan yang digunakan untuk mengetahui keterhubungan data dalah Kendall's Tau dan Spearman's Rho. Copula yang terbaik untuk menjelaskan keterhubungan data didasarkan pada p-value tertinggi dari uji kecocokan statistik Cram??r-von Mises yang diperoleh melalui simulasi parametric bootstrap. Paper ini memberikan hasil bahwa dua kelompok (kelompok 1 dan kelompok 4) yang mana keterhubungan data di dalamnya dapat digambarkan melalui copula, sedangkan dua kelompok yang lain tidak dapat digambarkan dari lima copula yang diujikan dalam paper ini. Kelompok 1 digambarkan melalui copula Gumbel dengan marginal-marginalnya adalah Normal- Normal, sedangkan kelompok 4 digambarkan melalui copula Ali-Mikhail-Haq (AMH) dengan marginal-marginalnya Weibull-Gamma.

Klasterisasi data rekam medis pasien pengguna layanan bpjs kesehatan menggunakan metode k-means[5]. Riwayat pasien yang menggunakan layanan Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan tersimpan pada data rekam medis. Setiap data rekam medis berisikan informasi penting yang sangat berharga dan dapat diolah untuk menggali pengetahuan baru menggunakan pendekatan data mining. Penelitian ini bertujuan untuk membantu rumah sakit Prof. Dr. Tabrani dalam mengelompokan data pasien pengguna BPJS Kesehatan, sehingga diketahui pola penyebaran penyakit berdasarkan kelas layanan. Data yang digunakan adalah data rekam medis pasien tahun 2019 periode bulan Oktober sampai dengan Desember, data tersebut akan diproses menggunakan algoritma K-Means Clustering dengan jumlah cluster sebanyak 3. Pada cluster 0 (H0) terdapat 3 pasien yang didominasi penyakit A09.9 (*Diare/Disentri*) pada Kelas 2 dan Kelas 3, untuk cluster 1 (H1) tedapat 5 pasien dengan jenis penyakit yang lebih beragam, sedangkan untuk cluster 2 (H2) terapat 5 pasien yang didominasi penyakit K30 (Dyspepsia) pada Kelas 1

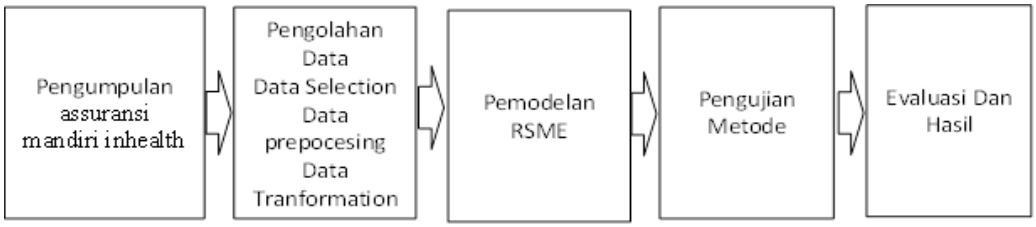
Berdasarkan uraian diatas penelitian ini menggunakan salah satu algoritma *regresi linier* untuk menentukan prediksi. Dengan tujuan memprediksi jumlah kasus klaim indemnity pada Asuransi Mandiri Inhealth. Algoritma regresi linear merupakan salah satu metode yang banyak digunakan dalam bidang data mining untuk memprediksi nilai suatu variabel tergantung pada nilai variabel lainnya. Dengan menggunakan algoritma regresi linear diharapkan dapat membantu Asuransi Mandiri Inhealth dalam mengelola dan mengantisipasi kasus klaim indemnity dimasa yang akan datang.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Tahapan penelitian untuk mencari hasil pada data retail maka dibuatlah langkah-langkah tersebut untuk meneliti dalam penelitian yaitu agar dapat mempermudah dan sesuai dengan tujuan yang diinginkan

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 4, No 2, September 2023, Hal 299 - 305
ISSN 2722-0524 (media online)
DOI 10.47065/ bit.v3i1. 733
https://journal.fkpt.org/index.php/BIT




**Gambar 1.** Tahapan Penelitian

### 2.2 Tahapan Data Mining

Dalam penelitian ini penulis menggunakan pendekatan Data Mining untuk melakukan analisis data. Data yang akan dijadikan dataset dalam penelitian ini adalah data assuransi mandiri inhealth yang di ambil dari data penyedia layanan kesehatan yang ada di bekasi kota, yang di dapatkan dalam bentuk file spreadsheet berformat excel dan data masih melakukan proses selection dan preprocesing berikut ini tahapan proses selection dan prepocesing.

### 2.3 Data Selection

Pada tahap data selection merupakan proses pembersihan dari data yang akan dipakai untuk penghapusan data dengan membuang missing value, duplikasi data, dan memeriksa inkonsistensi data dan memperbaiki kesalahan pada data. Proses pembersihan data dilakukan secara manual dengan bantuan software spreadsheet. Dalam proses ini, data awal yang ada memiliki 1000 record, namun terdapat 100 record yang missing value, dimana record ini hanya mencantumkan qty 0 dan mines

### 2.4 Data Preprocessing

Data *preprocessing* merupakan proses pemilihan data dari sekumpulan data operasional yang ada sebelum masuk ke tahap mining data maupun informasi. Pada tahap ini akan dilakukan langkah – langkah sebagai berikut :

1. Diambil data sebanyak 1000 record dari sekumpulan assuransi mandiri inhealth yang terdapat pada spreadsheet dan sudah selesai di cleaning sebelumnya.
2. Dari 1000 record dataset ini setelah dikategorikan berdasar nama , jumlah pasien dan faskes.
3. Dilakukan seleksi atribut yang akan dipakai dan dianalisis, karena pada
4. data awal terdapat beberapa atribut yang tidak dibutuhkan seperti atribut. Penghapusan atribut dari data utama dikarenakan tidak ada keterkaitan dalam perhitungan algoritma regresi liniear yang akan digunakan. Atribut yang akan dipakai hanya ada 3 yakni berdasar nama, jumlah dan faskes. Berikut adalah hasil pengolahan data awal setelah melawati tahapan diatas untuk dijadikan dataset pada tahap selanjutnya, ditunjukkan pada Tabel 1 **:**

Tabel 1 Data Selection

| Bulan Pelayanan | Jumalah Pasien | provider |
|---|---|---|
| 01-2022 | 48 | 1 |
| 02-2022 | 48 | 2 |
| 03-2022 | 87 | 3 |
| 04-2022 | 6 | 4 |
| 05-2022 | 33 | 2 |
| 06-2022 | 79 | 5 |
| 07-2022 | 72 | 5 |
| 08-2022 | 92 | 6 |
| 09-2022 | 16 | 3 |
| 10-2022 | 56 | 2 |

### 2.5 Data Transformation

Data Transformation merupakan proses mengubah format data awal menjadi sebuah format data standar untuk proses pembacaan data dengan algoritma regresi liniear pada program maupun tool yang digunakan. Berikut adalah hasil pengolahan data awal setelah melawati tahapan diatas untuk dijadikan dataset pada tahap selanjutnya, ditunjukkan pada

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 4, No 2, September 2023, Hal 299 - 305
ISSN 2722-0524 (media online)
DOI 10.47065/ bit.v3i1. 733
https://journal.fkpt.org/index.php/BIT

**Tabel 2.** Dataset

| Bulan Pelayanan | Jumalah Pasien | provider |
|---|---|---|
| 01-2022 | 48 | 1 |
| 02-2022 | 48 | 2 |
| 03-2022 | 87 | 3 |
| 04-2022 | 6 | 4 |
| 05-2022 | 33 | 2 |
| 06-2022 | 79 | 5 |
| 07-2022 | 72 | 5 |
| 08-2022 | 92 | 6 |
| 09-2022 | 16 | 3 |
| 10-2022 | 56 | 2 |

### 2.6 Pemodelan

Pengujian metode dilakukan untuk mengetahui hasil perhitungan yang dianalisa dan untuk mengetahui apakah fungsi bekerja dengan baik atau tidak. Setelah data dihitung secara manual, kemudian data diuji menggunaka *tools rapidminer* untuk memastikan apakah hasil perhitungan berjalan baik dalam menentukan prediksi dan mendapatkan hasil *RSME* dari pengujian data dari hasil yang diperoleh *rapidminer*, berikut ini adalah data sample
Evaluasi dapat dilakukan dengan cara mengamati dan menganalisa hasil dari algoritma yang digunakan untuk memastikan bahwa hasil pengujian benar-benar sesuai dengan pembahasan. Melakukan pengecekan terhadap setiap nilai atribut dan model yang sudah dibangun. Kemudian melakukan evaluasi dengan cara mengamati dan menganalisa hasil dari algoritma yang digunakan untuk memastikan bahwa hasil pengujian menghasilkan nilai *RSME* benar dan sesuai hasil pembahasan, pengujian dilakukan untuk mengukur keakuratan hasil dari tiap model yang diusulkan.
*Root Mean Squared Error* (RMSE) adalah aturan penilaian kuadrat yang juga mengukur besarnya ratarata kesalahan. RMSE adalah akar kuadrat dari rata-rata perbedaan kuadrat antara prediksi dan observasi aktual [11]. Cara yang juga biasanya digunakan untuk mengevaluasi model data mining adalah dengan RMSE. rumusnya ditulis sebagai berikut:

$$\mathrm{RMSE} = \left(\frac{\Sigma(y_i - \hat{y}_i)}{\mathrm{n}}\right)^{1/2} \quad [1]$$

Metode Kuadrat terkecil (*least square method*): metode yang paling populer untuk menetapkan persamaan regresi linear sederhana.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Proses Pengujian Data

Setelah proses pengumpulan data selesai dilakukan, maka tahapan selanjutnya yang dilakukan yaitu membuat pemodelan data pada rapid miner. Model yang digunakan pada rapid miner yaitu menggunakan algoritma regresi linear. Langkah pertama ialah memasukkan data hati kedalam rapidminer untuk diolah. Dalam penelitian ini dataset dibagi menjadi 2 bagian, yaitu 90% data *training* dan 10% data *testing* dengan menggunakan *cross validation*. *cross validation* adalah teknik validasi yang membagi data menjadi dua bagian secara acak, sebagian sebagai *data training* dan sebagian lainnya sebagai *data testing*.

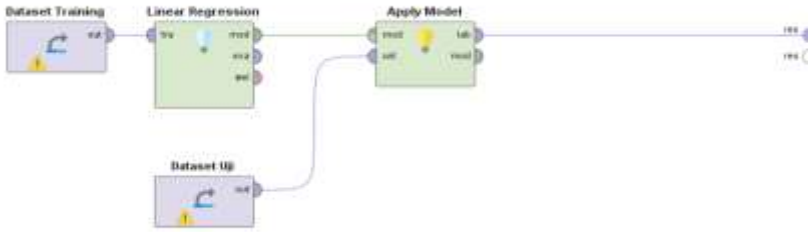
### 3.2 Evaluasi Data Uji pada Rapid Miner

Tahapan evaluasi terhadap data uji *(testing)* yang digunakan untuk menilai model persamaan algoritma *Regresi Linier* sederhana yang diterapkan dalam pembentukan prediksi kasus klaim indemnity pada assuansi inhealth dari suatu objek baru dengan keakurasian yang tepat. Penelitian ini penulis memanfaatkan aplikasi *rapidminer Studio*, hasil pengujian yang dilakukan menggunakan aplikasi *rapidminer Studio* adalah dengan menerapkan langkah-langkah sebagai berikut :

1. Melakukan *import* data yang diperlukan untuk proses pada *tools rapidminer*. Pada aplikasi Rapid Miner pilih dan klik **Import Data**, kemudian pilih data yang akan dipakai serta kemudian menentukan attribut dan label yang akan digunakan.
2. Klik menu **Design**, pada tampilan proses, tambahkan *dataset* latih *(training)* dan *dataset* uji *(testing)* pada folder ke layar tampilan proses.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 4, No 2, September 2023, Hal 299 - 305
ISSN 2722-0524 (media online)
DOI 10.47065/ bit.v3i1. 733
https://journal.fkpt.org/index.php/BIT

3. Berikutnya pada menu **Modelling**, dalam submenu **Predictive**, pilih operator fungsi *Linear Regression*, untuk menerapkan persamaan algoritma Regresi Linear sederhana terhadap proses prediksi objek yang akan dilakukan.
4. Selanjutnya pada menu **Scoring** pilih operator *Apply Model* dan drag ke layar tampilan proses, melalui fungsi ini dapat mengatur penerapan model dari *dataset* yang dipakai pada proses ini terhadap prediksi label yang akan diterapkan.
5. Koneksikan seluruh operator fungsi dari disain yang sudah disiapkan sehingga membentuk alur proses seperti pada gambar 2 berikut:




**Gambar 2** Implementasi Algoritma Regresi Linear pada aplikasi RapidMiner Studio

Setelah dilakukan ***Running Process*** pada *tools rapidminer*, didapatkan hasil prediksi objek testing dengan menggunakan algoritma Regresi Linier yang dapat dilihat pada gambar berikut dibawah ini

**Tabel 3.** Hasil prediksi pada dataset uji menggunakan aplikasi RapidMiner Studio

| | | | |
|---|---|---|---|
| 1 | ? | 30.090 | 36 |
| 2 | ? | 5.660 | 3 |
| 3 | ? | 4.919 | 2 |
| 4 | ? | 23.428 | 27 |
| 5 | ? | 6.400 | 4 |
| 6 | ? | 6.400 | 4 |
| 7 | ? | 7.140 | 5 |
| 8 | ? | 24.908 | 29 |
| 9 | ? | 6.400 | 4 |
| 10 | ? | 7.140 | 5 |
| 11 | ? | 6.400 | 4 |
| 12 | ? | 4.179 | 1 |
| 13 | ? | 7.140 | 5 |
| 14 | ? | 8.621 | 7 |
| 15 | ? | 10.102 | 9 |
| 16 | ? | 6.400 | 4 |
| 17 | ? | 15.284 | 16 |

Pengujian performa terhadap model dan algoritma dilakukan dengan maksud mengetahui hasil perhitungan yang dianalisa dan mengukur metode serta algoritma yang digunakan apakah berfungsi dengan baik atau tidak. Dengan menerapkan model tersebut untuk hasil dari perhitungan nilai RMSE *(Root Mean Square Error)* melalui aplikasi *rapidminer* Studio menghasilkan angka RMSE sebesar **0.273 +/- 0.000**, nilai yang cukup kecil sehingga hanya kecil perbedaan yang terjadi dari proses simulasi perhitungan manual dengan aplikasi *rapidminer* Studio terhadap data uji *(testing)* yang disiapkan.

### 3.3 Perbandingan Manual dan Rapid Miner

Persamaan regresi linier tersebut kemudian akan diimplementasikan dalam memprediksi pada data *testing*. Secara umum penerapan persamaan regresi linear tersebut diaplikasikan untuk memprediksi 20 *record* dataset sebagai data uji *(testing)* yang sudah ditentukan. Dari perhitungan yang dilakukan secara manual dan juga dibandingkan dengan proses pada aplikasi rapid miner, hasil yang ditunjukkan tidak memiliki perbedaan yang signifikan, dengan kata lain baik dari perhitungan manual maupun yang diproses dalam aplikasi menunjukkan kemiripan hasil. Dibawah ini adalah tabel perbandingan antara hasil perhitungan manual dan yang ada pada aplikasi rapid miner terhadap variabel Y.

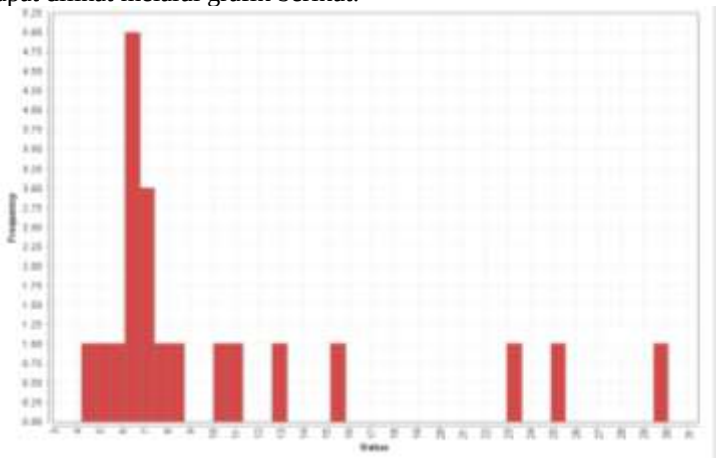
**Tabel 3. 3** Perbandingan Manual dan Rapid Miner

| Variael X | Variael Y | Hasil Rapidminer |
|---|---|---|
| 36 | 30 | 30.0 |
| 3 | 6 | 5.65 |

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 4, No 2, September 2023, Hal 299 - 305
ISSN 2722-0524 (media online)
DOI 10.47065/ bit.v3i1. 733
https://journal.fkpt.org/index.php/BIT

| | | |
|---|---|---|
| 2 | 5 | 4.91 |
| 27 | 23 | 23.4 |
| 4 | 6 | 6.40 |
| 4 | 6 | 6.40 |
| 5 | 7 | 7.14 |
| 29 | 25 | 24.9 |
| 4 | 6 | 6.40 |
| 5 | 7 | 7.14 |
| 4 | 6 | 6.40 |
| 1 | 4 | 4.17 |
| 5 | 7 | 7.14 |
| 7 | 9 | 8.62 |
| 9 | 10 | 10.10 |
| 4 | 6 | 6.40 |
| 16 | 15 | 15.28 |
| 13 | 13 | 13.06 |
| 10 | 11 | 10.84 |
| 6 | 8 | 7.88 |

Sedangkan untuk perbandingan nilai variabel Y aktual dari data *testing* (observasi) dengan nilai variabel Y prediksi pada aplikasi *rapidminer* dapat dilihat melalui grafik berikut.



**Gambar 3.** Grafik Perbandingan Nilai (Y) Observasi dengan (Y) Prediksi

Secara umum dapat dilihat melalui grafik, untuk nilai variabel Y aktual (observasi) ditandai dengan garis warna kuning, sedangkan nilai variabel Y prediksi dengan garis warna merah, dari data tersebut dapat dikatakan bahwa nilai prediksi memiliki rentang yang sama dengan nilai Y aktual (observasi). Grafik ini nantinya akan dievaluasi menggunakan metode RMSE untuk melihat seberapa besar nilai *error* tersebut. Selanjutnya evaluasi model ini adalah dengan mencari nilai *root mean squared error* atau RMSE, melalui aplikasi RapidMiner didapatkan nilai RMSE sebesar 0,273 dengan standar deviasi +/- 0,000.

Berdasarkan hasil pengujian yang telah dilakukan bahwa variabel atau atribut yang digunakan dalam penelitian ini variael X dan Y berpengaruh signifikan terhadap penelitian ini terbukti dengan menggunakan algoritma regresi linear mampu memberikan hasil yang baik dengan nilai *Root Mean Squared Error* 0.273 +/- 0.000. Hal ini dikarenakan adanya korelasi atau hubungan fungsional (sebab – akibat) antara variabel yang satu (dependen atau kriteria) dengan variabel yang lain *(independen atau predictor)*. Proses pengujian ini dilakukan untuk mengidentifikasi data klaim dengan algoritma regresi linear.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 4, No 2, September 2023, Hal 299 - 305
ISSN 2722-0524 (media online)
DOI 10.47065/ bit.v3i1. 733
https://journal.fkpt.org/index.php/BIT

# 4. KESIMPULAN

Pada penelitian ini dengan memanfaatkan beberapa data kasus klaim indemnity pada assuansi inhealth melalui pendekatan metode prediksi dan dapat diterapkan dalam menganalisis data guna melakukan prediksi data assuransi dimasa mendatang berdasar dari tingkat kebutuhan. Proses prediksi algoritma Regresi Linear sederhana dapat diimplementasikan dimana hasilnya juga menunjukkan sebuah wawasan baru bagi kebutuhan prediksi terhadap data klaim .Pengujian dengan menggunakan rapidminer menghasilkan performa yang relevan dengan skenario yang dimodelkan. Model persamaan *Regresi Linear* sederhana setelah dibandingkan hasil perhitungan secara manual dan juga dengan aplikasi *Rapid Miner* secara umum menunjukkan data yang sama. Nilai RMSE juga didapat saat melakukan evaluasi performa model yang diterapkan, dengan nilai RMSE sebesar 0,273 dengan standar deviasi +- 0,0.

**REFERENCES**

[1] M. Sholeh, Suraya, And D. Andayati, "Machine Linear Untuk Analisis Regresi Linier Biaya Asuransi Kesehatan Dengan Menggunakan Python Jupyter Notebook," *J. Edukasi Dan Penelit. Inform.*, Vol. 8, No. 1, Pp. 20–27, 2022.

[2] A. C. Nugraha And M. I. Irawan, "Komparasi Deteksi Kecurangan Pada Data Klaim Asuransi Pelayanan Kesehatan Menggunakan Metode Support Vector Machine (Svm) Dan Extreme Gradient Boosting (Xgboost)," *J. Sains Dan Seni Its*, Vol. 12, No. 1, 2023, Doi: 10.12962/J23373520.V12i1.107032.

[3] N. Riskya, S. Yuliana, T. Informatika, U. N. Jadid, And J. Timur, "Penerapan Data Mining Untuk Prediksi Perilaku Pelanggan Menggunakan Multiple," Vol. 11, No. 3, 2023.

[4] O. R. Kristyaningrum, A. Setiawan, And L. R. Sasongko, "Analisis Pengelompokan K-Means Untuk Data Bivariat Laju Kunjungan Dan Rasio Rujukan ( Studi Data Pada Seluruh Fasilitas Kesehatan Tingkat I Di Bpjs Kesehatan Surakarta )," Vol. 2, No. 1, 2018.

[5] J. Informasi, J. Wandana, And S. Defit, "Klasterisasi Data Rekam Medis Pasien Pengguna Layanan Bpjs Kesehatan Menggunakan Metode K-Means," Vol. 2, Pp. 4–9, 2020, Doi: 10.37034/Jidt.V2i4.73.

[6] D. Fithaloka, M. A. Budiman, And D. Rachmawati, "Perbandingan Algoritma Greedy Dan Hill Climbing Untuk Menentukan Fasilitas Kesehatan Tingkat Pertama ( Fktp ) Terdekat Bagi Peserta Bpjs Kesehatan," Vol. 1, No. 2, Pp. 13–24, 2017.

[7] S. S. Helma, R. R. R, And E. Normala, "Clustering Pada Data Fasilitas Pelayanan Kesehatan Kota Pekanbaru Menggunakan Algoritma K - Means," No. November, Pp. 131–137, 2019.

[8] G. R. P, A. P. Windarto, E. Irawan, And W. Saputra, "Penerapan Data Mining Menggunakan Algoritma C4 . 5 Dalam Mengukur Tingkat Kepuasan Pasien Bpjs," Vol. 2, Pp. 376–385, 2020.

[9] G. Widi N. Dicky Nofriansyah, *Algoritma Data Mining Dan Pengujian*. Yogyakarta: Cv Budi Utama, 2015.

[10] R. Fahlevi, "Penerapan Genetic Modified K-Nearest Neighbor Dalam Klasifikasi Penerima Beras Sejahtera," 2020.

[11] J. Tamaela, E. Sediyono, And A. Setiawan, "Implementasi Metode Association Rule Untuk Menganalisis Data Twitter Tentang Badan Penyelenggara Jaminan Sosial Dengan Algoritma Frequent Pattern - Growth," Vol. 01, Pp. 25–33, 2018.

[12] H. W. Herwanto, T. Widiyaningtyas, And P. Indriana, "Penerapan Algoritme Linear Regression Untuk Prediksi Hasil Panen Tanaman Padi," *J. Nas. Tek. Elektro Dan Teknol. Inf.*, Vol. 8, No. 4, P. 364, 2019, Doi: 10.22146/Jnteti.V8i4.537.

[13] S. Haryati, A. Sudarsono, And E. (2015) Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5," *J. Media Infotama*, Vol. 11, No. 2, Pp. 130–138, 2015.

[14] S. Sitalasari And M. Algoritma, "Penerapan Data Mining Dalam Menentukan Kelayakan Pemberian Kartu Indonesia Sehat ( Kis ) Pada Kecamatan," Vol. 1, Pp. 323–330, 2019.

[15] R. Mean *Et Al.*, "Evaluasi Dan Validasi Evaluasi."