

PENERAPAN ALGORITMA REGRESI LINIER UNTUK PREDIKSI JUMLAH KLAIM PADA ASURANSI KESEHATAN

Fahrul Nurzaman

Teknik Informatika Universitas Persada Indonesia Y.A.I

Jl. Salemba Raya 7/9A Jakarta Pusat

email : fnurzaman@gmail.com

ABSTRACT

The planning of service improvement must be continuously done by the company to satisfy the customer so that the business can run continuously and can compete well. The support of data, information and knowledge becomes an important factor in the planning done for the sake of the Company's business continuity. One of the decisive factors in planning service improvements especially in the business of Health Insurance is the estimated number of cases of claims to be handled. To support the data, information and knowledge related to the estimated number of claims cases, it is desirable to get predictions of the number of claims cases in each policy published in the next year. In this research, Data Mining process is made by using Linear Regression Method to get how many cases of claims that arise from the number of participants on each policy by dividing the three number of claims cases based on the total claims, types of services and benefits. The number of participants becomes the predictor variable while the number of claim cases becomes the response variable. From the calculation results obtained equation formula is $Y = - 63.19 2.34 X$, to get the prediction total number of claims total case, $Y = - 66.17 2.11 X$, to get predicted number of claims type of service cashless, $Y = - 2.97 0.23 X$, to get prediction number of claims type of reimbursement service, $Y = - 2.06 0.16 X$, to obtain prediction of number of inpatient benefit claim, and $Y = - 61.13 2.18 X$, to get predicted number of outpatient claim claim cases.

Keywords

Data mining Process, Linear Regression Method, predictor variables, response variabel

1. Pendahuluan

Pelayanan menjadi salah satu faktor yang harus diperhatikan dalam bisnis Asuransi Kesehatan. Pelayanan mencakup salah satu nya adalah penyelesaian penanganan

kasus klaim. Pelayanan dapat dikatakan baik jika penyelesaian penanganan kasus klaim dapat diselesaikan sesuai dengan Service Level Agreement (SLA) yang telah disepakati oleh Pihak Tertanggung dan Pihak Perusahaan Asuransi. Perencanaan peningkatan dan perbaikan layanan harus terus dilakukan pihak Asuransi demi memuaskan pelanggan sehingga bisnis dapat berjalan secara berkesinambungan dan dapat bersaing dengan baik. Dukungan data, informasi dan pengetahuan menjadi faktor penting dalam perencanaan yang dilakukan demi kelangsungan bisnis Asuransi Kesehatan. Salah satu faktor yang menentukan dalam membuat rencana peningkatan dan perbaikan layanan adalah perkiraan jumlah kasus klaim yang akan ditangani oleh Perusahaan Asuransi Kesehatan.

Dari latar belakang di atas terdapat permasalahan yaitu bagaimana mendapatkan prediksi jumlah kasus klaim yang akan ditangani. Prediksi Jumlah kasus klaim akan dibagi berdasarkan total keseluruhan klaim, jenis layanan dan jenis manfaat. Dengan mengetahui jumlah kasus klaim yang ditangani akan mempermudah perencanaan peningkatan dan perbaikan proses penyelesaian penanganan kasus klaim. Jumlah kasus klaim dilihat dari jumlah peserta yang ditanggung dan yang dilayani untuk setiap polis yang terbit. Dari banyak peserta tersebut maka ingin didapat berapa banyak kasus klaim yang muncul untuk setiap polis yang terbit dalam satu tahun kedepan dimana klaim tersebut harus diselesaikan dengan baik sesuai dengan Service Level Agreement (SLA) yang telah disepakati.

Untuk menyelesaikan permasalahan di atas maka dibuat proses *Data mining* untuk memprediksi jumlah kasus klaim dengan menggunakan Metode *Regresi Linear*. Metode *Regresi Linear* digunakan untuk mendapatkan berapa banyak kasus klaim yang muncul pada setiap polis yang terbit di tahun mendatang dari jumlah peserta yang dilayani. Konsep dasar metode Metode *Regresi Linear* adalah mencari apakah adanya korelasi atau tidak dari antar variabel. Analisis regresi lebih akurat dalam analisis korelasi karena tingkat perubahan suatu variabel

terhadap variabel lainnya dapat ditentukan. Jadi pada regresi, peramalan atau perkiraan nilai variabel terikat pada nilai variabel bebas lebih akurat pula. Regresi linier adalah regresi yang variabel bebasnya (variabel X) berpangkat paling tinggi satu. Untuk Regresi linier hanya melibatkan dua variabel (variabel X dan Y). Tujuan penelitian ini adalah mengembangkan model prediksi jumlah klaim berdasarkan perolehan jumlah peserta yang ditangani dengan menggunakan Metode Regresi Linier. Hasil dari penelitian adalah simulasi prediksi jumlah kasus klaim dalam bentuk penyajian data dan implementasi *database* menggunakan bahasa *Structure Query Language*. Prediksi jumlah kasus klaim terbagi atas total keseluruhan kasus klaim, jumlah kasus klaim berdasarkan jenis layanan dan jumlah kasus klaim berdasarkan manfaat rawat inap dan rawat jalan untuk setiap polis yang terbit dalam tahun yang akan diprediksi atau polis-polis yang akan diperpanjang untuk tahun berikutnya.

2. Landasan Teori

2.1. Prediksi

Prediksi pada dasarnya merupakan dugaan atau prediksi mengenai terjadinya suatu kejadian atau peristiwa di waktu yang akan datang. Prediksi bisa bersifat kualitatif (tidak berbentuk angka) maupun kuantitatif (berbentuk angka). Prediksi kualitatif sulit dilakukan untuk memperoleh hasil yang baik karena variabelnya sangat relatif sifatnya. Prediksi kuantitatif dibagi dua yaitu: prediksi tunggal (*point prediction*) dan prediksi selang (*interval prediction*). Prediksi tunggal terdiri dari satu nilai, sedangkan prediksi selang terdiri dari beberapa nilai, berupa suatu selang (*interval*) yang dibatasi oleh nilai batas bawah (prediksi batas bawah) dan batas atas (prediksi tinggi) [1].

Prediksi berfungsi untuk membuat suatu rencana kebutuhan (*demand*) yang harus dibuat yang dinyatakan dalam kuantitas (jumlah) sebagai fungsi dari waktu. Prediksi dilakukan dalam jangka panjang (*long term*). Prediksi yang berkaitan dengan pernyataan (1) *what will be demanded*, (2) *how many*, dan (3) *when it should be supplied* ? Prediksi sangat diperlukan dengan melakukan perbandingan antara kebutuhan yang diramalkan dengan yang sebenarnya.

2.2. Data Mining

Secara sederhana *data mining* adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar [4]. *Data mining* juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data [2]. *Data mining*, sering juga disebut

sebagai *Knowledge Discovery In Database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, histori untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar [6].

Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. *Data mining* berkaitan dengan bidang ilmu-ilmu lain, seperti *database system*, *data warehouse*, *statistic*, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu, *data mining* didukung oleh ilmu lain seperti *neural network*, pengenalan pola, *spatial data analysis*, *image database*, *signal processing* [7].

Data mining didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semi-otomatis [7]. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi.

Karakteristik *data mining* sebagai berikut

1. *Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
2. *Data mining* biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
3. *Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam Strategi [4].

Berdasarkan beberapa pengertian tersebut dapat ditarik kesimpulan bahwa *data mining* adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (*database*) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui. *Data mining* sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu *data mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan (*Artificial Intelligent*), *Machine Learning*, Statistik dan *Database*. Beberapa metode yang sering disebut-sebut dalam literature *data mining* antara lain *Clustering*, *Classification*, *Association Rules Mining*, *Neural Network*, *Generic Algorithm* dan lain-lain [2].

2.2.1. Tahap-tahap Data Mining

Sebagai suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantaraan knowledge base.

Tahap-tahap *data mining* ada 6 yaitu :

1. Pembersihan Data (Data Cleansing)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan.

Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa *data mining* yang dimiliki. Data yang tidak relevan itu juga lebih baik dibuang.

Pembersihan data juga akan mempengaruhi performansi dari teknik *data mining* karena data yang ditangani akan berkurang jumlah kompleksitasnya.

2. Integras Data (Data Integration)

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk *data mining* tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya.

3. Seleksi Data (Data Selection)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database.

4. Transformasi Data (Data Transformation)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numeric yang berlanut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

5. Proses Mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6. Evaluasi Pola (Pattern Evaluation)

Untuk mengidentifikasi pola-pola menarik kedalam knowledge based yang ditemukan. Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternative yang dapat diambil seperti menjadikannya

umpan balik untuk memperbaiki proses *data mining*, mencoba metode *data mining* lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

7. Presentasi Pengetahuan (Knowledge Presentation)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami *data mining*. Karenanya presentasi hasil *data mining* dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses *data mining*. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil Data Mining [7].

2.3 Metode Regresi Linier

Dalam Regresi Linear, data dimodelkan dalam bentuk grafik berbentuk garis *continues* dua dimensi. Oleh karena penggambaran data menggunakan dua dimensi, maka dibutuhkan variabel X dan Y [1]. Dalam regresi linear, variabel Y disebut sebagai *response variable* sedangkan variabel X disebut sebagai *predictor variable*. Kedua variabel diformulasikan secara statistik dengan rumus sebagai berikut :

$$Y = \alpha + \beta X \dots\dots\dots(1)$$

Nilai Y pada rumusan di atas dianggap sebagai nilai konstan, sedangkan nilai α dan β adalah nilai regression coefficient yang mempengaruhi penggambaran data dalam bentuk grafik dua dimensi. Nilai α dan β dapat dicari menggunakan metode least square yang berfungsi untuk meminimalkan nilai error antara data sebenarnya dan data hasil prediksi. Diberikan nilai sampel data S dengan titik-titik $(X_1, Y_1), (X_2, Y_2), \dots (X_3, Y_3)$, maka regression coefficient dapat dicari menggunakan rumus sebagai berikut: [1]

$$\beta = \frac{\sum_{i=1}^S (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^S (x_i - \bar{x})^2} \dots\dots\dots(2)$$

$$\alpha = \bar{y} - \beta \bar{x} \dots\dots\dots(3)$$

Dimana \bar{x} adalah rata-rata dari X_1, X_2, \dots, X_S dan \bar{y} adalah rata-rata dari Y_1, Y_2, \dots, Y_S .

3. Hasil Dan Pembahasan

3.1 Analisis Kebutuhan Data

Data yang dibutuhkan sebagai masukan dalam proses perhitungan Prediksi Jumlah kasus klaim dengan Metode Regresi Linier adalah data jumlah peserta yang terdaftar dan jumlah kasus klaim yang terjadi untuk setiap polis dalam 4 tahun terakhir. Jumlah kasus klaim terbagi atas tiga jenis kasus klaim yaitu kasus klaim secara menyeluruh, kasus klaim berdasarkan jenis layanan yaitu cashless dan reimburse, serta kasus klaim berdasarkan manfaat yaitu rawat inap dan rawat jalan.

Sumber data berasal dari Basis Data Transaksional atau Basis Data Operasional yang akan dimasukkan ke dalam Sistem *Datawarehouse* atau Basis Data *Staging*. Proses *Data mining* dan proses perhitungan prediksi akan menggunakan data yang terdapat pada Sistem *Datawarehouse* atau Basis Data *Staging* sebagai sumber data. Sumber data digunakan untuk mendapatkan jumlah peserta yang ditangani dan jumlah kasus klaim yang terjadi dimana kedua jumlah tersebut menjadi dua variabel dalam mendapatkan nilai persamaan regresinya. Di bawah ini adalah contoh Sumber data :

Tabel 1 Contoh Data Kepesertaan

MEMBER_ID	NOMOR_POLIS	TAHUN_LAYANAN
PST.14.001.0122D	PLS.14.002.0	2013
PST.13.005.0033A	PLS.13.013.0	2013
PST.13.009.0005D	PLS.13.007.0	2013
PST.13.001.0065A	PLS.13.011.0	2013
PST.14.001.0029A	PLS.14.002.0	2013
PST.13.009.0008D	PLS.13.007.0	2013
PST.14.001.0219A	PLS.14.002.0	2013
PST.13.026.0007E	PLS.13.006.0	2013
PST.13.001.0005D	PLS.13.005.0	2013
PST.14.001.0121A	PLS.14.002.0	2013
PST.14.001.0073A	PLS.14.002.0	2013
PST.13.001.0040B	PLS.13.011.0	2013

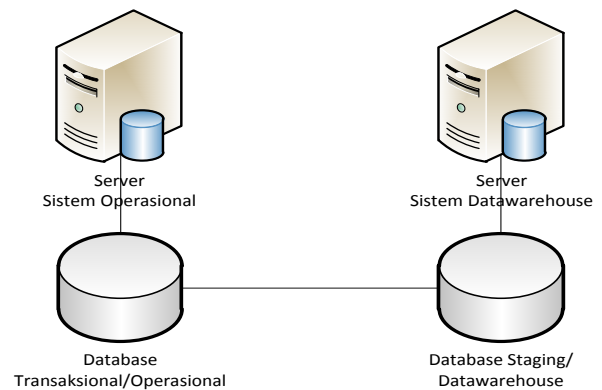
Tabel 2 Contoh Data Kasus Klaim

KLAIM_ID	MEMBER_ID	NOMOR_POLIS	JENIS	MANF_AAT	TAHUN_KEJADIAN
CLM.13.00276	PST.13.002.0014A	PLS.13.003.0	Reimburse	Rawat Jalan	2013
CLM.13.00363	PST.13.026.0095A	PLS.13.006.0	Reimburse	Rawat Inap	2013
CLM.14.02258	PST.13.026.0035E	PLS.13.006	Cashless	Rawat Inap	2013
CLM.14.00102	PST.13.006.0015B	PLS.13.014.0	Cashless	Rawat Jalan	2013
CLM.14.00015	PST.13.005.0008D	PLS.13.013	Cashless	Rawat Jalan	2013
CLM.13.00087	PST.13.008.0057A	PLS.13.002.0	Cashless	Rawat Jalan	2013
CLM.14.00107	PST.13.008.0059D	PLS.13.00	Cashless	Rawat	2013

KLAIM_ID	MEMBER_ID	NOMOR_POLIS	JENIS	MANF_AAT	TAHUN_KEJADIAN
		2.0		Jalan	
CLM.14.00369	PST.13.005.0024C	PLS.13.013.0	Cashless	Rawat Jalan	2013
CLM.13.00262	PST.13.005.0034D	PLS.13.013.0	Reimburse	Rawat Jalan	2013
CLM.13.00137	PST.13.008.0011B	PLS.13.002.0	Cashless	Rawat Jalan	2013
CLM.13.00230	PST.13.008.0011B	PLS.13.002.0	Cashless	Rawat Jalan	2013

3.2 Analisis Sistem Arsitektur Basis Data

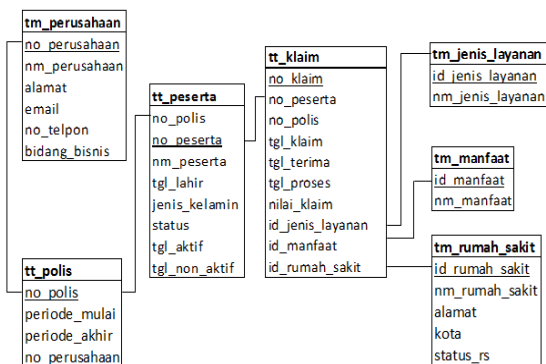
Sumber Data yang digunakan untuk proses *data mining* dan proses perhitungan prediksi adalah Sistem *Datawarehouse* atau Basis Data *Staging*. Data yang terdapat pada Basis data transaksional atau Basis Data operasional secara otomatis diambil sesuai dengan data yang dibutuhkan menggunakan implementasi kode *SQL (Select Query Language)* dimasukkan ke dalam Sistem *Datawarehouse* atau Basis Data *Staging*.



Gambar 1 Arsitektur Basis Data *Staging*

3.3 Analisis Skema Basis Data Transaksional

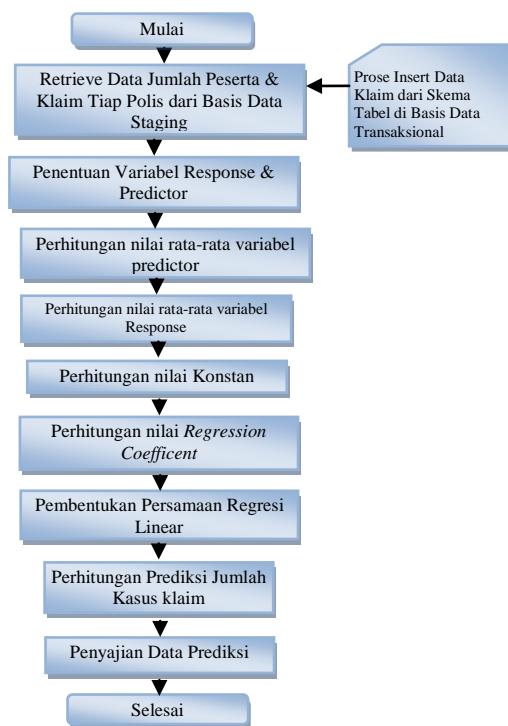
Data yang dibutuhkan untuk menjadi data masukan proses *data mining* dan proses perhitungan berasal dari database transaksional yang melibatkan banyak tabel. Tabel-tabel tersebut diantaranya Tabel Master Perusahaan, Tabel transaksi polis, tabel transaksi kepesertaan, Tabel master jenis manfaat klaim, Tabel master jenis layanan, Tabel master rumah sakit dan Tabel transaksi klaim. Berikut di bawah ini adalah Skema Basis Data Transaksional untuk transaksi data kepesertaan dan klaim :



Gambar 2 Skema Basis Data Operasional Klaim

3.4 Analisis Perhitungan Data

Proses Perhitungan data dimulai dari pengambilan data dari tabel peserta dan tabel klaim yang terdapat pada Sistem *Datawarehouse* atau Basis Data *Staging*. Data tabel peserta dan tabel klaim didapat dari proses insert data klaim dari Basis data transaksional ke dalam Sistem *Datawarehouse* atau Basis Data *Staging*. Dari kedua tabel tersebut, dilakukan proses pengambilan data jumlah peserta dan jumlah kasus klaim untuk setiap polis dalam 4 tahun terakhir. Jumlah kasus klaim dibagi menjadi tiga kategori total kasus klaim secara keseluruhan, kasus klaim berdasarkan jenis layanan dan kasus klaim berdasarkan manfaat. Di bawah ini adalah Alur Proses Perhitungan Prediksi :



Gambar 3 Alur Proses Perhitungan Prediksi dengan Metode Regresi Linear

3.5 Perancangan Basis Data

Untuk kebutuhan Basis Data *Staging* maka dibuat Tabel yang menampung data dari Basis Data Transaksional dan Tabel yang menampung data dari hasil proses perhitungan.

Tabel *ts_master_data_polis* digunakan untuk menampung data yang berasal dari basis data transaksional yang berisi data yang dijadikan sebagai variabel untuk mendapatkan persamaan regresi linear. Di bawah ini adalah struktur tabel dari tabel *ts_master_data_polis* :

nama kolom	type data
<u>no_polis</u>	char(10)
tahun_polis	nchar(4)
jumlah_peserta	int
jumlah_kasus	int
jumlah_kasus_cashless	int
jumlah_kasus_reimburse	int
jumlah_kasus_ri	int
jumlah_kasus_rj	int

Gambar 4 Struktur Data Tabel *ts_master_data_polis*

Tabel *ts_nilai_regresi* digunakan untuk menampung data hasil proses perhitungan jumlah dan rata-rata dari variabel *response* dan *predictor* serta nilai koefisien regresi. Di bawah ini adalah struktur tabel dari tabel *ts_nilai_regresi* :

nama kolom	type data
item	varchar(20)
n_data	int
jumlah_xi	decimal(18,2)
jumlah_yi	decimal(18,2)
jumlah_yixi	decimal(18,2)
jumlah_xi_xi	decimal(18,2)
rata2_xi	decimal(18,2)
rata_yi	decimal(18,2)
nilai_B0	decimal(18,2)
nilai_B1	decimal(18,2)

Gambar 5 Struktur Data Tabel *ts_nilai_regresi*

Tabel *ts_data_prediksi* digunakan untuk menampung data hasil perhitungan prediksi jumlah kasus klaim dimana data polis dan jumlah peserta didapat dari proses insert data dari basis data transaksional ke Basis data *staging*. Di bawah ini adalah struktur data dari tabel *ts_data_prediksi* :

nama kolom	tipe data
no_polis	char(10)
tahun_polis	nchar(4)
jumlah_peserta	int
prediksi_kasus	int
prediksi_kasus_cashless	int
prediksi_kasus_reimburse	int
prediksi_kasus_ri	int
prediksi_kasus_rj	int

Gambar 6 Struktur Data Tabel ts_data_prediksi

3.6 Implementasi Data mining

Nilai variabel *Response* dan *predictor* didapatkan dari data Jumlah peserta dan klaim dari tiap polis dalam 4 tahun terakhir dimana data diambil dari Basis Data transaksi dimasukkan ke dalam basis data staging, jumlah peserta ditentukan sebagai nilai variabel *response* dan jumlah klaim ditentukan sebagai nilai variabel *predictor*.

Berikut di bawah ini implementasi kode SQL nya :

```
insert into
[serverstaging].[dbstaging].[dbo].[ts_master_data
_polis]
select
no_polis,year(periode_mulai) as tahun_polis,
(select count(*) from tt_peserta where
no_polis=tt_polis.no_polis) as jumlah_peserta,
(select count(*) from tt_klaim where
no_polis=tt_polis.nopolis) as jumlah_kasus,
(select count(*) from tt_klaim where
no_polis=tt_polis.nopolis and
id_jenis_layanan=(select id_jenis_layanan from
tm_jenis_layanan where
nm_jenis_layanan='CASHLESS')) as
jumlah_klaim_cashless,
(select count(*) from tt_klaim where
no_polis=tt_polis.nopolis and
id_jenis_layanan=(select id_jenis_layanan from
tm_jenis_layanan where
nm_jenis_layanan='REIMBURSE')) as
jumlah_klaim_reimburse,
(select count(*) from tt_klaim where
no_polis=tt_polis.nopolis and id_manfaat=(select
id_manfaat from tm_manfaat where
nm_manfaat='RAWAT INAP')) as jumlah_klaim_RI,
(select count(*) from tt_klaim where
no_polis=tt_polis.nopolis and id_manfaat=(select
id_manfaat from tm_manfaat where
nm_manfaat='RAWAT JALAN')) as jumlah_klaim_RJ
from tt_polis
```

Dari hasil proses kode SQL tersebut didapat data nilai variabel *Response* dan *predictor* sebagai berikut :

Tabel 3 Contoh Data variabel *Response* dan *predictor*

no_polis	tahun	Jml peserta (Var X)	Jml Total (Var Y0)	Jml Casless (Var Y1)	Jml Reimburse (Var Y2)	Jml RI (Var Y3)	Jml RJ (Var Y4)
PLS.13.014.0	2013	67	107	93	14	7	100
PLS.13.010.0	2013	53	91	52	39	15	76
PLS.14.013.0	2013	8,936	11,062	9,038	2,024	1,237	9,825
PLS.13.002.0	2013	196	568	560	8	34	534
PLS.13.009.0	2013	59	103	103	0	6	97
PLS.13.016.0	2013	45	102	81	21	4	98
PLS.14.001.0	2013	70	204	167	37	12	192
PLS.13.013.0	2013	121	233	217	16	9	224
PLS.13.005.0	2013	118	103	49	54	12	91
PLS.13.008.0	2013	69	14	6	8	14	0
PLS.13.006.0	2013	607	32	23	9	32	0
PLS.13.011.0	2013	285	390	339	51	17	373
PLS.13.015.0	2013	25	45	41	4	8	37
PLS.13.007.0	2013	187	19	6	13	19	0
PLS.14.002.0	2013	679	75	45	30	75	0
PLS.13.004.0	2013	41	94	56	38	10	84
PLS.13.003.0	2013	28	65	46	19	8	57
PLS.13.017.0	2013	1,341	10,184	9,532	652	404	9,780

Data yang berada pada tabel ts_master_data_polis dijadikan sebagai data untuk mendapatkan nilai koefisien regresinya. Dilakukan perhitungan Nilai koefisien regresi dengan membagi menjadi tiga persamaan Regresi Linier yaitu untuk persamaan mendapatkan jumlah total kasus klaim secara keseluruhan, persamaan mendapatkan jumlah kasus klaim berdasarkan jenis layanan yaitu cashless dan reimburse serta persamaan mendapatkan jumlah kasus klaim berdasarkan manfaat yaitu rawat inap dan rawat jalan.

Penentuan variabel *Response* dan *Predictor* adalah sebagai berikut untuk mendapatkan prediksi jumlah total keseluruhan kasus klaim maka variabel *response* (y) nya adalah total jumlah kasus dan variabel *predictor* nya (x) adalah total jumlah peserta. Untuk mendapatkan prediksi jumlah kasus klaim dengan layanan cashless maka variabel *response* (y) nya adalah total jumlah kasus layanan cashless dan variabel *predictor* nya (x) adalah total jumlah peserta. Untuk mendapatkan prediksi jumlah kasus klaim dengan layanan reimburse maka variabel *response* (y) nya adalah total jumlah kasus layanan reimburse dan variabel *predictor* nya (x) adalah total jumlah peserta. Untuk mendapatkan prediksi jumlah kasus klaim dengan manfaat rawat inap maka variabel *response* (y) nya adalah total

jumlah kasus manfaat rawat inap dan variabel predictor nya (x) adalah total jumlah peserta. Untuk mendapatkan prediksi jumlah kasus klaim dengan manfaat rawat jalan maka variabel response (y) nya adalah total jumlah kasus manfaat rawat jalan dan variabel predictor nya (x) adalah total jumlah peserta

Untuk mendapatkan nilai koefisien untuk setiap persamaannya maka dibuat implementasi kode query. Di bawah ini adalah kode untuk membuat view untuk perhitungan nilai koefisien :

```
CREATE view [dbo].[vw_nilai_regresi] as
select
n_data, jumlah_x, jumlah_y_total,
jumlah_y_cashless, jumlah_y_reimburse, jumlah_y_ri,
jumlah_y_rj,
jumlah_xy_total, jumlah_xy_cashless, jumlah_xy_reimburse,
jumlah_xy_ri, jumlah_xy_rj, jumlah_xx,
rata2_x, rata2_y_total, rata2_y_cashless, rata2_y_reimburse,
rata2_y_ri, rata2_y_rj,
f1_total, f1_cashless, f1_reimburse, f1_ri, f1_rj, f2,
f1_total/f2 as B1_total, f1_cashless/f2 as B1_cashless,
f1_reimburse/f2 as B1_reimburse,
f1_ri/f2 as B1_ri, f1_rj/f2 as B1_rj,
rata2_y_total-(rata2_x*(f1_total/f2)) as B0_total,
rata2_y_cashless-(rata2_x*(f1_cashless/f2)) as B0_cashless,
rata2_y_reimburse-(rata2_x*(f1_reimburse/f2)) as B0_reimburse,
rata2_y_ri-(rata2_x*(f1_ri/f2)) as B0_ri,
rata2_y_rj-(rata2_x*(f1_rj/f2)) as B0_rj
from (
select
count(*) as n_data,
sum(jumlah_peserta) as jumlah_x,
sum(jumlah_kasus) as jumlah_y_total,
sum(jumlah_kasus_cashless) as jumlah_y_cashless,
sum(jumlah_kasus_reimburse) as jumlah_y_reimburse,
sum(jumlah_kasus_ri) as jumlah_y_ri,
sum(jumlah_kasus_rj) as jumlah_y_rj,
sum(jumlah_peserta*jumlah_kasus) as jumlah_xy_total,
sum(jumlah_peserta*jumlah_kasus_cashless) as jumlah_xy_cashless,
sum(jumlah_peserta*jumlah_kasus_reimburse) as jumlah_xy_reimburse,
sum(jumlah_peserta*jumlah_kasus_ri) as jumlah_xy_ri,
sum(jumlah_peserta*jumlah_kasus_rj) as jumlah_xy_rj,
sum(jumlah_peserta*jumlah_peserta) as jumlah_xx,
avg(jumlah_peserta) as rata2_x,
avg(jumlah_kasus) as rata2_y_total,
avg(jumlah_kasus_cashless) as rata2_y_cashless,
avg(jumlah_kasus_reimburse) as rata2_y_reimburse,
avg(jumlah_kasus_ri) as rata2_y_ri,
avg(jumlah_kasus_rj) as rata2_y_rj,
sum(jumlah_peserta*jumlah_kasus)-
((sum(jumlah_kasus)*sum(jumlah_peserta))/count(*))
as f1_total,
sum(jumlah_peserta*jumlah_kasus_cashless)-
((sum(jumlah_kasus_cashless)*sum(jumlah_peserta))/count(*)) as f1_cashless,
```

```
sum(jumlah_peserta*jumlah_kasus_reimburse)-
((sum(jumlah_kasus_reimburse)*sum(jumlah_peserta))/count(*)) as f1_reimburse,
sum(jumlah_peserta*jumlah_kasus_ri)-
((sum(jumlah_kasus_ri)*sum(jumlah_peserta))/count(*)) as f1_ri,
sum(jumlah_peserta*jumlah_kasus_rj)-
((sum(jumlah_kasus_rj)*sum(jumlah_peserta))/count(*)) as f1_rj,
sum(jumlah_peserta*jumlah_peserta)-
((sum(jumlah_peserta)*sum(jumlah_peserta))/count(*)) as f2
from ts_master_data_polis
) as tblA
```

Di bawah ini adalah kode untuk memasukkan hasil perhitungan nilai koefisien ke dalam tabel ts_nilai_regresi:

```
declare @i integer
declare @item varchar(15)
declare @n_data int
declare @jumlah_xi int
declare @jumlah_yi int
declare @jumlah_yixi int
declare @jumlah_xi_xi int
declare @rata2_xi decimal(18, 2)
declare @rata2_yi decimal(18, 2)
declare @nilai_B0 decimal(18, 2)
declare @nilai_B1 decimal(18, 2)

SET @i=0

select @n_data = n_data from vw_nilai_regresi
select @jumlah_xi = jumlah_x from vw_nilai_regresi
select @jumlah_yi = jumlah_y_total from vw_nilai_regresi
select @jumlah_yixi = jumlah_xy_total from vw_nilai_regresi
select @rata2_xi = rata2_x from vw_nilai_regresi

WHILE @i < 5
BEGIN
if @i=0
begin
set @item='TOTAL'
select @jumlah_yi=jumlah_y_total from vw_nilai_regresi
select @jumlah_yixi=jumlah_xy_total from vw_nilai_regresi
select @rata2_yi=rata2_y_total from vw_nilai_regresi
select @nilai_B0=B0_total from vw_nilai_regresi
select @nilai_B1=B1_total from vw_nilai_regresi
end
if @i=1
begin
set @item='CASHLESS'
select @jumlah_yi=jumlah_y_cashless from vw_nilai_regresi
select @jumlah_yixi=jumlah_xy_cashless from vw_nilai_regresi
select @rata2_yi=rata2_y_cashless from vw_nilai_regresi
select @nilai_B0=B0_cashless from vw_nilai_regresi
select @nilai_B1=B1_cashless from vw_nilai_regresi
```

```

end
if @i=2
begin
    set @item='REIMBURSE'
    select @jumlah_yi=jumlah_y_reimburse
from vw_nilai_regresi
    select @jumlah_yixi=jumlah_xy_reimburse
from vw_nilai_regresi
    select @rata2_yi=rata2_y_reimburse from
vw_nilai_regresi
    select @nilai_B0=B0_reimburse from
vw_nilai_regresi
    select @nilai_B1=B1_reimburse from
vw_nilai_regresi
end
if @i=3
begin set
    @item='RAWAT INAP'
    select @jumlah_yi=jumlah_y_ri from
vw_nilai_regresi
    select @jumlah_yixi=jumlah_xy_ri from
vw_nilai_regresi
    select @rata2_yi=rata2_y_ri from
vw_nilai_regresi
    select @nilai_B0=B0_ri from
vw_nilai_regresi
    select @nilai_B1=B1_ri from
vw_nilai_regresi
end
if @i=4
begin set
    @item='RAWAT JALAN'
    select @jumlah_yi=jumlah_y_rj from
vw_nilai_regresi
    select @jumlah_yixi=jumlah_xy_rj from
vw_nilai_regresi
    select @rata2_yi=rata2_y_rj from
vw_nilai_regresi
    select @nilai_B0=B0_rj from
vw_nilai_regresi
    select @nilai_B1=B1_rj from
vw_nilai_regresi
end
insert into ts_nilai_regresi values
(@item,@n_data,@jumlah_xi,@jumlah_yi,@jumlah_yixi,
@jumlah_xi_xi,@rata2_xi,@rata2_yi,@nilai_B0,@nilai_B1)
SET @i=@i+1
END

```

Dari hasil kode query tersebut didapat data semua nilai koefisien regresi nya sebagai berikut :

Tabel 4 Data hasil perhitungan persamaan Regresi

Var Y	jumlah_yi	jumlah_yixi	rata2_yi	nilai_B0	nilai_B1
TOTAL (Y0)	203,606	878,210,903	1,143.85	-63.19	2.34
CASHLESS (Y1)	182,134	791,769,888	1,023.22	-66.17	2.11
REIMBURSE (Y2)	21,472	86,441,015	120.63	2.97	0.23
RAWAT INAP (Y3)	14,795	62,193,932	83.12	-2.06	0.16
RAWAT JALAN (Y4)	188,811	816,016,971	1,060.74	-61.13	2.18

Data nilai regresi tersebut menjadi dasar untuk perhitungan prediksi jumlah klaim untuk polis-polis yang terbit pada tahun 2017. Berikut dibawah ini adalah persamaan regresi nya :

Untuk mencari prediksi jumlah kasus klaim untuk Total keseluruhan dengan menggunakan rumus persamaan regresi yaitu $Y0 = -63.19 + 2.34 X$.

Untuk mencari prediksi jumlah kasus klaim untuk jenis layanan cashless dengan menggunakan rumus persamaan regresi yaitu $Y1 = -66.17 + 2.11 X$.

Untuk mencari prediksi jumlah kasus klaim untuk jenis layanan reimburse dengan menggunakan rumus persamaan regresi yaitu $Y2 = -2.97 + 0.23 X$.

Untuk mencari prediksi jumlah kasus klaim untuk manfaat rawat inap dengan menggunakan rumus persamaan regresi yaitu $Y3 = -2.06 + 0.16 X$.

Untuk mencari prediksi jumlah kasus klaim untuk manfaat rawat jalan dengan menggunakan rumus persamaan regresi yaitu $Y4 = -61.13 + 2.18 X$.

Berikut di bawah ini adalah hasil perhitungan nilai Koefisien Korelasi dan Koefisien Determinasi nya untuk menyatakan derajat hubungan antara dua variabel.

Tabel 5 Hasil Perhitungan Koefisien Korelasi dan Determinasi

Variabel Response (Y)	Koefisien Korelasi (R)	Koefisien Determinasi(R2)
total_klaim (Y0)	0.76	0.57
cashless (Y1)	0.74	0.55
reimburse (Y2)	0.74	0.55
rawat_inap (Y3)	0.84	0.71
rawat_jalan (Y4)	0.75	0.56

Berikut di bawah ini adalah perhitungan dari Standar Error Estimate untuk mengetahui kekeliruan penyimpangan data terhadap regresi :

Tabel 6 Hasil Perhitungan Standar Error Estimate

Variabel Response (Y)	Standar Error Estimate
total_klaim (Y0)	2,763.81
cashless (Y1)	2,614.23
reimburse (Y2)	283.80
rawat_inap (Y3)	145.17
rawat_jalan (Y4)	2,638.94

Berikut ini adalah perhitungan Pengujian Koefisien Regresi dimana menggunakan nilai tingkat signifikansi nya adalah 5% :

Tabel 7 Hasil Perhitungan Pengujian Koefisien Regresi

Variabel Response (Y)	t tabel	t hitung	Keputusan
total_klaim (Y0)	1.97	23.58	nilai t-hitung = 23.58 > t-tabel = 1.97 sehingga Ho ditolak dan Ha diterima.
cashless (Y1)	1.97	21.92	nilai t-hitung = 21.92 > t-tabel = 1.97 sehingga Ho ditolak dan Ha diterima.
reimburse (Y2)	1.97	21.74	nilai t-hitung = 21.74 > t-tabel = 1.97 sehingga Ho ditolak dan Ha diterima.
rawat_inap (Y3)	1.97	38.31	nilai t-hitung = 38.31 > t-tabel = 1.97 sehingga Ho ditolak dan Ha diterima.
rawat_jalan (Y4)	1.97	22.61	nilai t-hitung = 22.61 > t-tabel = 1.97 sehingga Ho ditolak dan Ha diterima.

Berikut di bawah ini implementasi kode query untuk menghitung prediksi jumlah klaim nya :

```

select no_polis,
tahun_polis, jumlah_peserta,
nilai_B0_TOTAL + (nilai_B1_TOTAL*jumlah_peserta)
as prediksi_kasus_TOTAL,
nilai_B0_CASHLESS +
(nilai_B1_CASHLESS*jumlah_peserta) as
prediksi_kasus_CASHLESS,
nilai_B0_REIMBURSE +
(nilai_B1_REIMBURSE*jumlah_peserta) as
prediksi_kasus_REIMBURSE,
nilai_B0_RI + (nilai_B1_RI*jumlah_peserta) as
prediksi_kasus_RI,
nilai_B0_RJ + (nilai_B1_RJ*jumlah_peserta) as
prediksi_kasus_RJ
from (
select
no_polis, year(periode_mulai) as tahun_polis,
(select count(*) from tt_peserta where
no_polis=tt.polis.no_polis) as jumlah_peserta,
(select nilai_B0 from ts_nilai_regresi where
item='TOTAL' ) as nilai_B0_TOTAL,
(select nilai_B0 from ts_nilai_regresi where
item='CASHLESS' ) as nilai_B0_CASHLESS,
(select nilai_B0 from ts_nilai_regresi where
item='REIMBURSE' ) as nilai_B0_REIMBURSE,
(select nilai_B0 from ts_nilai_regresi where
item='RAWAT INAP' ) as nilai_B0_RI,
(select nilai_B0 from ts_nilai_regresi where
item='RAWAT JALAN' ) as nilai_B0_RJ,
(select nilai_B1 from ts_nilai_regresi where
item='TOTAL' ) as nilai_B1_TOTAL,
(select nilai_B1 from ts_nilai_regresi where
item='CASHLESS' ) as nilai_B1_CASHLESS,
(select nilai_B1 from ts_nilai_regresi where
item='REIMBURSE' ) as nilai_B1_REIMBURSE,
(select nilai_B1 from ts_nilai_regresi where
item='RAWAT INAP' ) as nilai_B1_RI,
(select nilai_B1 from ts_nilai_regresi where
item='RAWAT JALAN' ) as nilai_B1_RJ
from [servercore].[dbcore].[dbo].[tt_polis]
where year(periode_mulai)=2017
) as tbseleksi

```

Dari hasil kode query tersebut didapat data prediksi jumlah kasus klaim untuk polis-polis tahun 2017 adalah sebagai berikut :

Tabel 8 Hasil Perhitungan Prediksi Jumlah Kasus Klaim

no polis	tahun	peserta (Var X)	Total (Var Y0)	Cashless (Var Y1)	Reimburse (Var Y2)	RI (Var Y3)	RJ (Var Y4)
PLS.17.012.0	2017	18,083	42,251	38,089	4,162	2,891	39,360
PLS.16.040.0	2017	4,399	10,230	9,216	1,015	702	9,529
PLS.17.023.0	2017	4,341	10,095	9,093	1,001	693	9,402
PLS.14.043.3	2017	3,785	8,794	7,920	874	604	8,190
PLS.14.044.3	2017	2,814	6,522	5,871	650	448	6,073
PLS.17.008.0	2017	1,486	3,414	3,069	345	236	3,178
PLS.17.014.0	2017	1,210	2,768	2,487	281	192	2,577
PLS.15.004.2	2017	1,166	2,665	2,394	271	185	2,481
PLS.17.028.0	2017	915	2,078	1,864	213	144	1,934
PLS.17.011.0	2017	614	1,374	1,229	144	96	1,277
PLS.14.030.3	2017	545	1,212	1,084	128	85	1,127
PLS.17.027.0	2017	530	1,177	1,052	125	83	1,094
PLS.17.021.0	2017	528	1,172	1,048	124	82	1,090
PLS.17.032.0	2017	499	1,104	987	118	78	1,027
PLS.16.021.1	2017	453	997	890	107	70	926
PLS.17.015.0	2017	403	880	784	96	62	817
PLS.17.040.0	2017	378	821	731	90	58	763
PLS.17.001.0	2017	320	686	609	77	49	636
PLS.17.039.0	2017	309	660	586	74	47	612
PLS.16.027.1	2017	274	578	512	66	42	536
PLS.14.035.3	2017	267	562	497	64	41	521
PLS.17.004.0	2017	240	498	440	58	36	462
PLS.17.002.0	2017	238	494	436	58	36	458
PLS.17.016.0	2017	230	475	419	56	35	440
PLS.15.007.2	2017	177	351	307	44	26	325
PLS.17.013.0	2017	169	332	290	42	25	307
PLS.16.028.1	2017	161	314	274	40	24	290
PLS.17.026.0	2017	151	290	252	38	22	268
PLS.17.041.0	2017	147	281	244	37	21	259
PLS.16.031.1	2017	144	274	238	36	21	253
PLS.16.005.1	2017	117	211	181	30	17	194
PLS.17.005.0	2017	117	211	181	30	17	194
PLS.17.019.0	2017	115	206	176	29	16	190
PLS.17.038.0	2017	113	201	172	29	16	185
PLS.16.020.1	2017	106	185	157	27	15	170
PLS.15.003.2	2017	99	168	143	26	14	155
PLS.17.010.0	2017	96	161	136	25	13	148
PLS.17.022.0	2017	96	161	136	25	13	148
PLS.17.034.0	2017	93	154	130	24	13	142
PLS.17.042.0	2017	92	152	128	24	13	139
PLS.16.018.3	2017	84	133	111	22	11	122
PLS.16.018.4	2017	84	133	111	22	11	122
PLS.15.001.3	2017	77	117	96	21	10	107
PLS.17.003.0	2017	68	96	77	19	9	87
PLS.17.017.0	2017	64	87	69	18	8	78
PLS.17.009.0	2017	64	87	69	18	8	78
PLS.17.024.0	2017	59	75	58	17	7	67
PLS.17.025.0	2017	54	63	48	15	7	57
PLS.17.007.0	2017	51	56	41	15	6	50
PLS.14.032.3	2017	49	51	37	14	6	46
PLS.16.041.0	2017	47	47	33	14	5	41
PLS.17.020.0	2017	46	44	31	14	5	39
PLS.17.006.0	2017	39	28	16	12	4	24
PLS.16.023.1	2017	37	23	12	11	4	20

4. Kesimpulan

Dari hasil dan pembahasan Penerapan Algoritma Regresi Linier untuk prediksi jumlah kasus klaim pada Asuransi Kesehatan, dapat diambil kesimpulan sebagai berikut :

1. Sumber Data yang digunakan dalam proses pengolahan Data mining berasal dari Sistem Datawarehouse atau Basis Data Staging dimana datanya didapat dari Basis Data Transaksional yaitu Data jumlah Peserta dan Jumlah klaim yang terjadi pada Tiap Polis untuk 3 Tahun terakhir.
2. Berdasarkan hasil dari perhitungan koefisien korelasi menunjukkan bahwa hubungan positif antara jumlah peserta dengan jumlah klaim, jika dilihat dari nilai korelasi hubungan kedua variabel tersebut secara keseluruhan termasuk kategori tinggi. Dengan demikian berarti jumlah peserta memiliki hubungan tinggi terhadap jumlah klaim.
3. Berdasarkan hasil dari perhitungan koefisien determinasi menunjukkan kemampuan variabel jumlah peserta dalam mempengaruhi variabel jumlah klaim secara keseluruhan cukup besar.
4. Berdasarkan hasil perhitungan standar error estimate persamaan regresi menunjukkan penyimpangan data terhadap garis regresi, atau bagaimana penyimpangan data yang menyebar disekitar garis regresi cukup kecil.
5. Berdasarkan hasil perhitungan pengujian koefisien Regresi menunjukkan bahwa prediksi Jumlah kasus klaim dapat dipengaruhi dari jumlah peserta tiap polis. Dimana Jumlah peserta menjadi variabel prediktor (x) dan jumlah kasus klaim menjadi variabel response (y).
6. Data perhitungan Prediksi Jumlah kasus klaim merupakan hasil akhir proses penerapan data mining dengan menggunakan Algoritma Regresi Linier.

Terdapat beberapa saran untuk pengembangan penelitian tersebut diantaranya :

1. Untuk perhitungan Prediksi Jumlah Klaim dapat ditambahkan variabel predictor seperti Besar nya Premi, Banyak nya manfaat yang diambil peserta, jumlah peserta berdasarkan jenis kelamin, atau jumlah peserta berdasarkan usia
2. Penyajian data dapat disajikan dalam bentuk grafik melalui Aplikasi Dashboard.
3. Menggunakan beberapa Metode untuk perbandingan dalam mendapatkan hasil prediksi jumlah klaim seperti menggunakan Metode *Naïve Baise* atau Metode *Decision Tree* dengan Algoritma ID3 atau yang lainnya.

REFERENSI

- [1] Douglas C Montgomery & Team, 2013, "Applied Statistics and Probability for Engineers Edisi ke 6", John Willey & Sons, Inc, New Jersey.
- [2] Kusri, Emha Taufiq Lutfi, 2009, "Algoritma Data Mining", Penerbit Andi, Yogyakarta.
- [3] Larose, Daniel T., 2005, "Discovering Knowledge in Data: An Introduction to Data Mining", John Willey & Sons, Inc, New Jersey.
- [4] McLeod, Raymond, 1995, "Sistem Informasi Manajemen", PT. Tema Baru, Klaten
- [5] Santosa, Budi, 2007, "Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis", Graha Ilmu, Yogyakarta.
- [6] Susanto, Sani, dan Suryadi, Dedy, 2010, "Pengantar Data Mining", Penerbit Andi Yogyakarta, Yogyakarta.
- [7] Turban, Efraim., Aronso, Jay., Liang Peng Ting, 2005, "Decision Support System and Intelligence System (Versi Bahasa Indonesia), Edisi Ke-7", ANDI Offset, Yogyakarta.

Fahrul Nurzaman, memperoleh gelar ST dan MTI dari STT TELKOM Bandung dan Universitas Indonesia Jakarta tahun 2004 dan 2009. Saat ini sebagai Staf Pengajar program studi Teknik Informatika Universitas Persada Indonesia YAI.