

PEMODELAN DAN PROTOTYPE SISTEM INFORMASI UNTUK PREDIKSI PEMBAHARUAN POLIS ASURANSI MOBIL MENGGUNAKAN ALGORITMA C.45

William Frado Pattipeilohy^{1*}, Arief Wibowo², Dyah Retno Utari³

^{1,2,3} Universitas Budi Luhur

Jl. Cileduk Raya, Petukangan Utara, Jakarta Selatan, DKI Jakarta 12260

*Email: william_frado@yahoo.com

Abstrak

Teknologi data mining sangat berguna untuk membantu perusahaan asuransi menemukan informasi yang sangat penting dari gudang data mereka. Data mining mengeksplorasi basis data untuk menemukan pola-pola yang tersembunyi, meramalkan tren dan sifat-sifat perilaku bisnis yang sangat berguna untuk mendukung pengambilan keputusan, mencari informasi pemrediksi yang mungkin saja terlupakan oleh para pelaku bisnis karena terletak di luar ekspektasi mereka. Analisis yang diotomatisasi yang dilakukan oleh data mining menjawab pertanyaan-pertanyaan bisnis yang dengan cara tradisional memerlukan banyak waktu untuk menjawabnya. Hasil dari sistem adalah menggali informasi dari basis data pelanggan dan transaksi yang dilakukan pelanggan meliputi pembayaran premi asuransi dan maupun proses claim yang terjadi sehingga dapat memprediksi seberapa besar kemungkinan pelanggan tersebut akan melakukan renewal atas polis asuransi mereka. Pada penelitian ini, akan dilakukan analisis klasifikasi untuk menentukan keputusan pembaharuan (renewal) polis pelanggan dengan menggunakan data polis asuransi kendaraan roda empat. Algoritma yang akan digunakan adalah C4.5 untuk kemudian dapat menjadi aturan-aturan untuk menentukan keputusan renewal polis. Dengan algoritma C4.5 yang dievaluasi menggunakan cross validation $k=10$, menghasilkan confusion matrix dengan tingkat akurasi sebesar 82,14%. Evaluasi menggunakan data baru yang dilakukan dengan Prototipe Sistem Informasi Prediksi yang diimplementasi dari rule C4.5 menghasilkan nilai akurasi 86% dengan error rate 14%.

Kata kunci: Algoritma C4.5, Data Mining, Renewal Polis Asuransi

1. PENDAHULUAN

Persaingan dalam dunia bisnis asuransi menuntut terobosan dan strategi untuk tetap menjamin keberlangsungan bisnis. Salah satu aset utama perusahaan adalah data pelanggan dan riwayat polisnya yang tersedia dalam jumlah besar. Ketersediaan data tersebut menuntut akan adanya teknologi yang dapat memanfaatkan data tersebut untuk melakukan penguatan strategi bisnis. Prediksi terhadap minat pelanggan sangat penting bagi perusahaan asuransi, karena terhadap minat pelanggan untuk memperbaharui (renewal) terhadap polis asuransi mereka akan memberikan kontribusi bagi perusahaan untuk keberlangsungan usaha perusahaan asuransi.

Teknologi data mining dapat dimanfaatkan pada data pelanggan asuransi kendaraan dan data transaksi yang terjadi pada premi asuransi maupun data klaim. Dari data tersebut, akan diklasifikasi pelanggan-pelanggan yang potensial akan memperbaharui polis (renewal) atau memberhentikan polis yang dimiliki. Klasifikasi dapat diartikan sebagai sebuah proses menemukan suatu model atau fungsi yang menggambarkan dan membedakan kelas obyek data, dengan tujuan untuk menggunakan model yang dihasilkan dalam pembuatan prediksi terhadap kelas objek dimana kelas labelnya tidak diketahui. Algoritma yang dapat digunakan untuk klasifikasi antara lain C4.5 yang akan menghasilkan pola-pola dalam pengambilan keputusan para pelanggan dalam melakukan renewal atau pembaharuan polis asuransi. Dari uraian tersebut, maka penelitian ini menggunakan teknik klasifikasi dalam pendekatan Data Mining untuk memprediksi keputusan pelanggan asuransi kendaraan bermotor roda empat, dengan judul: "Pemodelan dan Prototipe Sistem Informasi Untuk Prediksi Pembaharuan Polis Asuransi Mobil Menggunakan Algoritma C.45". Hasil yang diharapkan dari penelitian ini yaitu suatu model prediksi untuk membantu bagian pemasaran untuk mengetahui lebih awal strategi untuk mempertahankan pelanggan yang sudah ada.

2. LANDASAN TEORI DAN KERANGKA PEMIKIRAN

2.1 Data Mining

Data mining merupakan suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan dalam bidang database. Data mining adalah proses yang menggunakan statistic, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Turban, 2005)

Menurut Larose (2005), data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dilakukan, yaitu:

1. Deskripsi, mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.
2. Estimasi, hampir sama dengan klasifikasi, namun variabel target estimasi lebih ke arah numerik dari pada ke arah kategori.
3. Prediksi, nilai dari hasil di masa mendatang akan dihasilkan
4. Klasifikasi, adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk mendapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui.
5. Pengklusteran (Clustering), merupakan pengelompokan *record*, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.
6. Asosiasi, adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja (Kusrini dan Luthfi, 2009).

Klasifikasi merupakan salah satu teknik dalam data mining. Klasifikasi (taksonomi) merupakan proses penempatan objek atau konsep tertentu ke dalam satu set kategori berdasarkan objek yang digunakan (Candraningsih dan Nurhadiyono, 2015). Salah satu teknik klasifikasi yang populer adalah menggunakan pohon keputusan.

2.2 Pohon Keputusan dan Algoritma C4.5

Pohon keputusan merupakan metode yang umum digunakan untuk melakukan klasifikasi pada data mining. Seperti yang telah dijelaskan sebelumnya, klasifikasi merupakan suatu teknik menemukan kumpulan pola atau fungsi yang mendeskripsikan serta memisahkan kelas data yang satu dengan yang lainnya untuk menyatakan objek tersebut masuk pada kategori tertentu dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Metode ini populer karena mampu melakukan klasifikasi sekaligus menunjukkan hubungan antar atribut. Banyak algoritma yang dapat digunakan untuk membangun suatu decision tree, salah satunya ialah algoritma C4.5 (Wiwit Supriyanti, 2016)

Algoritma C4.5 merupakan struktur pohon dimana terdapat simpul yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji dan setiap daun menggambarkan kelas. Algoritma ini memiliki kelebihan yaitu mudah dimengerti, fleksibel dan menarik karena dapat divisualisasikan dalam bentuk gambar (pohon keputusan). Algoritma C4.5 secara rekursif mengunjungi setiap simpul keputusan, memilih pembagian yang optimal, sampai tidak bisa dibagi lagi. Algoritma C4.5 menggunakan konsep information gain atau entropy reduction untuk memilih pembagian yang optimal. Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5, yaitu:

1. Menyiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung gain dari atribut, hitung dahulu nilai entropy. Entropy merupakan distribusi probabilitas dalam teori informasi dan diadopsi ke dalam algoritma C4.5 untuk mengukur tingkat homogenitas distribusi kelas dari sebuah himpunan (data set). Sebagai ilustrasi semakin tinggi tingkat entropy dari sebuah data set maka semakin homogen distribusi kelas pada data set tersebut. Untuk menghitung nilai entropy digunakan rumus:

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i$$

Keterangan:
 S = himpunan kasus
 n = jmlah partisi S
 pi = proporsi Si terhadap S

3. Kemudian hitung nilai gain menggunakan rumus:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Keterangan:
 S = himpunan kasus
 A = fitur
 n = jumlah partisi huruf A
 |Si| = proporsi Si terhadap S
 |S| = jumlah kasus dalam S

2.3 Evaluasi dan Validasi pada metode Klasifikasi Data Mining

Untuk menguji model pada penelitian ini, digunakan metode K-Fold Cross Validation dan Confusion Matrix. K-Fold Cross Validation merupakan salah satu dari variasi teknik pengujian *cross validation*. *k-fold cross validation* dilakukan dengan membagi *training set* dan *test set*. Keuntungan menggunakan *k-fold cross validation* dibandingkan dengan variasi *cross validation* seperti *repeated random sub-sampling validation* adalah semua data akan digunakan baik untuk data uji maupun untuk data latih. Sebagai ilustrasi dari *k-fold cross validation* dapat dilihat pada Gambar 1 (Payam et al, 2009).



Gambar 1. Ilustrasi k-fold cross validation

Dalam ilustrasi pada Gambar 1 ditunjukkan bahwa nilai fold adalah 6-fold yang akan dilakukan proses pengujian sejumlah nilai fold tersebut dengan fold ke 1 akan dijadikan data uji dan *fold (k-1)* dijadikan sebagai data latih. Dengan demikian masing-masing data sampel akan menjadi sebagai data latih dan data uji secara bergantian berdasarkan banyaknya fold yang ditentukan. Hal ini bertujuan untuk meminimalkan nilai akurasi yang dihasilkan oleh faktor kebetulan.

2.4 Confusion Matrix

Evaluasi dari kinerja model klasifikasi didasarkan pada banyaknya (*count*) *data test record* yang dapat diprediksi secara benar dan tidak benar oleh model. *Count* ini ditabulasikan dalam sebuah tabel yang dikenal sebagai *Confusion Matrix*. Dalam Weka *classifier* hasil klasifikasi yang diperoleh disertakan dengan beberapa alat ukur yang tersedia di dalamnya dan salah satunya juga *Confusion Matrix*. *Confusion matrix* adalah tabel yang dipergunakan sebagai alat ukur yang berguna untuk melakukan analisis seberapa baik pengklasifikasian benar dan salah dari prediksi yang dilakukan dalam kelas-kelas yang berbeda. Tabel ini dipergunakan untuk menentukan kinerja suatu model klasifikasi (Tan, 2005). Contoh *confusion matrix* dapat dilihat pada Tabel 1

Tabel 1. Confusion matrix untuk data tiga kelas

		Actual Class		
		C ₁	C ₂	C ₃
Predicted Class	C ₁	F ₁₁	F ₁₂	F ₁₃
	C ₂	F ₂₁	F ₂₂	F ₂₃
	C ₃	F ₃₁	F ₃₂	F ₃₃

Berdasarkan pada Tabel 1, maka untuk menghitung *Accuracy* dapat dipergunakan Persamaan

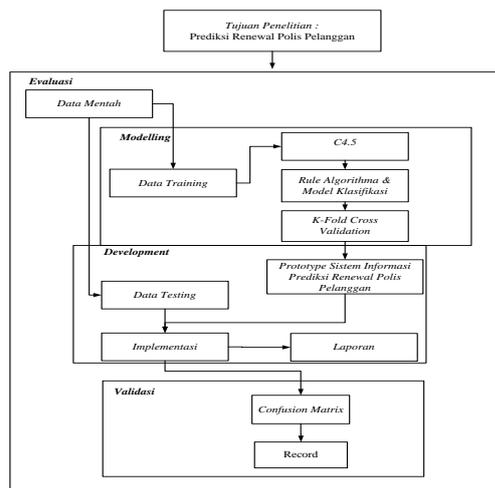
$$Accuracy = \frac{\sum \text{correct predictions}}{\sum \text{correct and incorrect prediction}} = \frac{F_{11} + F_{22} + F_{33}}{F_{11} + F_{12} + F_{13} + F_{21} + F_{22} + F_{23} + F_{31} + F_{32} + F_{33}}$$

Sementara untuk menghitung error rate dapat dipergunakan persamaan ;

$$Error\ rate = \frac{\sum \text{incorrect predictions}}{\sum \text{correct and incorrect prediction}} = \frac{F_{12} + F_{13} + F_{21} + F_{23} + F_{31} + F_{32}}{F_{11} + F_{12} + F_{13} + F_{21} + F_{22} + F_{23} + F_{31} + F_{32} + F_{33}}$$

2.5 Kerangka Konsep

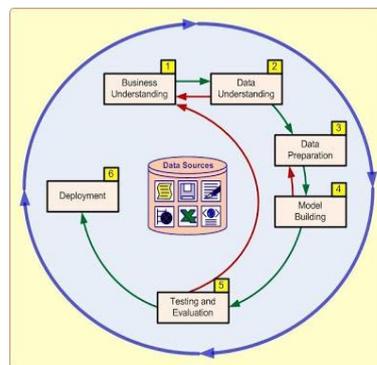
Dalam menyelesaikan penelitian ini, pola pikir yang digunakan untuk menyelesaikan rumusan masalah penelitian dapat dilihat pada gambar berikut:



Gambar 2. Kerangka Konsep

2.6 Desain Penelitian

Teknik analisis data menggunakan Data Kuantitatif, Analisa dilakukan melalui data polis menggunakan pengujian pada algoritma C4.5.



Gambar 3. Proses Cross Standard Industries for Data Mining (CRISP-DM.org)

Dalam penelitian ini menggunakan model CRISP-DM (Cross Standard Industries for Data Mining), yang terdiri dari 6 tahap (Maimon & Rokach, 2010). Adapun tahapan-tahapannya sebagai berikut:

1. Business/Research Understanding Phase

Ketika polis berada dalam periode tenggat waktu biasanya bagian marketing akan melakukan follow atas pelanggan tersebut agar bersedia memperpanjang atau melakukan *renewal* polis mereka. Hal ini dilakukan bagi setiap pelanggan tanpa mengkalsifikasi pelanggan yang bersedia melakukan *renewal* sehingga dirasakan kurang tepat sasaran dalam menentukan langka dan strategi bagi para pelanggan mereka.

2. Data Understanding Phase (Fase Pemahaman Data)

Dalam proses pengkategorian *renewal* polis, didapat data dari warehouse Production Asuransi Astra sebanyak 15.486.747 data polis, yang terdiri dari 11 atribut. Dimana 10 atribut prediktor dan 1 atribut hasil/target. Atribut-atribut yang menjadi parameter terlihat pada tabel III.2, yaitu:

Tabel 2. Data Atribut Polis

No	Atribut	Nilai	No	Atribut	Nilai
1	Year of Manufacturing	<=2010	5	Jenis Claim	<i>Partial Loss Accident</i>
		>2010			<i>Partial Loss Stolen</i>
2	Policy Period	1 tahun	6	Sum Insured	<i>No Claim</i>
		2 tahun			<=300 juta
		3 tahun			>300 juta - 1M
		4 tahun			Besar
3	Customer Type Description	5 tahun	7	Claim Amount	Kecil
		Personal			<i>No Claim</i>
		Company			8
4	Object Description	Truck	9	Claim Ratio	Rendah
		Sedan			Tinggi
		Multy Purpose Vehicle			Rendah
		Pick up			10
		Minibus	11	Renewal	>1x
					<i>No Claim</i>
					Ya
					Tidak

Berdasarkan data latih yang diperoleh, maka sebelum pemodelan dilakukan, diperlukan beberapa teknik pre-processing yang digunakan, yaitu:

a) Data validation

Untuk mengidentifikasi dan menghapus data yang ganjil (outlier/noise), data yang tidak konsistensi dan data yang tidak lengkap (missing value).

b) Data integration and transformation

Untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal.

c) Data size reduction and discretization

Untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informatif. Proses data training yang digunakan dalam penelitian ini, dilakukan seleksi atribut dan penghapusan data duplikasi menggunakan software Weka. Setelah dilakukan pengidentifikasian serta menghapus dan menggabungkan beberapa data yang lengkap maupun tidak lengkap juga dilakukan reduction and discretization untuk data training.

3. Modeling Phase (Fase Pemodelan)
Merupakan tahapan pemrosesan data training yang diklasifikasikan oleh model dan kemudian menghasilkan sejumlah aturan. Pada penelitian ini menggunakan algoritma yaitu C4.5 yang kemudian divalidasi dengan *K-Fold Cros-Validation*.
4. Evaluation Phase (Fase Evaluasi)
Melakukan pengujian terhadap model-model yang bertujuan untuk mendapatkan model yang paling akurat. Evaluasi dan validasi dilakukan dengan menggunakan *confusion matrix*.
5. Deployment Phase (Fase Penyebaran)
Pembentukan model selanjutnya melakukan analisa dan pengukuran pada tahap sebelumnya, pada tahap ini diterapkan model atau *rules* terbaik menjadi prototipe sistem informasi dan selanjutnya dapat digunakan untuk mengevaluasi data baru.

2.7 Teknik Pengujian Model

Metode pengujian yang dilakukan pada tahap awal adalah dengan menggunakan aplikasi yang dirancang berdasarkan rule algoritma yang memiliki tingkat akurasi tertinggi. Tahap dua dengan menggunakan *K-Fold Cros-Validation* dengan nilai $k=10$ serta metode sampling menggunakan metode Slovin

2.8 Hipotesis

Hipotesis pada penelitian ini adalah: Diduga metode klasifikasi dalam data mining dengan Algoritma C4.5 merupakan algoritma yang tepat dan sesuai untuk melakukan prediksi keputusan pelanggan yang melakukan pembaharuan atau *renewal* atas polis asuransi mereka.

3. HASIL DAN PEMBAHASAN

3.1 Pengelompokan dan Analisis

Dalam Dalam penelitian ini dilakukan proses validasi untuk menemukan dan mengkonversi data agar dapat digunakan dalam algoritma data mining dan memperoleh akurasi dan performa yang baik. Dalam dataset yang akan digunakan ini, validasi data yang digunakan adalah dengan menghapus data yang tidak lengkap atau kosong yang tidak memiliki nilai (*null*).

Setelah itu dilakukan seleksi atribut untuk dijadikan dataset yang dibutuhkan sesuai dengan algoritma yang digunakan dalam proses menganalisa *renewal* polis sehingga data yang diambil menjadi semakin sedikit.

Untuk informasi *renewal* polis, didapat data dari database production Asuransi Astra, sebanyak 190 data yang terdiri dari 11 atribut. Dimana 10 atribut predictor dan 1 atribut hasil. Penelitian ini bertujuan untuk memprediksi apakah pelanggan akan melakukan *renewal* polis atau tidak.

3.1.2 Perhitungan Algoritma C4.5

Langkah-langkah dalam perhitungan dengan menggunakan algoritma C4.5 dengan memakai data training yang berjumlah 140 data. Setelah menentukan atribut sebagai akar kemudian dilakukan perhitungan nilai Entropy, setelah didapatkan nilai Entropy kemudian menghitung nilai gain untuk setiap atribut, lalu pilih nilai gain tertinggi. Gain Atribut yang terpilih adalah yang mempunyai nilai gain tertinggi dari nilai gain-gain yang lainnya, kemudian dijadikan akar dari pohon. Perhitungan untuk menghasilkan nilai entropy dan gain menggunakan formula berikut ini:

$$\begin{aligned} \text{Entropy}(T) &= (-103/140 * \text{LOG}_2(103/140)) + (-37/140 * \text{LOG}_2(37/140)) \\ &= 0.833144954 \end{aligned}$$

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i)$$

$$\text{Gain Tertinggi} = 0.057996806$$

$$\text{Gain (total, SumInsured)} = 0.833144954 - ((123/140 * 0.882282445) + (17/140 * 0))$$

Hasil pemodelan membentuk Tree dengan susunan sebagai berikut:

```
SumInsured = >300juta - 1M: YA (17.0)
SumInsured = <=300juta
| YearOfManufacturing = <=2010
```

```

| | LossRatio = Rendah
| | | PeriodPolicy <= 4
| | | | PeriodPolicy <= 3: YA (2.0)
| | | | | PeriodPolicy > 3
| | | | | | ObjectDescription = TRUCK: NO (1.0)
| | | | | | ObjectDescription = MULTY PURPOSE VEHICLE: YA (16.0/4.0)
| | | | | | ObjectDescription = SEDAN: YA (13.0/5.0)
| | | | | | ObjectDescription = MINIBUS: YA (1.0)
| | | | | | ObjectDescription = PICK UP: YA (1.0)
| | | | PeriodPolicy > 4: YA (6.0)
| | LossRatio = Tinggi: YA (33.0/2.0)
| YearOfManufacturing = >2010
| | CustomerTypeDescription = Company: YA (7.0)
| | CustomerTypeDescription = Personal
| | | ObjectDescription = TRUCK: NO (0.0)
| | | ObjectDescription = MULTY PURPOSE VEHICLE
| | | | ClaimAmount = No Claim: NO (24.0/7.0)
| | | | ClaimAmount = Kecil
| | | | | PeriodPolicy <= 4: YA (3.0)
| | | | | PeriodPolicy > 4: NO (9.0/2.0)
| | | | ClaimAmount = Besar: YA (1.0)
| | | ObjectDescription = SEDAN: YA (6.0/1.0)
| | | ObjectDescription = MINIBUS: NO (0.0)
| | | ObjectDescription = PICK UP: NO (0.0)
    
```

3.2 Evaluasi dan Validasi

Dari proses pemodelan yang telah dilakukan dengan algoritma C4.5, dilakukan pengujian tingkat akurasi dengan menggunakan *confussion matrix*. Berikut ini adalah perhitungan nilai *confussion matrix* terhadap algoritma C4.5 dengan 11 atribut dan 140 record yang menghasilkan tingkat akurasi 82.14%.

```

=== Confusion Matrix ===

  a  b  <-- classified as
91 12 | a = YA
13 24 | b = NO
    
```

Gambar 4. Confussion Matrix Algoritma C4.5

- Baris pertama “91” dan “12” menunjukkan bahwa dari total 140 data, ada 91 polis yang melakukan renewal (YA) diklasifikasikan dalam kategori renewal (YA) dan terdapat 12 polis yang tidak melakukan renewal (NO) tetapi diklasifikasikan dalam kategori melakukan renewal (YA).
- Baris kedua “13” dan “24” menunjukkan bahwa dari total 140 data, ada 13 polis yang tidak melakukan renewal (NO) diklasifikasikan dalam kategori renewal (YA) dan terdapat 24 polis yang tidak melakukan renewal (NO) dan diklasifikasikan dalam kategori tidak melakukan renewal (NO).
- Dengan demikian maka jumlah prediksi yang dilakukan dengan hasil benar pada total 140 data, adalah $91 + 24 = 115$ dan jumlah prediksi yang dilakukan dengan hasil salah $13 + 12 = 25$. Maka akurasi model ini menjadi sebesar $(115/140) = 0,82$ atau 82,14%

3.3 Cross Validation

Hasil yang didapatkan dari pengolahan cross validation untuk algoritma C4.5 dengan menggunakan data training sebesar 82.14% yang dapat dilihat pada Gambar 5.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      115      82.1429 %
Incorrectly Classified Instances    25      17.8571 %
Kappa statistic                    0.5368
Mean absolute error                0.2711
Root mean squared error            0.3988
Relative absolute error            69.3514 %
Root relative squared error        90.3767 %
Total Number of Instances         140

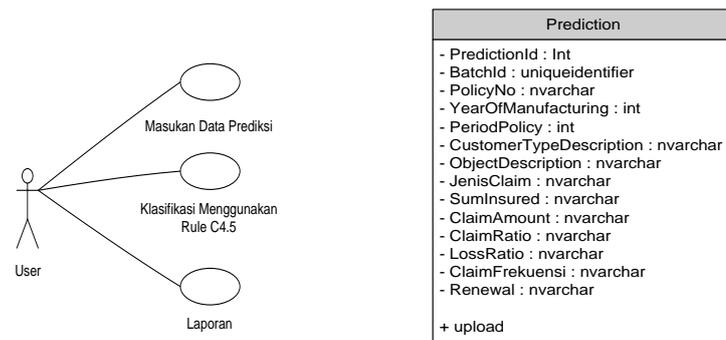
=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.883    0.351    0.875    0.883    0.879    0.537    0.773    0.895    YA
0.649    0.117    0.667    0.649    0.658    0.537    0.773    0.512    NO
Weighted Avg.   0.821    0.289    0.820    0.821    0.821    0.537    0.773    0.794
    
```

Gambar 5. Nilai Akurasi dari hasil Validasi k-fold

3.4 Implementasi Prototipe Sistem Informasi Prediksi

Pada tahapan *requirement analysis* didapat kebutuhan dari pengguna yang sudah terpetakan sesuai dengan tingkat kepentingan masing-masing, langkah selanjutnya adalah masuk kedalam tahapan logika perancangannya. Dalam logika perancangan ini menggunakan perancangan *Object Oriented Design* (OOD) yang menentukan aktor, membuat *use case* dan *class diagram*.



Gambar 6. Use Case dan Class Diagram

3.5 Evaluasi Model Prediksi Renewal Polis Pelanggan

Salah satu hal yang penting untuk menentukan kesalahan-kesalahan atau kekurangan-kekurangan pada prediksi renewal yang dikembangkan adalah melakukan pengujian (evaluasi). Pengujian dilakukan sebanyak lima kali ($k=5$) yang melibatkan 50 data selain data training dan menggunakan metode *Confusion Matrix*, yaitu tabel yang dipergunakan sebagai alat ukur yang berguna untuk melakukan analisis seberapa baik pengklasifikasian benar dan salah dari prediksi yang dilakukan.

Dari lima percobaan yang telah dilakukan maka didapatkan summary yaitu untuk tingkat akurasi sebesar $(90+90+80+90+80)/5=430/5=86\%$, dan untuk tingkat eror rate sebesar $(10+10+20+10+20)/5 = 70/5 = 14\%$.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan maka bagian penutup dari penelitian ini peneliti mengambil kesimpulan dan saran sebagai berikut:

4.1 Kesimpulan

Berdasarkan hasil pengujian, prediksi renewal polis yang dikembangkan dapat berjalan dengan baik dengan tingkat akurasi mencapai 86% dengan error rate 14%.

4.2 Saran

1. Perlu diadakan pelatihan yang dapat meningkatkan pengetahuan dalam penggunaan software sehingga berguna dan dapat menjadi solusi kedepannya.
2. Dari segi infrastruktur perlu dipersiapkan semaksimal mungkin untuk mendapatkan hasil dari implementasi prediksi renewal polis dan benar-benar sesuai dengan fungsinya.
3. Untuk pengembangan lebih lanjut dapat menambahkan jumlah sampel dan variable atau menggunakan feature selection untuk mendapatkan variable predictor yang lebih baik sehingga mempengaruhi hasil prediksi yang lebih akurat.

DAFTAR PUSTAKA

- Anik Andrian (2012), Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Drop Out, Jakarta.
- Dharwiyanti, Sri, (2003), Pengantar *Unified Modelling Language (UML)*, Ilmu Komputer
- Candraningsih, Nurhadiyono, B., (2015) Klasifikasi Calon Peserta Lomba Cerdas Cermat Siswa Menggunakan Algoritma C4.5, Universitas Dian Nuswantoro Semarang
- Gorunescu, F. (ed). (2010), *Data Mining Concepts, Model and Techniques Springer*.
- Han, Jiawei, dan Kamber, M. (ed). (2012), *Data Mining Concepts and Techniques*. Waltham, USA: Morgan Kaufmann, Elsevier.
- Kusrini, dan Luthfi E. T. (2009), Algoritma Data Mining, Yogyakarta.
- Larose, D. T. (2005), *Discovering Knowledge In Data*.
- Liliana Swastina (2013), Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa, Banjarmasin.
- Maimon, O., & Rokach, L. (2010), *Data Mining and Knowledge Discovery Handbook*, Israel:Springer.
- Payam, R., Tang L., and Liu, H. (ed). (2009), *In Encyclopedia of Database Systems*, Springer, USA.
- Sandy Kurniawan, Taufiq Hidayat, (2007), Penerapan Data Mining Dengan Metode Interpolasi Untuk Memprediksi Minat Konsumen Asuransi (Studi Kasus Asuransi Metlife), Yogyakarta.
- Supriyanti, W., Kusrini, Amborowati, A., (2016) Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes Untuk Ketepatan Pemilihan Konsentrasi Mahasiswa, Jurnal Informa, Surakarta
- Tan P,-N, (2006), *Introduction to Data Mining*, Addison-Wesley Longman PublishingCo.Inc., Boston, MA, USA.
- Turban, Efraim, Arison, Jay.E, dan Liang, T. P. (2005), *Decision Support System and Intelligent System*.
- Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., Steinberg, D. (2008), *Top 10 algorithms in data mining*. (ed.)Knowledge and Information, London:Springer-Verlag London Limited 2007.