
ANALISIS DENGAN METODE KLASIFIKASI MENGGUNAKAN *DECISION TREE* UNTUK PREDIKSI PENENTUAN RESIKO KREDIT PADA BANK BUKOPIN BATAM

Intan Utma Sari

Dosen Program Studi Sistem Informasi Universitas Putera Batam

ABSTRAK

Bank Bukopin Batam merupakan salah satu Bank yang menghimpun dana dan penyalur dana. Ada beberapa fasilitas dalam penyaluran dana ke nasabah yang di miliki Bank Bukopin. Salah satunya Kredit Pemilikan Rumah (KPR), selama ini pihak Bank memberikan KPR ke nasabah masih menggunakan prediksi resiko secara manual dalam memberikan kredit kepada nasabah yang banyak menyita waktu dan tenaga apalagi pada saat laporan nasabah dianalisa lebih lanjut oleh pihak Bank. Salah satu teknik yang dapat membantu pihak Bank dalam memprediksi Penentuan resiko kredit adalah teknik *Decision Tree* yang merupakan bagian dari teknik *Data Mining* untuk mengambil suatu keputusan dalam bentuk pohon. Dengan teknik *Decision Tree* diharapkan dapat membantu pihak bank agar lebih cepat dan mudah dalam memprediksi data dan menarik suatu kesimpulan dari data yang ada. Salah satu cara memprediksi data tersebut dengan menggunakan *software Dtree*. Pada *software* ini data yang digunakan hanya bisa dalam bentuk format “csv (*comma delimited*)”, jika tidak menggunakan format “csv (*comma delimited*)” maka data tersebut tidak bisa diproses oleh *software Dtree* dan selanjutnya jika format excel yang telah dirubah ke format “csv (*comma delimited*)”, maka akan dapat dilakukan proses analisa. *Dtree* dapat menghasilkan pohon keputusan, salah satu nya yaitu hasil keputusan resiko dari jumlah kredit pemilikan rumah berdasarkan jumlah nasabah.

Kata Kunci : Kredit, csv.

1. PENDAHULUAN

Saat ini sistem komputerisasi tidak hanya terbatas pada perusahaan bisnis, namun juga sampai pada perusahaan yang bergerak di bidang perbankan. Salah satu fungsi dari informasi hasil pengolahan data bagi perusahaan yang bergerak di bidang perbankan (dalam hal ini pada bank Bukopin) adalah dapat menghasilkan informasi yang bersifat klasifikasi yang terkait dengan perkreditan rumah. Untuk mendapatkan informasi tersebut dapat digunakan salah satu metode *Klasifikasi* yang ada di *data mining*.

Oleh sebab itu, peneliti mengajukan suatu penelitian untuk pemetaan alternatif pemecahan masalah tentang resiko kredit yang ada pada bank Bukopin batam dengan menggunakan aplikasi *data mining* yang telah ada (*DTREG*) agar data nasabah dapat diolah lebih lanjut dan menghasilkan informasi yang berguna untuk pengambilan keputusan oleh pihak Bank.

Metode *klasifikasi* yang akan digunakan adalah *decision tree*. Dengan menggunakan *decision tree* maka *user* dapat dengan mudah menganalisa suatu atribut. Misalnya *user* ingin menganalisa apa resiko kredit ‘A’ pada pengkreditan rumah yang akan

ditentukan dan dipilih sebagai atribut yang ingin dilihat keterkaitannya dengan resiko kredit 'A' adalah jumlah nasabah yang melakukan KPR, Jenis jangka waktu kredit, penyebab terjadinya resiko kredit, keterlambatan per bulan, resiko kredit yang dapat diambil. Dalam hal ini maka dapat ditentukan bahwa *field* yang ingin diidentifikasi adalah *field* resiko kredit, dan *field* yang akan dilihat keterkaitannya berdasarkan *field* penyebab terjadinya resiko kredit dan berupa *field* keterlambatan per bulan. Output yang akan dihasilkan adalah *tree* dengan *root*-nya adalah resiko kredit dan *branch node*-nya adalah penyebab terjadinya resiko kredit dan keterlambatan per bulan. Dari *tree* yang ada maka *user* dapat menyimpulkan 'A' bagaimana resiko kredit nasabah yang melakukan KPR.

Adapun perumusan masalah dari penelitian ini adalah mengklasifikasi diagnosa resiko kredit perumahan bagi nasabah, berdasarkan penyebab resiko kredit dan keterlambatan per bulan.

Tujuan dari penelitian ini adalah untuk mengklasifikasikan penentuan resiko kredit berdasarkan jumlah nasabah lebih kurang 46 sample nasabah dengan menggunakan *decision tree*. Yang mana melalui *decision tree* dapat dihasilkan informasi atau *knowledge* yang dapat digunakan untuk mendukung pengambilan keputusan dalam mengklasifikasikan penentuan resiko kredit yang diambil.

2. LANDASAN TEORI

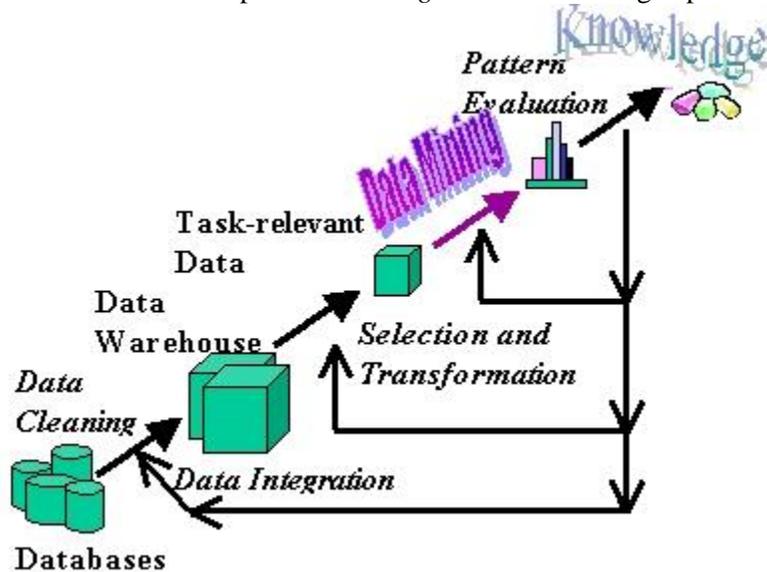
Metodologi tradisional untuk menganalisis data yang ada, tidak dapat menangani data dalam jumlah besar. Sementara para pelaku bisnis memiliki kebutuhan-kebutuhan untuk memanfaatkan gudang data yang sudah dimiliki, para peneliti melihat peluang untuk melahirkan sebuah teknologi baru yang menjawab kebutuhan ini, yaitu *Data mining*. Teknologi ini sekarang sudah ada dan diaplikasikan oleh perusahaan-perusahaan untuk memecahkan berbagai permasalahan bisnis.

Alasan utama mengapa *data mining* diperlukan adalah karena adanya sejumlah besar data yang dapat digunakan untuk menghasilkan informasi dan *knowledge* yang berguna (Han & Kamber, 2001). Informasi dan *knowledge* yang didapat tersebut dapat digunakan pada banyak bidang, mulai dari manajemen bisnis, kontrol produksi, kesehatan, dan lain-lain.

Secara sederhana, *data mining* dapat diartikan sebagai proses mengekstrak atau menggali *knowledge* yang ada pada sekumpulan data (Han & Kamber, 2001). Banyak orang yang setuju bahwa *data mining* adalah sinonim dari *Knowledge Discovery in Database*, atau yang biasa disebut *KDD*. Dari sudut pandang yang lain, *data mining* dianggap sebagai suatu langkah yang penting di dalam proses *KDD*. *KDD* terdiri dari langkah-langkah sebagai berikut (Han & Kamber, 2001):

1. *Data cleaning*: proses menghapus data yang tidak konsisten dan *noise*.
2. *Data ontegration*: penggabung *multiple data source*.
3. *Data selection*: pengambilan data yang relevan dari *database*.
4. *Data transformation*: proses dimana data ditransformasikan menjadi bentuk yang sesuai untuk diproses dalam *data mining*.
5. *Data mining*: suatu proses yang penting dengan melibatkan metode untuk menghasilkan suatu pola data.

6. *Pattern evaluation*: proses untuk menguji kebenaran dari pola data yang mewakili *knowledge* yang ada didalam data itu sendiri.
7. *Knowledge presentation*: proses visualisasi dan teknik penyajian *knowledge* digunakan untuk menampilkan *knowledge* hasil dari mining kepada *user*.



Gambar 1: Proses *Knowledge Discovery in Database (KDD)*

Decision tree merupakan salah satu metode klasifikasi pada *Text Mining*. Klasifikasi adalah proses menemukan kumpulan pola atau fungsi-fungsi yang mendeskripsikan dan memisahkan kelas data satu dengan lainnya, untuk dapat digunakan untuk memprediksi data yang belum memiliki kelas data tertentu (Jianwei Han, 2001).

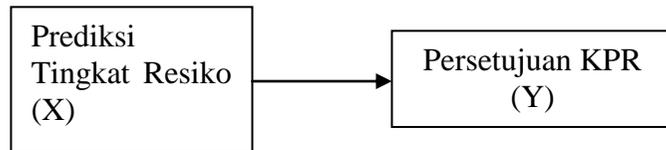
Decision tree adalah sebuah struktur pohon, dimana setiap *node* pohon merepresentasikan atribut yang telah diuji, setiap cabang merupakan suatu pembagian hasil uji, dan *node* daun (*leaf*) merepresentasikan kelompok kelas tertentu. Level *node* teratas dari sebuah *Decision tree* adalah *node* akar (*root*) yang biasanya berupa atribut yang paling memiliki pengaruh terbesar pada suatu kelas tertentu. Pada umumnya *Decision tree* melakukan strategi pencarian secara *top-down* untuk solusinya. Pada proses mengklasifikasi data yang tidak diketahui, nilai atribut akan diuji dengan cara melacak jalur dari *node* akar (*root*) sampai *node* akhir (daun) dan kemudian akan diprediksi kelas yang dimiliki oleh suatu data baru tertentu.

Aplikasi pendukung *Decision tree* yang digunakan adalah DTREG. DTREG adalah analisis program statistik yang kuat yang menghasilkan klasifikasi dan pohon keputusan bahwa model regresi data dan dapat digunakan untuk memprediksi nilai. Single-pohon, pohon *Tree Boots* dan Model Keputusan dapat dibuat.

DTREG adalah alat yang ideal untuk bisnis pemodelan, perbankan, dan medis data dengan variabel kategori seperti jumlah nasabah, jangka waktu, jaminan, resiko kredit, dan solusi yang dapat diambil. Pohon Keputusan ini logis, model yang

jelas yang dapat dimengerti dengan mudah oleh orang-orang yang tidak cenderung matematis.

Dalam penelitian ini kerangka pemikiran dapat digambarkan sebagai berikut:



Gambar 2: Kerangka Pemikiran

Berdasarkan latar belakang dan rumusan masalah maka dapat dikemukakan hipotesa yaitu, dengan mengklasifikasikan jangka waktu di Bank Bukopin Batam, diharapkan dapat membantu pihak Bank menentukan resiko kredit yang melakukan KPR berdasarkan penyebab terjadinya resiko kredit dan keterlambatan per bulan.

3. METODE PENELITIAN

Jenis penelitian ini adalah penelitian deskriptif kuantitatif. Menurut Arikunto (2007:29), penelitian deskriptif kuantitatif memiliki kejelasan unsur yang dirinci sejak awal, langkah penelitian yang sistematis menggunakan sampel yang hasil penelitiannya diberlakukan untuk populasi, memiliki hipotesis jika perlu, memiliki disain jelas dengan langkah-langkah penelitian dan hasil yang diharapkan, memerlukan pengumpulan data yang dapat mewakili serta ada analisis data yang dilakukan setelah semua data terkumpul.

Dalam penelitian kuantitatif, populasi diartikan sebagai wilayah generalisasi yang terdiri atas: obyek/subyek yang mempunyai kualitas dan karakteristik tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian ditarik kesimpulannya (Sugiyono, 2013:117). Yang merupakan populasi di dalam penelitian ini adalah pengujung data KPR Bank Bukopin Batam.

Didalam penelitian ini teknik sampling yang digunakan adalah *Nonprobability Sampling* yaitu teknik pengambilan sampel yang tidak memberi peluang/kesempatan sama bagi setiap unsur atau anggota populasi untuk dipilih menjadi sampel. Sedangkan penentuan pengambilan jumlah sampel (responden) dilakukan dengan menggunakan *Sampling Insidental/aksidental* yaitu teknik penentuan sampel berdasarkan kebetulan, yaitu siapa saja yang secara kebetulan/incidental bertemu dengan peneliti dapat digunakan sebagai sampel, bila dipandang orang yang kebetulan ditemui itu cocok sebagai sumber data. Dalam hal ini sampel (responden) adalah data KPR Bank Bukopin Batam.

Adapun metode yang digunakan dalam mengumpulkan data-data adalah sebagai berikut:

1. Riset Kepustakaan (*Library Research*)
Penelitian ini dilakukan, yakni dengan membaca buku-buku dan literatur-literatur yang berhubungan dengan penelitian ini.
2. Riset Perangkat Lunak (*Software Research*)

Penelitian ini dilakukan dengan menggunakan alat bantu komputer berupa *software data mining* DTREG untuk mengolah data dengan teknik *decision tree*.

Decision tree adalah metode prediksi yang digunakan pada klasifikasi, *clustering*, dan prediksi tugas. *Decision tree* menggunakan teknik “membagi dan menaklukkan” untuk membagi ruang pencarian masalah menjadi himpunan masalah. *Decision tree* merupakan salah satu teknik yang dapat digunakan untuk melakukan klasifikasi terhadap sekumpulan objek atau *record*. Teknik ini terdiri dari kumpulan *decision node*, dihubungkan oleh cabang, bergerak ke bawah dari root node sampai berakhir di *leaf node*.

4. HASIL PEMBAHASAN.

Dalam Membangun sebuah pohon keputusan, kami menggunakan Algoritma ID3 atau Algoritma C45, berikut merupakan penentuan resiko kredit dengan cara manualnya. Untuk memudahkan penjelasan mengenai algoritma C4.5 berikut ini disertakan contoh kasus yang dituangkan dalam tabel dibawah ini.

Tabel 1: Contoh data yang digunakan

<u>Jumlah Nasabah</u>	<u>jenis Jangka waktu KPR</u>	<u>penyebab terjadi Resiko Kredit</u>	<u>Keterlambatan/bln</u>	<u>Resiko Kredit</u>
1	<u>panjang</u>	<u>macet</u>	6	<u>penarikan sertifikat rumah</u>
2	<u>panjang</u>	<u>diragukan</u>	3	<u>Pemanggilan</u>
3	<u>panjang</u>	<u>diperhatikan khusus</u>	2	<u>kasih surat peringatan kedua</u>
4	<u>panjang</u>	<u>diragukan</u>	3	<u>Pemanggilan</u>
5	<u>panjang</u>	<u>diragukan</u>	3	<u>Pemanggilan</u>
6	<u>panjang</u>	<u>diragukan</u>	3	<u>Pemanggilan</u>
7	<u>panjang</u>	<u>Lancar</u>	0	<u>tidak ada</u>
8	<u>panjang</u>	<u>kurang lancar</u>	1	<u>kasih surat peringatan</u>
9	<u>panjang</u>	<u>diperhatikan khusus</u>	2	<u>kasih surat peringatan kedua</u>
10	<u>panjang</u>	<u>Macet</u>	6	<u>penarikan sertifikat rumah</u>
11	<u>pendek</u>	<u>Macet</u>	6	<u>penarikan sertifikat rumah</u>
12	<u>panjang</u>	<u>Macet</u>	6	<u>penarikan sertifikat rumah</u>
13	<u>panjang</u>	<u>Macet</u>	6	<u>penarikan sertifikat rumah</u>
14	<u>menengah</u>	<u>kurang lancar</u>	1	<u>kasih surat peringatan</u>
15	<u>menengah</u>	<u>diperhatikan khusus</u>	2	<u>kasih surat peringatan kedua</u>

Tabel 2: Perhitungan Hasil Kredit Pertama

Node	Attribute	Kategori	Jumlah kasus	Resiko Kredit					Entropy	Gain
				Tidak ada	Peringatan	Pemanggilan	P2	PS		
1	Jangka waktu KPR	Pendek	1	0	0	0	0	1	0	0
		Menengah	2	0	1	0	1	0	0	
		Panjang	12	1	1	4	2	4	0.19265	
	Penyebab terjadi Resiko Kredit	Macet	5	0	0	0	0	5	0.03344	0.052282
		Diragukan	4	0	0	4	0	0	0.02140	
		Diperhatikan khusus	3	0	0	0	3	0	0.01204	
		Lancar	1	1	0	0	0	0	0	
		Kurang lancar	2	0	2	0	0	0	0.00535	
	Keterlambatan per Bulan	Nol	1	1	0	0	0	0	0.052282	

Perhitungan diatas dilakukan dengan rumus :

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

Dengan :

- S : Himpunan kasus
- A : Atribut
- n : Jumlah partisi atribut A
- |S_i| : Jumlah kasus pada partisi ke i
- |S| : Jumlah kasus dalam S

$$Entropy(S) = \sum_{i=1} - p_i * \log_2 p_i \quad (2)$$

dengan :

- S : Himpunan Kasus
 A : Fitur
 n : Jumlah partisi S
 p_i : Proporsi dari S_i terhadap S

Dari perhitungan dengan menggunakan rumus diatas maka masing-masing :
 Entropy dan Gain nya adalah;

Entropy[Jangka Waktu KPR, Pendek]= 0

Entropy[Jangka Waktu KPR, Menengah]=0

Entropy[Jangka Waktu KPR, Panjang]=0.19265

Entropy[Penyebab terjadi resiko kredit, Macet]=0.03344

Entropy [Penyebab terjadi resiko kredit,[Diragukan]=0.02140

Entropy [Penyebab terjadi resiko kredit, Diperhatikan Khusus]=0.01204

Entropy [Penyebab terjadi resiko kredit, Lancar]=0

Entropy [Penyebab terjadi resiko kredit, Kurang Lancar]=0.00535

Entropy[Keterlambatan Per bualan, Nol]=0

Entropy[Keterlambatan Per bualan, Satu]=0.00535

Entropy[Keterlambatan Per bualan, Dua]=0.01204

Entropy[Keterlambatan Per bualan, Tiga]=0.02140

Entropy[Keterlambatan Per bualan, Enam]=0.03344

Baris TOTAL kolom Entropy pada Tabel 3.4 dihitung dengan rumus sebagai berikut:

Total Entropy(S)

$$= \left(-\frac{5}{15} * \log_2 \left(\frac{5}{15}\right)\right) + \left(-\frac{4}{15} * \log_2 \left(\frac{4}{15}\right)\right) + \left(-\frac{3}{15} * \log_2 \left(\frac{3}{15}\right)\right) + \left(\frac{1}{15} * \log_2 \left(\frac{1}{15}\right)\right) + \left(\frac{2}{15} * \log_2 \left(\frac{2}{15}\right)\right) = 0.07223$$

Gain [Total , Penyebab Terjadinya Resiko KREDIT]

=0.07223

$$-\left(\left(\frac{5}{15} * 0.03344\right) + \left(\frac{4}{15} * 0.02140\right) + \left(\frac{3}{15} * 0.01204\right) + \left(\frac{1}{15} * 0\right) + \left(\frac{2}{15} * 0.00535\right)\right)$$

= 0.052282

Dari hasil pada Tabel 2 dapat diketahui bahwa atribut dengan Gain tertinggi adalah **“Penyebab Terjadinya Resiko Kredit Dan Keterlambatan Per Bulan”** yaitu sebesar . Dengan demikian **“Penyebab Terjadinya Resiko Kredit dan Keterlambatan Per Bulan ”** dapat menjadi node akar. Ada 5 nilai atribut dari **“Penyebab Terjadinya Resiko Kredit”** yaitu Macet, Diragukan, Diperhatikan Khusus, Lancar, dan Kurang Lancar. Dari kelima nilai atribut tersebut, nilai atribut Lancar sudah mengklasifikasikan kasus menjadi 1 yaitu keputusan-nya **“Tidak**

Ada”, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut lainnya masih perlu dilakukan perhitungan lagi.

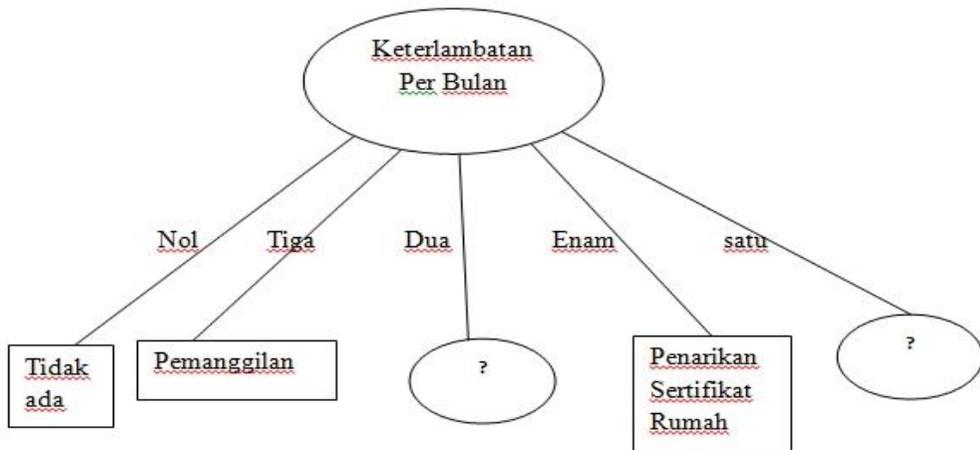
Kemudian “**Keterlambatan Per Bulan**” ada 5 nilai atribut juga yaitu Nol, Satu, Dua, Tiga, dan Enam. Dari kelima nilai atribut tersebut, nilai atribut 0 sudah mengklasifikasikan kasus menjadi 1 yaitu keputusan-nya “Tidak Ada”, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut lainnya masih perlu dilakukan perhitungan lagi juga.

Dari Perhitungan dengan menggunakan rumus diatas maka akan dihasilkan pohon keputusan sementara seperti gambar dibawah ini.



Gambar 3: Hasil pohon keputusan dari Node 1 berdasarkan akar penyebab terjadinya resiko kredit

Selanjutnya, Menghitung jumlah kasus, jumlah kasus untuk keputusan ”Tidak ada”, ”Kasih Surat Peringatan”, ”Penarikan Sertifikat Rumah” , ”Pemanggilan”, ”Kasih Surat Peringatan Kedua” dan Entropy dari semua kasus. Kasus yang dibagi berdasarkan atribut “**Keterlambatan Per Bulan**” yang dapat menjadi node akar dari nilai atribut Nol, Satu, Dua, Tiga, dan Enam. Setelah itu lakukan penghitungan Gain untuk masing-masing atribut. Bisa ditunjukkan sebagai berikut:

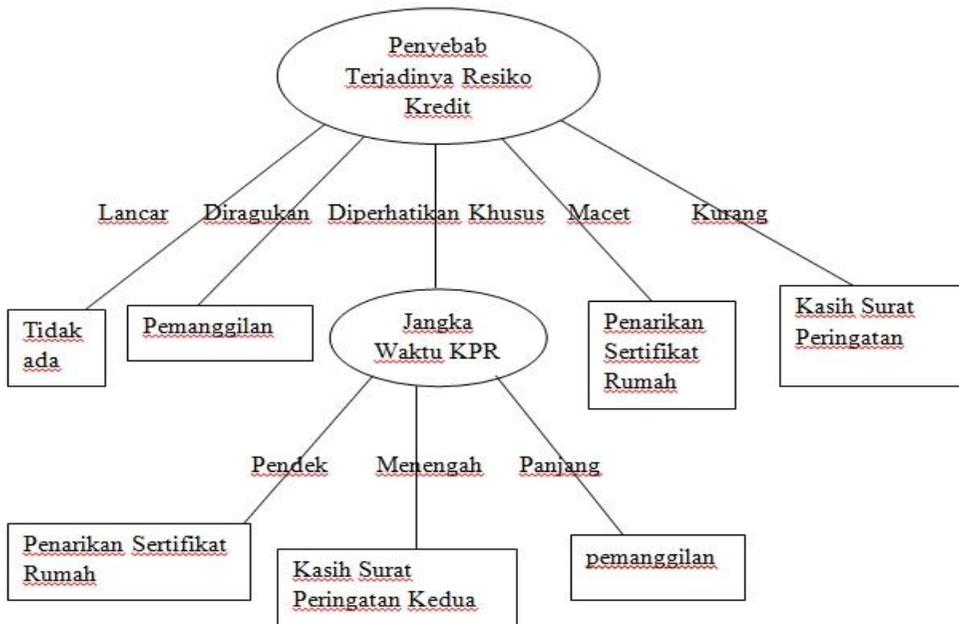


Gambar 4: Hasil pohon keputusan dari Node 1 berdasarkan akar keterlambatan per bulan

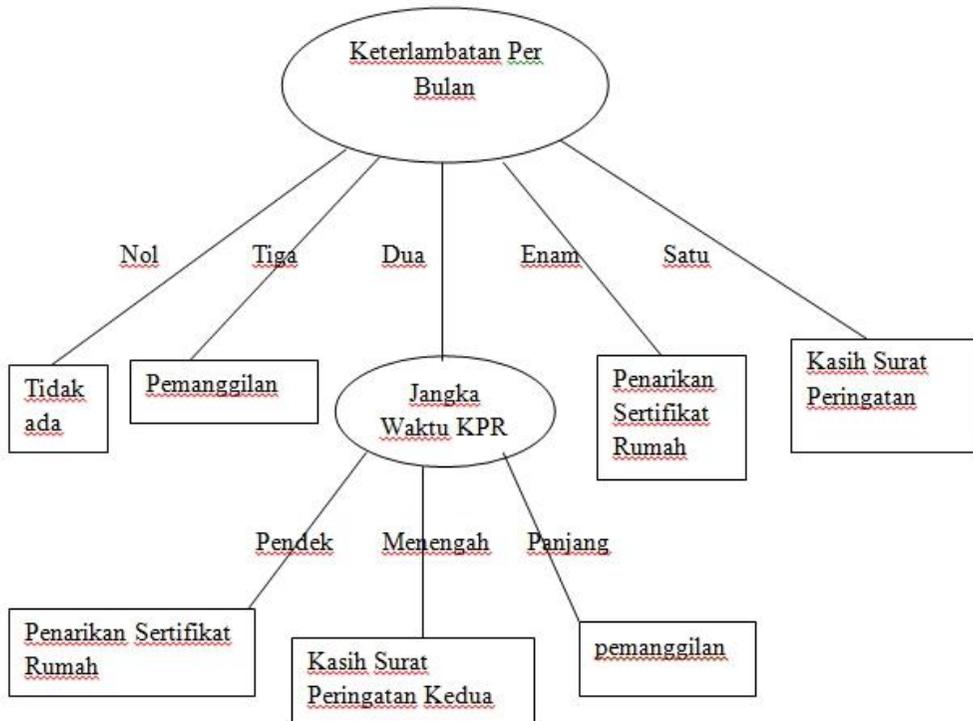
Tabel 3: Perhitungan Hasil Kredit Pertama

Node	Attribute	category	Jumlah kasus	Resiko Kredit					Entropy	Gain
				Tidak ada	Peringatan	Pemanggilan	P2	PS		
KPR	Jangka Waktu	Pendek	1	0	0	0	0	1	0	0.16723
		Menengah	2	0	1	0	1	0	0.06689	
		Panjang	3	0	1	0	2	0	0.10034	
Total kasus			9							

Dari hasil pada tabel 3 hanya ada satu atribut yaitu “**Jangka Waktu KPR**” . Dengan demikian “**Jangka Waktu KPR**” dapat menjadi node cabang dari nilai atribut Diperhatikan Khusus dan Kurang Lancar. Ada 3 nilai atribut dari “Pendek”, “Menengah” dan “Panjang” sehingga tidak perlu dilakukan perhitungan lebih lanjut untuk nilai atribut ini. Pohon keputusan yang terbentuk sampai tahap ini ditunjukkan pada Gambar berikut:



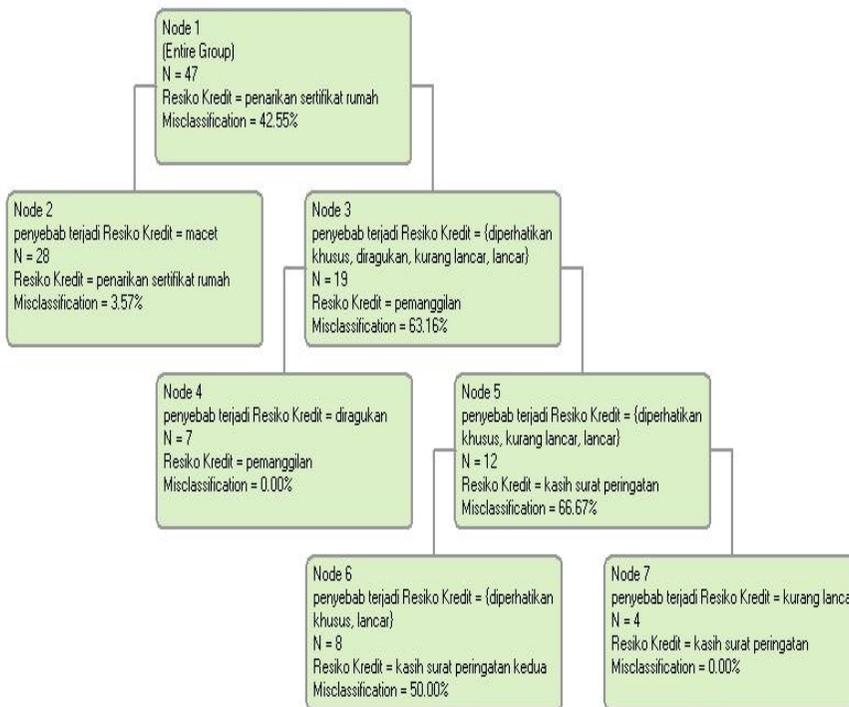
Gambar 5: hasil pohon keputusan akhir berdasarkan penyebab terjadinya resiko kredit



Gambar 6: Hasil pohon keputusan akhir berdasarkan keterlambatan per bulan

Untuk memperoleh data atau informasi dalam studi kasus, tentu perlu dilakukan kegiatan pengumpulan data yang berguna sebagai penunjang. Teknik pengumpulan data terdiri dari :

- a. Berdasarkan sumber data: (Sumber Primer: Sumber yang langsung memberikan data dan Sumber Sekunder: Sumber yang tidak langsung memberikan data).
- b. Berdasarkan Teknik Pengumpulan Data yaitu observasi
Dalam mengumpulkan data, penulis menggunakan teknik wawancara, yaitu pertemuan antara dua orang atau lebih untuk bertukar informasi dan ide melalui tanya jawab sehingga dapat dikonstruksikan makna dalam suatu topik tertentu dan sumber datanya merupakan sumber data primer, yaitu data yang diberikan secara langsung.



Gambar 7: Hasil Output Dari Pengolahan Data

Berdasarkan pohon keputusan nya, didapatkan hasil bahwa pada Bank Bukopin Batam dengan banyak sample 46 orang nasabah yang lebih spesifik lagi dan data mudah dimengerti maka dari 46 orang nasabah dipecah lagi berdasarkan prediksi node-node lebih kecil yaitu ;

1. prediksi pada 27 orang nasabah yang penyebab terjadinya resiko kredit macet, 3.57% *misclassification* nya dengan hasil resiko kredit berupa penarikan sertifikat rumah.
2. prediksi pada 19 orang nasabah yang penyebab terjadinya resiko kredit diperhatikan khusus, diragukan, kurang lancar, lancar. Dan dari 20 orang

- nasabah yang 63.16% *misclassification* nya dapat dibagi lagi menjadi kelompok prediksi node-node yang lebih kecil lagi yaitu
3. prediksi pada 7 orang nasabah yang penyebab terjadinya resiko kredit diragukan, 0.00% *misclassification* nya dengan hasil resiko kredit berupa pemnggilan.
 4. Prediksi pada 12 orang nasabah pasien yang penyebab terjadinya resiko kredit diperhatikan khusus, kurang lancar, dan lancar (tidak ada resiko kreditnya), 66.67 % *misclassification* nya dengan hasil resiko kredit berupa kasih surat peringatan, namun disini terbagi lagi node-node yang lebih kecil lagi, yaitu:
 5. prediksi pada 8 orang nasabah 50.00% *misclassification* nya dengan hasil resiko kredit berupa kasih surat peringatan kedua.
 6. prediksi pada 4 orang nasabah 0.00% *misclassification* nya dengan hasil resiko kredit berupa kasih surat peringatan pertama.

4. KESIMPULAN

Decision tree yang dihasilkan telah mampu menghasilkan knowledge (Resiko Kredit) dengan berdasarkan jumlah nasabah, jenis jangka waktu KPR, penyebab terjadi resiko kredit, keterlambatan per bulan, seperti yang tercatat pada data yang telah diambil. Pemecahan *field* penyebab terjadinya resiko kredit dan keterlambatan per bulan menjadi beberapa kelompok kecil membantu *user* dalam memahami informasi yang dihasilkan dari pohon keputusan yang dihasilkan. Dengan menggunakan software DTREG juga mampu menentukan resiko kredit pada Bank.

DAFTAR PUSTAKA

- Arhami, Muhammad. 2005. Konsep Dasar Sistem Pakar. Yogyakarta : Penerbit Andi
- Arikunto, 2007. Prosedur Suatu Pendekatan Praktik. Jakarta : Rineka Aksara
- Sugiyono, 2013. Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D. Bandung : ALFABETA
- Hartati, iswanti. 2008. Sistem Pakar dan Pengembangannya. Yogyakarta: Graha Ilmu.
- Kusrini, Emha Taufiq Luthfi. 2009. Algoritma Data Mining. Yogyakarta : Penerbit Andi.
- Kusrini. 2008. Aplikasi Sistem Pakar, Yogyakarta : Penerbit Andi.
- Han, Jiawei; Kamber, Micheline; Data Mining: Concepts and Techniques. Morgan Kaufmann, 2006
- Witten, Ian H.; Frank, Eibe; Data Mining Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann Publishers, 2005