

# PREDIKSI KEBUTUHAN NASABAH DENGAN TEKNIK DATA MINING DALAM MENDUKUNG STRATEGI PEMASARAN BANK

<sup>1</sup>Sanusi Amir, <sup>2</sup>Kusrini

Universitas Amikom Yogyakarta, Jurusan Teknik Informatika

e-mail : <sup>1</sup>uci\_mks2008@yahoo.com, <sup>2</sup>kusrini@amikom.ac.id

*Intisari*— Loyalitas nasabah dalam menggunakan produk perbankan merupakan salah satu faktor keberlangsungan sebuah Bank, sehingga menuntut pihak manajemen perbankan berinovasi untuk meng-*create* produk unggulan. produk-produk yang dihasilkan pun berevolusi dari produk klasik hingga produk digital yang memanfaatkan perkembangan teknologi. tujuan terpenting adalah meyakinkan nasabah untuk menggunakan berbagai produk perbankan yang tentunya sesuai dengan kebutuhan nasabah. *Database* nasabah merupakan pasar yang ideal dalam menawarkan produk. Pemanfaatan data nasabah dengan teknik *Data Mining* dan Algoritma C.45 dapat membantu pihak Bank dalam memprediksi kebutuhan nasabah, karena teknik *Data Mining* dapat mengungkapkan pola-pola tersembunyi dari perilaku nasabah. Informasi pola-pola perilaku nasabah dimaksud dapat dijadikan modal utama oleh pemasaran Bank untuk memprediksi kebutuhan transaksi keuangan dari nasabah .

*Kata Kunci*— Data Mining, Algoritma C4.5, Pohon Keputusan, Nasabah, Prediksi.

*Abstract*— Customer loyalty in using banking products is one of the sustainability factors of a Bank, therefore it requires banking management to innovate in creating superior products. The products produced have evolved from classic products to digital products that utilize technological developments. The most important goal is to convince customers to use a variety of banking products that are certainly in line with customer needs. The customer database is an ideal market in offering products. The use of customer data with Data Mining and C4.5 Algorithm techniques can help the Bank predict customer needs, because Data Mining techniques can reveal hidden patterns of customer behavior. Information on customer behavior patterns can be used as the main capital by Bank marketing to predict customer financial transaction needs.

*Keywords*— Data Mining, Algorithm C4.5, Decision Tree, Customer, Predict.

## I. PENDAHULUAN

Komitmen untuk terus menjadi yang terbaik dalam industri perbankan merupakan harapan yang ingin diwujudkan oleh setiap bank yang beroperasi di tanah air, tentunya dengan penerapan strategi yang tepat dalam mewujudkannya. Ekspansi kredit, peningkatan dana pihak ketiga serta peningkatan pendapatan merupakan indikator penilaian kinerja sebuah bank. Strategi pemasaran bank pun disusun oleh pihak manajemen pusat untuk dapat menjadi bank pilihan nasabah dalam melakukan transaksi keuangan. Strategi ekspansi kredit dilakukan dengan cara menerbitkan kebijakan suku bunga *single* digit untuk kredit produktif, promo berbagai pilihan suku bunga (*fix rate*) dengan *range* jangka waktu untuk kredit konsumtif dan biaya administrasi kredit yang bervariasi serta konsistensi waktu proses kredit sesuai SLA (*service level agreement*) yang telah ditentukan. Berbagai jenis produk simpanan dihasilkan untuk dapat meningkatkan dana pihak ketiga. Produk simpanan yang dihasilkan berevolusi dari produk transaksi klasik hingga transaksi digital yang bertujuan memudahkan nasabah dalam bertransaksi dimanapun dan kapanpun. Kecanggihan teknologi yang mendukung transaksi perbankan merupakan

nilai tambah dalam mensosialisasikan produk kepada nasabah. Promosi produk simpanan gencar dilakukan sebagai sarana mensosialisasikan produk simpanan ke nasabah dengan tujuan nasabah dapat menggunakan produk dimaksud sehingga dapat meningkatkan dana pihak ketiga.

Hasil ekspansi kredit akan mendatangkan pendapatan bunga, peningkatan dana pihak ketiga serta penggunaan kartu kredit akan menghasilkan pendapatan dari transaksi keuangan nasabah. Penyaluran kredit wajib *perudent* agar kualitas kredit berada dalam golongan lancar yang mempengaruhi pencadangan biaya. Pengumpulan dana pihak ketiga difokuskan pada dana murah (tabungan dan giro) sehingga beban bunga yang diberikan ke nasabah berada dalam presentase rendah. Penerapan strategi pemasaran bank seperti yang telah dijelaskan diatas dapat dikategorikan sebagai strategi *product centric*, dimana pemasaran lebih difokuskan kepada penawaran produk, baik itu produk pinjaman maupun produk simpanan. Kondisi yang ditimbulkan akibat penerapan strategi *product centric* adalah semakin terbukanya ruang persaingan dengan bank lain karena penerapan strategi ini akan menggerakkan tenaga pemasaran untuk mencari nasabah baru serta menawarkan produk bank dan tentunya akan berdampak pada kenaikan biaya operasional pemasaran.

Potensi pasar yang sesungguhnya berada dalam lingkungan bank itu sendiri berupa *database* nasabah, dengan *database* nasabah ini pihak manajemen dapat menganalisa pengetahuan yang tersembunyi untuk mendukung strategi pemasaran bank yang berorientasi *customer centric*, dimana pemasaran lebih difokuskan pada kebutuhan nasabah. Pemanfaatan *database* nasabah dapat menghasilkan informasi sebagai landasan pihak manajemen merumuskan strategi pemasaran bank.

dalam penelitian ini penulis mengeksplorasi 150 data nasabah salah satu bank (nama bank tidak disebutkan dalam penelitian ini) mencakup total saldo, agama, pendidikan, pekerjaan, jenis kelamin, status pernikahan, jumlah anak, dengan metode *data mining* disertai penerapan teknik pohon keputusan (*decision tree*) dengan algoritma C4.5 untuk mengetahui atribut yang paling berpengaruh dalam memprediksi kebutuhan nasabah *existing* terhadap jenis produk pinjaman dan simpanan.

#### A. Tinjauan Pustaka

Penelitian pemanfaatan *data mining* dalam dunia perbankan dilakukan oleh Zakirov dan Momtselidze yang menegaskan bahwa *data mining* dapat diterapkan untuk keperluan *Risk Management, Marketing, Mobile Banking Customer Relationship Management (CRM), Fraud Detection, Cross-Selling* [1].

Preethi dan Vijayalakshmi [2] mengungkapkan bahwa *data mining* memainkan peran penting dalam banyak organisasi. Sangat membantu dalam meneliti data yang dikumpulkan dan mengantarkannya ke dalam pola yang bisa dimengerti. Dalam skenario sekarang, perbankan merupakan sektor yang sedang berkembang dimana besarnya volume data elektronik tersimpan. Tugas penting di bidang perbankan adalah menangani transaksional yang besar dan pengambilan keputusan mengenai *customer retention, fraud detection, risk and marketing management*. Membuat keputusan secara manual tentunya memakan waktu dan rawan kesalahan. Dengan menggunakan beberapa teknik *data mining* pola dan basis pengetahuan yang tersembunyi dapat diketahui.

Zorić meneliti penggunaan *data mining* dalam industri perbankan dengan metode *neural network* untuk memprediksi nasabah yang beresiko meninggalkan layanan sebuah bank dan menganalisa apakah nasabah tertentu layak untuk dipertahankan. Dalam melakukan penelitian digunakan perangkat lunak *alyuda neurointelligence* yang mampu menjalankan proses metode *neural network*. Proses penelitian diawali dengan definisi masalah dimana masalah bisnis yang spesifik diterjemahkan ke dalam masalah *data mining*. Data yang telah dikumpulkan ditransformasikan menjadi format yang ditentukan untuk selanjutnya dilakukan proses pembersihan data (menghapus data yang tidak relevan dan tidak akurat). Persiapan data dilakukan beberapa kali dan tidak dalam urutan yang ditentukan. Fase persiapan data bisa memakan waktu hingga 80% dari semua waktu analisis karena kualitas data merupakan tantangan utama dalam *data mining*. Selanjutnya teknik permodelan dipilih dan diterapkan. Hasil *data mining* disajikan kepada pengguna secara visual, karena teknik visualisasi lebih efektif dalam memahami output. Hasil

penelitian menunjukkan bahwa nasabah yang menggunakan lebih banyak layanan bank (produk) lebih loyal [3].

Referensi [4] menunjukkan bahwa Khodamoradi dan Mosa meneliti resiko kecurangan dan resiko kredit pada bank dengan mempertimbangkan besarnya volume data yang dihasilkan oleh bank, dibutuhkan model analisa yang sesuai untuk pengambilan keputusan. Hasil penelitian ini menegaskan bahwa bank sebagai lembaga keuangan memerlukan proses manajemen resiko dan *data mining* adalah alat yang paling penting untuk mengevaluasi, mengidentifikasi dan memprediksi berbagai resiko di depan.

Strategi pemasaran bank tergantung pada pemanfaatan data elektronik nasabah yang sangat besar. Ukuran sumber data ini tidak mungkin bagi seorang analis manusia untuk menghasilkan informasi menarik yang akan membantu pengambilan keputusan. Model *data mining* sepenuhnya membantu dalam proses ini. Penelitian ini memperkenalkan analisis dan penerapan teknik yang paling penting dalam *data mining* : *multilayer perception neural network (MLPNN), tree augmented naïve bayes (TAN) known as bayesian networks, nominal regression or logistic regression (LR), and ross quinlan new decision tree model (C5.0)*. Tujuannya adalah meningkatkan efektifitas strategi pemasaran dengan mengidentifikasi karakteristik utama yang mempengaruhi keberhasilan (produk deposit yang digunakan oleh nasabah) berdasarkan MLPNN, TAN, LR dan C5.0. Hasil eksperimen menunjukkan, dengan akurasi yang lebih tinggi, keberhasilan model ini dalam memprediksi strategi pemasaran terbaik kepada nasabah untuk menggunakan produk deposit [5].

Pemasaran dengan pemanfaatan database merupakan strategi pemasaran untuk meningkatkan penjualan. Organisasi perusahaan harus menyimpan data pelanggan dan mengembangkan kapasitas untuk dianalisis. Salah satu transaksi yang telah otomatis adalah *customer relationship management (CRM)*. Dalam penelitian ini teknik *data mining* digunakan untuk menentukan pelanggan dan segmen pasar yang akan merespons penawaran oleh perusahaan. Penelitian ini mengandalkan data konsumen digabungkan secara elektronik melalui proses *customer relationship management (CRM)*. Penelitian ini mencoba untuk memperbaiki kinerja algoritma klasifikasi yang digunakan di bank berupa prediksi respons pemasaran nasabah dengan menggunakan *random forest ensemble*. Algoritma klasifikasi yang digunakan untuk memodelkan dataset bank meliputi *logistic regression, decision tree, naïve bayes and the random forest ensemble* menggunakan aplikasi *data mining orange 3.2*. hasil metode eksperimental menunjukkan bahwa *features duration, poutcome, contact, month and housing* adalah fitur yang paling penting berkontribusi pada suksesnya strategi pemasaran nasabah bank untuk menggunakan produk *deposit*. Penelitian ini juga mengungkapkan bahwa durasi panggilan ke nasabah, penggunaan ponsel berkontribusi positif terhadap keberhasilan strategi pemasaran [6].

#### 1) Data Mining

Referensi [7] menunjukkan bahwa, *data mining* adalah proses yang mempekerjakan satu atau lebih teknik

pembelajaran komputer (*machine learning*) untuk menganalisa dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari. *knowledge discovery in databases* (KDD) adalah penerapan metode saintifik pada *data mining*. Dalam konteks ini *data mining* merupakan satu langkah dari KDD. *Data mining* merupakan proses iteratif dan interaktif untuk menemukan pola atau model baru yang sah (sempurna), bermanfaat dan dapat dimengerti dalam suatu *database* yang sangat besar (*massive databases*).

- Sahih: dapat digeneralisasi untuk masa yang akan datang.
- Baru : apa yang sedang tidak diketahui.
- Bermanfaat: dapat digunakan untuk melakukan suatu tindakan.
- Iteratif: memerlukan sejumlah proses yang diulang.
- Interaktif: memerlukan interaksi manusia dalam prosesnya.

*Data Mining* berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola-pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lainnya.

Tahapan proses dalam penggunaan *data mining* yang merupakan *knowledge discovery in databases* (KDD) dapat diuraikan sebagai berikut:

- Memahami domain aplikasi untuk mengetahui dan menggali pengetahuan awal serta apa sasaran pengguna.
- Membuat target data-set yang meliputi pemilihan data dan fokus pada sub-set data.
- Pembersihan dan transformasi data meliputi eliminasi derau, *outliers*, *missing value* serta pemilihan fitur dan reduksi dimensi.
- Penggunaan algoritma *data mining* yang terdiri dari asosiasi, sekuensial, klasifikasi, klusterisasi, dll.
- Interpretasi, evaluasi dan visualisasi pola untuk melihat apakah ada sesuatu yang baru dan menarik dan dilakukan iterasi jika diperlukan.

Beberapa teknik dan sifat *data mining* adalah sebagai berikut :

- *Classification* [Predictive]

Klasifikasi adalah menentukan sebuah *record* data baru ke salah satu dari beberapa kategori (atau klas) yang telah didefinisikan sebelumnya. Disebut dengan '*supervised learning*'.

- *Clustering* [Descriptive]

Mempartisi data-set menjadi beberapa sub-set atau kelompok sedemikian rupa sehingga elemen-elemen dari suatu kelompok tertentu memiliki set properti yang dishare bersama, dengan tingkat similaritas yang tinggi dalam suatu

kelompok dan tingkat similaritas antar kelompok yang rendah. Disebut juga dengan '*unsupervised learning*'.

- *Association Rule Discovery* [Descriptive]

Mendeteksi kumpulan atribut-atribut yang muncul bersamaan dalam frekuensi yang sering, dan membentuk sejumlah kaidah dari kumpulan-kumpulan tersebut. Contoh : 9% orang yang berbelanja di suatu supermarket yang membeli roti juga membeli selai, dan 6% dari semua orang yang berbelanja membeli keduanya.

- *Sequential Pattern Discovery* [Descriptive]

Mencari sejumlah *event* yang secara umum terjadi bersama-sama. Contoh, dalam satu set urutan DNA, ACGTC diikuti oleh GTCA setelah suatu celah selebar 9 dengan probabilitas sebesar 3%..

- *Regression* [Predictive]

Memprediksi nilai dari suatu variabel kontinyu yang diberikan berdasarkan nilai dari variabel yang lain, dengan mengasumsikan sebuah model ketergantungan linier atau nonlinier. Teknik ini banyak dipelajari dalam statistika, bidang jaringan saraf tiruan (*neural network*).

- *Deviation Detection* [Predictive]

*Data mining* sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. *data mining* juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah *data mining* kadang disebut juga *knowledge discovery*. Pekerjaan yang berkaitan dengan *data mining* dapat dibagi menjadi : model prediksi (*prediction modelling*), analisis kelompok (*cluster analysis*), analisis asosiasi (*association analysis*), deteksi anomali (*anomaly detection*) [8].

Menurut (Fayyad et al., 1996) dalam [9], *data mining* adalah analisis terhadap proses penemuan pengetahuan di dalam basis data atau *knowledge discovery in databases* yang disingkat KDD. Pengetahuan bisa berupa pola data atau relasi antar data yang *valid* (yang tidak diketahui sebelumnya). Secara umum, kegunaan *data mining* dapat dibagi menjadi dua: deskriptif dan prediktif. Deskriptif berarti *data mining* digunakan untuk mencari pola-pola yang dapat dipahami manusia yang menjelaskan karakteristik data. Sedangkan prediktif berarti *data mining* digunakan untuk membentuk sebuah model pengetahuan yang akan digunakan untuk melakukan prediksi.

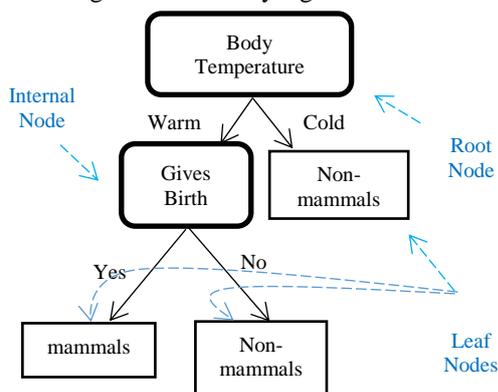
## 2) Pohon Keputusan

Haryanto dan Hansun [10] dalam bukunya Slamet, A., mengatakan Pohon keputusan adalah metode klasifikasi dan prediksi yang sudah terbukti *powerfull* dan sangat terkenal. Metode ini berfungsi untuk mengubah fakta menjadi pohon keputusan yang merepresentasikan aturan yang dapat mudah dimengerti dengan bahasa alami. Proses dari pohon keputusan ini dimulai dari *node* akar hingga *node* daun yang dilakukan secara rekursif dimana setiap percabangan menyatakan kondisi dan setiap ujung pohon akan menyatakan keputusan.

Berdasarkan apa yang ditulis oleh Triisant dalam jurnal Haryanto dan Hansun, pohon keputusan ini memiliki beberapa kekurangan dan kelebihan, yakni:

- Kelebihan Pohon Keputusan:
  - Daerah pengambilan keputusan yang kompleks dapat diubah menjadi sederhana.
  - Dapat menghilangkan perhitungan yang tidak penting karena proses pengujian hanya berdasarkan kriteria yang diperlukan saja.
  - Proses pemilihan fitur dari *internal node* yang berbeda lebih fleksibel. Fitur yang telah dipilih ini akan menjadi pembeda antara kriteria yang satu dengan kriteria lainnya.
  - Metode ini dapat menghindari munculnya permasalahan dengan cara menggunakan kriteria dengan jumlah yang sedikit pada *node internal* tanpa mengurangi kualitas keputusan yang dihasilkan.
- Kekurangan Pohon Keputusan :
  - Dapat terjadi *overlap* apabila hasil keputusan dan kriteria yang digunakan jumlahnya sangat banyak. Hal ini juga dapat berakibat bertambahnya waktu yang digunakan untuk pengambilan keputusan dan jumlah memori yang dibutuhkan semakin tinggi.
  - Akumulasi jumlah *error* dari setiap tingkat pohon keputusan besar.
  - Mendesain pohon keputusan yang optimal sulit.
  - Kualitas keputusan yang didapatkan sangat tergantung dengan bagaimana pohon tersebut didesain.

Referensi [7] menunjukkan bahwa Pohon keputusan merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas berhingga, dimana simpul internal maupun simpul akar ditandai dengan nama atribut, rusuk-rusuknya diberi label nilai atribut yang mungkin dan simpul daun ditandai dengan kelas-kelas yang berbeda.



Gambar 1. Contoh Pohon Keputusan Untuk Klasifikasi Mamalia

### 3) Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk melakukan proses klasifikasi data dengan menggunakan teknik pohon keputusan. Algoritma C4.5 merupakan ekstensi

dari algoritma ID3 dan menggunakan prinsip *decision tree*. Algoritma ini sudah sangat terkenal dan disukai karena memiliki banyak kelebihan. Kelebihan ini misalnya dapat mengolah data numerik dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan performanya merupakan salah satu yang tercepat dibandingkan dengan algoritma lain. Ide dasar dari algoritma ini adalah pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki nilai *gain* tertinggi berdasarkan nilai *entropy* atribut tersebut sebagai poros atribut klasifikasi. Kemudian secara rekursif cabang-cabang pohon diperluas sehingga seluruh pohon terbentuk. Menurut kamus IGI Global (*International Publisher of Progressive Academic*), *entropy* adalah jumlah data yang tidak relevan terhadap informasi dari suatu kumpulan data. *Gain* adalah informasi yang didapatkan dari perubahan *entropy* pada suatu kumpulan data, baik melalui observasi atau bisa juga disimpulkan dengan cara melakukan partisipasi terhadap suatu *set* data [10].

Menurut (Jefri, 2013) dalam jurnal [10], Terdapat empat langkah dalam proses pembuatan pohon keputusan pada algoritma C4.5, yaitu:

- Memilih atribut sebagai akar
- Membuat cabang untuk masing-masing nilai
- Membagi setiap kasus dalam cabang
- Mengulangi proses dalam setiap cabang sehingga semua kasus dalam cabang memiliki kelas yang sama.

Menurut (Jiandi,2016) dalam jurnal [10], data yang dimiliki harus disusun menjadi sebuah tabel berdasarkan kasus dan jumlah responden sebelum dilakukan perhitungan untuk mencari nilai *entropy* dan *gain*.

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

Rumus (1) merupakan rumus yang digunakan dalam perhitungan *entropy* yang digunakan untuk menentukan seberapa informatif atribut tersebut. Berikut keterangannya:

- $S$  : Himpunan kasus
- $n$  : Jumlah partisi  $S$
- $p_i$  : Jumlah kasus pada partisi ke- $i$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \left| \frac{S_i}{S} \right| * Entropy(S_i) \quad (2)$$

Rumus (2) merupakan rumus yang digunakan dalam perhitungan *gain* setelah melakukan perhitungan *entropy*. Berikut keterangannya:

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

Dengan mengetahui rumus-rumus diatas, data yang telah diperoleh dapat dimasukkan dan diproses dengan algoritma C4.5 untuk proses pembuatan pohon keputusan.

Algoritma C4.5 berfungsi untuk pembentukan pohon keputusan. Perhitungan dimulai dari menghitung banyaknya jumlah atribut dan menentukan atribut mana yang akan digunakan sebagai akar dari pohon keputusan. Selanjutnya akan dilakukan perhitungan *entropy* dan *gain* untuk menentukan *leaf* dari pohon keputusan tersebut. Setelah semua perhitungan selesai dilakukan, pohon keputusan dapat dibentuk berdasarkan nilai *gain* yang telah dihitung sebelumnya. Atribut dengan nilai *gain* tertinggi akan terletak pada prioritas yang lebih tinggi dan memiliki kedudukan yang lebih tinggi juga pada pohon keputusan.

## II. METODOLOGI PENELITIAN

### A. Jenis, Sifat, dan Pendekatan Penelitian

Jenis penelitian ini adalah tindakan, karena peneliti memonitoring alur kerja pada unit pemasaran hingga memperoleh data nasabah untuk dilakukan analisa. Sifat penelitian deskriptif, yaitu mendeskripsikan kegiatan penelitian dari tahap awal, analisa data nasabah hingga memprediksi jenis produk simpanan dan pinjaman yang dapat ditawarkan kepada nasabah berdasarkan data nasabah bank. Pendekatan penelitian yang digunakan adalah penelitian kualitatif yaitu penerapan metode Pohon Keputusan (*decision tree*) dengan algoritma C4.5 dalam memanfaatkan database nasabah untuk memprediksi jenis produk simpanan dan pinjaman yang dapat ditawarkan.

### B. Metode Pengumpulan Data

Data yang dibutuhkan dalam penelitian ini merupakan data nasabah yang diperoleh melalui pemanfaatan *tools* internal bank (nama bank tidak disebutkan dalam penelitian ini)

### C. Metode Analisis Data

Proses analisis data yang dilakukan dapat dirincikan sebagai berikut:

- Integrasi data

Mengkombinasikan atau mengintegrasikan data karena penggunaan *tools* internal bank menghasilkan data dengan kriteria-kriteria yang dibutuhkan terpisah antara satu data dengan yang lainnya.

- Seleksi data

Mengambil data-data sesuai dengan kriteria yang dijadikan dasar pengujian

- Analisa data

Analisa data dilakukan dengan teknik *data mining* model pohon keputusan C 4.5 untuk memprediksi jenis produk simpanan dan pinjaman.

Tahapan awal pengujian yaitu ditetapkan jenis-jenis produk simpanan dan pinjaman yang dijadikan acuan prediksi hasil pengujian data.

TABEL I

JENIS-JENIS PRODUK SIMPANAN DAN PINJAMAN

Jenis Produk	
Simpanan	Pinjaman
Deposito	perumahan
Pegawai	pegawai
Giro	pensiun
Anak	modal kerja / Investasi
Dolar	multiguna
Prioritas	Kredit usaha rakyat
Haji	
Pensiun	

TABEL II  
SKENARIO PENGUJIAN DATA

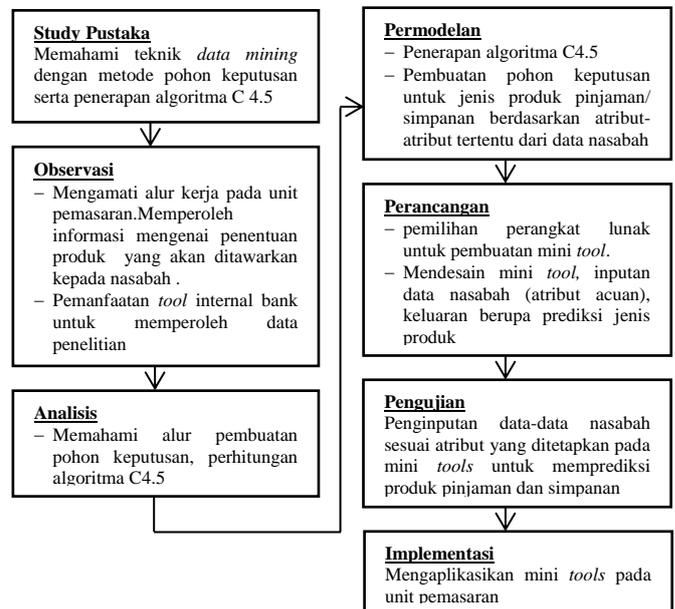
Nama	Status	Pekerjaan	Prediksi Produk	
			Simpanan	Pinjaman
AAAA	K	PNS	Anak pensiunan Giro	Pegawai Perumahan Pensiun
BBBB	TK	Swasta	Giro Deposito Pensiun	Kredit usaha rakyat Modal kerja/Investasi Perumahan

- Menarik kesimpulan

Kesimpulan didasari pada hasil akhir prediksi produk pinjaman dan simpanan, dimana kombinasi dari beberapa atribut data-data nasabah dapat menghasilkan prediksi *multi* produk pinjaman dan simpanan yang dapat ditawarkan

### D. Alur Penelitian

Alur penelitian dapat diuraikan sebagai berikut:



Gambar 2. Alur Penelitian

TABEL III  
DATA NASABAH

No	CIF	Nama	Total Saldo	Tgl Lahir	Agama	Pendidikan	Pekerjaan	Jenis Kelamin	...
1	222222 2	Sanusi	137.591.322	29/12/1972	Islam	S1	Pegawai BUMN/B UMD	W	...
...	...	...	...	...	...	...	...	...	...
150	xxxxxx	William	137.635.131	02/02/1964	Kristen	S1	Pegawai Negeri	W	...

TABEL III LANJUTAN

...	Status	Jml Anak	Jumlah Account	Telp	Alamat
...	L	0	1	0911352677	Kayu Putih – Ambon
...	...	...	...	...	...
...	L	0	1	0911352601	Karpan -Ambon

TABEL IV  
NILAI ATRIBUT DATA NASABAH

Atribut					
Total Saldo	Agama	Pendidikan	Pekerjaan	Jenis Kelamin	Status
<=250 Juta	Islam	SD	Pegawai Negeri	W (Wanita)	Menikah (M)
> 250 Juta s.d <=500 Juta	Kristen	SMP	Pegawai Swasta	P (Pria)	Lajang (L)
> 500 Juta s.d <=750 Juta	Katolik	SMA	Pegawai BUMN/BUMD		Janda (J)
>750 Juta s.d <=1M		D3	Wiraswasta		
>1M		S1	Ibu Rumah Tangga		
		S2	Dokter		
		S3	Dosen/Guru Swasta		
			Dosen/Guru Negeri		
			Pensiunan		
			Pejabat Negara		
			Pengusaha		
			Pedagang		
			TNI/Polri		
			Petani/Nelayan		
			Profesi		
			Pelajar/Mahasiswa		

### III. HASIL DAN PEMBAHASAN

*Database* nasabah Bank yang menjadi data dalam penelitian ini terdiri dari 150 data nasabah dengan rincian *field* berupa CIF, nama nasabah, total saldo, tanggal lahir, agama, pendidikan, pekerjaan, jenis kelamin, status, jumlah anak, jumlah *account*, nomor telepon, alamat. dalam pengujian data menggunakan algoritma C4.5 disertai teknik pohon keputusan. *field* data nasabah yang digunakan untuk memprediksi kebutuhan nasabah adalah : total saldo, agama, pendidikan, pekerjaan, jenis kelamin, status.

#### A. Perhitungan Nilai Entropy

Perhitungan *Entropy* sesuai persamaan (1) untuk produk pinjaman dan simpanan sebagai berikut:

TABEL V  
PERHITUNGAN NILAI ENTROPY UNTUK PRODUK PINJAMAN

Total Kasus	Perumahan		Entropy
	Sum (Ya)	Sum (Tidak)	
150	53	97	0.9370098
$=(-53/150)*\text{LOG}_2(53/150) + (-97/150)*\text{LOG}_2(97/150)$ $=(-0.3533333)*\text{LOG}_2(0.3533333) + (-0.6466667)*\text{LOG}_2(0.6466667)$ $=(-0.3533333)*(-1.5008982) + (-0.6466667)*(-0.6289058)$ $=0.9370098$			
150	Pegawai		0.9967916
	70	80	
$=(-70/150)*\text{LOG}_2(70/150) + (-80/150)*\text{LOG}_2(80/150)$ $=(-0.4666666)*\text{LOG}_2(0.4666666) + (-0.5333333)*\text{LOG}_2(0.5333333)$ $=(-0.4666666)*(-1.0995356) + (-0.5333333)*(-0.9068905)$ $=0.9967916$			

TABEL V  
LANJUTAN

Total Kasus	Pensiun		Entropy
	Sum (Ya)	Sum (Tidak)	
150	30	120	0.7219281
$=(-30/150)*\text{LOG2}(30/150) + (-120/150)*\text{LOG2}(120/150)$ $=(-0.2)*\text{LOG2}(0.2) + (-0.8)*\text{LOG2}(0.8)$ $=(-0.2)*(-2.3219280) + (-0.8)*(-0.3219280)$ $=0.7219281$			
Modal Kerja & Investasi			
150	65	85	0.9871378
$=(-65/150)*\text{LOG2}(65/150) + (-85/150)*\text{LOG2}(85/150)$ $=(-0.4333333)*\text{LOG2}(0.4333333) + (-0.5666666)*\text{LOG2}(0.5666666)$ $=(-0.4333333)*(-1.2064508) + (-0.5666666)*(-0.8194277)$ $=0.9871378$			
Multiguna			
150	139	11	0.3782416
$=(-139/150)*\text{LOG2}(139/150) + (-11/150)*\text{LOG2}(11/150)$ $=(-0.9266666)*\text{LOG2}(0.9266666) + (-0.7333333)*\text{LOG2}(0.7333333)$ $=(-0.9266666)*(-0.1098776) + (-0.7333333)*(-3.7693870)$ $=0.3782416$			
KUR			
150	58	92	0.9626147
$=(-58/150)*\text{LOG2}(58/150) + (-92/150)*\text{LOG2}(92/150)$ $=(-0.3866666)*\text{LOG2}(0.3866666) + (-0.6133333)*\text{LOG2}(0.6133333)$ $=(-0.3866666)*(-1.3708376) + (-0.6133333)*(-0.7052567)$ $=0.9626147$			

TABEL VI  
PERHITUNGAN NILAI ENTROPY UNTUK PRODUK SIMPANAN

Total Kasus	Deposito		Entropy
	Sum (Ya)	Sum (Tidak)	
150	99	51	0.9248187
$=(-99/150)*\text{LOG2}(99/150) + (-51/150)*\text{LOG2}(51/150)$ $=(-0.66)*\text{LOG2}(0.66) + (-0.34)*\text{LOG2}(0.34)$ $=(-0.66)*(-0.5994620) + (-0.34)*(-1.5363933)$ $=0.9248187$			
Pegawai			
150	70	80	0.9967916
$=(-70/150)*\text{LOG2}(70/150) + (-80/150)*\text{LOG2}(80/150)$ $=(-0.4666666)*\text{LOG2}(0.4666666) + (-0.5333333)*\text{LOG2}(0.5333333)$ $=(-0.4666666)*(-1.0995356) + (-0.5333333)*(-0.9068905)$ $=0.9967916$			
Giro			
150	99	51	0.9248187
$=(-99/150)*\text{LOG2}(99/150) + (-51/150)*\text{LOG2}(51/150)$ $=(-0.66)*\text{LOG2}(0.66) + (-0.34)*\text{LOG2}(0.34)$ $=(-0.66)*(-0.5994620) + (-0.34)*(-1.5363933)$ $=0.9248187$			
Anak			
150	135	15	0.4689956
$=(-135/150)*\text{LOG2}(135/150) + (-15/150)*\text{LOG2}(15/150)$ $=(-0.9)*\text{LOG2}(0.9) + (-0.1)*\text{LOG2}(0.1)$ $=(-0.9)*(-0.1520030) + (-0.1)*(-3.3219280)$ $=0.4689956$			
Dolar			
150	98	52	0.9310558
$=(-98/150)*\text{LOG2}(98/150) + (-52/150)*\text{LOG2}(52/150)$ $=(-0.65)*\text{LOG2}(0.65) + (-0.35)*\text{LOG2}(0.35)$ $=(-0.65)*(-0.6141088) + (-0.35)*(-1.5283789)$ $=0.9310558$			
Prioritas			
150	52	98	0.9310558
$=(-52/150)*\text{LOG2}(52/150) + (-98/150)*\text{LOG2}(98/150)$ $=(-0.35)*\text{LOG2}(0.35) + (-0.65)*\text{LOG2}(0.65)$ $=(-0.35)*(-1.5283789) + (-0.65)*(-0.6141088)$ $=0.9310558$			

TABEL VI  
LANJUTAN

Total Kasus	Haji		Entropy
	Sum (Ya)	Sum (Tidak)	
150	73	77	0.999487
$=(-73/150)*\text{LOG2}(73/150) + (-77/150)*\text{LOG2}(77/150)$ $=(-0.486666667)*\text{LOG2}(0.486666667) +$ $(-0.513333333)*\text{LOG2}(0.513333333)$ $=(-0.486666667)*(-1.0389941) + (-0.513333333)*(-0.9620321)$ $=0.999487$			
Pensiun			
150	72	78	0.9988455
$=(-72/150)*\text{LOG2}(72/150) + (-78/150)*\text{LOG2}(78/150)$ $=(-0.48)*\text{LOG2}(0.48) + (-0.513333333)*\text{LOG2}(0.513333333)$ $=(-0.52)*(-1.0588936) + (-0.52)*(-0.9434164)$ $=0.9988455$			

B. Hasil Perhitungan Entropi dan Gain Seluruh Nilai Atribut

TABEL VII  
PERBANDINGAN NILAI GAIN TERBESAR

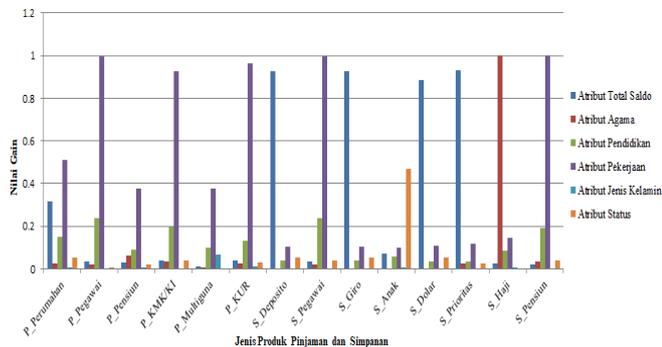
Jenis Produk	Atribut			
	Total Saldo	Agama	Pendidikan	.....
<b>Pinjaman:</b>				
Perumahan	0.3147238	0.0274458	0.1505943	.....
Pegawai	0.0358525	0.0205552	0.2383941	.....
Pensiun	0.0281809	0.0626893	0.0895250	.....
KMK/KI	0.0409646	0.0350216	0.1976897	.....
Multiguna	0.0114408	0.0067973	0.0989723	.....
KUR	0.0377902	0.0252171	0.1313053	.....
<b>Deposito</b>				
Deposito	<b>0.9248187</b>	0.0031641	0.0372458	.....
Pegawai	0.0358525	0.0205552	0.2383941	.....
Giro	<b>0.9248187</b>	0.0031641	0.0372458	.....
Anak	0.0707965	0.0046859	0.0585494	.....
Dolar	<b>0.8845103</b>	0.0045579	0.0333379	.....
Prioritas	<b>0.9310558</b>	0.0261602	0.0353117	.....
Haji	0.0246883	<b>0.9994870</b>	0.0846128	.....
Pensiun	0.0188859	0.0346801	0.1919196	.....

TABEL VII LANJUTAN

.....	Jenis Produk	Atribut		
		Pekerjaan	Jenis Kelamin	Status
<b>Pinjaman:</b>				
.....	Perumahan	<b>0.5112589</b>	0.0064872	0.0515646
.....	Pegawai	<b>0.9967916</b>	0.0007470	0.0079294
.....	Pensiun	<b>0.3780466</b>	0.0082407	0.0208079
.....	KMK/KI	<b>0.9283850</b>	0.0006358	0.0391760
.....	Multiguna	<b>0.3782416</b>	0.0687464	0.0038280
.....	KUR	<b>0.9626147</b>	0.0101559	0.0302746
<b>Simpanan:</b>				
.....	Deposito	0.1044338	0.0013625	0.0536031
.....	Pegawai	<b>0.9967916</b>	0.0007470	0.0402945
.....	Giro	0.1044338	0.0013625	0.0536031
.....	Anak	0.0997059	0.0052951	<b>0.4689956</b>
.....	Dolar	0.1106709	0.0024625	0.0525562
.....	Prioritas	0.1186253	0.0024625	0.0239474
.....	Haji	0.1470731	0.0066336	0.0033602
.....	Pensiun	<b>0.9988455</b>	0.0004004	0.0378342

Hasil perhitungan atribut, nilai, banyaknya kejadian nilai, entropi dan gain Untuk setiap produk Pinjaman dan simpanan sesuai persamaan (2) akan menghasilkan nilai gain terbesar

yang menjadi node dalam pembentukan pohon keputusan. Berdasarkan data perbandingan nilai gain pada tabel VII dan grafik perbandingan pada gambar 3 terlihat atribut pekerjaan merupakan atribut yang memiliki jumlah frekuensi terbanyak sebagai node dalam pembentukan pohon keputusan.



Gambar 3. Grafik Perbandingan Nilai Gain Terbesar

### C. Tools Prediksi Kebutuhan Nasabah

Tools prediksi dibuat dengan menggunakan VBA Excel (*Visual Basic For Application*). Penentuan fasilitas produk pinjaman dan simpanan didasarkan pada hasil penerapan algoritma C4.5 dan pohon keputusan.

Sebagai contoh, dilakukan penginputan data pada atribut yang menjadi dasar pengujian : total saldo  $\rightarrow$  756.000.000, agama  $\rightarrow$  islam, pekerjaan  $\rightarrow$  pegawai swasta, jenis kelamin  $\rightarrow$  pria, status  $\rightarrow$  menikah, jumlah anak  $\rightarrow$  2. Tools prediksi akan menghasilkan produk yang dapat ditawarkan kepada nasabah berupa pinjaman  $\rightarrow$  perumahan, pegawai, multiguna, simpanan  $\rightarrow$  deposito, pegawai, giro, anak, dolar, prioritas, haji, pensiun. Hasil prediksi terlihat pada gambar 4.

Gambar 4. Contoh Simulasi Prediksi Kebutuhan Nasabah

## IV. KESIMPULAN

Setelah dilakukan penelitian terhadap 150 data nasabah dengan menggunakan algoritma C4.5 disertai pembentukan pohon keputusan, terlihat bahwa atribut pekerjaan merupakan atribut yang paling berpengaruh (nilai gain terbesar) diantara atribut-atribut lainnya, karena memiliki jumlah frekuensi terbanyak sebagai *node* dalam pembentukan pohon keputusan. Mini *tools* prediksi kebutuhan nasabah yang didesain berdasarkan hasil pengujian menggunakan algoritma C4.5

sangat berguna bagi staff pemasaran dalam menentukan jenis produk yang ditawarkan kepada nasabah.

Untuk pengembangan lebih lanjut alangkah baiknya digunakan data nasabah dalam jumlah yang lebih besar agar lebih maksimal dalam proses pengujian data. Proses *cleansing* data dilakukan secara berkala agar tidak terdapat nilai atribut yang kosong serta *tools* prediksi dikembangkan mengikuti perkembangan teknologi dan terkoneksi langsung pada database nasabah *core banking system* agar memperoleh data yang *update* dalam proses prediksi.

## UCAPAN TERIMA KASIH

Terima kasih penulis ucapkan kepada pihak APTIKOM yang telah menyelenggarakan pendidikan jarak jauh (PJJ), Ibu Kusri sebagai dosen pembimbing, teman-teman PJJ angkatan 4 dan STIKOM Ambon atas kerjasamanya.

## REFERENSI

- [1] Zakirov D., Momtselidze N., "Application of Data Mining in the Banking Sector", *Journal of Technical Science and Technologies*, ISSN 2298-0032, Vol. 4 Issue 1, 2015.
- [2] Preethi M., Vijayalakshmi M., "Data Mining In Banking Sector", *International Journal of Advanced Networking & Applications (IJANA)*, Vol. 08 Issue: 05, pp. 1-4, 2017.
- [3] Zorić B.A., "Predicting Customer Churn In Banking Industry Using Neural Networks", *Interdisciplinary Description of Complex Systems* 14(2), pp. 116-124, 2016.
- [4] Khodamoradi M., Mosa N.R., "The application of data mining techniques in risk management in banking industry", *Extensive Journal of Applied Sciences*, pp. 163-167, 2016.
- [5] Elsalamony A.H., "Bank Direct Marketing Analysis of Data Mining Techniques", *International Journal of Computer Applications (0975 – 8887)*, Vol. 85 No. 7, Jan. 2014.
- [6] Apampa O., "Evaluation of Classification and Ensemble Algorithms for Bank Customer Marketing Response Prediction", *Journal of International Technology and Information Management*, Vol. 25 Issue 4 Article 6, 2016.
- [7] Hermawati A.F, *Data Mining*, Andi, Yogyakarta, 2013.
- [8] Prasetyo E., *Data Mining Konsep dan Aplikasi Menggunakan Matlab*, Andi, Yogyakarta, 2012.
- [9] Suyanto, *Data Mining untuk Klasifikasi dan Klusterisasi Data*, Informatika, Bandung, 2017.
- [10] Harryanto F.F., Hansun S., "Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE", *JatISI*, Vol. 3 No. 2, Mar. 2017.

