

Penerapan *Data Mining* dalam Mengklasifikasi Nasabah Sebagai Rekomendasi Mengikuti Program Pesta Hadiah dan Program Labbaik *Special Gift* Menggunakan Algoritma C4.5 Di Bank Syariah Indonesia KCP Bintaro Kebayoran Arcade 1

Muhammad Asyroful Nur Maulana Yusuf¹
 Institut Teknologi Sumatera
 e-mail: muhammad.119140026@student.itera.ac.id

Diajukan: 2 Agustus 2022; Direvisi: .. Agustus 2022; Diterima: .. Agustus 2022

Abstrak

PT. Bank Syariah Indonesia KCP Bintaro Kebayoran Arcade 1 merupakan salah satu bank yang menyediakan layanan program pesta hadiah dan program labbaik special gift terhadap nasabah dengan melakukan pembukaan buku rekening serta melakukan deposito dengan nominal jumlah tertentu dan nantinya akan ditukarkan dengan hadiah dalam bentuk barang kepada nasabah. Pada PT. Bank Syariah Indonesia KCP Bintaro Kebayoran Arcade 1 penawaran kedua program tersebut terhadap nasabah terdapat masalah dikarenakan kedua program tersebut memerlukan deposito dengan nominal jumlah tertentu. Sedangkan, pihak nasabah yang berdatangan memiliki nominal saldo rekening, background pekerjaan, dan pendapatan yang berbeda-beda. Dengan menerapkan data mining dan menggunakan algoritma C4.5 data nasabah Bank Syariah Indonesia dapat diprediksi dengan accuracy yang cukup tinggi dengan menggunakan data training dan data testing sebesar 94.65% dan 88.18% dengan dua kasus split data yang berbeda.

Kata Kunci: *Data Mining, Algoritma C4.5, RapidMiner*

Abstract

Pt. Bank Syariah Indonesia KCP Bintaro Kebayoran Arcade 1 is one of the banks that provides gift party program services and special gift lab programs for customers by opening an account book and making deposits with a certain amount and later will be exchanged for prizes in kind to customers. On PT. Bank Syariah Indonesia KCP Bintaro Kebayoran Arcade 1 offers both programs to customers there is a problem because both programs require deposits with a certain nominal amount. Meanwhile, the customers who arrive have different nominal account balances, work backgrounds, and incomes. By applying data mining and using the C4.5 algorithm, Bank Syariah Indonesia customer data can be predicted with fairly high accuracy using training data and testing data of 94.65% and 88.18% with two different split data cases.

Keywords: *Data Mining, Algorithm C4.5, RapidMiner*

1. Pendahuluan

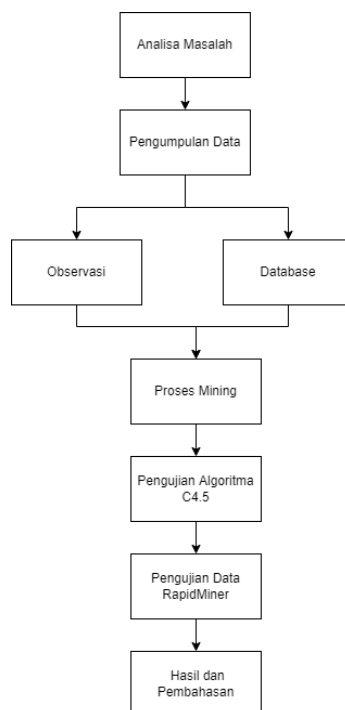
Bank memiliki data besar yang disimpan dalam bentuk database dan dikelola untuk menghasilkan sebuah informasi mengenai nasabah, serta data tersebut dapat digunakan untuk menjaga hubungan antara pihak bank dan nasabah dalam melakukan penawaran produk bank. Perkembangan teknologi saat ini, menyebabkan persaingan antar perusahaan semakin ketat untuk menjadi yang terbaik antara bank syariah dan konvensional. Penerapan *Data Mining* pada industri perbankan menjadi salah satu kunci dalam melakukan identifikasi dan keefektifan dalam mendapatkan, mengelola, dan menganalisa data nasabah, produk, dan layanan pada kegiatan per-bank. PT. Bank Syariah Indonesia KCP Bintaro Kebayoran Arcade 1 merupakan salah satu bank yang menyediakan layanan program pesta hadiah dan program labbaik *special gift* terhadap nasabah dengan melakukan pembukaan buku rekening serta melakukan deposito dengan nominal jumlah tertentu dan nantinya akan ditukarkan dengan hadiah dalam bentuk barang kepada nasabah.

Pada PT. Bank Syariah Indonesia KCP Bintaro Kebayoran Arcade 1 penawaran kedua program tersebut terhadap nasabah terdapat masalah dikarenakan kedua program tersebut memerlukan deposito dengan nominal jumlah tertentu. Sedangkan, pihak nasabah yang berdatangan memiliki nominal saldo rekening, background pekerjaan, dan pendapatan yang berbeda-beda. Sehingga proyek ini bisa dijadikan acuan pihak Bank untuk merekomendasikan nasabah mengikuti program pesta hadiah dan program labbaik *special gift* berdasarkan *record* data yang ada. Oleh karena itu, data nasabah dari Bank Syariah Indonesia KCP Bintaro Kebayoran Arcade 1 menjadi bahan acuan untuk menganalisa nasabah dalam mengikuti

program pesta hadiah dan program labbaik *special gift*. Rekomendasi tersebut berkategori nasabah dapat direkomendasikan dalam mengikuti hanya program pesta hadiah, hanya program labbaik *special gift*, mengikuti kedua program tersebut, atau tidak direkomendasikan mengikuti kedua program tersebut. Dalam kasus ini digunakan teknik data mining decision tree untuk merekomendasikan program terhadap nasabah dengan menggunakan pohon keputusan algoritma C4.5. Penerapan data mining ini nantinya akan saya buat sebagai project dan akan dimasukkan kedalam laporan dan jurnal kerja praktek yang berjudul “Penerapan Data Mining dalam Mengklasifikasi Nasabah Sebagai Rekomendasi Mengikuti Program Pesta Hadiah dan Program Labbaik *Special Gift* Menggunakan Algoritma C4.5”. Batasan pengerjaan laporan ini hanya menganalisa data nasabah dan merekomendasikan nasabah hanya dengan dua program Bank Syariah Indonesia yaitu program pesta hadiah dan program labbaik *special gift*. Semoga dengan analisa tersebut dapat memudahkan pihak Bank dalam merekomendasikan nasabah untuk mengikuti kegiatan program pesta hadiah dan program labbaik *special gift*.

2. Metode Penelitian

Adapun metode yang digunakan dalam menyelesaikan proyek penerapan data mining dalam mengklasifikasi nasabah sebagai rekomendasi mengikuti program pesta hadiah dan program labbaik *special gift* menggunakan algoritma C4.5 sebagai berikut :



Dapat dilihat pada gambar di atas bahwa penyelesaian proyek dilakukan secara bertahap. Mulai dari tahap Analisa masalah di lapangan lebih tepatnya di Bank Syariah Indonesia KCP Bintaro Kebayoran Arcade 1. Kemudian dilanjutkan dengan tahapan pengumpulan data yang dilakukan dalam observasi dan pengambilan data dari database Bank Syariah Indonesia melalui website T24 BSI. Tahapan selanjutnya melakukan proses mining dimana melakukan *data selection*, *data preprocessing* atau *data cleaning*, *data transformation*. Tahap selanjutnya melakukan pengujian algoritma C4.5 dengan menghitung nilai entropy dan gain setelah itu menganalisa data dengan melakukan visualisasi data. Tahapan selanjutnya melakukan pengujian data dengan RapidMiner dimana menghasilkan pemodelan pohon keputusan (*decision tree*) dimana akan menghasilkan aturan-aturan dari pohon keputusan dan menghasilkan hasil *accuracy* dari pengujian algoritma C4.5 dalam mengklasifikasi data nasabah. Dari pengujian yang sudah dilakukan maka hasil tersebut dijabarkan dalam penulisan laporan.

2.1 Data Mining

Data mining merupakan proses ekstraksi pengetahuan dari data yang besar. Berdasarkan fungsinya, data mining merupakan proses yang digunakan untuk membantu dalam mengambil suatu keputusan. Adapun pengertian lainnya mengenai data mining adalah sebuah proses untuk menemukan pola, hubungan, dan trend berdasarkan data yang besar dengan menggunakan teknik pengenalan pola seperti teknik statistika

dan matematika[1].

Tahapan proses dalam data mining sebagai berikut:

A. *Data Selection*

Pemilihan data baru dari sekumpulan data operasional dan dilakukan sebelum tahap pemanggilan informasi KDD (Knowledge Discovery in Database). Data hasil seleksi yang dilakukan disimpan dalam suatu file secara terpisah dari basis data operational.

B. *Preprocessing / Cleaning*

Proses cleaning merupakan proses dimana data dipersiapkan untuk dianalisis pada proses selanjutnya, dengan cara memodifikasi atau menghapus suatu data yang tidak berkaitan, tidak lengkap, dan tidak dibutuhkan.

C. *Transformation*

Melakukan pengodingan ataupun bahasa pemrograman terhadap data yang telah dipilih, sehingga data tersebut sesuai dengan proses data mining. Proses coding dalam pengolahan data merupakan proses dalam menentukan pola informasi yang akan dicari dalam basis data.

D. *Data Mining*

Data mining merupakan proses mencari suatu pola ataupun informasi dalam data dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode ataupun algoritma berdasarkan tujuan dan proses dari pengolahan data secara keseluruhan.

E. *Interpretation / Evaluation*

Mengevaluasi dari empat tahapan hasil prediksi yang dihasilkan berdasarkan metode algoritma yang dipilih serta menghasilkan nilai yang mendekati klasifikasi data sebenarnya.

2.2 Algoritma C4.5

Algoritma C4.5 merupakan algoritma pengembangan dari algoritma ID3. Algoritma C4.5 menjadi penyusun pohon keputusan berdasarkan set data training yang berisi kasus-kasus atau tupel dalam basis data, dimana tiap tupel memiliki atribut kontinyu diskrit atau keduanya[2].

Secara umum algoritma C4.5 menjadi penyusun atau pondasi pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar
2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua kasus memiliki kelas yang sama.

Untuk memilih atribut akar berdasarkan pada nilai gain tertinggi dari atribut yang ada[3]. Menghitung nilai gain menggunakan rumus sebagai berikut :

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

Sedangkan perhitungan dari nilai entropy menggunakan rumus sebagai berikut :

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

2.3 Classification

Data Classification memiliki dua tahap dimana tahap pertama adalah membangun suatu model berdasarkan rangkaian data class yang disebut interval model dengan tujuan menganalisa *database tuple*. Dengan adanya atribut class label maka tahapan ini bisa disebut sebagai *supervised learning*, berbeda dengan *unsupervised learning* atau yang dikenal dengan clustering[4].

Dalam klasifikasi terdapat target variabel kategori. Dimana pada proyek ini mengenai penerapan data mining dalam mengklasifikasi nasabah sebagai rekomendasi mengikuti kegiatan program Bank Syariah Indonesia terdapat 4 variabel kategori yaitu nasabah direkomendasikan untuk mengikuti program pesta hadiah, nasabah direkomendasikan untuk mengikuti program labbaik special gift, nasabah direkomendasikan untuk mengikuti dua program BSI, atau nasabah tidak direkomendasikan sama sekali untuk mengikuti kegiatan program pesta hadiah ataupun program labbaik *special gift*.

2.4 Pohon Keputusan (*Decision Tree*)

Pohon keputusan merupakan salah satu metode dalam mengklasifikasi pada penerapan data mining. Pada dasarnya decision tree merubah data menjadi pohon keputusan dengan adanya aturan-aturan keputusan[5].

Pohon keputusan berfungsi untuk mengeksplorasi data serta menemukan pengklasifikasian berdasarkan variabel kategori. Salah satu metode yang sering digunakan dalam membuat pohon keputusan adalah algoritma C4.5 dikarenakan merupakan algoritma terpopuler dalam penyusunan pohon keputusan. Pohon yang dibentuk tidak selalu merupakan pohon biner, namun jika dalam data set menggunakan 2 macam nilai kategorikal maka bentuk dari pohon keputusan merupakan pohon biner[6]. Jikalau dalam dataset berisi lebih dari 2 macam kategorikal maka bentuk pohon keputusan tidak berupa pohon biner.

Kelebihan-kelebihan *decision tree* sebagai berikut:

1. Menyediakan *visual result*
2. Dibangun berdasarkan aturan-aturan yang dapat dimengerti dan dapat dipahami
3. Bersifat prediktif
4. Memungkinkan untuk melakukan prediksi
5. Menamipkan pengklasifikasian dari data yang penting

2.5 *Rapid Miner*

Rapid Miner merupakan suatu perangkat lunak yang dibuat oleh Dr. Markus Hifmann dan Ralf Klinkenberd dari rapid-i.com. Perangkat lunak ini bersifat terbuka (*open source*). *Rapid Miner* merupakan solusi untuk melakukan analisa ataupun prediksi dari data yang besar sehingga menghasilkan suatu informasi atau pengetahuan untuk pengambilan sebuah keputusan dengan menggunakan berbagai macam teknik deskriptif dan prediksi[7]. Penggunaan *Rapid Miner* tidak membutuhkan kodingan khusus dikarenakan semua fasilitas sudah disiapkan pada perangkat lunak tersebut dan dikhususkan untuk penggunaan data mining.

2.6 Bank Syariah

Bank merupakan sebuah institusi keuangan yang menangani debit atau kredit dengan syariat yang diajarkan dalam agama Islam. Bank bisa meminjamkan, menerima serta menyimpan uang nasabah. Bank Syariah Indonesia memberikan pelayanan jasa keuangan dan berbagai produk Bank Syariah seperti tabungan bisnis, program ABATANA (Ayo Berhaji Aman dan Terencana), BSI KUR, Giro Valas, tabungan easy Mudharabah dan Wadiah, gadai emas, dan BSI Tabungan Haji Indonesia.

Terdapat dua program dimana program tersebut dapat menukarkan jumlah deposito nasabah kepada bank dengan produk barang, program tersebut adalah :

A. Program Pesta Hadiah

Program Pesta Hadiah merupakan program Bank Syariah Indonesia dimana nasabah membuka rekening baru dan top up rekening eksisting. Slogan program pesta hadiah Bank Syariah Indonesia “Tambah Dananya, Berbagi Berkahnya. Nabung berkah, bonus hadiah!”

Terdapat beberapa persyaratan bagi nasabah dalam mengikuti program pesta hadiah, diantaranya:

1. Penempatan dana fresh Fund untuk pembukaan Tabungan Easy Mudharabah baru atau eksisting minimal Top Up Rp 10 Juta dan maksimal Top Up Rp 200 Juta.
2. Jangka waktu blokir dapat dipilih 6 atau 12 bulan.
3. Nominal hadiah sesuai tabel dan telah ditambahkan kontribusi Wakaf senilai Rp 10 Ribu/NoA nya.
4. Hadiah dalam bentuk barang / voucher yang telah ditetapkan.
5. Hanya berlaku untuk nasabah individu (non priority) dan non institusi.

Berikut list tabel hadiah program pesta hadiah Bank Syariah Indonesia:

Nominal Blokir Dana	Produk Barang Pesta Hadiah
Rp 10.000.000	Maspion Saucepan 16 Cm
Rp 15.000.000	Oxone Knife
Rp 20.000.000	Bantal Pijat atau 24 Inch Suitcase
Rp 25.000.000	Happycall Pan
Rp 30.000.000	Moegan Wok Pan
Rp 35.000.000	Cookware Set 4 Pcs
Rp 40.000.000	Supra Presto 8 Liter
Rp 45.000.000	Cosmos Juicer

Rp 50.000.000	Bolde Oven Diamond
Rp 100.000.000	Mito Air Fryer
Rp 200.000.000	Kurumi Vacuum Cleaner

B. Program Labbaik *Special Gift*

Program Labbaik *Special Gift* merupakan program Bank Syariah Indonesia mengenai BSI Tabungan Haji Indonesia. Slogan dari program labbaik *special gift* Bank Syariah Indonesia “Top up BSI Tabungan Haji Indonesia dan dapatkan gimmick menarik”.

Deskripsi program labbaik special gift:

1. Top up Dana/Fresh Fund.
2. Diblokir 1 tahun.
3. Mendapatkan gimmick sesuai tiering.

Berikut list tabel hadiah program labbaik *special gift* Bank Syariah Indonesia:

Top Up Nominal	Produk Barang Labbaik <i>Special Gift</i>
Rp 5.000.000 - 10.000.000	Sajadah
Rp 10.000.000	Mukena

2.7 Bank Syariah

Confusion matrix merupakan matrix yang menggambarkan hasil pengukuran performa pada permasalahan klasifikasi machine learning. Nilai yang dihasilkan pada tabel confusion matrix dibagi menjadi 4, diantaranya True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN)[8]. Tabel confusion matrix dapat dilihat pada tabel di bawah berikut:

Tabel 3 Confusion Matrix

	<i>Predicted Positif</i>	<i>Predicted Negative</i>
<i>Actual Positive</i>	TP	FN
<i>Actual Negative</i>	FP	TN

Keterangan untuk tabel 2.3 dinyatakan sebagai berikut:

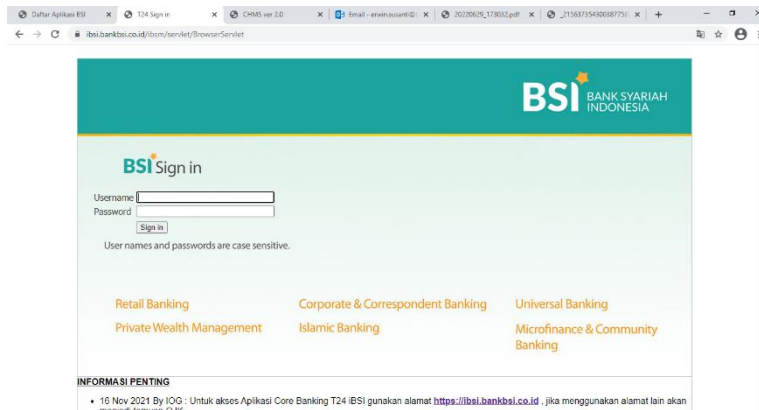
- a. True Positive (TP) : Jumlah data positif yang terdeteksi benar sebagai positif.
- b. True Negative (TF) : Jumlah data negatif yang terdeteksi benar sebagai negatif.
- c. False Positif (FP) : Jumlah data negatif yang terdeteksi positif.
- d. Fale Negative (FN) : Jumlah data positif yang terdeteksi negatif.

Pada proses evaluasi kinerja ini akan menampilkan serta membandingkan nilai sebenarnya dengan nilai hasil prediksi yang dapat digunakan dalam menghasilkan matriks evaluasi. Matriks evaluasi tersebut berupa nilai akurasi, precision, recall, dan F-measure.

3. Hasil dan Pembahasan

3.1 Data Selection

Pada proses data selection, pengambilan data nasabah Bank Syariah Indonesia di KCP Bintaro Kebayoran Arcade 1 menggunakan aplikasi BSI pada web back office T24. Penampilan web back office T24 Bank Syariah Indonesia sebagai berikut :



Data Nasabah Bank Syariah Indonesia di KCP Bintaro Kebayoran Arcade 1 sebanyak 11.442 data dimana data tersebut berdasarkan periode 30 Juni 2022. Pada dataset tersebut 17 atribut data. Dataset nasabah BSI terdapat data pembukaan rekening sesudah dan sebelum Merger BSI. Dataset tersebut sebagai berikut:

No	Tanggal Buka Rekening	Kode Cabang	---	Rekomendasi Program BSI
1	30/06/2022	ID0018158	---	Tidak Direkomendasikan
2	29/06/2022	ID0018158	---	Tidak Direkomendasikan
3	29/06/2022	ID0018158	---	Tidak Direkomendasikan
:	:	:	---	:
11.442	17/01/2011	ID0018158	---	Tidak Direkomendasikan

Tabel di atas menunjukkan bahwa data nasabah berawal dari 17 Januari 2011 sampai 30 Juni 2022, dimana sebelum merger menjadi Bank Syariah Indonesia bank cabang dengan kode ID001858 merupakan BNI Syariah. Dikarenakan terdapat atribut yang tidak digunakan dalam pengolahan data nasabah dan pembuatan pohon keputusan maka diperlukannya data cleaning dengan bertujuan untuk membersihkan dataset dari beberapa data yang tidak dibutuhkan dalam pengolahan data.

3.2 Data Cleaning

Pada proses data cleaning hal yang dilakukan adalah memeriksa dataset apakah terdapat missing value serta menghapus atribut data yang tidak digunakan. Pada tahapan ini menggunakan pemrograman python untuk mempermudah dan mempersingkat waktu dalam pemrosesan. Proses data cleaning sebagai berikut :

Pada tahapan pertama yang dilakukan adalah mengimport modul pandas dan numpy pada python dengan bertujuan untuk memproses dan mengelola data file dengan ekstensi .xlsx atau .csv di python. Setelah itu import data excel ke dalam python dengan menggunakan syntax **pd.read_excel ()**.

Pada tahapan kedua melakukan pengecekan informasi dataset data nasabah BSI dengan menggunakan syntax **data.info** serta **data.columns** untuk mengetahui jumlah serta rincian atribut pada dataset.

Pada tahapan ketiga melakukan pengecekan missing value ataupun data kosong pada dataset menggunakan syntax **data.isnull().sum()** bertujuan untuk mengetahui jumlah data missing value berdasarkan atribut pada dataset.

Pada tahapan keempat melakukan penghapusan atribut dataset menggunakan syntax **data.drop()** dimana bertujuan untuk menghilangkan atribut yang tidak digunakan dalam pengolahan data. Setelah itu melakukan pengecekan dataset, apakah data tersebut sudah memiliki atribut yang dibutuhkan dalam pengolahan. Dengan menggunakan syntax **data.to_excel ()** maka data cleaning berhasil diexport dan dilakukan.

Berdasarkan data cleaning yang sudah dilakukan maka dataset nasabah Bank Syariah Indonesia terdapat 5 atribut yang terdiri dari Mt.Uang, Gender, Kategori Saldo, Pekerjaan, dan Rekomendasi Program BSI sebanyak 11.442 data nasabah BSI.

3.3 Data Transformation

Pada proses data transformation melakukan split data pada dataset nasabah Bank Syariah Indonesia. Pada tahapan ini terdapat dua kasus split dataset nasabah Bank Syariah Indonesia : split data yang pertama menggunakan rasio data training dan data testing 60:40, dimana data training sebanyak 6.865

data dan data testing sebanyak 4.577. Sedangkan split data yang kedua dimana memisahkan data training dan data testing berdasarkan waktu merger Bank Syariah Indonesia tanggal 1 Februari 2021, dimana banyak dataset setelah merger Bank Syariah Indonesia sebanyak 702 data nasabah dan sebelum merger sebanyak 10.740 data. Apabila menggunakan rasio perbandingan presentase data nasabah setelah merger dan sebelum merger sebesar 6,14% : 93,86%.

3.4 Pengujian Algoritma C4.5

Pengujian algoritma C4.5 dilakukan berdasarkan dataset nasabah Bank Syariah Indonesia KCP Bintaro Kebayoran Arcade 1 yang sudah melalui tahapan data cleaning. Pengujian data dengan algoritma C4.5 dapat dilakukan dengan menghitung nilai Entropy dan Gain berdasarkan atribut-atribut yang dimiliki. Dimana dari hasil perhitungan nilai Gain tertinggi akan digunakan sebagai simpul akar atau atribut akar pada pohon keputusan. Perhitungan nilai Entropy dan Gain pada dataset nasabah BSI sebagai berikut:

Atribut	Value	Jumlah	K1	K2	K3	K4	Entropy	Gain
Total	-	11.442	10.177	387	492	386	0,676	-
Gender								
	L	5.771	5.044	229	289	209	0,744	0,002
	P	5.671	5.133	158	203	177	0,602	
Kategori Saldo								
	<5 Juta	9.943	9.929	14	0	0	0,000	0,547
	5 Juta - 20 Juta	874	248	338	270	18	1,685	
	20 Juta - 50 Juta	342	0	35	222	85	0,000	
	>50 Juta	283	0	0	0	283	0,000	
Pekerjaan								
	Mahasiswa / Pelajar	3.455	3.455	0	0	0	0,000	0,141
	Ibu Rumah Tangga	1.033	1.019	8	0	6	0,000	
	Wiraswasta	2.251	2.083	87	30	51	0,492	
	Karyawan Swasta	4.621	3.620	292	462	247	1,086	
	BUMN	82	0	0	0	82	0,000	

Keterangan :

K1 : Tidak Direkomendasikan

K2 : Labbaik Special Gift

K3 : Pesta Hadiah

K4 : Pesta Hadiah dan Labbaik Special Gift

Perhitungan Entropy :

$$1. \text{Entropy Total} = ((-K1/Jumlah) * \log_2(K1/Jumlah)) + ((-K2/Jumlah) * \log_2(K2/Jumlah)) + ((-K3/Jumlah) * \log_2(K3/Jumlah)) + ((-K4/Jumlah) * \log_2(K4/Jumlah))$$

$$\text{Entropy Total} = ((-10177/11442) * \log_2(10177/11442)) + ((-387/11442) * \log_2(387/11442)) + ((-492/11442) * \log_2(492/11442)) + ((-386/11442) * \log_2(386/11442))$$

$$\text{Entropy Total} = 0,676$$

$$2. \text{Entropy Gender (L)} = ((-K1/Jumlah) * \log_2(K1/Jumlah)) + ((-K2/Jumlah) * \log_2(K2/Jumlah)) + ((-K3/Jumlah) * \log_2(K3/Jumlah)) + ((-K4/Jumlah) * \log_2(K4/Jumlah))$$

$$\text{Entropy Gender (L)} = ((-5044/5771) * \log_2(5044/5771)) + ((-229/5771) * \log_2(229/5771)) + ((-289/5771) * \log_2(289/5771)) + ((-209/5771) * \log_2(209/5771))$$

$$\text{Entropy Gender (L)} = 0,744$$

$$3. \text{Entropy Gender (P)} = ((-K1/Jumlah) * \log_2(K1/Jumlah)) + ((-K2/Jumlah) * \log_2(K2/Jumlah)) + ((-K3/Jumlah) * \log_2(K3/Jumlah)) + ((-K4/Jumlah) * \log_2(K4/Jumlah))$$

$$\text{Entropy Gender (P)} = ((-5133/5671) * \log_2(5133/5671)) + ((-158/5671) * \log_2(158/5671)) + ((-203/5671) * \log_2(203/5671)) + ((-177/5671) * \log_2(177/5671))$$

$$\text{Entropy Gender (L)} = 0,602$$

$$4. \text{Entropy } <5 \text{ Juta} = ((-K1/\text{Jumlah}) * \log_2(K1/\text{Jumlah})) + ((-K2/\text{Jumlah}) * \log_2(K2/\text{Jumlah})) + ((-K3/\text{Jumlah}) * \log_2(K3/\text{Jumlah})) + ((-K4/\text{Jumlah}) * \log_2(K4/\text{Jumlah}))$$

$$\text{Entropy } <5 \text{ Juta} = ((-9929/9943) * \log_2(9929/9943)) + ((-14/9943) * \log_2(14/9943)) + ((-0/9943) * \log_2(0/9943)) + ((-0/9943) * \log_2(0/9943))$$

$$\text{Entropy } <5 \text{ Juta} = 0,744$$

$$5. \text{Entropy } 5 \text{ Juta} - 20 \text{ Juta} = ((-K1/\text{Jumlah}) * \log_2(K1/\text{Jumlah})) + ((-K2/\text{Jumlah}) * \log_2(K2/\text{Jumlah})) + ((-K3/\text{Jumlah}) * \log_2(K3/\text{Jumlah})) + ((-K4/\text{Jumlah}) * \log_2(K4/\text{Jumlah}))$$

$$\text{Entropy } 5 \text{ Juta} - 20 \text{ Juta} = ((-248/874) * \log_2(248/874)) + ((-338/874) * \log_2(338/874)) + ((-270/874) * \log_2(270/874)) + ((-18/874) * \log_2(18/874))$$

$$\text{Entropy } 5 \text{ Juta} - 20 \text{ Juta} = 1,685$$

$$6. \text{Entropy } 20 \text{ Juta} - 50 \text{ Juta} = ((-K1/\text{Jumlah}) * \log_2(K1/\text{Jumlah})) + ((-K2/\text{Jumlah}) * \log_2(K2/\text{Jumlah})) + ((-K3/\text{Jumlah}) * \log_2(K3/\text{Jumlah})) + ((-K4/\text{Jumlah}) * \log_2(K4/\text{Jumlah}))$$

$$\text{Entropy } 20 \text{ Juta} - 50 \text{ Juta} = ((-0/342) * \log_2(0/342)) + ((-35/342) * \log_2(35/342)) + ((-222/342) * \log_2(222/342)) + ((-85/342) * \log_2(85/342))$$

$$\text{Entropy } 20 \text{ Juta} - 50 \text{ Juta} = 0$$

$$7. \text{Entropy } >50 \text{ Juta} = ((-K1/\text{Jumlah}) * \log_2(K1/\text{Jumlah})) + ((-K2/\text{Jumlah}) * \log_2(K2/\text{Jumlah})) + ((-K3/\text{Jumlah}) * \log_2(K3/\text{Jumlah})) + ((-K4/\text{Jumlah}) * \log_2(K4/\text{Jumlah}))$$

$$\text{Entropy } >50 \text{ Juta} = ((-0/283) * \log_2(0/283)) + ((-0/283) * \log_2(0/283)) + ((-0/283) * \log_2(0/283)) + ((-0/283) * \log_2(0/283))$$

$$\text{Entropy } >50 \text{ Juta} = 0$$

$$8. \text{Entropy Mahasiswa / Pelajar} = ((-K1/\text{Jumlah}) * \log_2(K1/\text{Jumlah})) + ((-K2/\text{Jumlah}) * \log_2(K2/\text{Jumlah})) + ((-K3/\text{Jumlah}) * \log_2(K3/\text{Jumlah})) + ((-K4/\text{Jumlah}) * \log_2(K4/\text{Jumlah}))$$

$$\text{Entropy Mahasiswa / Pelajar} = ((-3455/3455) * \log_2(3455/3455)) + ((-0/3455) * \log_2(0/3455)) + ((-0/3455) * \log_2(0/3455)) + ((-0/3455) * \log_2(0/3455))$$

$$\text{Entropy Mahasiswa / Pelajar} = 0$$

$$9. \text{Entropy Ibu Rumah Tangga} = ((-K1/\text{Jumlah}) * \log_2(K1/\text{Jumlah})) + ((-K2/\text{Jumlah}) * \log_2(K2/\text{Jumlah})) + ((-K3/\text{Jumlah}) * \log_2(K3/\text{Jumlah})) + ((-K4/\text{Jumlah}) * \log_2(K4/\text{Jumlah}))$$

$$\text{Entropy Ibu Rumah Tangga} = ((-1019/1033) * \log_2(1019/1033)) + ((-8/1033) * \log_2(8/1033)) + ((-0/1033) * \log_2(0/1033)) + ((-6/1033) * \log_2(6/1033))$$

$$\text{Entropy Ibu Rumah Tangga} = 0$$

$$10. \text{Entropy Wiraswasta} = ((-K1/\text{Jumlah}) * \log_2(K1/\text{Jumlah})) + ((-K2/\text{Jumlah}) * \log_2(K2/\text{Jumlah})) + ((-K3/\text{Jumlah}) * \log_2(K3/\text{Jumlah})) + ((-K4/\text{Jumlah}) * \log_2(K4/\text{Jumlah}))$$

$$\text{Entropy Wiraswasta} = ((-2083/2251) * \log_2(2083/2251)) + ((-87/2251) * \log_2(87/2251)) + ((-30/2251) * \log_2(30/2251)) + ((-51/2251) * \log_2(51/2251))$$

$$\text{Entropy Wiraswasta} = 0,492$$

$$11. \text{Entropy Karyawan Swasta} = ((-K1/\text{Jumlah}) * \log_2(K1/\text{Jumlah})) + ((-K2/\text{Jumlah}) * \log_2(K2/\text{Jumlah})) + ((-K3/\text{Jumlah}) * \log_2(K3/\text{Jumlah})) + ((-K4/\text{Jumlah}) * \log_2(K4/\text{Jumlah}))$$

$$\text{Entropy Karyawan Swasta} = ((-3620/4621) * \log_2(3620/4621)) + ((-292/4621) * \log_2(292/4621)) + ((-462/4621) * \log_2(462/4621)) + ((-247/4621) * \log_2(247/4621))$$

$$\text{Entropy Karyawan Swasta} = 1,086$$

$$12. \text{Entropy BUMN} = ((-K1/\text{Jumlah}) * \log_2(K1/\text{Jumlah})) + ((-K2/\text{Jumlah}) * \log_2(K2/\text{Jumlah})) + ((-K3/\text{Jumlah}) * \log_2(K3/\text{Jumlah})) + ((-K4/\text{Jumlah}) * \log_2(K4/\text{Jumlah}))$$

$$\text{Entropy BUMN} = ((-0/82) * \log_2(0/82)) + ((-0/82) * \log_2(0/82)) + ((-0/82) * \log_2(0/82)) + ((-82/82) * \log_2(82/82))$$

$$\text{Entropy BUMN} = 0$$

Perhitungan Gain :

$$1. \text{Gain Gender} = (\text{Entropy Total}) - (((\text{Jumlah(L)}/\text{Jumlah Total}) * \text{Entropy(L)}) + ((\text{Jumlah(P)}/\text{Jumlah$$

Total)*Entropy(P)))

Gain Gender = (0,676) - (((5771/11442)*0,744) + ((5671/11442)*0,602)))

Gain Gender = 0,002

2. Gain Kategori Saldo = (Entropy Total) - (((Jumlah (<5 Juta)/Jumlah Total)*Entropy(<5 Juta)) + ((Jumlah(5 Juta – 20 Juta)/Jumlah Total)*Entropy(5 Juta – 20 Juta)) + ((Jumlah (20 Juta – 50 Juta)/Jumlah Total)*Entropy(20 Juta – 50 Juta)) + ((Jumlah (>50 Juta)/Jumlah Total)*Entropy(>50 Juta)))

Gain Kategori Saldo = (0,676) - (((9943/11442)*0) + ((874/11442)*1,685) + ((342/11442)*0) + ((283/11442)*0)))

Gain Kategori Saldo = 0,547

3. Gain Pekerjaan = (Entropy Total) - (((Jumlah (Mahasiswa/Pelajar)/Jumlah Total)*Entropy(mahasiswa/pelajar) + ((Jumlah(Ibu Rumah Tangga)/Jumlah Total)*Entropy(Ibu Rumah Tangga) + ((Jumlah (Wiraswasta)/Jumlah Total)*Entropy(Wiraswasta)) + ((Jumlah(Karyawan Swasta)/Jumlah Total)*Entropy(Karyawan Swasta)) + ((Jumlah(BUMN)/Jumlah Total)*Entropy(BUMN)))

Gain Pekerjaan = (0,676) - (((3455/11442)*0) + ((1033/11442)*0) + ((2251/11442)*0,492) + ((4621/11442)*1,086) + ((82/11442)*0)))

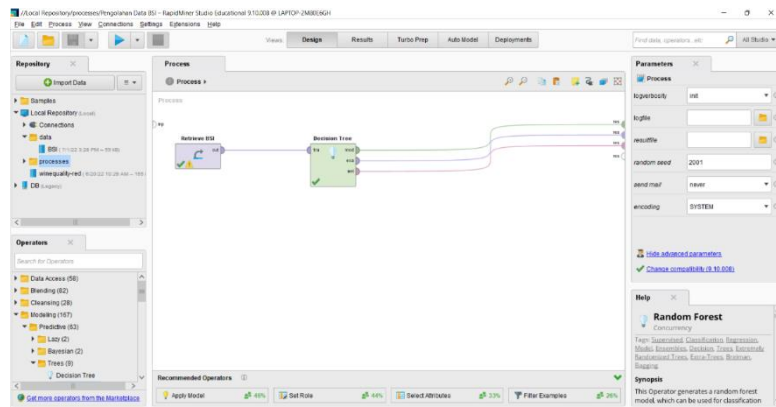
Gain Pekerjaan = 0,141

Berikut hasil visualisasi data dari dataset Nasabah Bank Syariah Indonesia KCP Bintaro Kebayoran Arcade 1:

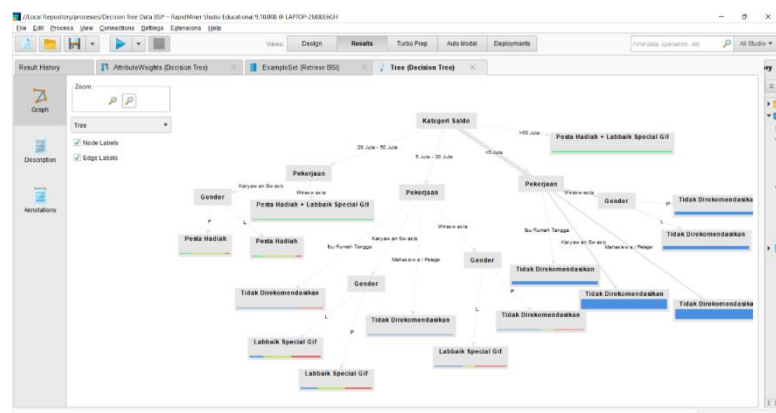


3.5 Pengujian Data Dengan RapidMiner

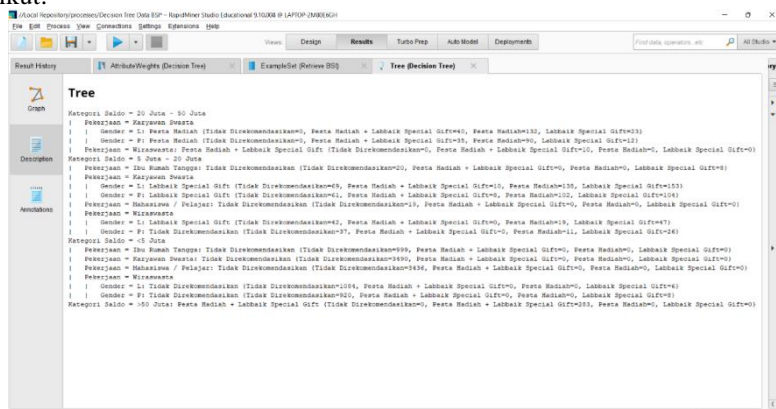
Pada pengujian dataset nasabah Bank Syariah Indonesia KCP Bintaro Kebayoran Arcade 1 sebanyak 11.442 data dengan 5 atribut data dan 1 atribut label. Pengujian menggunakan aplikasi RapidMiner seperti pada gambar di bawah berikut:



3.6 Hasil Pemodelan Pohon Keputusan



Pohon keputusan C4.5 pada gambar di atas dibaca dari atas ke bawah atau dari akar sampai ke daun. Cara membacanya dengan mengacu pada tiap simpulnya. Jika kategori saldo 20 Juta – 50 Juta dengan Pekerjaan Karyawan Swasta dan gender adalah laki-laki maka direkomendasikan untuk mengikuti program pesta hadiah. Jika kategori saldo sebesar 20 Juta – 50 Juta dengan pekerjaan karyawan swasta dan gender perempuan maka direkomendasikan mengikuti program Pesta Hadiah. Jika kategori saldo sebesar 20 Juta - 50 Juta dengan pekerjaan wiraswasta maka direkomendasikan mengikuti pesta hadiah dan labbaik special gift dan seterusnya. Hasil pemodelan pohon keputusan C4.5 akan menghasilkan sebuah aturan dalam pengambilan keputusan merekomendasikan nasabah untuk mengikuti program pesta hadiah dan program labbaik special gift Bank Syariah Indonesia. Aturan decision tree dalam mengklasifikasi nasabah pada gambar dibawah berikut:

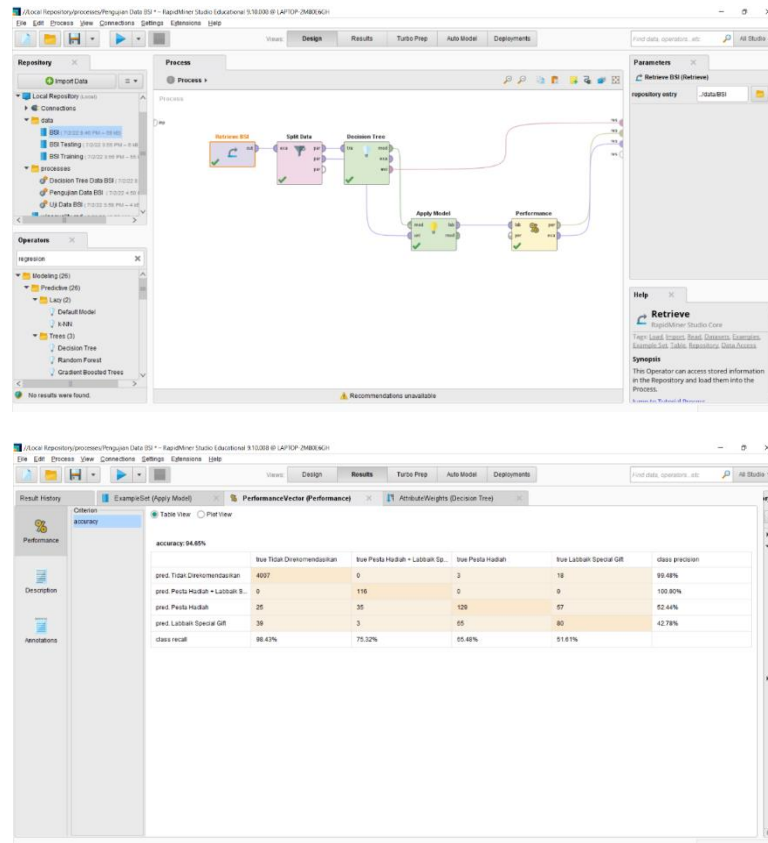


3.7 Hasil Accuracy

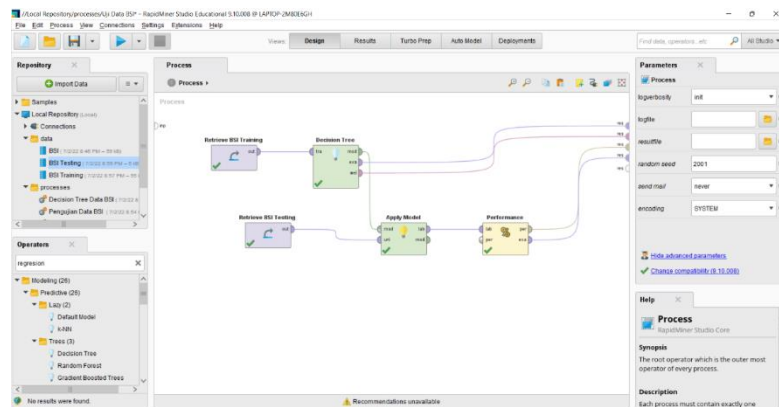
Pada hasil pengukuran hasil accuracy yang diperoleh berdasarkan data training dan data testing dataset nasabah Bank Syariah Indonesia. Terdapat dua kasus dalam melakukan split data pada pengujian

dataset nasabah BSI.

Hasil accuracy pada split data dengan perbandingan rasio data training dan data testing sebesar 60:40 dimana data training sebanyak 6.865 data dan data testing sebanyak 4.577 data memperoleh hasil accuracy sebesar 94.65%. Class recall terendah terdapat pada true labbaik special gift sebesar 51.61 % dan class recall tertinggi terdapat pada true tidak direkomendasikan sebesar 98.43%. Bentuk arsitektur rapidminer dan hasil accuracy ditunjukkan pada gambar di bawah berikut:



Hasil accuracy pada split data dengan perbandingan data training dan data testing sebelum merger dan setelah merger sebesar 93.86%:6,14% dimana data training sebelum merger sebanyak 10.740 data dan data testing setelah merger BSI pada tanggal 1 Februari 2022 sebanyak 702 data. Pada pengujian tersebut hasil accuracy yang diperoleh sebesar 88.18% dengan class recall terendah pada true pesta hadiah sebesar 35.59% dan class recall tertinggi pada true tidak direkomendasikan sebesar 95.26%. Bentuk arsitektur rapidminer dan hasil accuracy ditunjukkan pada gambar di bawah berikut:



	True Total Rekomendasi	True Pesta Hadiah + Labbaik Sp.	True Pesta Hadiah	True Labbaik Special Gift	Class precision
pred. Total Rekomendasi	545	0	0	7	98.72%
pred. Pesta Hadiah + Labbaik S.	0	20	0	0	100.00%
pred. Pesta Hadiah	0	10	21	0	67.74%
pred. Labbaik Special Gift	27	1	38	30	32.65%
Class recall	95.23%	64.52%	35.59%	62.50%	

4. Kesimpulan

Kesimpulan dari pelaksanaan kerja praktik yang dilakukan di Bank Syariah Indonesia KCP Bintaro Kebayoran Arcade 1 ialah:

1. Algoritma C4.5 sangat baik digunakan dalam mengklasifikasi serta merekomendasikan nasabah dengan pohon keputusan (decision tree).
2. Mengklasifikasikan data nasabah BSI dengan menggunakan algoritma C4.5 dapat menggunakan RapidMiner Studio.
3. Terdapatnya aturan-aturan pohon keputusan (decision tree) untuk merekomendasikan program pesta hadiah dan program labbaik special gift terhadap nasabah dengan algoritma C4.5
4. Data nasabah BSI dapat diprediksi dengan accuracy yang cukup tinggi dengan menggunakan data training dan data testing sebesar 94.65% dan 88.18% dengan dua kasus split data yang berbeda.

Daftar Pustaka

- [1] D. C. Aprilla Donny Aji Baskoro Lia Ambarwati I Wayan Simri Wicaksana Editor and R. Sanjaya, "Identitas Belajar Data Mining dengan RapidMiner Hak Cipta © pada Penulis Hak Guna mengikuti Open Content model Desain sampul: Dennis Aprilla C."
- [2] H. Marcos and I. Hidayah, "IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI NASABAH KREDIT BANK "X" MENGGUNAKAN CLASSIFICATION RULE," 2014.
- [3] J. Sistem Komputer dan Sistem Informasi and P. Studi Teknologi Komputasi dan Informatika Stmik Bina Bangsa Kendari, "Router Research Penerapan Data Mining untuk Mengklasifikasi Pola Nasabah Menggunakan Algoritma C4.5 pada Bank Bri Unit Anduonohu Kendari," 2019. [Online]. Available: <http://ejournal.stipwunaraha.ac.id/index.php/router>
- [4] A. Ginanjar Mabur and R. Lubis, "PENERAPAN DATA MINING UNTUK MEMPREDIKSI KRITERIA NASABAH KREDIT," 2012.
- [5] J. A. Ginting, "DATA MINING UNTUK ANALISA PENGAJUAN KREDIT DENGAN MENGGUNAKAN METODE LOGISTIK REGRESI," *Jurnal Algoritma, Logika dan Komputasi*, vol. 2, no. 2, Nov. 2019, doi: 10.30813/j-alu.v2i2.1845.
- [6] M. Yunus, H. Ramadhan, D. R. Aji, and A. Yulianto, "Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indonesia Pintar (KIP)," *Paradigma - Jurnal Komputer dan Informatika*, vol. 23, no. 2, Oct. 2021, doi: 10.31294/p.v23i2.11395.
- [7] A. A. Damanik, Z. Azhar, and A. Sapta, "Penerapan Data Mining Mengklasifikasi Pola Nasabah Menggunakan Algoritma C4.5 Pada Pegadaian Tanjungbalai," *JUTSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 2, no. 1, pp. 1–6, Feb. 2022, doi: 10.33330/jutsi.v2i1.1510.
- [8] Isy Karima Fauzia, Budi Arif Dermawan, and Tesa Nur Padilah, "Penerapan K-Means Clustering pada Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Kabupaten Karawang," *Jurnal Sistem dan Informatika (JSI)*, vol. 15, no. 1, pp. 81–87, Nov. 2020, doi: 10.30864/jsi.v15i1.350.