

**KLASIFIKASI PEMBIAYAAN WARUNG MIKRO  
MENGUNAKAN METODE *RANDOM FOREST* DENGAN TEKNIK  
*SAMPLING KELAS IMBALANCED*  
(Studi Kasus: Data Nasabah Pembiayaan Warung Mikro Bank Syariah  
Mandiri KC Jambi)**

**TUGAS AKHIR**



Disusun oleh:

Julia Widiastuti  
14 611 240

**PROGRAM STUDI STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS ISLAM INDONESIA  
YOGYAKARTA  
2018**

**KLASIFIKASI PEMBIAYAAN WARUNG MIKRO MENGGUNAKAN  
METODE *RANDOM FOREST* DENGAN TEKNIK *SAMPLING* KELAS  
*IMBALANCED***

**(Studi Kasus: Data Nasabah Pembiayaan Warung Mikro Bank Syariah  
Mandiri KC Jambi)**

**TUGAS AKHIR**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana  
Jurusan Statistika**



Disusun oleh:

Julia Widiastuti

14 611 240

**PROGRAM STUDI STATISTIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS ISLAM INDONESIA**

**YOGYAKARTA**

**2018**

## HALAMAN PERSETUJUAN PEMBIMBING

### TUGAS AKHIR

Judul : Klasifikasi Pembiayaan Warung Mikro Menggunakan Metode *Random Forest* Dengan Teknik *Sampling* Kelas *Imbalanced* (Studi Kasus: Data Nasabah Pembiayaan Warung Mikro Bank Syariah Mandiri KC Jambi)

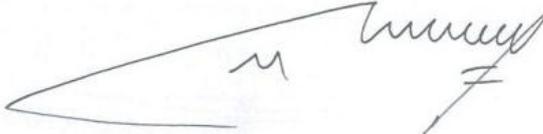
Nama Mahasiswa : Julia Widiastuti

Nomor Mahasiswa : 14 611 240

TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK  
DIUJIKAN

Yogyakarta, 11 Mei 2018

Pembimbing



Muhammad Muhajir, S.Si., M.Sc.

## HALAMAN PENGESAHAN

### TUGAS AKHIR

#### KLASIFIKASI PEMBIAYAAN WARUNG MIKRO MENGUNAKAN METODE *RANDOM FOREST* DENGAN TEKNIK *SAMPLING KELAS IMBALANCED*

(Studi Kasus: Data Nasabah Pembiayaan Warung Mikro Bank Syariah  
Mandiri KC Jambi)

Nama Mahasiswa : Julia Widiastuti

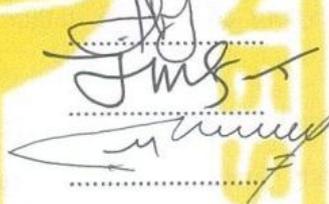
Nomor Mahasiswa : 14 611 240

TUGAS AKHIR INI TELAH DIUJIKAN  
PADA TANGGAL 11 MEI 2018

Nama Penguji

1. Ir. Suparna, M.Si.
2. Prof. Akhmad Fauzy, S.Si., M.Si., Ph.D.
3. Muhammad Muhajir, S.Si., M.Sc.

Tanda Tangan



Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Drs. Alwar, M.Sc., Ph.D.

## KATA PENGANTAR



*Assalamu'alaikum Wr. Wb.*

*Alhamdulillah* rabbil 'alamin, puji syukur penulis ucapkan kepada Allah SWT atas limpahan rahmat, taufik, hidayah, dan inayat-Nya serta tidak lupa shalawat dan salam selalu tercurahkan kepada Nabi Muhammad SAW atas petunjuk untuk selalu berada di jalan yang diridhoi-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul **“Klasifikasi Pembiayaan Warung Mikro Menggunakan Metode *Random Forest* Dengan Teknik *Sampling* Kelas *Imbalanced* (Studi Kasus: Data Nasabah Pembiayaan Warung Mikro Bank Syariah Mandiri KC Jambi)”**.

Penulis menyadari bahwa penulisan tugas akhir ini banyak memperoleh bantuan dari berbagai pihak, baik yang berupa saran, kritik, bimbingan maupun bantuan lainnya. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih kepada:

1. Bapak Drs. Allwar, M.Sc., Ph.D., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia yang telah memberikan izin penelitian.
2. Bapak Dr. RB Fajriya Hakim, S.Si., M.Si., selaku Ketua Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia yang telah banyak membantu dan memberi ilmu serta wawasan baru kepada penulis.
3. Bapak Muhammad Muhajir, S.Si., M.Sc., selaku Dosen Pembimbing Tugas Akhir atas arahan dan bimbingan beliau selama ini, dari laporan kerja praktik hingga tugas akhir yang selalu sabar mendengar keluh kesah penulis.
4. Seluruh Dosen dan Staff Program Studi Statistika yang telah banyak

memberikan bimbingan dan bantuan kepada penulis.

memberikan bimbingan dan bantuan kepada penulis.

5. Papa, mama, dan adik-adik tersayang yang telah memberikan do'a, dukungan, dan bantuan dalam bentuk apapun sehingga kuliah dan tugas akhir penulis dapat dilancarkan serta dimudahkan.
6. Cikma (Sri Rahma) yang telah memfasilitasi dan bersedia dimintai tolong tentang informasi tugas akhir penulis.
7. Sahabat dan juga keluarga kedua selama di Yogyakarta, yaitu Nanda, Ellak, Tistun, Mamih Septi, Tiwi, Handut, Yusi, Mei, Dhila, Feby, Udin, Husnik, Febrian, Fizhan, Sendhy, Alan, Aufa yang selalu ada ketika penulis sedang dalam kesusahan dari awal hingga akhir kuliah selama empat tahun ini.
8. Furqon Tri Kurniawan yang selalu sabar memberikan semangat, dukungan, dan menemani penulis hingga tugas akhir ini selesai.
9. Anak-anak kost Nganggrung yaitu Nanda, Mba mamal, Dhesya, Pretisya, dan lainnya yang selalu membawa keceriaan dan kebahagiaan selama di Yogyakarta.
10. Teman-teman XIX, Statistika UII angkatan 2014, yang telah banyak membantu penulis serta memberikan dukungan dalam usaha penyelesaian tugas akhir ini.
11. Semua pihak yang telah membantu dalam penyusunan tugas akhir ini yang tidak dapat penulis sebutkan satu persatu.

Semoga Allah SWT selalu memberikan rahmat dan hidayah-Nya kepada mereka. Penulis menyadari bahwa tugas akhir ini masih jauh dari kata sempurna, oleh karena itu segala kritik dan saran yang bersifat membangun selalu penulis harapkan. Semoga tugas akhir ini dapat bermanfaat bagi penulis khususnya dan pihak-pihak lain yang membutuhkan.

Wassalamu'alaikum Wr. Wb.

Yogyakarta, 11 Mei 2018

  
Penulis



## DAFTAR ISI

<b>HALAMAN PERSETUJUAN PEMBIMBING</b> ....Error! Bookmark not defined.	
<b>HALAMAN PENGESAHAN</b> .....	<b>iii</b>
<b>KATA PENGANTAR</b> .....	<b>iv</b>
<b>DAFTAR ISI</b> .....	<b>vi</b>
<b>DAFTAR TABEL</b> .....	<b>ix</b>
<b>DAFTAR GAMBAR</b> .....	<b>xi</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>xiii</b>
<b>PERNYATAAN</b> .....Error! Bookmark not defined.	
<b>INTISARI</b> .....	<b>xv</b>
<b>ABSTRACT</b> .....	<b>xvi</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	6
1.3. Batasan Masalah .....	6
1.4. Tujuan Penelitian .....	6
1.5. Manfaat Penelitian .....	7
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>8</b>
<b>BAB III LANDASAN TEORI</b> .....	<b>11</b>
3.1. Usaha Mikro, Kecil, dan Menengah (UMKM).....	11
<b>3.1.1. Pengertian UMKM</b> .....	11
<b>3.1.2. Karakteristik UMKM</b> .....	12
<b>3.1.3. Profil UMKM di Indonesia</b> .....	12
3.2. Bank Syariah.....	13
3.2. Pembiayaan .....	16
<b>3.2.1. Pengertian Pembiayaan</b> .....	16
<b>3.2.2. Tujuan dan Manfaat Pembiayaan</b> .....	17
<b>3.2.3. Akad-akad Pembiayaan Syariah</b> .....	18

3.3.	Bank Syariah Mandiri (BSM) .....	20
	<b>3.3.1. Produk-produk Pembiayaan BSM</b> .....	20
	<b>3.3.2. Pembiayaan Warung Mikro BSM</b> .....	23
	<b>3.3.3. Syarat dan Prosedur Pembiayaan Warung Mikro</b> .....	24
3.4.	Analisis Deskriptif .....	26
3.5.	<i>Data Mining</i> .....	31
3.6.	<i>Machine Learning</i> .....	35
3.7.	Klasifikasi .....	36
3.8.	<i>Decision Tree</i> .....	38
3.9.	<i>Random Forest</i> .....	38
3.10.	Teknik Sampling Kelas <i>Imbalanced</i> .....	45
<b>BAB IV METODOLOGI PENELITIAN</b> .....		<b>46</b>
4.1.	Populasi Penelitian .....	46
<b>4.2.</b>	<b>Metode Penelitian</b> .....	46
<b>4.3.</b>	<b>Variabel Penelitian</b> .....	46
<b>4.4.</b>	<b>Teknik Pengumpulan Data</b> .....	47
<b>4.5.</b>	<b>Alat dan Cara Organisir Data</b> .....	48
<b>4.6.</b>	<b>Diagram Alir Penelitian</b> .....	49
<b>BAB V HASIL DAN PEMBAHASAN</b> .....		<b>50</b>
5.1.	Analisis Deskriptif .....	50
5.2.	<i>Random Forest</i> .....	57
5.3.	<i>Random Forest</i> Menggunakan Teknik <i>Sampling</i> Kelas <i>Imbalanced</i> ....	61
5.3.1.	<i>Random Forest</i> Menggunakan <i>Undersampling</i> .....	61
5.3.2.	<i>Random Forest</i> Menggunakan <i>Oversampling</i> .....	64
5.3.3.	<i>Random Forest</i> Menggunakan SMOTE .....	66
5.4.	Ukuran Tingkat Kepentingan .....	69

6.1 Kesimpulan.....	73
6.2 Saran.....	74
<b>DAFTAR PUSTAKA .....</b>	<b>75</b>
<b>Paper Tugas Akhir .....</b>	<b>80</b>
<b>LAMPIRAN.....</b>	<b>91</b>

## DAFTAR TABEL

Nomor	Judul	Halaman
1	Kriteria UMKM	12
2	Matriks <i>Confusion</i>	43
3	Definisi Operasional Variabel	46
4	Analisis Deskriptif Variabel Umur	51
5	Analisis Deskriptif Variabel Harga Pokok Pembiayaan	53
6	Analisis Deskriptif Variabel Margin	53
7	Analisis Deskriptif Variabel Jumlah Angsuran (Bulan)	54
8	Analisis Deskriptif Harga Jaminan	55
9	Tabel Silang Variabel Jenis Nasabah dan Keterangan Saldo	55
10	Tabel Silang Variabel Jenis Pembiayaan Mikro dan Keterangan Saldo	55
11	Tabel Silang Variabel Jenis Nasabah, Jenis Pembiayaan Mikro, dan Keterangan Saldo	56
12	Data Latih ( <i>Train</i> ) dan Data Uji ( <i>Test</i> )	57
13	Pengujian Nilai <i>Error OOB</i>	58
14	Pengujian Nilai <i>Error OOB</i> Dengan <i>Ntree</i> Berbeda	58
15	Hasil Prediksi Data Latih ( <i>Training</i> )	59
16	Hasil Prediksi Data Uji ( <i>Test</i> )	60
17	Data Latih ( <i>Train</i> )	62
18	Hasil Prediksi Data Latih ( <i>Training</i> ) <i>Random Forest Undersampling</i>	62
19	<i>Confusion Matrix</i> Metode <i>Random Forest Undersampling</i>	62
20	Hasil Prediksi Data Latih ( <i>Training</i> ) <i>Random Forest Oversampling</i>	64
21	<i>Confusion Matrix</i> Metode <i>Random Forest</i>	65

	<i>Oversampling</i>	
22	Hasil Prediksi Data Latih ( <i>Training</i> ) <i>Random Forest</i>	66
	<i>SMOTE</i>	
23	<i>Confusion Matrix</i> Metode <i>Random Forest SMOTE</i>	67
24	Perbedaan Keempat Metode <i>Random Forest</i> yang Digunakan	68

## DAFTAR GAMBAR

Nomor	Judul	Halaman
1	Contoh diagram batang-daun	28
2	Contoh diagram batang	29
3	Contoh diagram garis	29
4	Contoh <i>piechart</i>	30
5	Contoh Piktogram	30
6	Contoh Histogram dan Poligon	31
7	Contoh <i>Ogive</i>	31
8	Contoh <i>Classification (Decision Tree)</i>	33
9	Contoh <i>Clustering</i>	34
10	Contoh <i>Association</i>	34
11	Contoh <i>Forecasting</i>	35
12	Fase <i>Learning</i>	37
13	Fase <i>Testing</i>	37
14	Contoh <i>Random Forest</i>	40
15	Diagram Alir Penelitian	49
16	Tipe Data Variabel-variabel Penelitian	50
17	<i>Piechart</i> Presentase Variabel Keterangan Saldo	51
18	<i>Piechart</i> Presentase Variabel Jenis Nasabah	52
20	<i>Piechart</i> Presentase Variabel Jenis Pembiayaan Mikro	52
21	Grafik Jangka Waktu Angsuran (Bulan)	53
22	Grafik Jenis Pekerjaan	54
23	Grafik Nilai <i>Error OOB</i> Menurut <i>Mtry</i> yang di Dapat	58
24	Hasil klasifikasi <i>random forest</i> menggunakan data latihan ( <i>training</i> )	59
25	Hasil klasifikasi <i>random forest</i> menggunakan data latihan ( <i>training</i> ) Hasil klasifikasi <i>random forest oversampling</i>	62

26	menggunakan data latih ( <i>training</i> )	64
	Hasil klasifikasi <i>random forest SMOTE</i> menggunakan	
27	data latih ( <i>training</i> )	66
	Ukuran Tingkat Kepentingan	
28	<i>Decision Tree</i> yang Dibentuk <i>Random Forest</i>	69
29	<i>Oversampling</i>	70

## DAFTAR LAMPIRAN

- |          |   |  |
|----------|---|--|
| Lampiran | 1 | Sintaks <i>Random Forest</i>   |
| Lampiran | 2 | Sintaks <i>Random Forest</i> Menggunakan Teknik <i>Sampling</i> Kelas <i>Imbalanced</i> serta <i>Variable Importance</i> |

## PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 11 Mei 2018



**KLASIFIKASI PEMBIAYAAN WARUNG MIKRO  
MENGUNAKAN METODE *RANDOM FOREST* DENGAN TEKNIK  
*SAMPLING KELAS IMBALANCED*  
(Studi Kasus: Data Nasabah Pembiayaan Warung Mikro Bank Syariah  
Mandiri KC Jambi)**

Oleh: Julia Widiastuti

*Program Studi Statistika, Fakultas MIPA*

*Universitas Islam Indonesia*

E-mail: [14611240@students.uii.ac.id](mailto:14611240@students.uii.ac.id)

**INTISARI**

*Perekonomian rakyat merupakan prioritas utama dalam pembangunan sosial dengan upaya menciptakan lapangan kerja dan meningkatkan pendapatan masyarakat. Salah satu upaya untuk menciptakan lapangan kerja dan meningkatkan pendapatan masyarakat yaitu Usaha Mikro, Kecil dan Menengah (UMKM). UMKM telah menjadi tonggak perekonomian nasional karena mampu bertahan dan menguasai sebagian besar usaha di Indonesia. Namun, masih terdapat permasalahan pada sektor UMKM, salah satunya adalah kurangnya permodalan. Pinjaman uang dari bank merupakan solusi pada sektor UMKM dalam memenuhi kebutuhan modalnya, salah satu bank yang membantu akses pembiayaan dan pengembangan UMKM yaitu Bank Syariah Mandiri dengan produk Warung Mikro. Namun, dalam pembiayaan terdapat permasalahan seperti kredit atau pembiayaan macet, maka dari itu diperlukan analisis untuk melihat nasabah yang berpotensi untuk mengalami pembiayaan bermasalah. Random forest merupakan teknik data mining untuk melakukan klasifikasi. Namun jika dataset yang digunakan merupakan sebuah data yang imbalanced (tidak seimbang) maka diperlukan pendekatan dengan melakukan resample pada data asli menggunakan teknik sampling kelas imbalanced. Data yang digunakan yaitu data nasabah pembiayaan Warung Mikro tahun 2014-2017. Berdasarkan hasil penelitian, didapatkan bahwa dengan menggunakan random forest nilai sensitivty bernilai 0 yang berarti menghasilkan akurasi yang bias. Setelah menggunakan teknik sampling kelas imbalanced, nilai sensitivity, precision, dan G-mean meningkat dibandingkan menggunakan data asli. Random forest dengan teknik oversampling mempunyai nilai AUC tertinggi yaitu sebesar 66.69%.*

**Kata Kunci:** *UMKM, Warung Mikro, Klasifikasi, Random Forest, Data Imbalanced*

# CLASSIFICATION OF WARUNG MIKRO FINANCING USING RANDOM FOREST METHOD WITH SAMPLING TECHINQUES FOR IMBALANCED CLASS

By: Julia Widiastuti

*Faculty of Mathematics and Science, Department of Statistics*

*Universitas Islam Indonesia*

*E-mail: [14611240@students.uii.ac.id](mailto:14611240@students.uii.ac.id)*

## ABSTRACT

*Citizens economy is a top priority in social development by creating jobs and increasing people's incomes. One way to create employment and increase the income of the community that is Micro, Small and Medium Enterprises (UMKM). UMKM are able to survive and control most of the business in Indonesia and become a national economic milestone. However, there are still problems in the sector of UMKM, one of which is the lack of capital. Loan money from banks is a solution in the sector of UMKM in fulfilling their capital needs. One of the banks that help access financing and development of UMKM is Syariah Mandiri Bank with Warung Micro products. However, in the case of financing, there are problems such as credit or financing bogged down, therefore an analysis is required to see potential customers in experiencing problem financing. Random forest is a data mining technique for classification. However, if the dataset that is being used is an imbalanced data, it is necessary to approach by resample on the original data using the imbalanced class sampling technique. The data that is being used are customer financing data of Warung Micro in 2014-2017. Based on the result of the research, it is found that by using random forest the value of sensitivity is 0, which means to produce bias accuracy. After using the imbalanced class sampling technique, the sensitivity, precision, and G-mean value increase compared to using the original data. Random forest with oversampling technique has the highest AUC score that is equal to 66.69%.*

**Keywords:** *UMKM, Warung Mikro, Classification, Random Forest, Imbalanced Data*

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Prioritas utama dalam pembangunan sosial di Indonesia adalah kesejahteraan rakyat, dimana dapat diwujudkan dengan membangun perekonomian rakyat. Menurut Tedjasuksmana (2014), hal tersebut dapat dilakukan dengan cara menciptakan lapangan kerja yang memadai, meningkatkan pendapatan masyarakat, dan mengurangi kemiskinan. Salah satu upaya untuk menciptakan lapangan kerja dan meningkatkan pendapatan masyarakat yaitu Usaha Mikro, Kecil dan Menengah (UMKM). Jika perekonomian Indonesia diletakkan dalam suatu kuadran dapat dilihat bahwa sebagian besar kegiatan perekonomian Indonesia berada pada perekonomian tradisional yang ditandai dengan sektor usaha kerakyatan yang bersifat *labor intensive*, yaitu sektor UMKM (Abdullah, 2006).

Pemberdayaan UMKM merupakan langkah yang strategis dalam meningkatkan dan memperkuat dasar kehidupan perekonomian dari sebagian besar rakyat Indonesia, khususnya melalui penyediaan lapangan kerja dan mengurangi kesenjangan dan tingkat kemiskinan (Nikmah dkk, 2014). Peran UMKM bagi perekonomian Indonesia tidak diragukan lagi, dimana tingkat penyerapan tenaga kerja sekitar 97% dari seluruh tenaga kerja nasional serta 57% kontribusi terhadap Produk Domestik Bruto (PDB).

Dalam Undang-undang Republik Indonesia Nomor 20 Tahun 2008 Bab I Pasal 1 tentang Usaha Mikro, Kecil dan Menengah dimana dunia usaha adalah usaha mikro, usaha kecil, usaha menengah, dan usaha besar yang melakukan kegiatan ekonomi di Indonesia dan berdomisili di Indonesia. Pada Undang-undang Republik Indonesia Nomor 20 Tahun 2008 Bab III Pasal 5 disebutkan bahwa tujuan pemberdayaan Usaha Mikro, Kecil, dan Menengah untuk mewujudkan struktur perkenomian nasional yang seimbang, berkembang, dan berkeadilan. Usaha Mikro berdasarkan Undang-undang Republik Indonesia Nomor 20 Tahun 2008 adalah usaha produktif milik orang perorangan dan atau

badan usaha yang memenuhi kriteria usaha mikro sebagaimana diatur dalam Undang-undang ini.

UMKM telah menjadi tonggak perekonomian nasional karena mampu bertahan dan menguasai sebagian besar usaha di Indonesia. Hal tersebut dapat dibuktikan dari data Badan Pusat Statistik (BPS) yang memperlihatkan bahwa pasca krisis ekonomi 1997 – 1998 jumlah UMKM tidak berkurang, justru semakin meningkat hingga tahun 2012 UMKM mampu menyerap 85 juta hingga 107 juta tenaga kerja. Jumlah usaha di Indonesia pada tahun 2012 menurut Departemen Koperasi dan Usaha Kecil dan Menengah sebanyak 56.539.560, dimana usaha tersebut didominasi dengan usaha mikro dengan presentase 98.79% atau sebanyak 55.856.179 unit, diikuti oleh usaha kecil dengan presentase 1.11% atau sebanyak 48.997 unit, lalu usaha menengah sebesar 0.01% atau sebanyak 4.968 unit, sedangkan sisanya yaitu 0.01% atau sebanyak 4.968 unit adalah usaha besar (Bank Indonesia, 2015).

Oleh karena itu UMKM merupakan salah satu sektor yang mendapat perhatian khusus dari pemerintah Provinsi Jambi. Menurut data per Juli 2015 yang dikemukakan oleh Kepala dinas Koperasi dan UMKM Provinsi Jambi tercatat sebanyak 81.959 unit UMKM dari 11 kabupaten dan kota dalam Provinsi Jambi. Dari 11 kabupaten kota tersebut Kabupaten Muaro Jambi adalah kabupaten dengan jumlah unit UMKM terbanyak sedangkan kabupaten dengan jumlah unit terkecil adalah Kabupaten Batanghari (Muzakkir, 2016).

Walaupun UMKM telah menjadi tonggak perekonomian nasional, namun masih terdapat permasalahan pada sektor UMKM yang sangat penting bagi pertumbuhan sektor tersebut, salah satu faktor internalnya adalah kurangnya permodalan dan terbatasnya akses pembiayaan bagi pemilik usaha (Suhendar, 2010). Tanpa adanya dana atau modal, usaha tidak akan dapat berjalan dengan baik.

Pinjaman uang dari bank merupakan solusi pada sektor UMKM dalam memenuhi kebutuhan modalnya dibandingkan mengumpulkan uang dengan menabung yang memakan waktu yang cukup lama. Bank merupakan suatu lembaga yang mempunyai peranan penting sebagai lembaga yang dapat

menghimpun dan menyalurkan dana masyarakat secara efektif dan efisien. Pinjaman yang diberikan untuk Provinsi Jambi yang tercatat pada periode Desember 2017 sebesar Rp 41.627.161 dengan kredit UMKM sebesar Rp 12.329.541 (Bank Indonesia, 2017).

Selain bank konvensional, bank syariah juga menawarkan pinjaman sektor ekonomi khususnya sektor UMKM. Mengingat Indonesia adalah negara dengan penduduk muslim terbesar di dunia, perbankan syariah memiliki potensi yang sangat besar dalam perekonomian. Terbentuknya perbankan dengan sistem syariah diharapkan dapat memberikan kontribusi dalam pemenuhan jasa perbankan bagi masyarakat. Bank syariah adalah lembaga intermediasi dan penyedia jasa keuangan yang bekerja berdasarkan etika dan sistem nilai Islam yang mempunyai sifat khusus yakni bebas dari kegiatan spekulatif yang nonproduktif seperti perjudian, bebas dari hal-hal yang tidak jelas dan meragukan (tidak pasti), berprinsip pada keadilan dan hanya membiayai kegiatan usaha yang halal (Ascarya & Diana, 2005).

Melihat ruang lingkup kegiatan usahanya dapat dilihat bahwa produk perbankan syariah lebih variatif dibandingkan dengan produk bank konvensional, dimana bank syariah memberikan peluang yang lebih luas dalam memenuhi kebutuhan nasabah sesuai dengan kebutuhannya (Kara, 2013). Pada Bank Syariah perkreditan dikenal dengan nama pembiayaan, pembiayaan adalah menyediakan dana guna membiayai kebutuhan nasabah yang memerlukannya dan layak memperolehnya (Arifin, 2002). UU No.21 Tahun 2008 tentang Perbankan Syariah Pasal 1 Poin 25 menerangkan bahwa pembiayaan adalah penyediaan dana atau tagihan yang berdasarkan persetujuan atau kesepakatan antara Bank Syariah dan atau Unit Usaha Syariah (UUS) dan pihak lain yang mewajibkan pihak yang dibiayai dan atau diberi fasilitas dana untuk mengembalikan dana tersebut setelah jangka waktu tertentu dengan imbalan *ujrah*, tanpa imbalan, atau bagi hasil. Sistem keuangan syariah hanya membolehkan memberikan pembiayaan berdasarkan aset jelas dengan nilai pembiayaan wajar.

Seperti yang diketahui, di Indonesia terdapat banyak bank syariah dimana salah satu Bank Syariah yang peduli akan pembiayaan untuk UMKM yaitu Bank

Syariah Mandiri (BSM). Pada bank syariah, Kredit Tanpa Angunan (KTA) disebut dengan produk pembiayaan dimana bunga tidak diterapkan pada seluruh aktivitas pembiayaan melainkan bagi hasil atau sistem kemitraan. Salah satu produk BSM yang bertujuan untuk membantu akses pembiayaan dan pengembangan UMKM adalah warung mikro. Dengan adanya warung mikro, maka dapat memaksimalkan peran UMKM dalam menciptakan stabilitas perekonomian rakyat. Warung mikro sendiri memiliki target wirausaha mikro yaitu wirausaha pada sektor unggulan dengan golongan berpenghasilan tetap maupun bukan golongan berpenghasilan tetap.

Pembiayaan usaha mikro pada tahun 2016 menurut laporan keberlanjutan Bank Syariah Mandiri tahun 2016 meningkat dibandingkan tahun 2015 yaitu sebesar Rp 0.65 triliun. Untuk dapat mengajukan peminjaman uang atau kredit warung mikro BSM dapat dilakukan oleh peminjam yang telah memenuhi syarat dari pihak bank. Bagi kebanyakan bank, syarat-syarat perkreditan yaitu kelayakan peminjam, kelayakan hukum, kelayakan bisnis, kelayakan keuangan, dan kelayakan jaminan. Dalam kredit atau pembiayaan tentu harus berhati-hati dalam memilih nasabah, pembiayaan bermasalah adalah sebuah resiko yang tidak dapat dihindari oleh suatu bank atau lembaga keuangan lainnya. Kredit macet atau macetnya pembiayaan adalah salah satu masalah utama, dimana masalah tersebut dapat dipecahkan dengan cara mengelola pembiayaan bermasalah tersebut.

Dalam hal tersebut, pihak bank harus melakukan tindakan agar tidak menimbulkan kerugian, misalnya memberikan keringanan berupa jangka waktu atau angsuran bagi nasabah yang terkena musibah atau penyitaan bagi penundaan pembiayaan nasabah yang sengaja lalai untuk membayar. Maka dari itu, agar menjadi acuan untuk menganalisa karakteristik nasabah pada warung mikro BSM digunakan data nasabah warung mikro Bank Syariah Mandiri KC Jambi.

Berdasarkan uraian di atas diperlukan analisis untuk melihat nasabah yang berpotensi untuk mengalami pembiayaan bermasalah. Pada permasalahan ini dapat diterapkan teknik *data mining* untuk melakukan klasifikasi menggunakan metode *Random Forest*. Metode klasifikasi *Machine Learning* telah lama digunakan di bidang *data mining*. Menurut Han (2001), klasifikasi adalah proses

menemukan kumpulan pola data satu dengan yang lainnya untuk dapat digunakan memprediksi data yang belum memiliki kelas data tertentu. Sebagai algoritma klasifikasi yang melibatkan metode dari statistik, kecerdasan buatan dan manajemen database, juga dapat di kategorikan sebagai kunci elemen dalam interpretasi data dan visualisasi data (Berry, 2004). *Decision tree* atau pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal, selain itu juga berguna untuk mengeksplor data dan menemukan hubungan yang tersembunyi (Witten dan Eibe Frank, 2005).

Salah satu klasifikasi *data mining* tersebut adalah *random forest*. *Random forest* merupakan metode pohon gabungan yang berasal dari pengembangan metode *Classification and Regression Tree* (CART), yaitu dengan menerapkan metode *bootstrap aggregating* (*bagging*) dan *random feature selection* (Breiman, 2001). Metode *random forest* telah diaplikasikan pada berbagai permasalahan dalam bidang kesehatan, bisnis, pendidikan, dan lain-lain.

Akurasi merupakan ukuran kinerja ketepatan klasifikasi dari sebuah algoritma *machine learning*, namun jika dataset yang digunakan merupakan sebuah data yang *imbalanced* (tidak seimbang) maka ukuran akurasi menjadi tidak tepat (Chawla dkk, 2002). Maka dari itu, peneliti menanggapi permasalahan kelas yang *imbalanced* dengan pendekatan melakukan *resample* pada data asli menggunakan tiga teknik *sampling* kelas *imbalanced* yaitu *undersampling*, *oversampling*, dan *Synthetic Minority Oversampling Technique* (SMOTE).

Berdasarkan hal tersebut, maka penulis melakukan penelitian dengan memilih judul **“Klasifikasi Pembiayaan Warung Mikro Menggunakan Metode *Random Forest* Dengan Teknik *Sampling* Kelas *Imbalanced*”** dimana penelitian ini dilakukan untuk mengetahui perbandingan klasifikasi data nasabah menggunakan metode *random forest* sebelum dan sesudah penerapan teknik *sampling* pada kelas *imbalanced* sehingga hasil penelitian ini dapat dijadikan salah satu alternatif untuk pengambilan kebijakan bagi Warung Mikro BSM KC Jambi.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang tersebut, maka dapat diambil rumusan masalah sebagai berikut:

1. Bagaimana gambaran umum pembiayaan Warung Mikro Bank Syariah Mandiri KC Jambi?
2. Bagaimana klasifikasi pembiayaan Warung Mikro Bank Syariah Mandiri KC Jambi menggunakan metode *random forest*?
3. Bagaimana perbandingan klasifikasi pembiayaan Warung Mikro Bank Syariah Mandiri KC Jambi setelah diterapkan teknik *sampling* kelas *imbalanced* pada metode *random forest* ?

## 1.3. Batasan Masalah

Agar pembahasan dalam penelitian ini tidak terlalu meluas, maka terdapat batasan masalah dalam penelitian ini yaitu:

1. Data yang digunakan adalah data sekunder yaitu Dataset Nasabah Warung Mikro Bank Syariah Mandiri KC Jambi tahun 2014 – 2017.
2. Variabel yang digunakan adalah umur, jenis nasabah, jenis pembiayaan mikro, harga pokok pembiayaan, margin, jangka waktu angsuran, besarnya angsuran, pekerjaan, harga jaminan, dan keterangan saldo mencukupi atau tidak.
3. Variabel harga pokok pinjaman diasumsikan sebagai modal awal sehingga dapat digunakan untuk penetapan jenis usaha. Jika harga pokok pinjaman sebesar Rp 15.000.000,00 – Rp 50.000.000,00 dikategorikan sebagai usaha mikro dan jika harga pokok pinjaman sebesar Rp 51.000.000,00 – Rp 200.000.000,00 dikategorikan sebagai usaha kecil.
4. Data diolah dengan menggunakan program *Microsoft Excel* dan R dengan metode analisa yang digunakan adalah statistika deskriptif , *Random Forest* dan teknik *sampling* kelas *imbalanced*.

## 1.4. Tujuan Penelitian

Adapun tujuan penelitian ini adalah sebagai berikut:

1. Mengetahui gambaran umum pembiayaan Warung Mikro Bank Syariah Mandiri KC Jambi.
2. Mengklasifikasikan pembiayaan Warung Mikro Bank Syariah Mandiri KC Jambi menggunakan metode *Random Forest*.
3. Mengetahui perbandingan klasifikasi pembiayaan Warung Mikro Bank Syariah Mandiri KC Jambi setelah diterapkan teknik *sampling* kelas *imbalanced* pada metode *random forest*.

### **1.5. Manfaat Penelitian**

Manfaat dari penelitian ini adalah sebagai berikut:

1. Menambah pengetahuan tentang pengklasifikasian menggunakan metode *Random Forest*.
2. Menambah pengetahuan mengatasi kelas *imbalanced* menggunakan teknik *sampling* kelas *imbalanced* pada *Machine Learning*.
3. Menjadi bahan analisis Bank Syariah Mandiri dalam menentukan tingkat kepentingan calon nasabah Warung Mikro di masa yang akan datang.
4. Membantu meningkatkan kualitas dari pelayanan Bank Syariah Mandiri di masa yang akan datang dalam hal pembiayaan Warung Mikro.

## **BAB II**

### **TINJAUAN PUSTAKA**

Terdapat beberapa referensi jurnal dan skripsi yang berhubungan dalam penelitian yang dilakukan oleh penulis. Penelitian pertama berjudul “*Random Forest Based Imbalanced Data Cleaning and Classification*” oleh Gu (2007) yang menggunakan dataset PAKDD 2007, dimana dataset tersebut adalah *imbalanced data* (data tidak seimbang). Tujuan dari penelitian ini adalah mengeliminasi instansi yang menyebabkan penurunan kinerja model serta membentuk model yang menghasilkan akurasi tertinggi pada masing-masing instansi dengan perbandingan lima metode.

Dari penelitian tersebut, setelah membandingkan lima metode yang dapat mengatasi *imbalanced data* (data yang tidak seimbang) yaitu SMOTE, *under-sampling*, *over-sampling*, *tomek links*, *Condensed Nearest Neighbour Rule* (CNN), dan *Random forest* didapatkan hasil bahwa dari delapan dataset, empat dataset dengan akurasi tertinggi menggunakan metode *Random Forest*.

Kedua yaitu penelitian yang dilakukan oleh Dewi (2011) yang melakukan penelitian dengan judul “Penerapan Metode *Random Forest* dalam *Driver Analysis*”. Tujuan penelitian tersebut yaitu untuk mengetahui ukuran *random forest* dan ukuran contoh peubah penjelas yang menghasilkan *random forest* berakurasi prediksi tinggi dan stabil dan menghasilkan *driver analysis* yang stabil. Data yang digunakan adalah data sekunder yang diperoleh dari perusahaan riset pemasaran di Indonesia yang terdiri atas sejumlah merek berbeda dengan amatan sebanyak 1200 amatan.

Hasil penelitian tersebut menunjukkan akurasi prediksi yang tinggi dan stabil diperoleh saat *random forest* dibangun menggunakan ukuran contoh peubah penjelas sebesar empat dan ukuran *random forest* lebih dari 500 dengan rata-rata tingkat misklasifikasi sebesar 34.5%. Pada penerapan *random forest* dalam *driver analysis* berdasarkan MDG menghasilkan *driver analysis* yang stabil jika ukuran *random forest* lebih dari 500, namun jika berdasarkan rata-rata MDG dari 100

*random forest* tetap dihasilkan *driver analysis* stabil meskipun ukuran *random forest* yang digunakan cukup kecil.

Penelitian ketiga, penelitian berjudul “*Business Analytics using Random Forest Trees for Credit Risk Prediction: A Comparison Study*” oleh Ghatasheh (2014). Penelitian ini bertujuan untuk mempelajari kinerja *random forest*, membandingkan pendekatan yang berbeda, menemukan konfigurasi terbaik untuk akurasi yang lebih tinggi, dan menemukan kelemahan dari pendekatan prediksi yang dipilih. Data yang digunakan yaitu dataset kredit Jerman yang dipublikasikan oleh University of California, Irvine (UCI) *Machine Learning Repository*, dimana terdapat 1000 instansi yang terbagi menjadi dua kelas ; 700 “*good credit*” dan 300 “*bad/refused credit*”.

Hasil yang didapatkan yaitu metode *random forest trees* menjadi salah satu peluang yang menjanjikan untuk analisis bisnis dalam memprediksi risiko kredit, dimana keuntungan utama nya adalah ketepatan klasifikasi dan sederhana. Tingkat akurasi metode *random forest* pada data yang dipakai adalah 78.4%, sehingga klasifikasi menggunakan *random forest* menjadi fokus banyak peneliti tanpa menghilangkan kelebihan dan kekurangan dari penelitian yang dilakukan.

Selanjutnya, penelitian keempat yaitu penelitian yang dilakukan oleh Adnyana (2015) yaitu Prediksi Lama Studi Mahasiswa dengan Metode *Random Forest* dengan studi kasus STIKOM Bali. Tujuan dari penelitian ini adalah menerapkan metode *random forest* pada permasalahan prediksi lama studi mahasiswa yang diharapkan membantu bidang akademik dalam menyusun strategi untuk mengurangi rata-rata lama studi mahasiswa di STIKOM Bali untuk periode berikutnya. Data yang digunakan adalah data mahasiswa yang sudah lulus periode tahun 2014/2015 dari STIKOM Bali. Hasil yang dapat disimpulkan dari lima buah fitur data yang digunakan sebagai *input* dan prediksi lama studi sebagai *output* yang terdiri dari dua klasifikasi (lulus tepat waktu dan lulus batas waktu studi) diperoleh akurasi 83.54%.

Pada penelitian kelima dengan judul “Perbandingan Teknik *Sampling* Dalam *Random Forest* Pada Kelas *Imbalanced*” oleh Jatmiko dkk (2017) menggunakan dataset hasil SDKI 2012 sebanyak 1500 sampel yang diambil

secara *simple random sampling*. Tujuan dari penelitian ini yaitu membandingkan beberapa teknik *sampling* dalam *machine learning* menggunakan metode *random forest* pada kelas *imbalanced* perilaku seksual remaja di Indonesia hasil SDKI 2012.

Hasil dari penelitian ini adalah metode *random forest* pada data remaja yang pernah melakukan hubungan seksual pranikah hasil SDKI 2012 menghasilkan akurasi yang bias. Setelah menggunakan teknik *sampling*, pada nilai *sensitivity*, *precision*, dan *recall* meningkat dari data asli. Teknik *undersampling* memberikan nilai *sensitivity* dan *recall* terbesar yaitu 84.62% dan 97.97% sedangkan *precision* terbaik diberikan oleh teknik *sampling* SMOTE dengan nilai sebesar 28.21%.

## **BAB III**

### **LANDASAN TEORI**

#### **3.1. Usaha Mikro, Kecil, dan Menengah (UMKM)**

Usaha Mikro, Kecil, dan Menengah (UMKM) merupakan kelompok pelaku ekonomi terbesar dalam perekonomian Indonesia dan terbukti menjadi katup pengaman perekonomian nasional dalam masa krisis, serta menjadi dinamisator pertumbuhan ekonomi pasca krisis ekonomi (Aslinda, 2014). UMKM menciptakan peluang kerja yang cukup besar bagi tenaga kerja di Indonesia, sehingga sangat membantu mengurangi pengangguran. Di Indonesia, hari UMKM Nasional diperingati setiap tanggal 31 Maret (Geevv, 2017).

##### **3.1.1. Pengertian UMKM**

Pengertian UMKM menurut UU RI No.20 Tahun 2008 tentang Usaha Mikro, Kecil, dan Menengah (UMKM) Pasal 1, yaitu:

1. Usaha Mikro adalah usaha produktif milik orang perorangan dan/atau badan usaha perorangan yang memenuhi kriteria Usaha Mikro sebagaimana diatur dalam Undang-undang.
2. Usaha Kecil adalah usaha ekonomi produktif yang berdiri sendiri, yang dilakukan oleh orang perorangan atau badan usaha yang bukan merupakan anak perusahaan atau bukan cabang perusahaan yang dimiliki, dikuasai, atau menjadi bagian baik langsung maupun tidak langsung dari Usaha Menengah atau Usaha Besar yang memenuhi kriteria Usaha Kecil sebagaimana dimaksud dalam Undang-undang.
3. Usaha Menengah adalah usaha ekonomi produktif yang berdiri sendiri, yang dilakukan oleh orang perorangan atau badan usaha yang bukan merupakan anak perusahaan atau bukcabang perusahaan yang dimiliki, dikuasai, atau menjadi bagian baik langsung maupun tidak langsung dengan Usaha Kecil dalam Undang-undang.

Menurut Rudjito (2003), Usaha Kecil dan Menengah (UKM) di Indonesia yang memiliki peranan yang penting dalam perekonomian Indonesia, baik ditinjau dari segi jumlah usaha maupun dari segi penciptaan lapangan kerja. Asas dan tujuan UMKM menurut UU RI No.20 Tahun 2008 tentang Usaha Mikro, Kecil,

dan Menengah (UMKM) Pasal 2 adalah kekeluargaan, demokrasi ekonomi, kebersamaan, efisiensi berkeadilan, berkelanjutan, berwawasan lingkungan, kemandirian, keseimbangan kemajuan, dan kesatuan ekonomi sosial.

### 3.1.2. Karakteristik UMKM

Menurut UU RI No.20 Tahun 2008 tentang Usaha Mikro, Kecil, dan Menengah (UMKM) Pasal 6, kriteria UMKM terdiri dari:

1. Kriteria Usaha Mikro adalah sebagai berikut:
  - a. memiliki kekayaan bersih paling banyak Rp 50.000.000,00 (lima puluh juta rupiah) tidak termasuk tanah dan bangunan tempat usaha; atau
  - b. memiliki hasil penjualan tahunan paling banyak Rp 300.000.000,00 (tiga ratus juta rupiah).
2. Kriteria Usaha Kecil adalah sebagai berikut:
  - a. memiliki kekayaan bersih lebih dari Rp 50.000.000,00 (lima puluh juta rupiah) sampai dengan paling banyak Rp 500.000.000,00 (lima ratus juta rupiah) tidak termasuk tanah dan bangunan tempat usaha; atau
  - b. memiliki hasil penjualan tahunan lebih dari Rp 300.000.000,00 (tiga ratus juta rupiah) sampai dengan paling banyak Rp 2.500.000.000,00 (dua milyar lima ratus juta rupiah).
3. Kriteria Usaha Menengah adalah sebagai berikut:
  - a. memiliki kekayaan bersih lebih dari Rp 500.000.000,00 (lima ratus juta rupiah) sampai dengan paling banyak Rp 10.000.000.000,00 (sepuluh milyar rupiah) tidak termasuk tanah dan bangunan tempat usaha; atau
  - b. memiliki hasil penjualan tahunan lebih dari Rp 2.500.000.000,00 (dua milyar lima ratus juta rupiah) sampai dengan paling banyak Rp 50.000.000.000,00 (lima milyar rupiah).

**Tabel 3. 1** Kriteria UMKM

No	Uraian Kriteria	Aset	Omzet
1	Usaha Mikro	Maksimal 50 Juta	Maksimal 300 Juta
2	Usaha Kecil	> 50 Juta – 500 Juta	> 300 Juta – 2.5 Miliar
3	Usaha Menengah	> 500 Juta – 10 Miliar	> 2.5 Miliar – 50 Miliar

### 3.1.3. Profil UMKM di Indonesia

Menurut Tambunan (2009), pada data Badan Pusat Statistik (BPS) jumlah UMKM terus meningkat dan mendominasi jumlah perusahaan dimana pada tahun 2006 terdapat sekitar 48 juta UMKM dibandingkan dengan 7200 usaha berskala

besar yang berarti kesempatan kerja UMKM menyumbang sekitar 97 persen dari jumlah pekerja di Indonesia. Namun, sebelum krisis pecah pada tahun 1997 UKM kurang mendapatkan perhatian di Indonesia karena UKM mampu bertahan pada saat itu bahkan jumlahnya meningkat dengan pesat setelah krisis ekonomi melanda, perhatian pada UKM pun menjadi lebih besar (Prasetyo, 2011).

Tidak dapat dipungkiri bahwa UMKM memiliki daya tahan tangguh dalam menghadapi gejolak sejak terjadinya krisis moneter yang diikuti oleh krisis ekonomi dan berbagai krisis lainnya, UMKM telah menopang ketahanan perekonomian nasional (Kompas, 2008). Pemerintah menargetkan penurunan pengangguran dengan menggerakkan sektor riil yang komponennya didominasi oleh UMKM yang mencapai 99.9%, dimana UMKM mengambil peran yang sangat strategis dalam menggerakkan aktivitas perekonomian Indonesia dengan menyediakan 99.5% kesempatan kerja penduduk (Bappenas, 2011).

### **3.2. Bank Syariah**

Untuk menghindari pengoperasian bank dengan sistem bunga, Islam memperkenalkan prinsip-prinsip muamalah Islam dengan solusi alternatif terhadap persoalan pertentangan antara bunga bank dan riba yaitu Bank Syariah (Wilardjo, 2005). Pada tahun 1991 telah berdiri dua bank syariah, yaitu BPR Syariah Dna Mardhotillah dan BPR Syariah Berkah Amal Sejahtera yang berada di Bandung. Namun setelah terbitnya UU RI Nomor 7 Tahun 1992 yang direvisi dengan UU RI Nomor 10 Tahun 1998 tentang Perbankan dalam bentuk sebuah bank yang beroperasi dengan sistem bagi hasil atau Bank Syariah, maka saat itu pula berdiri Bank Muamalat Indonesia.

Menurut Siamat (2005), Bank Syariah merupakan bank yang dalam menjalankan usahanya berdasarkan prinsip-prinsip hukum atau syariah dengan selalu mengacu pada Al-Qur'an dan Al-Hadist. Pengertian Bank Syariah menurut UU RI Nomor 21 Tahun 2008 tentang Perbankan Syariah yaitu bank yang menjalankan kegiatan usahanya berdasarkan prinsip syariah dan menurut jenisnya terdiri atas Bank Umum Syariah (BUS) dan Bank Pembiayaan Rakyat Syariah. Bank Syariah (BPRS) merupakan lembaga intermediasi dan penyedia jasa keuangan yang bekerja berdasarkan etika dan sistem nilai islam, khususnya yang bebas dari bunga (*riba*), bebas dari kegiatan spekulatif yang nonproduktif seperti

perjudian (*maysir*), bebas dari hal-hal yang tidak jelas dan meragukan (*gharar*), berprinsip keadilan, dan hanya membiayai kegiatan usaha yang halal (Ascarya & Yumanita, 2005).

Dengan adanya ketentuan-ketentuan baku yang harus ditaati oleh bank-bank yang menjalankan syariah secara benar, ternyata telah membuktikan bahwa bank syariah telah teruji keberadaannya dengan keadaan pada saat krisis keuangan global (Claudia, 2010). Dimana prinsip-prinsip dasar perbankan syariah mengikuti aturan dan norma Islam yaitu bebas dari bunga (*riba*), bebas dari kegiatan spekulatif yang non produktif seperti perjudian (*masyir*), tidak jelas dan meragukan (*gharar*), rusak atau tidak sah (*bathil*), dan hanya membiayai kegiatan usaha yang halal (Yumanita, 2005). Bank Syariah harus mematuhi prinsip syariah serta Fatwa Dewan Syariah Nasional (DSN), dimana DSN merupakan satu-satunya dewan yang mempunyai kewenangan mengeluarkan fatwa atas jenis-jenis kegiatan, produk dan jasa keuangan syariah, serta mengawasi penerapan fatwa yang dimaksud oleh lembaga-lembaga keuangan syariah di Indonesia (Minarni, 2013).

Dalam menjalankan pekerjaannya, bank Islam menggunakan berbagai teknik dan metode investasi seperti kontrak *mudharabah* (Wilardjo, 2005). Secara umum jumlah kantor perbankan syariah mengalami peningkatan pesat dari tahun 2008 hingga 2014 (OJK, 2014). Jika dilihat dari penduduk Indonesia yang mayoritas beragama Islam, bank syariah mempunyai peluang besar untuk mendapatkan nasabah. Bank syariah yang efisien menunjukkan jasa keuangan yang lebih baik, sehingga akan menarik perhatian lebih deposan dan investor terhadap bank, serta akan meningkatkan industri perbankan dan keuangan, dan mendorong pertumbuhan ekonomi (Zainal, 2012).

Fungsi bank syariah dibagi empat, yaitu (Antonio, 2007) :

1. Sebagai Manajemen Investasi

Fungsi ini berdasarkan kontrak *mudharabah* atau kontrak perwakilan. Menurut kontrak *mudharabah*, bank (dalam kapasitasnya sebagai *mudharib* atau pihak pelaksana investasi dana dari pihak lain) menerima presentase suatu keuntungan hanya dalam kasus untung. Bank syariah bertindak sebagai manajer investasi dari pemilik dana (*shahibul*

*maal*) dari dana yang dihimpun (dalam perbankan lazim disebut deposit atau penabung), karena besar kecilnya pendapatan (bagi hasil) yang diterima pemilik dana sangat bergantung pada pendapatan yang diterima oleh bank syariah dalam pengelolaan dana *mudharabah* sehingga tergantung kepada kehati-hatian, keahlian, dan sikap profesionalisme.

Fungsi ini dapat dilihat dari segi penghimpun dana bank syariah, khususnya dana *mudharabah*, bank syariah dalam posisi ini bertindak sebagai manajer investasi dalam arti dana tersebut harus dapat disalurkan pada penyalur produktif, sehingga dana yang terhimpun tersebut dapat menghasilkan dan hasilnya akan dibagi hasil dengan pemilik dana. Bank syariah tidaklah sepatutnya menghimpun dana *mudharabah* jika tidak dapat menyalurkan dana tersebut pada hal yang produktif, karena hasil yang diperoleh akan tetap dibagikan kepada pemilik dana sehingga hal tersebut jelas akan berakibat merugikan pemilik dana yang sudah ada.

## 2. Sebagai investasi

Bank-bank Islam menginvestasikan dana yang disimpan pada bank tersebut (dana pemilik bank maupun dana rekening investasi) dengan menggunakan alat investasi yang sesuai dengan syariah. Investasi yang sesuai dengan syariah tersebut meliputi akad *murabahah*, sewa menyewa, *musyarakah*, akad *mudharabah*, akad *salam* atau *istishna'*, pembentukan perusahaan atau akuisisi, pengendalian atau kepentingan lain dalam rangka mendirikan perusahaan, memperdagangkan produk, dan investasi atau memperdagangkan saham yang dapat diperjualbelikan atau *real estate*.

Keuntungan dibandingkan kepada pihak yang memberikan *mudharib*-nya yang sudah diepakati antara pemilik rekening investasi dan bank sebelum pelaksanaan akad. Fungsi ini dapat dilihat dalam hal penyaluran dana yang dilakukan bank syariah, baik yang dilakukan dengan mempergunakan prinsip jual beli maupun dengan prinsip bagi hasil.

## 3. Sebagai Jasa Keuangan

Bank syariah dalam fungsi ini juga dapat menawarkan berbagai jasa keuangan lainnya berdasarkan *wupah* (*fee based*) dalam sebuah

kontrak perwakilan atau penyewaan. Contohnya garansi, transfer kawat, L/C, dan lain-lain.

#### 4. Sebagai Jasa Keuangan

Dalam prinsip perbankan Islam mengharuskan bank Islam melaksanakan jasa sosial, jasa tersebut bisa melalui dana *qardg* (pinjaman kebajikan), dana zakat, atau dana sosial yang sesuai dengan ajaran Islam. Lebih jauh lagi, konsep perbankan dalam Islam juga mengharuskan bank Islam berperan untuk mengembangkan sumber daya insani dan menumbang dana untuk pemeliharaan dan pengembangan lingkungan hidup.

Menurut Wilardjo (2005), adanya bank Islam diharapkan dapat memberikan sumbangan terhadap pertumbuhan ekonomi masyarakat melalui pembiayaan-pembiayaan yang dikeluarkan oleh bank Islam sehingga dapat menjadi mitra dengan nasabah. Bank syariah memberikan pembiayaan dengan tujuan untuk memperoleh pendapatan dan membantu nasabah atau masyarakat lain yang membutuhkan dan pembiayaan tersebut dalam mengembangkan berbagai usahanya sehingga perekonomiannya menjadi sejahtera

### 3.2. Pembiayaan

#### 3.2.1. Pengertian Pembiayaan

Pembiayaan dalam arti luas berarti *financing* atau pembelanjaan, yaitu pendanaan yang dikeluarkan untuk mendukung investasi yang telah direncanakan, baik dilakukan sendiri maupun dijalankan dengan orang lain, sedangkan dalam arti sempit pembiayaan ialah pendanaan yang dilakukan oleh lembaga pembiayaan seperti bank syariah kepada nasabah (Muhammad, 2005).

Menurut Kasmir (2003), pengertian pembiayaan secara umum adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan antara bank dengan pihak lain yang mewajibkan pihak yang dibiayai untuk mengembalikan uang atau tagihan tersebut setelah jangka waktu tertentu dengan imbalan atau bagi hasil.

Dalam UU RI No.21 Bab 1 Pasal 1 tentang Perbankan Syariah, pembiayaan adalah penyediaan dana atau tagihan yang dipersamakan berupa:

- a. transaksi bagi hasil dalam bentuk *mudharabah* dan *musyarakah*;

- b. transaksi sewa-menyewa dalam bentuk ijarah atau sewa beli dalam bentuk *ijarah muntahiya bittamlik*;
- c. transaksi jual beli dalam bentuk piutang *murabahah, salam, dan istishna'*;
- d. transaksi sewa-menyewa jasa dalam bentuk ijarah untuk transaksi multijasa.

Menurut Veitzhal (2008), istilah pembiayaan pada intinya berarti *I Believe, I trust*, “saya percaya” atau “saya menaruh kepercayaan”. Perkataan pembiayaan yang artinya kepercayaan (*trust*), berarti lembaga pembiayaan selaku *shahibul mal* menaruh kepercayaan kepada seseorang untuk melaksanakan amanah yang diberikan. Dana tersebut harus digunakan dengan benar, adil, dan harus disertai ikatan dan syarat-syarat yang jelas, dan saling menguntungkan kedua belah pihak.

Penilaian kredit atau pembiayaan oleh bank dapat dilakukan dengan berbagai cara untuk mendapatkan keyakinan tentang nasabahnya, seperti melalui prosedur yang benar dan sungguh-sungguh (Kasmir, 2002). Islam tidak melarang hubungan pinjam-meminjam dalam suatu kegiatan ekonomi, bahkan kegiatan tersebut sangat dianjurkan karena dapat saling membantu antara sesama manusia. Adapun dasar hukum dibolehkannya pinjam-meminjam dalam pembiayaan tersebut berdasarkan firman Allah SWT dalam surat Al-Baqarah (2) ayat 245.

Menurut Kasmir (2003), perbedaan antara kredit yang diberikan oleh bank berdasarkan konvensional dengan pembiayaan yang diberikan oleh bank syariah terletak pada keuntungan yang diharapkan. Dimana keuntungan bagi bank konvensional diperoleh melalui bunga, sedangkan keuntungan bagi bank syariah diperoleh melalui bagi hasil atau imbalan. Perbedaan lainnya terletak pada analisis pemberian pembiayaan (kredit) beserta persyaratannya.

### **3.2.2. Tujuan dan Manfaat Pembiayaan**

Menurut Kasmir (2010), tujuan utama pemberian pembiayaan antara lain:

1. Mencari keuntungan, yaitu memperoleh hasil dari pemberian pembiayaan tersebut terutama dalam bentuk keuntungan yang diterima dari usaha yang dikelola oleh bank dan nasabah. Keuntungan ini berguna untuk kelangsungan hidup bank, dimana jika bank menderita kerugian maka besar kemungkinan bank tersebut akan dilikuidir (dibubarkan).
2. Membantu usaha nasabah, dimana bank akan membantu usaha nasabah yang memerlukan dana baik dana investasi maupun dana modal kerja yang

diharapkan dapat membantu mengembangkan dan memperluas usaha debitur.

3. Membantu pemerintah, semakin banyak pembiayaan yang disalurkan oleh pihak perbankan maka semakin baik karena adanya peningkatan pembangunan diberbagai sektor. Keuntungan penyebaran pemberian pembiayaan adalah penerimaan pajak dari keuntungan yang diperoleh nasabah dari bank, membuka kesempatan kerja, meningkatkan jumlah barang dan jasa, menghematkan devvisa negara, dan meningkatkan devisa negara.

Sedangkan menurut Sinungan (1997), pembiayaan memiliki fungsi dalam kehidupan perekonomian yang modern saat ini yaitu:

1. Pembiayaan dapat meningkatkan daya guna *utility* dari uang.
2. Pembiayaan dapat meningkatkan daya guna dari barang.
3. Pembiayaan meningkatkan peredaran dan lalu lintas uang.
4. Pembiayaan adalah salah satu alat stabilitas ekonomi.
5. Pembiayaan dapat meningkatkan kegairahan berusaha masyarakat.

### 3.2.3. Akad-akad Pembiayaan Syariah

Akad (*al-'aqd*) dalam Bahasa Arab artinya perikatan, perjanjian, dan pemufakatan (Yunus, 1990). Akad dalam artian umum (Zuhaili, 2002) adalah segala sesuatu yang dikehendaki seseorang untuk dikerjakan, baik yang muncul dari kehendaknya sendiri maupun yang membutuhkan kehendak dua pihak dalam melakukannya. Sedangkan menurut Lathif (1997) akad dalam artian khusus adalah pertalian atau keterikatan antara ijab dan qabul sesuai dengan kehendak syariah yang menimbulkan akibat hukum pada obyek akad.

Menurut Veithzal (2008), lazimnya dalam pembiayaan terdapat tiga ketentuan dalam melakukan akad pada bank syariah antara lain:

1. Prinsip Bagi Hasil atau Syikrah (*Profit Sharing*)
  - a. *Mudharabah (Trust Financing, Trust Investment)*

Sistem kerjasama usaha antara dua pihak atau lebih dimana pihak pertama (*shahib al-ma*) menyediakan seluruh kebutuhan modal sedangkan *customer* sebagai pengelola (*mudharib*) mengajukan permohonan pembiayaan dan untuk ini *customer* sebagai pengelola menyediakan keahliannya.

b. *Musyarakah (Partnership, Project Financing Participation)*

Adanya keinginan dari dua pihak atau lebih untuk melakukan kerjasama untuk suatu usaha tertentu, dimana masing-masing menyertakan dan menyetorkan modalnya dengan pembagian keuntungan dikemudian hari sesuai kesepakatan. Kesepakatan tersebut bisa berupa dana, keahlian, kepemilikan, peralatan, barang perdagangan ataupun aset.

c. *Al-Muzara'ah (Harvest Yield Profit Sharing)*

Kerjasama pengolahan pertanian antara pemilik lahan dan penggarap dimana pemilik lahan memberikan lahan pertanian kepada penggarap untuk ditanami dan dipelihara dengan imbalan tertentu dari hasil panen.

2. Prinsip Jual Beli atau *Ba'i (Sale and Purchase)*

a. *Murabahah*

Transaksi jual beli dimana lembaga pembiayaan menyebutkan sejumlah keuntungan tertentu dimana bank bertindak sebagai penjual dan pihak *costumer* sebagai pembeli sehingga harga beli dari *supplier* atau pemasok ditambah dengan keuntungan lembaga sebelum dijual kepada *costumer*.

b. *Al-Ba'i Naqdan*

Akad jual beli biasa yang dilakukan secara tunai, dimana *Al-Ba'i* artinya jual beli sedangkan *Naqdan* artinya tunai.

c. *Al-Ba'i Muajjal*

Akad jual beli secara tidak tunai atau cicilan, dimana barang diserahkan pada awal periode sedangkan uang dapat diserahkan pada periode berikutnya.

d. *Al-Ba'i Salam*

Pada akad jual beli ini, barang yang ingin dibeli biasanya belum ada (masih harus diproduksi atau dipesan) dimana uang langsung diserahkan sekaligus dimuka sedangkan barangnya diserahkan pada akhir periode pembiayaan.

e. *Al-Istishna*

Kontrak penjualan antara pembeli dengan produsen atau *supplier* dimana produsen menerima pesanan dari pembeli. Produsen tersebut meminta orang lain untuk membuat atau membeli barang menurut spesifikasi yang

disepakati (sejak awal) dan menjualnya kembali kepada pembeli akhir. Sehingga pembeli dan produsen sepakat atas harga serta sistem pembayaran (dimuka, dicicil atau ditangguhkan sampai waktu tertentu).

### 3. Prinsip Sewa Menyewa (*Ijarah*)

Akad untuk memanfaatkan jasa, baik jasa atas barang yang disebut sewa menyewa maupun jasa atas tenaga kerja yang disebut dengan upah mengupah.

### 3.3. Bank Syariah Mandiri (BSM)

Bank Syariah Mandiri (BSM) merupakan salah satu bank yang berkonsep syariah Indonesia dan bank syariah terbesar di Indonesia saat ini yang didirikan pada tanggal 25 Oktober 1999 serta mulai beroperasi tanggal 1 November 1999.

#### 3.3.1. Produk-produk Pembiayaan BSM

Produk-produk pembiayaan BSM adalah sebagai berikut (Rosyidah, 2017):

##### 1. BSM Oto

BSM Oto merupakan pembiayaan untuk pembelian kendaraan bermotor dengan sistem *murabahah*. Pembiayaan yang dapat dikategorikan sebagai berikut:

- a. Jenis kendaraan ; mobil dan motor.
- b. Kondisi kendaraan : baru dan bekas.

Untuk kendaraan baru, jangka waktu pembiayaan hingga lima tahun sedangkan kendaraan bekas hingga sepuluh tahun (dihitung termasuk usia kendaraan dan jangka waktu pembiayaan). Selanjutnya, syarat dan ketentuan BSM Oto, yaitu:

- 1) Pemohon harus mempunyai pekerjaan dan pendapatan yang tetap.
- 2) Usia pemohon pada saat pengajuan Pembiayaan Kendaraan Bermotor (PKB) minimal 21 tahun dan maksimal 55 tahun pada saat jatuh tempo fasilitas PKB.
- 3) Pengajuan PKB dapat dilakukan sendiri-sendiri atau koordinir secara kolektif oleh instansi dimana pemohon bekerja.

##### 2. BSM Griya

Pembiayaan Griya BSM adalah pembiayaan jangka pendek, menengah atau panjang untuk membiayai pembelian rumah tinggal (konsumer), baik baru maupun bekas di lingkungan developer dengan sistem *murabahah* (dibawah tangan). Karakteristik BSM Griya adalah:

- a. Angsuran tetap hingga jatuh tempo pembiayaan.
- b. Proses permohonan yang mudah dan cepat.
- c. Fleksibel untuk membeli rumah baru dengan Rp 5 miliar.
- d. Jangka waktu pembiayaan yang panjang.
- e. Fasilitas autodebet dari tabungan BSM.

Dimana persyaratan yang diperlukan adalah sebagai berikut:

1. Warga Negara Indonesia (WNI) cukup hukum.
  2. Usia minimal 21 tahun dan maksimal 55 tahun pada saat jatuh tempo.
  3. Besar angsuran tidak melebihi 40% dari penghasilan bulanan bersih.
  4. Fasilitas pembiayaan untuk unit yang belum selesai dibangun atau inden dapat diberikan untuk fasilitas pembiayaan yang pertama.
  5. Pencairan pembiayaan dapat diberikan apabila progres pembangunan telah mencapai 50% dengan total pencairan maksimal sebesar 50%.
  6. Untuk pencairan unit yang belum selesai dibangun atau inden, harus melalui perjanjian kerjasama antara developer dan BSM kantor pusat.
3. BSM Gadai Emas

Gadai Emas BSM merupakan produk pembiayaan atas dasar jaminan berupa emas sebagai salah satu alternatif memperoleh uang tunai dengan cepat. Akad yang dilakukan adalah sebagai berikut:

- a. *Qardh* dalam rangka *rahn* (gadai) adalah akad perjanjian pinjaman dari bank untuk nasabah yang disertai dengan penyerahan tugas agar bank menjaga barang jaminan yang diserahkan.
- b. Biaya pemeliharaan menggunakan akad *ijarah*.

Syarat BSM Gadai Emas, yaitu:

- 1) Kartu identitas nasabah.
- 2) Pembiayaan mulai dari Rp 500.000,00.
- 3) Jaminan berupa emas perhiasan atau lantakan (batangan).

- 4) Jangka waktu empat bulan dan dapat diperpanjang atau dapat digadai ulang (setelah dilakukan penaksiran dan melunasi biaya gadai).

#### 4. BSM Cicil Emas

Fasilitas yang disediakan oleh BSM untuk membantu nasabah untuk membiayai pembelian atau kepemilikan emas berupa lantakan (batangan) dengan cara mudah punya emas dan menguntungkan pembiayaan menggunakan akad *murabahah*. Pengikatan agunan dengan menggunakan akad *rahn*.

Dimana uang muka minimal 20% dari harga perolehan emas dan dibayar secara tunai (tidak dicicil) oleh nasabah kepada bank. Sumber uang muka harus berasal dari dana nasabah sendiri (*self financing*) dan bukan berasal dari pembiayaan yang diberikan oleh bank. Syarat dari pembiayaan ini adalah:

- 1) WNI cukup umur.
- 2) Pegawai dengan usia minimal 21 tahun s.d usia maksimal 55 tahun.
- 3) Pensiunan berusia maksimal 70 tahun pada saat pembiayaan jatuh tempo.
- 4) Profesional dan wiraswasta berusia maksimal 60 tahun.
- 5) Menyerahkan Kartu Identitas seperti Kartu Tanda Penduduk (KTP).

#### 5. BSM Warung Mikro

Pembiayaan Warung Mikro dengan pokok pembiayaan mulai dari Rp 15.000.000,00 — Rp 200.000.000,00. Pembiayaan ini diperuntukkan (Wawancara):

- a. Perorangan: Golongan berpenghasilan tetap (Golbertap) yang tidak bekerjasama dengan instansi lainnya seperti Pegawai Negero Sipil (PNS), pegawai, swasta, dsb.
- b. Kerjasama Koperasi Unit Desa (KUD): Untuk nasabah yang bekerjasama dengan KUD, misalnya petani sawit.
- c. Kerjasama Perusahaan: Untuk nasabah yang bekerjasama dengan perusahaan dimana nasabah tersebut beraliansi atau golbertap sehingga pinjaman tersebut dipotong langsung oleh bendahara perusahaan.

Terdapat dua produk pembiayaan yang telah ditentukan baku sesuai dengan manual produk pembiayaan mikro dan petunjuk teknis yang berlaku, yaitu:

- 1) Pembiayaan Usaha Mikro Madya adalah pembiayaan usaha mikro dari bank kepada nasabah dengan pembiayaan Rp 15.000.000,00 – Rp 50.000.000,00.
- 2) Pembiayaan Usaha Mikro Utama adalah pembiayaan usaha mikro dari bank kepada nasabah dengan pembiayaan Rp 50.000.000,00 – Rp 200.000.000,00.

### **3.3.2. Pembiayaan Warung Mikro BSM**

BSM memberikan pembiayaan bagi pengusaha mikro, kecil dan menengah melalui produk pembiayaan mikro yang ditawarkan kepada nasabah yang mempunyai usaha dan mengembangkannya dengan akad jual beli (*murabahah*), baik nasabah Golongan Berpenghasilan Tetap (Golbertap) maupun nasabah Golongan Berpenghasilan Tidak Tetap (Non-Golbertap). Warung Mikro menggunakan akad pembiayaan, dimana terdapat pembiayaan Usaha Mikro Madya, pembiayaan Usaha Mikro Utama (wawancara). Sesuai dengan surat edaran Direksi Bank Syariah Mandiri No. 12/009/PEM tanggal 13 Februari 2010 tentang Pembiayaan Warung Mikro, program pelayanan pembiayaan Warung Mikro di BSM Cabang Jambi dimulai pada tanggal 6 Mei 2009 (Samsuddin, 2015).

Warung mikro merupakan salah satu produk BSM yang bertujuan membantu akses pembiayaan dan mengembangkan UMKM dalam rangka penanggulangan kemiskinan dan perluasan kesempatan kerja, dimana hal ini terbukti dari banyaknya pengusaha yang membutuhkan tambahan modal khususnya usaha kecil dan mikro yang terus menunjukkan peningkatan (Samsuddin, 2015). Alasan pembiayaan warung mikro menggunakan akad *mudharabah* dimana bank membiayai kebutuhan nasabah yang kekurangan dana sebesar harga barang (keperluan yang dibutuhkan) karena sangat mudah diaplikasikan dan bank dapat menentukan *margin* di awal. Hal itu disebabkan oleh transaksi jual beli yang diterapkan oleh bank sehingga bank lebih mudah dalam mendapatkan keuntungan.

Menurut wawancara dengan Sri Rahma, pegawai BSM KC Jambi, penetapan harga per jenis pembiayaan mikro di BSM antara lain:

1. *Murabahah*, dimana penetapan harga seoptimal mungkin mampu memberikan keuntungan yang wajar bagi bank tanpa bersifat memberatkan nasabah dan bersaing dengan harga pembiayaan yang wajar di pasar. Tingkat margin *Murabahah* ditetapkan oleh kantor pusat sebagai harga jual yang berlaku untuk pembiayaan *Murabahah*.
2. *Ijarah* yaitu akad sewa-menyewa antara pemilik obyek sewa dan penyewa untuk mendapatkan imbalan atas manfaat obyek sewa yang disewakannya dan *Ijarah Muntahiyah Bittamlik* (IMBT) yaitu fasilitas pembiayaan dengan skim sewa atas suatu atas obyek sewa antara bank dan nasabah dalam periode yang ditentukan yang diakhiri dengan kepemilikan barang di tangan nasabah.

### **3.3.3. Syarat dan Prosedur Pembiayaan Warung Mikro**

Persyaratan atau ketentuan umum untuk pembiayaan warung mikro, yaitu (Rosyidah, 2017):

1. Wiraswasta atau profesi
  - a. Fotokopi KTP, KK, surat nikah pemohon suami istri.
  - b. Fotokopi rekening tabungan selama tiga bulan terakhir.
  - c. Surat Ijin Usaha Perorangan (SIUP) dan Tanda Daftar Perusahaan (TDP).
  - d. Usaha telah berjalan minimal dua tahun.
  - e. Usia minimal 21 tahun atau sudah menikah dan maksimal 55 tahun saat pembiayaan lunas.
  - f. Surat keterangan atau ijin usaha RT atau RW, nota belanja, khusus pedagang pasar cukup melampirkan fotokopi surat keterangan dari pengelola pasar setempat.
  - g. Nomor Pokok Wajib Pajak (NPWP).
2. Perorangan Golbertap
  - a. Fotokopi KTP, KK, surat nikah pemohon suami istri.
  - b. Status pegawai tetap dengan masa dinas minimal satu tahun.
  - c. Usia minimal 21 tahun atau sudah menikah dan maksimal 55 tahun saat jatuh tempo fasilitas pembiayaan.
  - d. Slip gaji tiga bulan terakhir.

- e. Surat keterangan kerja atau SK pegawai dan ID *card*.
  - f. NPWP.
3. Badan Usaha
- a. Usaha telah berjalan minimal dua tahun.
  - b. Surat keterangan atau ijin usaha.
  - c. Akta pendirian atau perubahan perusahaan.

Menurut Samsuddin (2015), terdapat empat tahapan yang akan dilalui oleh nasabah yang hendak mengajukan pembiayaan warung mikro, yaitu:

1. Permohonan pengajuan pembiayaan, dimana nasabah mengajukan jumlah pembiayaan yang diinginkan kepada bank dengan mengisi aplikasi permohonan. Selanjutnya, nasabah mengumpulkan kelengkapan data persyaratan pembiayaan.
2. Tim analis Warung Mikro menganalisa pembiayaan yang diajukan oleh nasabah. Terdapat tiga pilar analisa, yaitu kemampuan nasabah, aspek legalitas, dan objek akad. Analisa kemampuan dapat dilihat melalui fotokopi rekening tabungan (mutasi tabungan rekening perbulan), slip gaji, BI *checking* untuk mengetahui apakah calon nasabah memiliki pinjaman di bank lain atau tidak. Analisa legalitas data-data dapat diketahui melalui hasil wawancara nasabah dan memverifikasi data-data calon nasabah yang sudah masuk, baik melalui telepon dan juga survei ke lapangan (*on the spot*). Selain itu bank juga akan memeriksa melalui Sistem Informasi Debitur (SID) untuk mengetahui apakah calon nasabah masuk daftar hitam Bank Indonesia atau tidak.
3. Bila masih ada kekurangan yang belum dilengkapi oleh nasabah, maka nasabah harus melengkapi persyaratan. Setelah persyaratan lengkap, tim analis Warung Mikro akan membuat proposal pembiayaan untuk dilaporkan kepada komite pembiayaan dan kepala cabang sehingga proposal tersebut akan dibawa ke rapat komite pembiayaan. Jika komite pembiayaan setuju, maka dapat lanjut ke tahapan berikutnya.
4. Terakhir, melakukan akad antara pihak bank dan nasabah. Setelah akad dilaksanakan, dana pembiayaan akan langsung ditransfer oleh bank ke rekening nasabah. Sebelumnya nasabah tentunya telah melunasi biaya

administrasi yang menjadi kewajiban pihak nasabah. Selanjutnya nasabah bisa menyetorkan angsuran pembayaran pertama sebulan setelah ditandatanganinya akad dengan cara menyetorkan angsuran yang telah disepakatai dalam kontrak.

### **3.4. Analisis Deskriptif**

Menurut Kamus Besar Bahasa Indonesia (2001), deskripsi adalah pemaparan atau penggambaran dengan kata-kata secara jelas dan terperinci. Analisis data merupakan kegiatan setelah data dari seluruh responden terkumpul. Kegiatan dalam analisis data adalah mengelompokkan data berdasarkan variabel dan jenis responden, menstabilisasi data berdasarkan variabel dari seluruh responden, menyajikan data tiap variabel yang diteliti, melakukan perhitungan untuk menguji hipotesis yang diajukan (Sugiyono, 2017). Menurut Sugiyono (2009), metode deskriptif adalah penelitian yang dilakukan untuk mengetahui keberadaan variabel mandiri, baik hanya pada suatu variabel atau lebih (variabel yang berdiri sendiri) tanpa membuat perbandingan dan mencari hubungan variabel itu dengan variabel yang lain.

Hasan (2001) menjelaskan bahwa statistik deskriptif adalah bagian dari statistika yang mempelajari cara pengumpulan data dan penyajian data sehingga mudah dipahami. Statistika deskriptif hanya berhubungan dengan hal menguraikan atau memberikan keterangan-keterangan mengenai suatu data atau keadaan. Statistika deskriptif berfungsi menerangkan keadaan, gejala, atau persoalan. Penarikan kesimpulan pada statistika deskriptif hanya ditujukan pada kumpulan data yang ada.

Statistika deskriptif hanya untuk memperoleh gambaran (*description*) atau ukuran-ukuran tentang data yang ada di tangan (Furqon, 1993). Statistika deskriptif ialah bagian dari statistik yang membicarakan mengenai penyusunan data ke dalam daftar-daftar atau jadwal, pembuatan grafik-grafik, dan lain-lain yang sama sekali tidak menyangkut penarikan kesimpulan (Pasaribu, 1975). Contoh dari penyajian data dalam statistika deskriptif adalah tabel, diagram, dan grafik (Walpole, 1995).

Menurut Mattjik & Sumertajaya (2006), terdapat peringkasan data pada analisis deskriptif. Peringkasan data tersebut terdiri dari:

## 1. Ukuran Pemusatan

Ukuran pemusatan merupakan suatu gambaran (informasi) yang memberikan penjelasan bahwa data memiliki satu (mungkin lebih) titik dimana dia memusat atau terkumpul. Ukuran-ukuran pemusatan yang sering digunakan antara lain:

### a. Median

Median adalah suatu nilai data yang membagi dua sama banyak kumpulan data yang telah diurutkan. Apabila banyaknya data ganjil, median adalah data yang tepat ditengah-tengah sedangkan bila banyaknya data genap, median adalah rata-rata dua data yang ada ditengah.

### b. Modus

Suatu nilai data yang paling sering terjadi atau yang mempunyai frekuensi paling tinggi disebut modus. Jika dalam suatu data terdapat tiga nilai pengamatan yang memiliki frekuensi pemunculan tertinggi maka modulusnya adalah ketiga nilai pengamatan tersebut dan bukan rata-rata dari ketiganya.

### c. Kuartil

Kuartil adalah nilai-nilai yang menyekat gugus data menjadi empat kelompok data yang masing-masing terdiri dari 25% amatan. Nilai-nilai yang menyekat data menjadi empat kelompok data tersebut dikenal dengan sebutan kuartil 1 (Q1), kuartil 2 (Q2), dan kuartil 3 (Q3).

### d. Nilai Tengah (Rataan)

Nilai tengah merupakan ukuran pemusatan data yang menimbang data menjadi dua kelompok data yang memiliki massa yang sama. Apabila  $x_1, x_2, \dots, x_n$  adalah anggota suatu populasi terhingga berukuran  $N$ , maka nilai tengah populasinya adalah:

$$\mu = \frac{1}{N} \sum_{i=1}^N X_i \quad (3.1)$$

dimana:

$\mu$  = nilai rata-rata

$N$  = banyak data populasi

$X_i$  = data ke- $i$

Sedangkan jika  $x_1, x_2, \dots, x_n$  adalah anggota suatu contoh berukuran  $n$ , maka nilai tengah contoh tersebut adalah:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n X_i \quad (3.2)$$

dimana:

$\bar{x}$  = nilai rata-rata

$n$  = banyak data sampel

$X_i$  = data ke- $i$

Macam-macam diagram beserta karakteristiknya antara lain:

### 1. Diagram Batang-daun (*Steam and Leaf*)

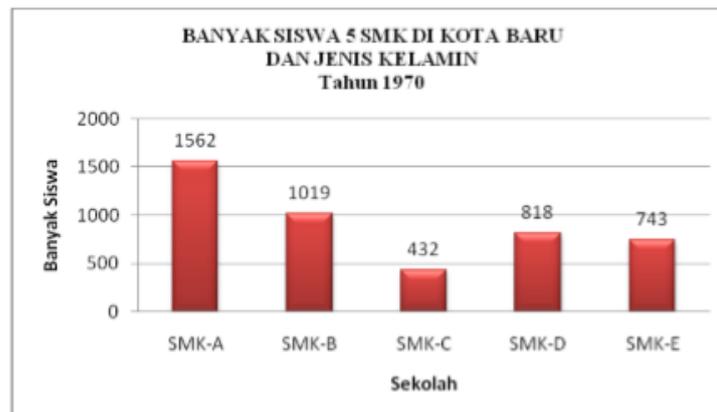
Menurut Somantri (2006), penyajian data dengan diagram batang daun selain dapat memperoleh informasi mengenai distribusi dari gugus data juga dapat dilihat nilai-nilai pengamatan aslinya.

Batang	Daun
4	4 7
5	6 9
6	1 3 4 5 6 8
7	0 1 1 1 1 2 3 3 3 4 5 5 6 6 6 9
8	3 4 4 5 5 9
9	1 4 5 7

**Gambar 3. 1** Contoh diagram batang-daun

### 2. Diagram Batang

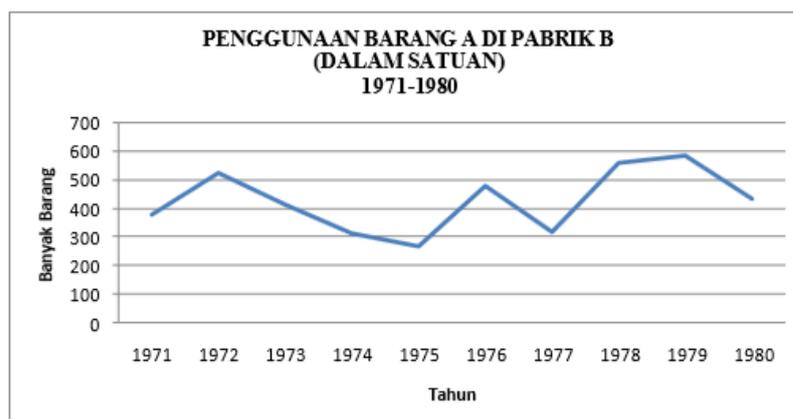
Grafik batang atau balok adalah grafik data berbentuk persegi panjang yang lebarnya sama dan dilengkapi dengan skala atau ukuran sesuai dengan data yang bersangkutan (Hasan, 2009). Menurut Riduwan (2003) mengemukakan penyajian data berbentuk diagram batang banyak modelnya yaitu diagram batang satu komponen atau lebih, diagram batang dua arah, diagram batang tiga dimensi, dan lain-lain sesuai dengan variasinya atau tergantung kepada keahlian pembuat diagram.



**Gambar 3. 2** Contoh diagram batang

### 3. Diagram Garis

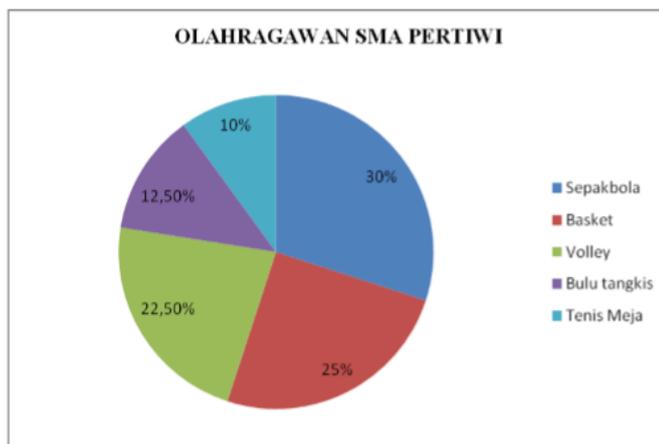
Diagram garis digunakan untuk menggambarkan keadaan yang serba terus atau berkesinambungan, misalnya produksi minyak tiap tahun, jumlah penduduk tiap tahun, keadaan temperatur badan tiap jam, dan lain-lain (Riduwan, 2003). Menurut Hasan (2009), grafik garis adalah grafik data berupa garis, diperoleh dari beberapa ruas garis yang menghubungkan titik-titik pada bidang bilangan atau sistem salib sumbu.



**Gambar 3. 3** Contoh diagram garis

### 4. Diagram Lingkaran atau *Piechart*

Penyajian data dalam bentuk diagram lingkaran (*piechart*) didasarkan pada sebuah lingkaran yang dibagi menjadi beberapa bagian sesuai dengan banyaknya kelas penyusunan (Somantri, 2006). Untuk membuat diagram lingkaran (*piechart*), gambarkan sebuah lingkaran, lalu dibagi-bagi menjadi beberapa sektor dimana tiap sektor melukiskan kategori data yang terlebih dahulu diubah kedalam derajat (Sudjana, 2005).



**Gambar 3. 4** Contoh *piechart*

#### 5. Diagram Gambar (Piktogram)

Piktogram adalah grafik data yang menggunakan gambar atau lambang dari data itu sendiri dengan skala tertentu (Hasan, 2009). Sering digunakan untuk mendapatkan gambaran kasar sesuatu hal dan sebagai alat visual bagi orang awam dimana setiap satuan dijadikan lambang yang disesuaikan dengan macam datanya.

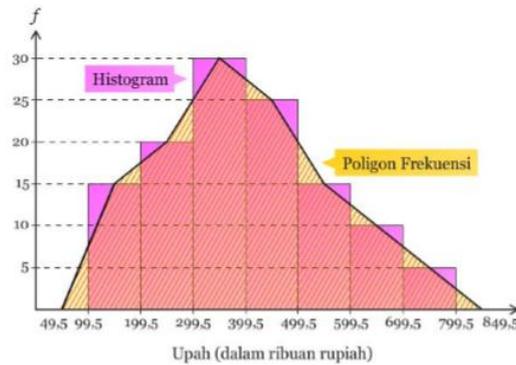
PROGRAM STUDI	JUMLAH SISWA	LAMBANG
Konstruksi Bangunan	60	👤👤👤👤👤👤👤👤
Elektronika	65	👤👤👤👤👤👤👤👤👤
Listrik Instalasi	35	👤👤👤👤👤
Mesin Produksi	60	👤👤👤👤👤👤👤👤
Mekanik Otomotif	75	👤👤👤👤👤👤👤👤👤👤

Keterangan : 👤 = 10 siswa

**Gambar 3. 5** Contoh piktogram

#### 6. Histogram dan Poligon Frekuensi

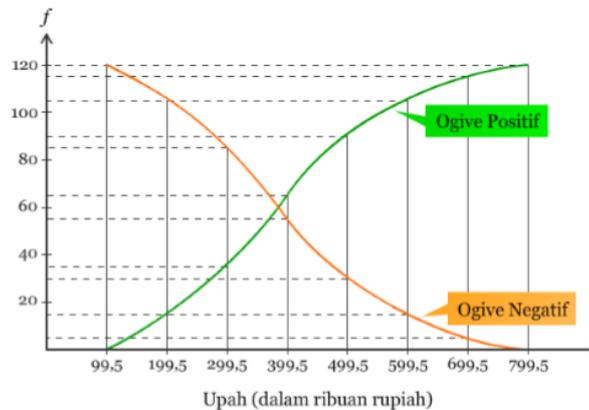
Menurut Riduwan (2003), histogram adalah grafik yang menggambarkan suatu distribusi frekuensi dengan bentuk beberapa segi empat sedangkan poligon frekuensi ialah grafik garis yang menghubungkan nilai tengah tiap sisi atas yang berdekatan dengan nilai tengah jarak frekuensi mutlak masing-masing.



**Gambar 3. 6** Contoh histogram dan poligon

### 7. *Ogive*

Menurut Siregar (2010), grafik *ogive* dibuat dari daftar sebaran frekuensi kumulatif kurang dari dan frekuensi kumulatif lebih dari dimana untuk membuat grafik *ogive* terlebih dahulu mencari nilai frekuensi kumulatif.



**Gambar 3. 7** Contoh *ogive*

### 3.5. *Data Mining*

Menurut Davies (2004), secara sederhana *data mining* adalah penambangan atau informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar. *Data mining* sering juga disebut sebagai *knowledge discovery in database* (KDD) yaitu kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santoso, 2007). Sedangkan menurut Turban (2005), *data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar.

Menurut Han dan Kamber (2011), secara garis besar *data mining* dapat dikelompokkan menjadi dua kategori utama, yaitu:

### 1. *Predictive*

*Predictive* merupakan proses untuk menemukan pola dari data dengan menggunakan beberapa variabel lain di masa depan. Salah satu teknik yang terdapat dalam *predictive mining* adalah klasifikasi. Tujuan dari tugas *predictive* adalah untuk memprediksi nilai dari atribut tertentu berdasarkan pada nilai atribut-atribut lain. Atribut yang diprediksi umumnya dikenal sebagai target atau variabel tak bebas, sedangkan atribut-atribut yang digunakan untuk membuat prediksi dikenal sebagai *explanatory* atau variabel bebas. Contohnya, perusahaan retail dapat menggunakan *data mining* untuk memprediksikan penjualan dari produk mereka di masa depan dengan menggunakan data-data yang telah didapatkan dari beberapa minggu.

### 2. *Descriptive*

*Descriptive* dalam *data mining* merupakan proses untuk menemukan karakteristik penting dari data dalam suatu basis data. Tujuan dari tugas *descriptive* adalah untuk menurunkan pola-pola (korelasi, *trend*, *cluster*, teritori, dan anomali) yang meringkas hubungan yang pokok dalam data. Tugas *data mining descriptive* sering merupakan penyelidikan dan seringkali memerlukan teknik *post-processing* untuk validasi dan penjelasan hasil.

Menurut Larose (2005), *data mining* dilakukan dengan *tool* khusus yang mengeksekusi operasi *data mining* yang telah didefinisikan berdasarkan model analisis. Terhadap data dengan penekanan menemukan informasi yang tersembunyi pada sejumlah data besar yang disimpan ketika menjalankan bisnis perusahaan. Faktor yang mendorong kemajuan dalam *data mining* antara lain:

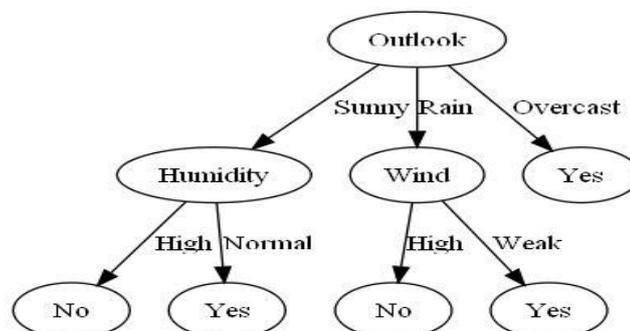
1. Pertumbuhan yang cepat dalam kumpulan data
2. Penyimpanan data dalam *data warehouse*, sehingga seluruh perusahaan memiliki akses ke dalam *database* yang andal
3. Adanya peningkatan akses data melalui navigasi *web* dan internet
4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi
5. Perkembangan teknologi perangkat lunak untuk *data mining* (ketersediaan teknologi)

6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Menurut Maclennan, Tang, & Crivat (2009), fungsi *data mining* secara umum adalah sebagai berikut:

### 1. *Classification*

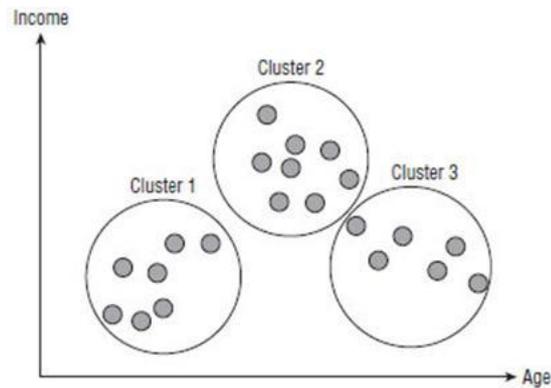
Proses untuk mencari model atau fungsi yang menggambarkan dan membedakan kelas-kelas atau konsep data untuk mengklasifikasikan target *class* ke dalam kategori yang dipilih disebut *classification*. Salah satu contohnya adalah *desicion tree* yaitu sebuah *flowchart* yang menyerupai struktur pohon, dimana tiap node menunjukkan sebuah hasil dari test yang mempresentasikan kelas-kelas atau distribusi kelas. Dalam klasifikasi, terdapat target variabel kategori, misalnya pendapatan tinggi, sedang, dan rendah (Larose, 2005). Ada banyak metode untuk membangun klasifikasi seperti *naïve-bayesian*, *Support Vector Machine (SVM)*, *random forest*, dan *neighbor classification*.



**Gambar 3. 8** Contoh *classification (decision tree)*

### 2. *Clustering*

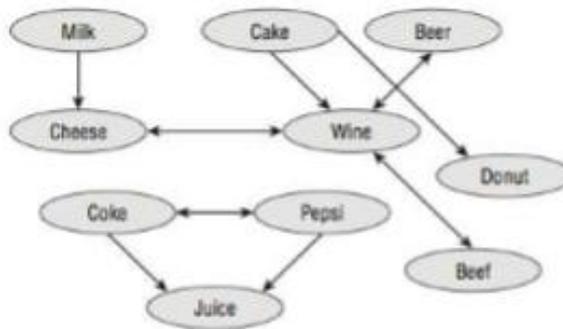
Fungsi dari *clustering* adalah untuk mencari pengelompokkan atribut ke dalam segmentasi-segmentasi berdasarkan similaritas. Menurut Larose (2005), *cluster* berbeda dengan *classification* karena tidak adanya variabel target dalam *cluster*. Algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan *record* dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.



**Gambar 3. 9** Contoh *clustering*

### 3. *Association*

Fungsi dari *association* adalah untuk mencari keterkaitan antara atribut atau *item set*, berdasarkan jumlah *item* yang muncul dari *association rule* yang ada. Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu (Larose, 2005).



**Gambar 3. 10** Contoh *association*

### 4. *Regression*

Fungsi dari *regression* hampir sama dengan klasifikasi, yaitu bertujuan untuk mencari prediksi dari suatu pola yang ada.

### 5. *Forecasting*

Fungsi dari *forecasting* adalah untuk peramalan waktu yang akan datang berdasarkan *trend* yang telah terjadi di waktu sebelumnya.



**Gambar 3. 11** Contoh *forecasting*

#### 6. *Sequence Analysis*

Fungsi dari *sequence analysis* adalah untuk mencari pola urutan dari rangkaian kejadian.

#### 7. *Deviation Analysis*

Fungsi dari *deviation analysis* adalah untuk mencari kejadian langka yang sangat berbeda dari keadaan normal (kejadian abnormal).

Kelebihan *data mining* sebagai alat analisis, yaitu:

- Data mining* mampu menangani data dalam jumlah besar dan kompleks.
- Data mining* dapat menangani data dengan berbagai macam tipe atribut.
- Data mining* mampu mencari dan mengolah data secara otomatis, karena dalam beberapa teknik *data mining* diperlukan parameter yang harus di-*input* oleh *user* secara manual.
- Data mining* dapat menggunakan pengalaman ataupun kesalahan terdahulu untuk meningkatkan kualitas dan hasil analisa sehingga mendapat hasil yang terbaik.

### 3.6. *Machine Learning*

Metode klasifikasi *machine learning* telah lama digunakan di bidang *data mining* dan banyak bidang lain dari ilmu komputer sebagai algoritma klasifikasi yang melibatkan metode dari statistik, kecerdasan buatan dan manajemen *database*, juga dapat dikategorikan sebagai kunci elemen dalam interpretasi data dan visualisasi data (Berry & Linoff, 2004). *Machine learning* membutuhkan data untuk belajar sehingga juga diistilahkan dengan *learn from data* (Alpaydin, 2010).

*Machine Learning* adalah cabang dari kecerdasan buatan, merupakan disiplin ilmu yang mencakup perancangan dan pengembangan algoritma yang memungkinkan komputer untuk mengembangkan perilaku yang didasarkan kepada data empiris, seperti dari sensor data pada basis data.

Sistem pembelajaran dapat memanfaatkan contoh atau data untuk menangkap ciri yang diperlukan dari probabilitas yang mendasarinya atau yang tidak diketahui. Salah satu tugas dari *Machine Learning* adalah klasifikasi. Jika semua informasi penting sudah terkumpul, hal terakhir yang akan dilakukan pastinya adalah proses klasifikasi, yang nantinya akan menghasilkan output berupa jenis klasifikasi yang kita inginkan. Klasifikasi pada *Machine Learning* biasanya dilakukan dengan menggunakan algoritma yang sudah sangat baik untuk digunakan dalam proses tersebut.

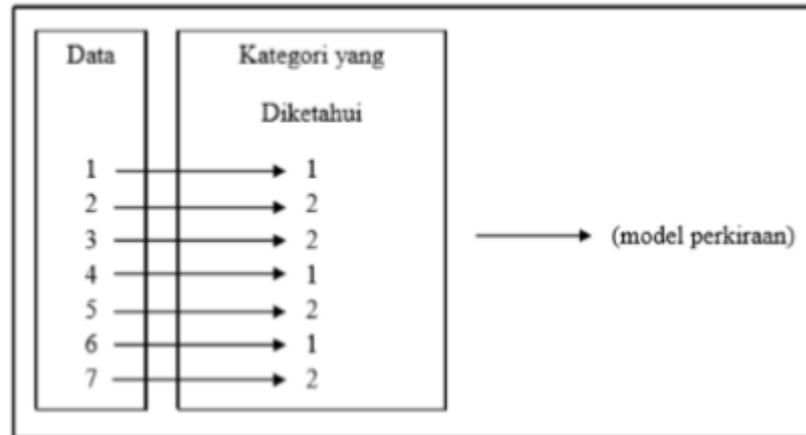
Tugas utama dari *Machine Learning* adalah klasifikasi. Untuk melakukan proses klasifikasi semua informasi penting harus dikumpulkan, yang nantinya akan menghasilkan *output* berupa jenis klasifikasi yang kita inginkan. Klasifikasi pada *Machine Learning* biasanya dilakukan dengan menggunakan algoritma yang sudah sangat baik untuk digunakan dalam proses tersebut. Pada bagian klasifikasi proses yang terjadi adalah menentukan pada class apa sebuah instance itu berada. Tugas lain dari *Machine Learning* adalah regresi. Regresi adalah prediksi dari nilai numerik.

### **3.7. Klasifikasi**

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui (Han & Kamber, 2011). Sedangkan menurut Olson & Shi (2013), metode-metode klasifikasi ditunjukkan untuk pembelajaran fungsi-fungsi berbeda yang memetakan masing-masing data terpilih ke dalam salah satu dari kelompok kelas yang telah ditetapkan sebelumnya.

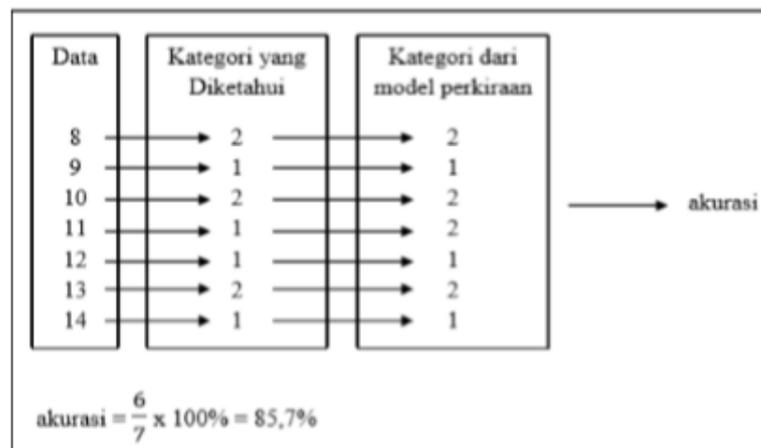
Menurut Aliady dkk (2017), proses klasifikasi dapat dibagi menjadi dua fase. Fase pertama adalah fase *learning* dimana sebagian data yang sudah memiliki label kelas data akan diumpamakan untuk membentuk model perkiraan.

Pada gambar dibawah ini merupakan fase *learning* dimana suatu metode mencari pemecahan masalah dengan *input* berupa data dan hasil kategori yang diketahui.



**Gambar 3. 12** Fase *Learning*

Fase kedua yaitu fase *testing* dimana model yang sudah terbentuk diuji dengan sebagian data lainnya (selain data yang digunakan untuk fase *learning*) sehingga didapatkan nilai akurasi dari model yang dihasilkan pada fase *learning*. Setelah didapatkan dan nilai akurasinya telah mencukupi maka model ini dapat dipakai untuk memprediksi kelas data lainnya yang belum diketahui (selain data yang digunakan untuk fase *learning* dan *testing*). Dapat dilihat pada gambar dibawah ini, fase *testing* mencari tingkat akurasi dari sebuah algoritma yang telah dibuat oleh metod tersebut dimana jika tingkat akurasi terbesar merupakan metode terbaik yang didapatkan.



**Gambar 3. 13** Fase *Testing*

### 3.8. *Decision Tree*

*Decision tree* atau pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal, selain itu juga berguna untuk mengeksplor data dan menemukan hubungan yang tersembunyi (Witten & Frank, 2005). Menurut Saputra (2014), *decision tree* merupakan metode yang paling efisien untuk menyaring sesuatu lewat pohon keputusan apakah suatu data lolos atau tidak terhadap saringan dengan proses yang cukup cepat dengan tahapan dalam membuat sebuah pohon keputusan sebagai berikut:

1. Menyiapkan data *training* yang sudah dikelompokkan ke dalam kelas tertentu.
2. Menentukan akar dari atribut, dimana akar akan diambil dari atribut yang terpilih yaitu dengan cara menghitung nilai *gain* dari masing-masing atribut. Nilai *gain* yang tertinggi akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung nilai *entropy*.

Untuk menghitung nilai *entropy* menggunakan rumus :

$$Entropy(y) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - \dots - p_n \log_2 p_n \quad (3.3)$$

dimana:  $p_1, p_2, \dots, p_n$  = proporsi kelas 1, kelas 2, ..., dan kelas n dalam *output*.

### 3.9. *Random Forest*

Istilah *Random Forest* diusulkan pertama kali oleh Tin Kam Ho pada tahun 1995, lalu dikembangkan oleh Leo Breiman pada tahun 2001. *Random Forest* merupakan salah satu metode *ensemble* untuk meningkatkan akurasi suatu klasifikasi data dari sebuah pemilah tunggal yang tidak stabil melalui kombinasi banyak pemilah dari suatu metode yang sama dengan proses *voting* untuk memperoleh prediksi klasifikasi akhir (Wezel & Potharst, 2007).

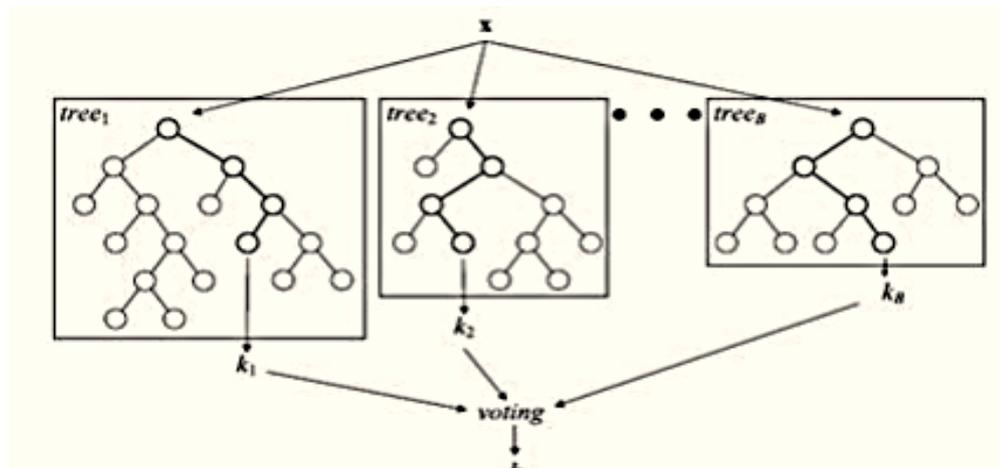
Metode *random forest* adalah pengembangan dari metode *Classification and Regression Tree* (CART), yaitu dengan menerapkan metode *bootstrap aggregating (bagging)* dan *random feature selection* (Breiman, 2001). CART sendiri merupakan metode eksplorasi data yang didasarkan pada teknik pohon keputusan dimana pohon klasifikasi dihasilkan saat peubah respon berupa data kategorik sedangkan pohon regresi dihasilkan saat peubah respons berupa data

numerik. Pembangunan pohon klasifikasi CART meliputi tiga hal, yaitu (Breiman, 1984):

1. Pemilihan pemilah (*split*).
2. Penentuan simpul terminal.
3. Penandaan label kelas.

*Bootstrap aggregating (Bagging)* adalah teknik yang dapat digunakan untuk membentuk sampel *bootstrap* dimana setiap pohon keputusan dibangun dengan menggunakan sampel *bootstrap* dari data kandidat atribut untuk dibagi di setiap node yang berasal dari himpunan atribut acak dari hasil data yang dipilih dan disimpan (Mambang & Byna, 2017). Menurut Alfaro dk (2008), *Bagging* merupakan metode *ensemble* yang banyak diterapkan pada algoritma klasifikasi yang bertujuan meningkatkan akurasi pengklasifikasi dengan menggabungkan pengklasifikasi tunggal dan hasilnya lebih baik daripada *random sampling*. *Bagging* dan *Boosting* merupakan metode *ensemble* yang relatif baru namun telah menjadi populer (Nidhomuddin dkk, 2015).

Menurut Liaw & Wiener (2002), *Random forest* melalui proses pengacakan yang tidak hanya dilakukan pada data sampel saja melainkan juga pada pengambilan variabel bebas sehingga pohon klasifikasi yang dibangun akan memiliki ukuran dan bentuk yang berbeda-beda. *Random forest* adalah klasifikasi yang terdiri dari beberapa pohon keputusan yang dibangun dengan menggunakan vektor acak (Mambang & Byna, 2017). Menurut Adnyana (2015), *random forest* merupakan pengembangan dari *decision tree* dengan menggunakan beberapa *decision tree* dimana setiap *decision tree* telah dilakukan *training* menggunakan sampel individu dan setiap atribut dipecah pada *tree* yang dipilih antara atribut *subset* yang bersifat acak dan pada proses klasifikasi, individunya didasarkan pada *vote* dari suara terbanyak pada kumpulan populasi *tree*.



**Gambar 3. 14** Contoh *Random Forest*

Menurut Nugroho & Emiliyawati (2017) pada saat pemilihan menentukan klasifikasi secara keseluruhan, pohon-pohon yang buruk akan membuat prediksi yang merupakan prediktor yang baik akan muncul sebagai jawaban. Menurut Dewo (2011) agar ukuran kepentingan semakin stabil disarankan untuk menggunakan banyak pohon ketika penelitian mempertimbangkan ukuran kepentingan dan saat dihadapkan pada variabel bebas yang banyak. Terdapat ukuran kepentingan pada *random forest*, yaitu *Mean Decrease Accuracy* (MCA) dan *Mean Decrease Gini* (MDG).

Operator *random forest* menghasilkan satu set pohon acak, kelas yang dihasilkan dari proses klasifikasi dipilih dari kelas yang paling banyak (modus) yang dihasilkan oleh pohon acak yang ada (Biau, 2012). Banyak pohon ditumbuhkan dalam metode *random forest*, sehingga terbentuk hutan (*forest*) yang akan dianalisis. Pada gugus data yang terdiri atas  $n$  amatan dan  $p$  peubah penjelas, *random forest* dilakukan dengan cara (Breiman, 2001; Breiman & Cutler, 2003):

1. Melakukan penarikan contoh acak berukuran  $n$  dengan pemulihan pada gugus data dimana tahapan ini merupakan tahapan *bootstrap*.
2. Dengan menggunakan contoh *bootstrap*, pohon dibangun sampai mencapai ukuran maksimum (tanpa pemangkasan). Pada setiap simpul, pemilihan pemilah dilakukan dengan memilih  $m$  peubah penjelas secara acak, dimana  $m \ll p$ , lalu pemilah terbaik dipilih berdasarkan  $m$  peubah penjelas tersebut dimana tahapan ini disebut dengan tahapan *random feature selection*.
3. Ulangi langkah 1 dan 2 sebanyak  $k$  kali, sehingga terbentuk sebuah hutan yang terdiri atas  $k$  pohon.

Metode *random forest* harus menentukan  $m$  jumlah variabel prediktor yang diambil secara acak dan  $k$  pohon yang akan dibentuk agar mendapatkan hasil yang optimal. Menurut Breiman (1996), nilai  $k$  yang disarankan untuk digunakan pada metode *bagging* juga dicobakan yaitu  $k = 50$  sudah memberikan hasil yang memuaskan untuk masalah klasifikasi sedangkan menurut Sutton (2005)  $k \geq 100$  cenderung menghasilkan tingkat misklasifikasi yang rendah.

Ukuran contoh peubah penjelas ( $m$ ) saat menggunakan metode *random forest* sangat mempengaruhi korelasi dan kekuatan masing-masing pohon. Untuk menentukan  $m$  yaitu jumlah variabel prediktor yang diambil secara acak dengan nilai  $p$  adalah banyak variabel *independent* (bebas), seperti berikut ini (Hastie dkk, 2001):

1. Untuk klasifikasi, penentuan nilai  $m$  adalah dengan cara  $\lfloor \sqrt{p} \rfloor$  dengan nilai *node* atau simpul terkecil adalah 1.
2. Untuk regresi, penentuan nilai  $m$  adalah dengan cara  $\lfloor p/3 \rfloor$  dengan nilai *node* atau simpul terkecil adalah 5.

Sedangkan menurut Breiman & Cutler (2003), terdapat tiga cara untuk mendapatkan nilai  $m$  untuk mengamati *error* OOB yaitu:

$$m = \frac{1}{2} \lfloor \sqrt{p} \rfloor \quad (3.4)$$

$$m = \lfloor \sqrt{p} \rfloor \quad (3.5)$$

$$m = 2 \times \lfloor \sqrt{p} \rfloor \quad (3.6)$$

dimana:  $p$  = total variabel .

Menurut Breiman (2001), penggunaan  $m$  yang tepat akan menghasilkan *random forest* dengan korelasi antar pohon cukup kecil namun kekuatan setiap pohon cukup besar yang ditunjukkan dengan perolehan *error* OOB bernilai kecil. Terdapat respons suatu amatan diprediksi dengan menggabungkan (*aggregating*) hasil prediksi  $k$  pohon. Pada masalah klasifikasi dilakukan berdasarkan *majority vote* atau kategori atau kelas yang paling sering muncul sebagai hasil prediksi dari  $k$  pohon klasifikasi. Terdapat *data out-of-bag* (OOB), yaitu sepertiga amatan gugus data asli yang tidak termuat dalam contoh *bootstrap* pada setiap iterasinya (Breiman, 2001).

Data OOB digunakan untuk membangun pohon, melainkan menjadi data validasi pada pohon yang bersesuaian. Nilai salah klasifikasi *random forest* diduga melalui *error* OOB yang diperoleh dengan cara (Breiman, 2001; Breiman & Cutler, 2003; Liaw & Wiener, 2002):

1. Lakukan prediksi terhadap setiap data OOB pada pohon yang bersesuaian.
2. Umumnya setiap amatan gugus data asli akan menjadi data OOB sebanyak sekitar 36% atau sepertiga dari banyak pohon yang dibentuk. Oleh karena itu, pada langkah 1 masing-masing amatan gugus data asli mengalami prediksi sebanyak sekitar sepertiga kali dari banyaknya pohon. Jika  $a$  adalah sebuah amatan dari gugus data asli, maka hasil prediksi *random forest* terhadap  $a$  adalah gabungan dari hasil prediksi setiap kali  $a$  menjadi data OOB.

*Error OOB* bergantung pada korelasi antar pohon dan kekuatan (strength) masing-masing pohon dalam *random forest* dimana peningkatan korelasi dapat meningkatkan *error OOB* sedangkan peningkatan pohon dapat menurunkan *error OOB* (Breiman, 2001). *Error OOB* dihitung dari proporsi misklasifikasi hasil prediksi *random forest* dari seluruh amatan gugus data asli. Menurut Breiman & Cutler (2003), menggunakan banyak pohon misalnya 1000 pohon atau lebih agar menghasilkan *variable importance* semakin stabil.

### 3.9.1 Ukuran Ketepatan Klasifikasi

*Imbalanced Dataset* adalah dataset yang memiliki contoh kelas negatif (kelas mayoritas) jauh lebih banyak daripada contoh kelas positif (kelas minoritas) (Sembiring, 2007). Menurut Cai (2009), dalam mempelajari data yang tidak seimbang, akurasi klasifikasi secara keseluruhan seringkali bukan ukuran kinerja yang tepat. Untuk data tidak seimbang, akurasi lebih didominasi oleh ketepatan pada data kelas minoritas, maka matriks yang tepat adalah AUC (*Area Under the ROC Curve*), *F-Measure*, *G-mean*, *Apparent Error Rate* (APER), *Total Accuracy Rate* (1-APER), dan akurasi kelas minoritas (Zhang & Wang, 2011). Evaluasi kinerja model klasifikasi didasarkan pada pengujian objek yang diprediksi dengan benar dan salah, hitungan ini ditabulasikan *confusion matrix* (Alfisahrin, 2014).

Untuk mengukur kinerja klasifikasi digunakan *confusion matrix* yang memberikan keputusan yang diperoleh dalam pelatihan dan pengujian (Bramer, 2007). *Confusion Matrix* adalah alat yang berguna untuk menganalisis seberapa

baik *classifier* dapat mengenali tupel dari kelas yang berbeda dimana kelas yang diprediksi akan ditampilkan dibagian atas *matrix* dan kelas yang diobservasi ditampilkan dibagian kiri (Han & Kamber, 2006). Berikut adalah tabel matriks *confusion* seperti **Tabel 3.2** di bawah ini.

**Tabel 3. 2** Matriks Confusion

	<i>Predicted Positive Class</i>	<i>Predicted Negative Class</i>
<i>Actual Positive Class</i>	<i>TP (True Positive)</i>	<i>FN (False Negative)</i>
<i>Actual Negative Class</i>	<i>FP (False Positive)</i>	<i>TN (True Negative)</i>

dimana:

TP (*True Positive*) : Jumlah observasi positif yang tepat prediksi.

TN (*True Negative*) : Jumlah observasi negatif yang tepat prediksi.

FP (*False Positive*) : Jumlah obesrvasi positif yang salah diprediksi sebagai negatif.

FN (*False Negative*) : Jumlah observasi negatif yang salah diprediksi sebagai positif.

*False Positive* dikenal sebagai *error* tipe 1, terjadi ketika kasus yang seharusnya diklasifikasikan sebagai negatif diklasifikasikan positif sedangkan *false negative* dikenal sebagai *error* tipe 2 terjadi jika kasus yang seharusnya diklasifikasikan sebagai positif diklasifikasikan negatif (Bramer, 2013).

Menurut Wang & Yao (2013), akurasi kelas minoritas dapat menggunakan matriks *True Positive Rate* atau *recall* (sensitivitas), *G-mean* dan *AUC (Area Under the ROC Curve)* merupakan evaluasi prediktor yang lebih komprehensif dalam konteks ketidakseimbangan. Dengan menggunakan matriks seperti *true negative rate (specificity)*, *true positive rate (sensitivity)*, *Apparent Error Rate (APER)*, *total accuracy rate (1-APER)*, *G-mean*, *precision*, dan *F-measure* untuk dapat mengevaluasi kinerja *Machine learning* pada saat data yang tidak seimbang. seperti **Tabel 3.2** maka didapatkan rumus sebagai berikut:

$$\text{True Negative Rate (Acc}^{\bar{)} \text{ atau specificity} = \frac{TN}{TN+FP} \quad (3.7)$$

$$\text{True Positive Rate (Acc}^{+} \text{) atau sensitivity / Recall} = \frac{TP}{FN+TP} \quad (3.8)$$

$$\text{Precision atau PPV} = \frac{TP}{TP+FP} \quad (3.9)$$

$$\text{Apparent Error Rate (APER)} = \frac{FP+FN}{N} \quad (3.10)$$

$$\text{Total Accuracy Rate (1-APER)} = \frac{TP+TN}{N} \quad (3.11)$$

$$G\text{-mean} = \sqrt{\text{specificity} \times \text{sensitivity}} \quad (3.12)$$

$$F\text{-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{(\text{Precision} + \text{Recall})} \quad (3.13)$$

$$\text{False Positive Rate} = \frac{FP}{FP + TN} \quad (3.14)$$

$$\text{Area Under the ROC Curve} = \frac{1 + \text{Sensitivity} - \text{FPrate}}{2} \quad (3.15)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.16)$$

*Sensitivity* (sensitivitas) dan *specificity* (spesifisitas) berguna sebagai ukuran statistik dari kinerja klasifikasi *biner*, mengukur model terbaik, dan memilih model yang paling efisien (Prasetio & Pratiwi, 2015). Menurut Bramer (2013), sensitivitas mengukur proporsi *true positive* yang diidentifikasi dengan benar sedangkan spesifitas mengukur proporsi *true negative* yang diidentifikasi dengan benar. Untuk meningkatkan *recall* atau *sensitivity* tanpa mempengaruhi *precision* adalah tujuan utama data *training* dari dataset *imbalanced*, namun tujuan dari *precision* dan *recall* sering bersebrangan karena ketika meningkatkan nilai *true positive* kelas minoritas, jumlah *false positive* juga meningkat dan menyebabkan nilai *precision* berkurang (Jatmiko dkk, 2017).

*Precision* adalah tingkat ketepatan antara Nilai *G-mean* bertujuan untuk keseimbangan akurasi prediksi yang biasa digunakan untuk mengukur permasalahan data *imbalanced*. Menurut Jatmiko dkk (2017), hal tersebut diperlukan karena metode klasifikasi cenderung baik dalam memprediksi kelas dengan data sampel yang lebih banyak namun buruk dalam memprediksi kelas dengan data sampel yang sedikit. Menurut Ramadhani dkk (2017), *F-measure* berguna untuk mengukur beberapa kriteria kinerja klasifikasi untuk mendukung kesimpulan dan penyelesaian masalah dalam penelitian yang dilakukan.

*Area Under the ROC Curve* (AUC) merupakan ukuran numerik untuk membedakan kinerja model dan menunjukkan seberapa sukses dan benar peringkat model dengan memisahkan pengamatan positif dan negatif (Attenberg & Ertekin, 2013). Menurut Weiss (2013), AUC merangkum informasi kinerja pengklasifikasi ke dalam satu angka yang memperoleh perbandingan model ketika tidak ada kurva ROC yang mendominasi. AUC merupakan cara yang baik untuk mendapatkan nilai kinerja pengklasifikasi secara umum dan untuk membandingkannya dengan pengklasifikasian yang lain (Japkowicz, 2013).

Menurut Liu & Zhou (2013), AUC adalah ukuran kinerja yang populer dalam ketidakseimbangan kelas dimana jika nilai AUC tinggi menunjukkan kinerja yang lebih baik sehingga untuk memilih model terbaik dengan cara menganalisa nilai AUC.

### 3.9.2 Ukuran Tingkat Kepentingan

Salah satu ukuran tingkat kepentingan variabel bebas yang dihasilkan oleh *random forest* adalah *Mean Decrease Accuracy* (MDA), dimana MDA menampilkan besar tambahan obeservasi yang mengalami misklasifikasi jika satu persatu variabel bebas tidak diikutsertakan dalam pengujian. Menurut Breiman & Cutler (2003) dan Sandri & Zuccolotto (2006), *Mean Decrease Gini* (MDG) adalah salah satu ukuran tingkat kepentingan (*variable importance*) peubah penjelas yang dihasilkan oleh metode *random forest*. Misalkan terdapat  $p$  peubah penjelas dengan  $h = 1, 2, \dots, p$  maka MDG mengukur tingkat kepentingan peubah penjelas  $X_h$  dengan cara:

$$MDG_h = \frac{1}{k} \sum_t [d(h, t) I(h, t)] \quad (3.15)$$

dimana:  $d(h, t)$  : besar penurunan indeks Gini untuk peubah penjelas  $X_h$

$I(h, t)$  : memilah simpul  $t$

$k$  : banyaknya pohon dalam *random forest* (ukuran *random forest*)

### 3.10. Teknik Sampling Kelas *Imbalanced*

Data *imbalanced* adalah suatu kondisi dimana kategori klasifikasi tidak terwakili secara seimbang (Chawla, 2002). Menurut Barandela dkk (2003), salah satu metode yang paling populer untuk masalah ketidakseimbangan kelas adalah teknik *sampling*. Dataset *imbalanced* dapat menyebabkan masalah seperti *accuracy paradox* dimana untuk analisis prediktif menyatakan model prediktif dengan tingkat akurasi tertentu mungkin memiliki daya prediksi lebih besar daripada model dengan tingkat akurasi yang lebih tinggi, dimana ukuran ketepatan klasifikasi yang lebih baik dipakai seperti presisi dan *recall* (Zhu, 2007).

Menurut Chawla (2002), permasalahan kelas *imbalanced* dapat ditangani melalui dua pendekatan. Pertama adalah menetapkan biaya yang berbeda untuk contoh-contoh *training* atau cara kedua yaitu melakukan *resample* pada data asli, baik melalui *oversampling* pada kelas minoritas, *undersampling* pada kelas

mayoritas maupun memadukan *undersampling* pada kelas mayoritas dengan bentuk khusus dari *oversampling* pada kelas minoritas yang dikenal dengan *Synthetic Minority Over Sampling Technique* (SMOTE).

Teknik *oversampling* dan *undersampling* merupakan teknik yang paling sering digunakan untuk mengubah distribusi kelas asli dari dataset yang tidak seimbang dengan menghilangkan sebagian besar data kelas mayoritas atau meningkatkan data dari kelas minoritas (Prasetio & Pratiwi, 2015).

Hasil pengujian metode klasifikasi di dataset yang tidak seimbang biasanya memiliki ciri khas berupa nilai *instance* yang terklasifikasi (*misclassification cost*) di kelas minoritas lebih tinggi dibandingkan dengan *missclassification cost* di kelas mayoritas. Untuk mengatasi hal tersebut yaitu mengklasifikasi kelas minoritas secara lebih akurat, untuk mengatasi ketidakseimbangan kelas adalah dengan melakukan sampel ulang (*resample*) dataset asli, baik pada kelas minoritas (*oversampling*) ataupun kelas mayoritas (*undersampling*). Beberapa teknik sampling untuk mengatasi kelas *imbalanced* dalam *machine learning* antara lain (Japkowicz, 2000):

#### 1. *Undersampling*

Teknik *undersampling* ini merupakan metode sampling dimana metode yang digunakan dalam kategori ini berisi *re-sampling* eliminasi dari kelas mayoritas secara acak sampai jumlahnya sebanyak contoh dari kelas lainnya atau kelas minoritas.

#### 2. *Oversampling*

Teknik *oversampling* ini merupakan metode sampling dimana metode yang digunakan dalam kategori ini berisi *re-sampling* dari kelas minoritas secara acak sampai jumlahnya sebanyak contoh dari kelas lainnya atau kelas mayoritas.

#### 3. *Synthetic Minority Over-sampling Technique* (SMOTE)

Teknik SMOTE ini merupakan metode *sampling* dimana kelas minoritas dilakukan *oversampling* dengan membuat sampel “sintetis” daripada dilakukan *oversampling* dengan melakukan duplikasi pada data sebenarnya dan tergantung jumlah *oversampling* yang dibutuhkan.



## BAB IV

### METODOLOGI PENELITIAN

#### 4.1. Populasi Penelitian

Populasi dalam penelitian ini adalah data nasabah pembiayaan Warung Mikro Bank Syariah Mandiri KC pada tahun 2014 – 2017.

#### 4.2. Metode Penelitian

Metode pertama yang digunakan dalam penelitian ini adalah analisis deskriptif, dimana metode ini bertujuan untuk memberikan gambaran umum dari data yang akan diteliti. Selanjutnya, metode kedua yaitu metode *Random Forest* dimana metode tersebut yang akan digunakan untuk mengklasifikasi pembiayaan bermasalah pada data nasabah pembiayaan Warung Mikro Bank Syariah Mandiri KC Jambi.

#### 4.3. Variabel Penelitian

Variabel penelitian ini terdiri dari variabel dependen (terikat) dan variabel independen (bebas). Variabel dependen yang digunakan yaitu keterangan saldo mencukupi atau tidak mencukupi. Keterangan saldo mencukupi atau tidak mencukupi tersebut berpengaruh terhadap masalah pembiayaan Warung Mikro, dimana jika saldo tidak mencukupi akan mempengaruhi macet tidaknya pembiayaan tersebut. Terdapat enam variabel independen pada penelitian ini yaitu jenis kelamin, pekerjaan, jenis nasabah, jenis pembiayaan mikro, lama angsuran, jumlah pinjaman, margin, dan besarnya angsuran.

Berikut adalah uraian dari definisi operasional variabel dependen dan independen di atas.

**Tabel 4. 1** Definisi Operasional Variabel

No	Variabel	Definisi	Satuan
1	Keterangan saldo mencukupi dan tidak mencukupi (Y)	Keterangan apakah saldo nasabah masuk ke dalam kategori mencukupi atau tidak mencukupi.	Ya / Tidak
2	Umur (X1)	Umur nasabah pembiayaan Warung Mikro Bank Syariah Mandiri KC Jambi.	Tahun

No	Variabel	Definisi	Satuan
3	Jenis Nasabah (X2)	Jenis pihak yang membutuhkan fasilitas pembiayaan dari bank untuk menyelesaikan atau mengembangkan usahanya.	Perorangan / Kerjasama KUD / Kerjasama Perusahaan
4	Jenis Pembiayaan Mikro (X3)	Jenis Pembiayaan Mikro yang dinilai berdasarkan kategori harga pokok pembiayaan.	PM Madya / PM Utama
5	Harga Pokok Pinjaman (X4)	Jumlah pembiayaan yang diajukan oleh nasabah.	Rupiah
6	Margin (X5)	Selisih antara harga jual dan harga pokok pembiayaan dengan skema jual beli ( <i>Murabahah</i> dan <i>Ijarah</i> ).	%
7	Lama Angsuran (X6)	Lama angsuran sesuai dengan jumlah pinjaman yang diajukan oleh nasabah tersebut.	Bulan
8	Besarnya Angsuran (X7)	Besarnya angsuran yang harus dibayar oleh nasabah perbulannya.	Rupiah
9	Pekerjaan (X8)	Pekerjaan nasabah pembiayaan Warung Mikro Bank Syariah Mandiri KC Jambi.	-
10	Harga Jaminan (X9)	Harga dari jaminan yang digunakan oleh nasabah sebagai jaminan pembiayaan.	Rupiah

#### 4.4. Teknik Pengumpulan Data

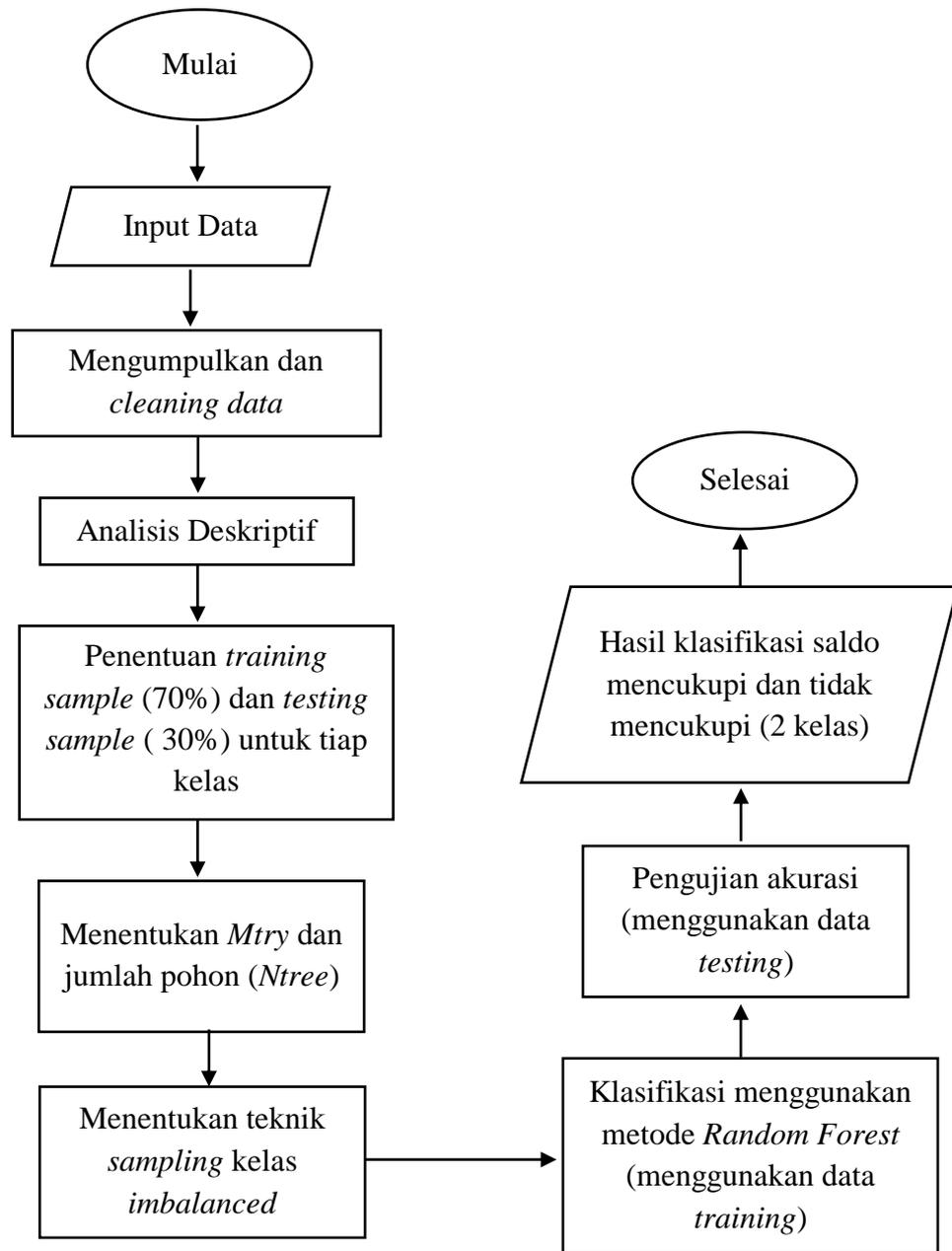
Dalam penelitian ini, data yang digunakan adalah data sekunder yang diperoleh dari Bank Syariah Mandiri KC Jambi. Data tersebut direkapitulasi oleh pegawai bagian Produk Pembiayaan Mikro di Bank Syariah Mandiri KC Jambi. Data tersebut berupa keterangan tentang nasabah yang melakukan pembiayaan Warung Mikro tahun 2014-2017.

#### 4.5. Alat dan Cara Organisir Data

Alat pengolahan data yang digunakan yaitu program *Ms. Excel* dan *R*. Dalam tahapan ini dilakukan pengkajian data yang telah diperoleh berdasarkan teori, yaitu analisis data secara deskriptif dan menggunakan metode *random forest*, kemudian dilakukan penerapan langkah-langkah sebagai berikut:

1. Melakukan analisis deskriptif semua variabel yang digunakan, baik variabel *dependent* (terikat) maupun variabel *independent* (tidak terikat).
2. Menentukan data latih (*training*) yang akan digunakan untuk melakukan metode *random forest*.
3. Menentukan nilai *Mtry* terbaik.
4. Memilih nilai *n* yang menunjukkan jumlah pohon.
5. Membuat plot untuk melihat proses pelatihan data latih agar mendapatkan hasil yang lebih optimal.
6. Memilih teknik *sampling* yang akan digunakan pada metode *random forest*.
7. Mengklasifikasi data dengan metode *random forest* menggunakan data latih (*training*) yang telah ditentukan sebelumnya.
8. Melakukan pengujian akurasi menggunakan data uji (*testing*) untuk melihat seberapa besar akurasi metode *random forest* terhadap data penelitian.
9. Membuat plot ukuran tingkat kepentingan *variable importance* dari variabel *independent* (tidak terikat).

#### 4.6. Diagram Alir Penelitian



**Gambar 4. 1** Diagram Alir Penelitian

## BAB V

### HASIL DAN PEMBAHASAN

#### 5.1. Analisis Deskriptif

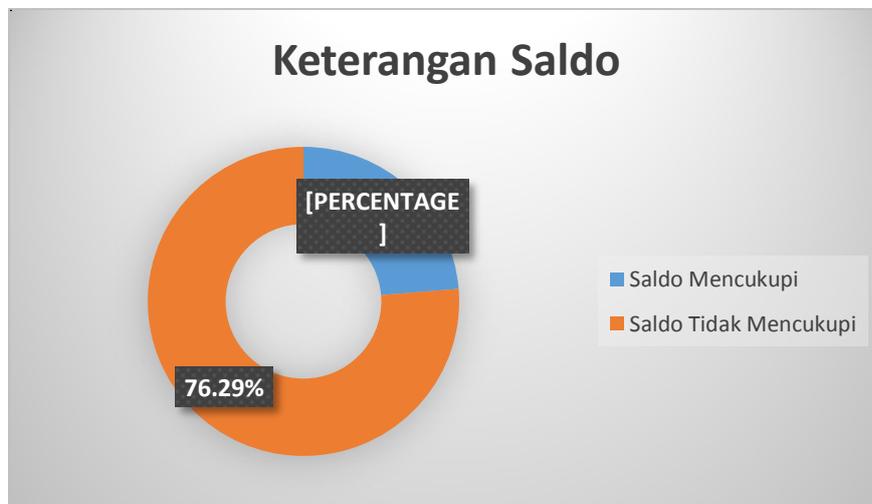
Peneliti akan melihat gambaran umum dengan membuat analisis deskriptif dari masing-masing variabel data nasabah Warung Mikro Bank Syariah Mandiri KC Jambi tahun 2014-2017. Pertama, peneliti akan melihat tipe data setiap variabel yang digunakan pada penelitian.

```
str(datanasabah)
'data.frame': 232 obs. of 10 variables:
 $ x1: int  50 50 48 62 26 44 47 36 51 29 ...
 $ x2: Factor w/ 2 levels "Kerjasama KUD",...: 2 2 1 1 2 1 2 2
 2 2 ...
 $ x3: Factor w/ 2 levels "PM MADYA","PM UTAMA": 2 1 1 1 1 1
 1 1 1 1 ...
 $ x4: int  60000000 50000000 50000000 50000000 17000000 3500
 0000 50000000 25000000 50000000 15000000 ...
 $ x5: num  26 30 23 23 22 23 26 26 24 22 ...
 $ x6: int  48 36 36 36 36 36 36 36 36 36 ...
 $ x7: int  2023038 2122579 1935486 1953490 649238 1354840 20
 14531 1007266 1961643 572857 ...
 $ x8: Factor w/ 10 levels "Buruh Harian Lepas",...: 10 10 6 6
 10 6 10 10 10 4 ...
 $ x9: int  75000000 62500000 62500000 62500000 21250000 4375
 0000 62500000 31250000 62500000 18750000 ...
 $ y : Factor w/ 2 levels "N","Y": 1 1 1 1 1 1 1 1 1 1 ...
```

**Gambar 5.1** Tipe Data Variabel-variabel Penelitian

Pada **Gambar 5.1** terdapat 232 observasi dan 10 variabel yang digunakan pada penelitian. Variabel *y* adalah variabel *dependent* (terikat) yaitu keterangan saldo, dengan tipe data kategorik dengan 2 kategori yaitu saldo mencukupi (Y) dan saldo tidak mencukupi (N). Selanjutnya, variabel *independent* yang terdiri dari umur (x1), jenis nasabah (x2), jenis pembiayaan mikro (X3), harga pokok pembiayaan(x4), margin (x5), jangka waktu pembiayaan (x6), angsuran perbulan (x7), pekerjaan (x8), dan harga jaminan (x9). Dimana x2, x3, dan x8 bertipe data kategorik lalu x4, x6, x7, dan x9 bertipe data *integer* (tipe data yang terdiri dari angka bulat dan tidak mengandung nilai pecahan atau nilai desimal). Sedangkan x5 dengan tipe numerik (tipe data yang terdiri dari bentuk bilangan atau angka baik angka atau bilangan tersebut merupakan bilangan bulat ataupun bilangan

real). Selanjutnya, peneliti akan melihat gambaran umum dari masing-masing variabel.



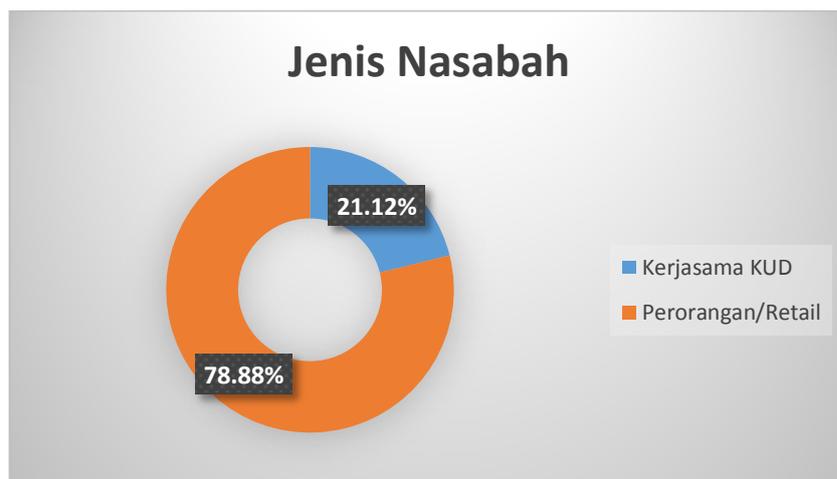
**Gambar 5. 2** Piechart Presentase Variabel Keterangan Saldo

Dapat dilihat pada **Gambar 5.2** terdapat *piechart* presentase variabel keterangan saldo yang dibagi menjadi dua kategori yaitu saldo mencukupi dan saldo tidak mencukupi. Terdapat 23.71% dari 232 nasabah dengan saldo mencukupi sedangkan 76.29% dari 232 nasabah dengan saldo tidak mencukupi. Selanjutnya, yaitu variabel umur dengan gambaran umum seperti tabel di bawah.

**Tabel 5. 1** Analisis Deskriptif Variabel Umur

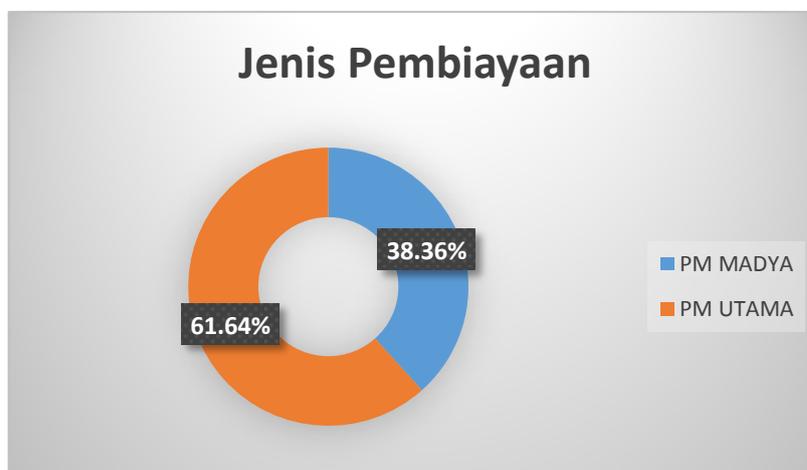
Min	1st Qu.	Median	Mean	3rd Qu.	Max
25	38	46	44.93	52	65

Pada **Tabel 5.1** terdapat analisis deskriptif dari variabel umur, dimana minimal umur data nasabah Warung Mikro BSM KC Jambi adalah 25 tahun sedangkan umur maksimal nasabah Warung Mikro BSM KC Jambi adalah 65 tahun dan rata-rata umur nasabah adalah 45 tahun.



**Gambar 5. 3** Piechart Presentase Variabel Jenis Nasabah

Pada **Gambar 5.3** dapat dilihat bahwa lebih banyak jenis nasabah yang menjalankan usahanya perorangan atau retail dibandingkan dengan kerjasama KUD, dimana 21.12% dari 232 nasabah menjalankan usaha nya perorangan atau retail dan sisanya sebesar 78.88% dari 232 nasabah menjalankan usahanya dengan kerjasama KUD. Variabel selanjutnya yaitu jenis pembiayaan mikro, yang dibagi menjadi dua kategori yaitu PM Madya dan PM Utama. PM Madya ditujukan bagi nasabah yang melakukan pembiayaan dengan harga pokok pembiayaan sebesar Rp 15.000.000,00 – Rp 50.000.000,00 dimana dapat diasumsikan PM Madya adalah jenis usaha mikro dan PM Utama ditujukan bagi nasabah yang melakukan pembiayaan dengan harga pokok pembiayaan sebesar Rp 51.000.000,00 – Rp 200.000.000,00 dimana dapat diasumsikan PM Madya adalah jenis usaha kecil.



**Gambar 5. 4** Piechart Presentase Variabel Jenis Pembiayaan Mikro

Berdasarkan **Gambar 5.4**, dapat disimpulkan bahwa terdapat 38.36% dari 232 nasabah yang mengajukan PM Madya dan 61.64% dari 232 nasabah yang mengajukan PM Utama. Itu berarti bahwa lebih banyak nasabah yang melakukan pembiayaan warung mikro > Rp 50.000.000,00. Selanjutnya yaitu variabel harga pokok pembiayaan seperti tabel di bawah ini.

**Tabel 5.2** Analisis Deskriptif Variabel Harga Pokok Pembiayaan

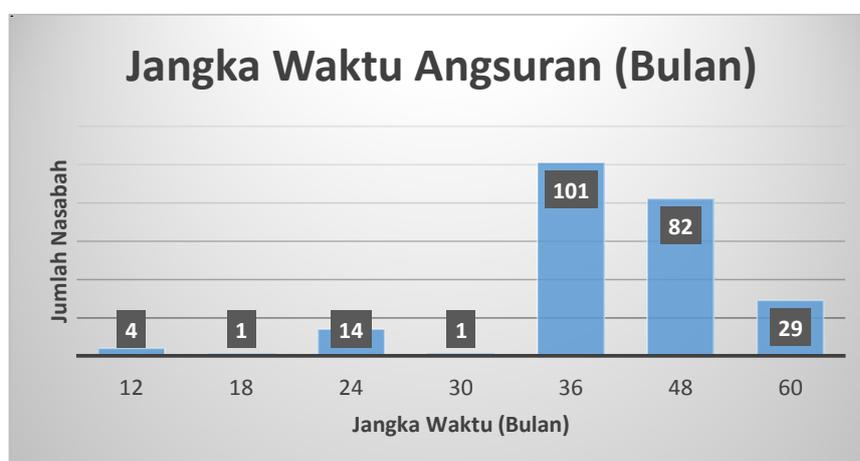
Min	1st Qu.	Median	Mean	3rd Qu.	Max
15000000	49000000	60000000	79100000	100000000	200000000

Berdasarkan **Tabel 5.2** dapat dilihat bahwa minimal harga pokok pembiayaan data nasabah Warung Mikro BSM KC Jambi tahun 2014-2017 sebesar Rp 15.000.000,00 sedangkan maksimal harga pokok pembiayaannya sebesar Rp 200.000.000,00. Rata-rata harga pokok pembiayaan sebesar Rp 79.100.000,00. Variabel selanjutnya adalah persentase margin nasabah.

**Tabel 5.3** Analisis Deskriptif Variabel Margin

Min	1st Qu.	Median	Mean	3rd Qu.	Max
16%	19.1%	22%	21.3%	23%	36%

Pada **Tabel 5.3** dapat dilihat bahwa margin terendah nasabah sebesar 16% sedangkan margin tertinggi sebesar 36%. Margin disesuaikan berdasarkan harga pokok pembiayaan, jangka waktu angsuran, dan harga jaminan. Rata-rata variabel margin sebesar 21.3%. Selanjutnya adalah variabel jangka waktu angsuran pembiayaan dalam jangka waktu bulan.



**Gambar 5.5** Grafik Jangka Waktu Angsuran (Bulan)

Pada **Gambar 5.5** terdapat variabel jangka waktu angsuran dalam bulan. Dimana pada grafik tersebut terdapat 101 nasabah dengan jangka waktu angsuran 36 bulan atau setara dengan 3 tahun, 82 nasabah dengan jangka waktu angsuran 48 bulan atau setara dengan 4 tahun, 29 nasabah dengan jangka waktu angsuran 60 bulan atau setara dengan 5 tahun, 14 nasabah dengan jangka waktu angsuran 24 bulan atau setara dengan 2 tahun, 4 nasabah dengan jangka waktu angsuran 12 bulan atau setara dengan 1 tahun, dan masing-masing 1 orang nasabah dengan jangka waktu angsuran 18 bulan (1.5 tahun) dan 30 bulan (2.5 tahun). Selanjutnya adalah variabel pekerjaan nasabah Warung Mikro BSM KC Jambi tahun 2014-2017.

**Tabel 5. 4** Analisis Deskriptif Variabel Jumlah Angsuran (Bulan)

Min	1st Qu.	Median	Mean	3rd Qu.	Max
572857	1647410	2090025	2595443	3205147	9714480

Jumlah angsuran minimal yang harus dibayarkan nasabah tiap bulannya yaitu Rp 572.857,00 sedangkan jumlah angsuran maksimal yang harus dibayarkan nasabah tiap bulannya yaitu Rp 9.714.480,00. Rata-rata jumlah angsuran yang harus dibayar nasabah tiap bulannya sebesar Rp 2.595.443,00. Selanjutnya adalah gambaran umum variabel pekerjaan nasabah Warung Mikro BSM KC Jambi.



**Gambar 5. 6** Grafik Jenis Pekerjaan

Dapat dilihat pada grafik jenis pekerjaan, jenis pekerjaan terbanyak pada data nasabah Warung Mikro BSM KC Jambi adalah wiraswasta dengan jumlah

sebanyak 152 nasabah, diikuti oleh petani dengan jumlah 42 nasabah, karyawan dengan jumlah 14 nasabah, Pegawai Negeri Sipil (PNS) dengan jumlah 13 nasabah, buruh harian lepas sebanyak 2 nasabah, pegawai sebanyak 2 nasabah, dan pekerjaan lainnya sebanyak 7 nasabah. Selanjutnya adalah variabel terakhir, yaitu variabel harga jaminan dari jaminan yang diajukan oleh nasabah.

**Tabel 5. 5** Analisis Deskriptif Harga Jaminan

Min	1st Qu.	Median	Mean	3rd Qu.	Max
16000000	61250000	75000000	90921336	12500000	250000000

Harga jaminan terendah senilai Rp 16.000.000,00 sedangkan harga jaminan tertinggi Rp 250.000.000,00 yang dapat dilihat pada **Tabel 5.5**. Rata-rata harga jaminan adalah sebesar Rp 90.921.336,00. Harga jaminan harus bisa menutup harga pokok pembiayaan yang diajukan oleh nasabah Warung Mikro tersebut. Selanjutnya, peneliti akan melihat gambaran nasabah berdasarkan variabel  $x_2$  (jenis nasabah) dan  $y$  (keterangan saldo).

**Tabel 5. 6** Tabel Silang Variabel Jenis Nasabah dan Keterangan Saldo

Jenis Nasabah	Keterangan Saldo		TOTAL
	Mencukupi	Tidak Mencukupi	
Kerjasama KUD	7	42	49
Perorangan/Retail	48	135	183
TOTAL	55	177	232

Dapat dilihat dari **Tabel 5.6** jenis nasabah perorangan/retail dengan saldo keterangan tidak mencukupi sebesar 135 nasabah. Ini berarti bahwa pada data yang digunakan, nasabah dengan keterangan saldo tidak mencukupi paling banyak terdapat pada jenis nasabah perorangan/retail. Selanjutnya, peneliti akan melihat gambaran nasabah berdasarkan variabel  $x_2$  (jenis nasabah) dan  $y$  (keterangan saldo).

**Tabel 5. 7** Tabel Silang Variabel Jenis Pembiayaan Mikro dan Keterangan Saldo

Jenis Pembiayaan	Keterangan Saldo	TOTAL
------------------	------------------	-------

Mikro	Mencukupi	Tidak Mencukupi	
PM Madya	22	67	89
PM Utama	33	110	143
<b>TOTAL</b>	<b>55</b>	<b>177</b>	<b>232</b>

Pada **Tabel 5.7** terdapat tabel silang antara variabel jenis pembiayaan mikro dan keterangan saldo, dimana dapat dilihat nasabah terbanyak terdapat pada saat PM Utama dan keterangan saldo tidak mencukupi. Ini berarti bahwa pada data yang digunakan, terdapat 110 nasabah dengan keterangan saldo tidak mencukupi paling banyak terdapat pada jenis pembiayaan PM Utama.

**Tabel 5. 8** Tabel Silang Variabel Jenis Nasabah, Jenis Pembiayaan Mikro, dan Keterangan Saldo

Jenis Nasabah	Jenis Pembiayaan Mikro	Keterangan Saldo		TOTAL
		Mencukupi	Tidak Mencukupi	
Kerjasama KUD	PM Madya	3	20	23
	PM Utama	4	22	26
<b>TOTAL</b>		<b>7</b>	<b>42</b>	<b>49</b>
Perorangan/Retail	PM Madya	19	47	66
	PM Utama	29	88	117
<b>TOTAL</b>		<b>48</b>	<b>135</b>	<b>183</b>
<b>JUMLAH TOTAL</b>		<b>55</b>	<b>177</b>	<b>232</b>

Dapat dilihat pada **Tabel 5.8**, nasabah dengan keterangan saldo tidak mencukupi paling banyak terdapat pada jenis nasabah perorangan/retail dan jenis pembiayaan PM Utama yaitu sejumlah 88 nasabah. Pada keterangan saldo mencukupi, nasabah paling banyak saat jenis nasabah perorangan/retail dan jenis pembiayaan PM Utama sejumlah 29 nasabah. Ini berarti bahwa pada data yang digunakan, nasabah dengan keterangan saldo tidak mencukupi paling banyak terdapat pada jenis nasabah perorangan/retail dan jenis pembiayaan PM Utama dan untuk nasabah dengan keterangan saldo mencukupi paling banyak juga terdapat pada jenis nasabah perorangan/retail dan jenis pembiayaan PM Utama.

## 5.2. *Random Forest*

### 5.2.1. Penentuan Data Latih (*Training*) & Data Uji (*Test*)

Data latih sangat diperlukan jika menggunakan algoritma *Machine learning*, dimana data latih digunakan untuk melatih kinerja metode pada *machine learning* tersebut. Jumlah data yang digunakan pada penelitian ini adalah 232 data, dengan keterangan 177 nasabah dengan saldo tidak mencukupi dan 55 nasabah dengan saldo mencukupi. Peneliti menggunakan 70% untuk data latih (*train*) dan 30% untuk data uji (*test*), pembagian data latih dan uji adalah sebagai berikut:

**Tabel 5.9** Data Latih (*Train*) dan Data Uji (*Test*)

	Train	Test	TOTAL
Mencukupi	44	11	55
Tidak Mencukupi	125	52	177
TOTAL	169	63	232

Karena data yang dipakai tidak terlalu banyak, maka dari itu peneliti memakai semua data yang ada. Pada **Tabel 5.9** setelah peneliti menentukan jumlah data untuk data latih, didapatkan hasil untuk data latih (*train*) sebesar 70% kategori mencukupi sebanyak 44 nasabah dan kategori tidak mencukupi sebanyak 125 nasabah sedangkan untuk data uji (*test*) sebesar 30% kategori mencukupi sebanyak 11 nasabah dan kategori tidak mencukupi sebanyak 52 nasabah.

### 5.2.2. Penentuan *Mtry* dan Jumlah Pohon (*Ntree*) terbaik

Sebelum melakukan metode *random forest* dan setelah menentukan data latih (*training*) dan data uji (*test*), langkah selanjutnya adalah menentukan nilai *Mtry* atau peubah penjelas dan *Ntree* atau jumlah pohon agar mendapatkan model yang optimal dan nilai *error* OOB yang bernilai kecil. Penentuan *Mtry* agar didapatkan nilai yang optimal dapat dilakukan dengan tiga cara, yaitu:

1. 
$$Mtry = \frac{\sqrt{\text{total variabel}}}{2}$$

$$= \frac{\sqrt{9}}{2} = 1.5 \approx 1$$
2. 
$$Mtry = \sqrt{\text{total variabel}}$$

$$= \sqrt{9} = 3$$
3. 
$$Mtry = \sqrt{\text{total variabel}} \times 2$$

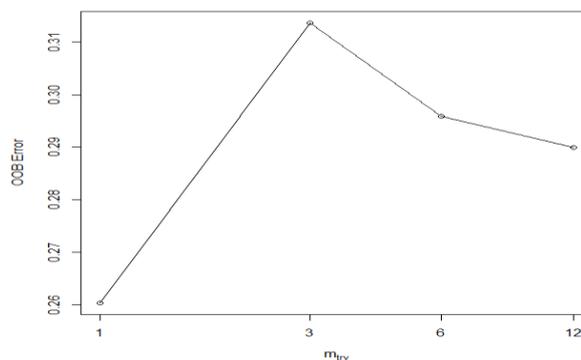
$$= \sqrt{9} \times 2 = 3 \times 2 = 6$$

Setelah menentukan  $Mtry$ , selanjutnya yaitu mencoba  $Mtry$  yang telah didapatkan untuk melakukan klasifikasi. Setelah diujikan dengan masing-masing nilai  $Mtry$  yang didapatkan menggunakan *default* jumlah pohon ( $Ntree$ ) sebanyak 500, maka didapatkan nilai *error* OOB sebagai berikut:

**Tabel 5.10** Pengujian Nilai *Error* OOB

$Mtry$	<i>Error</i> OOB
1	26.63%
3	31.95%
6	29.59%

Dapat dilihat dari **Tabel 5.10** didapatkan nilai *error* OOB terendah yaitu 26.63% pada saat nilai  $Mtry$  sebesar 1 sedangkan nilai *error* OOB tertinggi pada saat nilai  $Mtry$  sebesar 6. Berikut adalah grafik nilai *error* OOB menurut nilai  $Mtry$  yang telah dicari sebelumnya.



**Gambar 5.7** Grafik Nilai *Error* OOB Menurut  $Mtry$  yang di Dapat

Karena nilai *error* OOB terendah didapatkan pada saat  $Mtry$  sebesar 1, maka peneliti mengambil keputusan untuk menggunakan  $Mtry = 1$ . Kemudian, langkah selanjutnya yaitu menentukan jumlah pohon ( $Ntree$ ) yang akan digunakan menggunakan nilai  $Mtry$  yang didapatkan sebelumnya. Jumlah pohon ( $Ntree$ ) yang akan dicoba oleh peneliti dimulai dari 25, 50, 100, 250, 300, 500, dan 1000. Berikut adalah nilai *error* yang dihasilkan oleh setiap jumlah pohon dengan nilai  $Mtry$  sebesar 1 :

**Tabel 5.11** Pengujian Nilai *Error* OOB Dengan  $Ntree$  Berbeda

$Ntree$	<i>Error</i> OOB
25	28.4%
50	28.99%
100	26.63%
250	25.44%

300	26.04%
500	26.63%
1000	26.63%

Setelah melakukan pengujian dengan jumlah pohon (*Mtry*) berbeda, didapatkan nilai *error* OOB terendah pada saat jumlah pohon (*Ntree*) sebesar 250. Nilai *error* OOB terendah pada saat jumlah pohon (*Ntree*) sebesar 250 adalah 25.44%. Maka dari itu, untuk melihat model dan juga akurasi dari metode *random forest* data nasabah Warung Mikro Bank Syariah Mandiri KC Jambi tahun 2014-2017.

### 5.2.3. Proses Pelatihan dan Pembentukan Model

Dalam proses pelatihan metode *random forest* dibuat menjadi 250 iterasi agar mendapatkan hasil yang lebih optimal. Untuk mendapatkan model *random forest* terbaik, hasil iterasi harus optimal karena hasil iterasi adalah jumlah pohon yang akan digunakan dalam *learning*.

```
randomForest(formula = y ~ ., data = train, ntree =250)
      Type of random forest: classification
      Number of trees: 250
No. of variables tried at each split: 1

      OOB estimate of error rate: 25.44%
```

**Gambar 5. 8** Hasil klasifikasi *random forest* menggunakan data latih (*training*)

Pada **Gambar 5.8** dapat dilihat bahwa tipe *reandom forest* yang digunakan adalah klasifikasi, dimana jumlah pohon yang digunakan sebanyak 250 yang diambil dari plot jumlah iterasi sebelumnya. *No of variables tried at each split* didapatkan dari hasil *m* yang telah didapatkan pada pengujian *Mtry* sebelumnya. Nilai estimasi *error* OOB yang didapatkan menggunakan data latih (*training*) sebesar 25.44%.

**Tabel 5. 12** Hasil Prediksi Data Latih (*Training*)

	Saldo Mencukupi	Saldo Tidak Mencukupi	<i>Class Error</i>
Saldo Mencukupi	1	43	0.977
Saldo Tidak Mencukupi	0	125	0.000

Pada prediksi kelas saldo mencukupi diketahui tidak terlalu bagus dimana dari jumlah 44 nasabah hanya 1 nasabah yang tepat prediksi sedangkan 43 nasabah lainnya tidak tepat prediksi dengan error yang didapatkan sebesar 97.7%.

Kelas saldo tidak mencukupi diketahui sangat bagus karena dari 125 nasabah, semuanya tepat prediksi dengan nilai error 0.00%. Dapat dilihat pada **Tabel 5.12**, kelas keterangan saldo mencukupi sangat mempengaruhi *error* yang dihasilkan.

#### 5.2.4. Pengujian Akurasi Metode *Random Forest*

Setelah membuat model pada metode *random forest*, langkah selanjutnya yaitu melakukan pengujian akurasi menggunakan data uji (*test*).

**Tabel 5.13** Hasil Prediksi Data Uji (*Test*)

	Saldo Mencukupi	Saldo Tidak Mencukupi	<i>Class Error</i>
Saldo Mencukupi	0	11	1.000
Saldo Tidak Mencukupi	0	52	0.000

Selanjutnya akan dilakukan pengujian akurasi menggunakan data uji (*test*) dimana telah dijelaskan sebelumnya terdapat 11 nasabah dengan keterangan saldo mencukupi dan 52 nasabah dengan keterangan saldo tidak mencukupi. Pada prediksi kelas saldo mencukupi diketahui semua nasabah tidak tepat prediksi sedangkan kelas saldo tidak mencukupi diketahui semua nasabah tepat prediksi. Seperti diketahui, perbandingan antara *class error* pada masing-masing kelas sangat tidak seimbang. Dari *confussion matrix* didapatkan hasil sebagai berikut ini:

$$\begin{aligned}
 \textit{Specificity} &= \frac{52}{(52+0)} = 1 \times 100\% = 100\% \\
 \textit{Sensitivity/Recall} &= \frac{0}{(11+0)} = 0 \times 100\% = 0\% \\
 \textit{Precision} &= \frac{0}{(0+0)} = 0 \times 100\% = 0\% \\
 \textit{APER} &= \frac{(0+11)}{63} = 0.1746 \times 100\% = 17.46\% \\
 1\text{-APER} &= \frac{(0+52)}{63} = 0.8254 \times 100\% = 82.54\% \\
 \textit{F-measure} &= \frac{2 \times 0 \times 0}{(0+0)} = 0\% \\
 \textit{G-mean} &= \sqrt{1 \times 0} = 0\% \\
 \textit{AUC} &= \frac{1+0-0}{2} = 0.5 \times 100\% = 50\%
 \end{aligned}$$

Dapat dilihat pada hasil diatas, nilai *specificity* atau pengukuran proporsi *true negative* (saldo tidak mencukupi) yang diidentifikasi dengan benar

sebesar 100% sedangkan nilai *sensitivity* atau pengukuran proporsi *true positive* (saldo mencukupi) yang diidentifikasi dengan benar sebesar 0%. Pada perhitungan 1-APER pada metode *random forest* bernilai 82.54% yang berarti terdapat 82.54% data sampel yang tepat diklasifikasikan. Dikarenakan data yang digunakan merupakan dataset yang *imbalanced*, maka akurasi klasifikasi lainnya harus diperhatikan juga. *G-mean* berguna untuk menentukan keseimbangan akurasi prediksi untuk mengukur permasalahan *imbalanced*, dimana *G-mean* yang didapatkan jika menggunakan metode *random forest* adalah 0 yang menandakan ketidakseimbangan prediksi. *G-mean* diperlukan karena metode klasifikasi cenderung baik dalam memprediksi kelas dengan data sampel yang lebih banyak namun buruk dalam memprediksi kelas dengan data sampel yang sedikit. AUC merupakan cara yang baik untuk mendapatkan nilai kinerja pengklasifikasi secara umum dimana AUC adalah ukuran kinerja yang populer dalam ketidakseimbangan kelas .

Hasil akurasi yang didapatkan menunjukkan bahwa metode *random forest* bagus dalam mengklasifikasikan kelas saldo tidak mencukupi, namun tidak dapat mengklasifikasikan kelas saldo mencukupi dengan baik. Hal ini menandakan hasil akurasi yang bias atau disebut dengan *accuracy paradox*. Jika hal ini terjadi, maka lebih baik menggunakan ukuran ketepatan klasifikasi yang lain selain akurasi seperti *precision* dan *sensitivity*. Akibat ketidakseimbangan prediksi yang ada, maka dari itu peneliti akan menangani kelas *imbalanced* dengan menggunakan teknik *sampling* kelas *imbalanced* diantaranya *oversampling*, *undersampling*, dan SMOTE.

### **5.3. Random Forest Menggunakan Teknik Sampling Kelas Imbalanced**

#### **5.3.1. Random Forest Menggunakan Undersampling**

Teknik *undersampling* akan digunakan untuk mengatasi data *imbalanced* dengan cara *re-sampling* eliminasi kelas mayoritas (saldo tidak mencukupi) secara

acak sampai jumlahnya sebanyak kelas minoritas (saldo mencukupi). Jumlah data latih (*training*) adalah sebagai berikut :

**Tabel 5. 14** Data Latih (*Train*)

	Saldo Mencukupi	Saldo Tidak Mencukupi
Jumlah data	44	125

Dimana teknik *undersampling* menggunakan data latih (*train*) dengan jumlah data paling sedikit yaitu kelas saldo mencukupi dengan jumlah data 44. Maka dari itu, peneliti akan menyamaratakan kelas dengan masing-masing jumlah data yaitu 44.

```
randomForest(formula = y ~ ., data = train, ntree =250)
      Type of random forest: classification
      Number of trees: 250
No. of variables tried at each split: 1

      OOB estimate of error rate: 54.55%
```

**Gambar 5. 9** Hasil klasifikasi *random forest undersampling* menggunakan data latih (*training*)

Pada **Gambar 5.9** dapat dilihat bahwa nilai *error* OOB metode *random forest* menggunakan *undersampling* meningkat yaitu 54.45%. Dimana selisih metode *random forest* yang tidak menggunakan teknik *undersampling* dan menggunakan teknik *undersampling* sebesar 29.01%. Itu berarti bahwa terdapat 54.45% data yang tidak masuk ke dalam data *bootstrap* yang dibentuk.

**Tabel 5. 15** Hasil Prediksi Data Latih (*Training*) *Random Forest Undersampling*

	Saldo Mencukupi	Saldo Tidak Mencukupi	<i>Class Error</i>
Saldo Mencukupi	17	27	0.614
Saldo Tidak Mencukupi	23	21	0.477

Pada prediksi kelas saldo mencukupi diketahui tidak terlalu bagus dimana dari jumlah 44 nasabah hanya 17 nasabah yang tepat prediksi sedangkan 27 nasabah lainnya tidak tepat prediksi dengan error yang didapatkan sebesar 61.4%. Kelas saldo tidak mencukupi diketahui terdapat 21 nasabah yang tepat prediksi sedangkan 23 nasabah lainnya tidak tepat prediksi dengan error yang didapatkan sebesar 47.7%. Dapat dilihat pada **Tabel 5.15**, kelas keterangan saldo mencukupi sangat mempengaruhi *error* yang dihasilkan.

Selanjutnya, peneliti akan melakukan pengujian akurasi menggunakan data uji (*test*) maka didapatkan *confussion matriks* seperti **Tabel 5.13**.

**Tabel 5. 16** *Confusion Matrix* Metode *Random Forest Undersampling*

	Saldo Mencukupi	Saldo Tidak Mencukupi
Saldo Mencukupi	5	6
Saldo Tidak Mencukupi	19	33

Pengujian akurasi menggunakan data uji (*test*) dimana telah dijelaskan sebelumnya terdapat 11 nasabah dengan keterangan saldo mencukupi dan 52 nasabah dengan keterangan saldo tidak mencukupi. Pada metode *random forest undersampling* prediksi kelas saldo mencukupi diketahui 5 nasabah tepat prediksi dan 6 nasabah tidak tepat prediksi. Untuk kelas saldo tidak mencukupi diketahui 33 nasabah tepat prediksi sedangkan 19 nasabah tidak tepat prediksi. Dari *confusion matrix* didapatkan hasil sebagai berikut ini:

$$\begin{aligned}
 \textit{Specificity} &= \frac{33}{(33+19)} = 0.6346 \times 100\% = 63.46\% \\
 \textit{Sensitivity/Recall} &= \frac{5}{(5+6)} = 0.4545 \times 100\% = 45.45\% \\
 \textit{Precision} &= \frac{5}{(5+19)} = 0.2083 \times 100\% = 20.83\% \\
 \textit{APER} &= \frac{(19+6)}{63} = 0.3968 \times 100\% = 39.68\% \\
 \textit{1-APER} &= \frac{(5+33)}{63} = 0.6032 \times 100\% = 60.32\% \\
 \textit{F-measure} &= \frac{2 \times 0.2083 \times 0.4545}{(0.2083 + 0.4545)} = 0.2857 \times 100\% = 28.57\% \\
 \textit{G-mean} &= \sqrt{0.6346 \times 0.4545} = 0.5408 \times 100\% = 54.08\% \\
 \textit{Fprate} &= \frac{19}{(19+33)} = 0.3654 \times 100\% = 36.54\% \\
 \textit{AUC} &= \frac{1 + 0.4545 - 0.3654}{2} = 0.5446 \times 100\% = 54.46\%
 \end{aligned}$$

Dapat dilihat pada hasil diatas, nilai *specificity* atau pengukuran proporsi *true negative* (saldo tidak mencukupi) yang diidentifikasi dengan benar menurun dengan nilai sebesar 63.46% sedangkan nilai *sensitivity* atau pengukuran proporsi *true positive* (saldo mencukupi) yang diidentifikasi dengan benar meningkat sebesar 45.45%. Nilai 1-APER dari metode *random forest undersampling* yang didapatkan turun menjadi 54.46% dimana pada metode *random forest* tanpa menggunakan teknik *sampling* memiliki tingkat akurasi yang didapatkan sebesar 82.54%.

Keseimbangan akurasi prediksi untuk mengukur permasalahan data *imbalanced* diukur dari nilai *G-means* yaitu sebesar 54.08% dimana nilai *G-means* ini naik dibandingkan hanya menggunakan metode *random forest* biasa. Lalu pada nilai AUC yang dihasilkan dari metode *random forest undersampling* sebesar 54.46%.

### 5.3.2. *Random Forest Menggunakan Oversampling*

Teknik *oversampling* akan digunakan untuk mengatasi data *imbalanced* dengan cara *re-sampling* eliminasi kelas minoritas (saldo mencukupi) secara acak sampai jumlahnya sebanyak kelas mayoritas (saldo tidak mencukupi). Dimana menggunakan data latih (*train*) dengan jumlah data paling banyak yaitu kelas saldo tidak mencukupi dengan jumlah 125. Peneliti akan menyamaratakan kelas dengan masing-masing jumlah data yaitu 125. Peneliti akan melakukan pengujian akurasi menggunakan data uji (*test*) maka didapatkan *confussion matrix* seperti dibawah ini.

```
randomForest(formula = y ~ ., data = train, ntree =250)
      Type of random forest: classification
      Number of trees: 250
No. of variables tried at each split: 1

      OOB estimate of error rate: 14.80%
```

**Gambar 5. 10** Hasil klasifikasi *random forest oversampling* menggunakan data latih (*training*)

Pada **Gambar 5.10** dapat dilihat bahwa nilai *error* OOB metode *random forest* menggunakan *undersampling* sebesar 14.80%. Dimana selisih metode *random forest* teknik *undersampling* dan *random forest* teknik *oversampling* sebesar 39.65%. Itu berarti bahwa terdapat 14.80% data yang tidak masuk ke dalam data *bootstrap* yang dibentuk.

**Tabel 5. 17** Hasil Prediksi Data Latih (*Training*) *Random Forest Oversampling*

	Saldo Mencukupi	Saldo Tidak Mencukupi	<i>Class Error</i>
Saldo Mencukupi	111	14	0.112
Saldo Tidak Mencukupi	23	101	0.184

Pada prediksi kelas saldo mencukupi diketahui dari jumlah 125 nasabah terdapat 111 nasabah yang tepat prediksi sedangkan 14 nasabah lainnya tidak tepat prediksi dengan error yang didapatkan sebesar 11.2%. Kelas saldo tidak

mencukupi diketahui dari 125 nasabah, terdapat 101 nasabah yang tepat prediksi sedangkan 23 nasabah lainnya tidak tepat prediksi dengan error yang didapatkan sebesar 18.4%.

**Tabel 5. 18** *Confusion Matrix* Metode *Random Forest Oversampling*

	Saldo Mencukupi	Saldo Tidak Mencukupi
Saldo Mencukupi	6	5
Saldo Tidak Mencukupi	11	41

Pada metode *random forest oversampling* prediksi kelas saldo mencukupi diketahui 6 nasabah tepat prediksi dan 5 nasabah tidak tepat prediksi. Untuk kelas saldo tidak mencukupi diketahui 41 nasabah tepat prediksi sedangkan 11 nasabah tidak tepat prediksi. Dari *confusion matrix* didapatkan hasil sebagai berikut ini:

$$\begin{aligned}
 \textit{Specificity} &= \frac{41}{(41+11)} = 0.7885 \times 100\% = 78.85\% \\
 \textit{Sensitivity/Recall} &= \frac{6}{(5+6)} = 0.5455 \times 100\% = 54.55\% \\
 \textit{Precision} &= \frac{6}{(6+11)} = 0.3529 \times 100\% = 35.29\% \\
 \textit{APER} &= \frac{(11+5)}{63} = 0.2539 \times 100\% = 25.39\% \\
 \textit{1-APER} &= \frac{(6+41)}{63} = 0.7460 \times 100\% = 74.60\% \\
 \textit{F-measure} &= \frac{2 \times 0.3529 \times 0.5455}{(0.3529 + 0.5455)} = 0.4286 \times 100\% = 42.86\% \\
 \textit{G-mean} &= \sqrt{0.7885 \times 0.5455} = 0.5609 \times 100\% = 65.58\% \\
 \textit{FPrate} &= \frac{11}{(11+41)} = 0.2115 \times 100\% = 21.15\% \\
 \textit{AUC} &= \frac{1 + 0.5455 - 0.2115}{2} = 0.6669 \times 100\% = 66.69\%
 \end{aligned}$$

Dapat dilihat pada hasil diatas, nilai *specificity* atau pengukuran proporsi *true negative* (saldo tidak mencukupi) yang diidentifikasi dengan benar meningkat dengan nilai sebesar 78.85% sedangkan nilai *sensitivity* atau pengukuran proporsi *true positive* (saldo mencukupi) yang diidentifikasi dengan benar meningkat sebesar 54.55%. Nilai 1-APER dari metode *random forest oversampling* yang didapatkan meningkat menjadi 74.60%.

Keseimbangan akurasi prediksi untuk mengukur permasalahan data *imbalanced* diukur dari nilai *G-means* yaitu sebesar 65.58% dimana nilai *G-*

*means* ini lebih tinggi dibandingkan menggunakan metode *random forest undersampling*. Nilai AUC dari metode *random forest oversampling* lebih tinggi juga jika dibandingkan dengan *random forest undersampling* yaitu sebesar 66.69%.

### 5.3.3. *Random Forest* Menggunakan SMOTE

Teknik *Synthetic Minority Over-Sampling Technique* (SMOTE) akan digunakan untuk mengatasi data *imbalanced* dengan cara *re-sampling* dimana kelas minoritas dilakukan *oversampling* dengan melakukan duplikasi pada data sebenarnya dan tergantung kepada jumlah *oversampling* dibutuhkan. Dimana peneliti membuat data tersebut menjadi 500 dimana masing-masing kelas terdiri dari 250 data.

```
randomForest(formula = y ~ ., data = train, ntree =250)
      Type of random forest: classification
      Number of trees: 250
No. of variables tried at each split: 1

      OOB estimate of error rate: 31.40%
```

**Gambar 5. 11** Hasil klasifikasi *random forest* SMOTE menggunakan data latihan (*training*)

Pada **Gambar 5.11** dapat dilihat bahwa nilai *error* OOB metode *random forest* menggunakan SMOTE sebesar 31.40%. Dimana error OOB tersebut lebih rendah dibandingkan *random forest undersampling* dan lebih tinggi dibandingkan dengan *random forest oversampling*. Itu berarti bahwa terdapat 31.40% data yang tidak masuk ke dalam data *bootstrap* yang dibentuk.

**Tabel 5. 19** Hasil Prediksi Data Latih (*Training*) *Random Forest* SMOTE

	Saldo Mencukupi	Saldo Tidak Mencukupi	<i>Class Error</i>
Saldo Mencukupi	186	80	0.301
Saldo Tidak Mencukupi	77	157	0.329

Pada prediksi kelas saldo mencukupi diketahui terdapat 186 nasabah yang tepat prediksi sedangkan 77 nasabah lainnya tidak tepat prediksi dengan error yang didapatkan sebesar 30.1%. Kelas saldo tidak mencukupi diketahui terdapat 157 nasabah yang tepat prediksi sedangkan 77 nasabah lainnya tidak tepat prediksi dengan error yang didapatkan sebesar 30.1%.

Peneliti akan melakukan pengujian akurasi menggunakan data uji (*test*) maka didapatkan *confusion matrix* seperti dibawah ini.

**Tabel 5. 20** *Confusion Matrix* Metode *Random Forest* SMOTE

	Saldo Mencukupi	Saldo Tidak Mencukupi
Saldo Mencukupi	5	6
Saldo Tidak Mencukupi	16	36

Pada metode *random forest* SMOTE prediksi kelas saldo mencukupi diketahui 5 nasabah tepat prediksi dan 6 nasabah tidak tepat prediksi. Untuk kelas saldo tidak mencukupi diketahui 36 nasabah tepat prediksi sedangkan 16 nasabah tidak tepat prediksi. Dari *confussion matrix* didapatkan hasil sebagai berikut ini:

$$\begin{aligned}
 \textit{Specificity} &= \frac{36}{(16+36)} = 0.6923 \times 100\% = 69.23\% \\
 \textit{Sensitivity/Recall} &= \frac{5}{(6+5)} = 0.4546 \times 100\% = 45.46\% \\
 \textit{Precision} &= \frac{5}{(5+16)} = 0.2381 \times 100\% = 23.81\% \\
 \textit{APER} &= \frac{(16+6)}{63} = 0.3492 \times 100\% = 34.92\% \\
 \textit{1-APER} &= \frac{(5+36)}{63} = 0.6508 \times 100\% = 65.08\% \\
 \textit{F-measure} &= \frac{2 \times 0.2381 \times 0.4546}{(0.2381 + 0.4546)} = 0.3125 \times 100\% = 31.25\% \\
 \textit{G-mean} &= \sqrt{0.6923 \times 0.4546} = 0.5609 \times 100\% = 56.09\% \\
 \textit{FPrate} &= \frac{16}{(16+36)} = 0.3079 \times 100\% = 30.79\% \\
 \textit{AUC} &= \frac{1+0.4546-0.3079}{2} = 0.5734 \times 100\% = 57.34\%
 \end{aligned}$$

Nilai *specificity* atau pengukuran proporsi *true negative* (saldo tidak mencukupi) yang diidentifikasi dengan benar menurun dengan nilai sebesar 69.23% sedangkan nilai *sensitivity* atau pengukuran proporsi *true positive* (saldo mencukupi) yang diidentifikasi dengan benar sebesar 45.46%. Nilai 1-APER dari metode *random forest* SMOTE yang didapatkan juga menjadi 65.08%.

Keseimbangan akurasi prediksi untuk mengukur permasalahan data *imbalanced* diukur dari nilai *G-means* yaitu sebesar 56.09% dimana nilai *G-means* ini lebih tinggi dibandingkan hanya menggunakan metode *random forest undersampling*. Sedangkan untuk nilai AUC dari *random forest* menggunakan

SMOTE lebih rendah dibandingkan dengan teknik *oversampling*, yaitu sebesar 57.34%.

Berikut adalah perbandingan dari keempat metode yang dipakai oleh peneliti untuk melakukan klasifikasi pada data nasabah Warung Mikro Bank Syariah KC Jambi. Peneliti akan memilih metode apa yang cocok untuk data tersebut dengan melakukan perbandingan terhadap metode yang dipakai, yaitu *random forest*, *random forest* menggunakan *oversampling*, *random forest* menggunakan *undersampling*, dan *random forest* menggunakan SMOTE.

**Tabel 5. 21** Perbedaan Keempat Metode *Random Forest* yang Digunakan

	<i>Random Forest</i>	<i>Undersampling</i>	<i>Oversampling</i>	SMOTE
<i>Specificity</i>	100%	63.46%	<b>78.85%</b>	69.23%
<i>Sensitivity</i>	0%	45.45%	<b>54.55%</b>	45.46%
<i>Precision</i>	0%	20.83%	<b>35.29%</b>	23.81%
<i>APER</i>	17.46%	39.68%	25.39%	34.92%
<i>1-APER</i>	82.54%	60.32%	<b>74.60%</b>	65.08%
<i>G-mean</i>	0%	54.08%	<b>65.58%</b>	56.09%
<i>AUC</i>	50%	54.46%	<b>66.69%</b>	57.34%

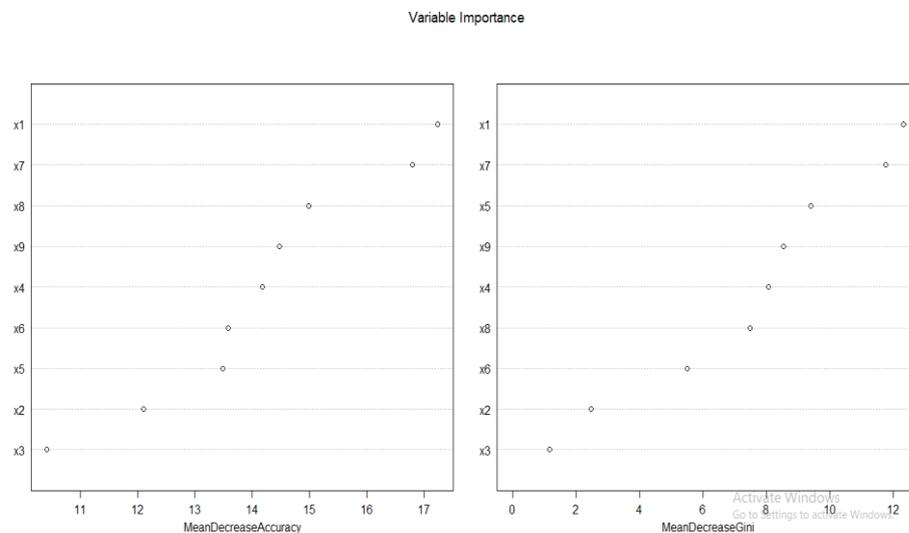
Menurut **Tabel 5.21** nilai 1-APER (*Total Accucary Rate*) pada ketiga teknik *sampling* memberikan nilai akurasi lebih kecil dibandingkan dengan metode *random forest* tanpa teknik *sampling imbalanced data*. Walaupun nilai akurasi dari metode *random forest* paling tinggi, untuk pengklasifikasian kelas *true positive* tidak ada yang tepat. Nilai akurasi *oversampling* lebih tinggi dibandingkan dengan *undersampling* dan SMOTE. Nilai *sensitivity* tertinggi yaitu pada saat menggunakan teknik *oversampling* yaitu sebesar 54.55% yang berarti bahwa sebesar 54.55% data nasabah dengan keterangan saldo mencukupi yang tepat diklasifikasikan dengan keterangan saldo mencukupi.

Selanjutnya adalah nilai *specificity* tertinggi saat menggunakan teknik *oversampling* yaitu sebesar 78.85% yang berarti bahwa data nasabah dengan keterangan saldo tidak mencukupi yang tepat diklasifikasikan dengan keterangan saldo tidak mencukupi adalah 78.85%. Dari ketiga teknik *sampling* kelas data *imbalanced* memberikan nilai *precision* dan *sensitivty* yang lebih baik dibandingkan dengan menggunakan metode *random forest* tanpa teknik *sampling data imbalanced*.

Nilai *precision* dengan menggunakan teknik *oversampling* lebih tinggi dibandingkan kedua teknik lainnya yaitu sebesar 35.29%. Selanjutnya, yaitu nilai AUC yang merupakan cara mendapatkan nilai kinerja pengklasifikasian secara umum dan untuk membandingkannya dengan pengklasifikasian yang lain serta ukuran kinerja yang populer dalam ketidakseimbangan kelas dimana nilai AUC tertinggi yaitu jika menggunakan *random forest* dengan teknik *oversampling* yaitu sebesar 66.69% sedangkan teknik *sampling* dengan nilai AUC terendah yaitu *random forest* dengan teknik *undersampling* yaitu sebesar 54.46%. Maka dari itu, peneliti akan menggunakan metode *random forest* dengan teknik *oversampling* untuk melihat tingkat kepentingan dari variabel *independent* (tidak terikat).

#### 5.4. Ukuran Tingkat Kepentingan

Selanjutnya adalah melihat ukuran tingkat kepentingan (*variable importance*) dari setiap variabel *independent* (tidak terikat) yang diteliti.



**Gambar 5.12** Ukuran Tingkat Kepentingan

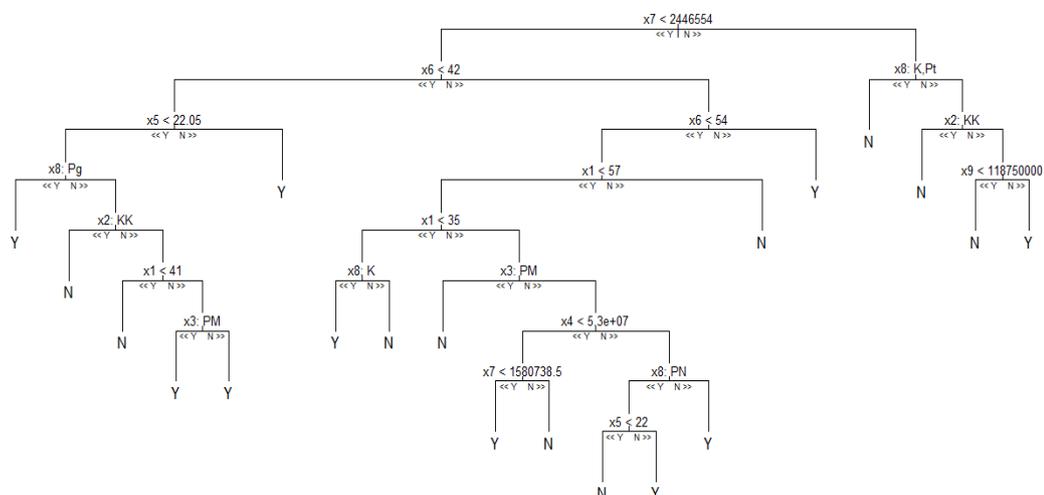
Metode *random forest* menghasilkan ukuran tingkat kepentingan atau *variable importance* pada masing-masing variabel bebas (*independent*), yaitu *Mean Decrease Accuracy* (MDA) dan *Mean Decrease Gini* (MDG) seperti **Gambar 5.12**. Dimana MDA berguna untuk melihat seberapa besar tambahan observasi mengalami misklasifikasi jika salah satu variabel bebas tidak diikutsertakan pada pengujian sedangkan MDG digunakan untuk melihat

kestabilan variabel bebas dari  $X_1$  sampai  $X_9$ . Tingkat kepentingan disusun berdasarkan peringkat.

Pada ukuran tingkat kepentingan MDA, variabel dengan dua tingkat kepentingan tertinggi yaitu variabel  $x_1$  (umur) dan  $x_7$  (besar angsuran perbulan) yang berarti variabel tersebut memiliki peranan penting sebagai variabel bebas yang mempengaruhi variabel *dependent* (terikat) pada metode tersebut sedangkan untuk dua peringkat terakhir yaitu variabel  $x_2$  (jenis nasabah) dan  $x_3$  (jenis pembiayaan mikro) yang berarti bahwa variabel tersebut menandakan tingkat kepentingan yang kecil.

Pada ukuran tingkat kepentingan MDG sama seperti tingkat kepentingan MDA, dimana variabel bebas yang memiliki tingkat kepentingan tertinggi secara berurutan adalah  $x_1$  (umur) dan  $x_7$  (besar angsuran perbulan), dan seterusnya hingga variabel terakhir yaitu  $x_3$  (jenis pembiayaan mikro). Jadi, dapat disimpulkan bahwa dua variabel terbawah dalam ukuran tingkat kepentingan MDA dan MDG yaitu variabel  $x_2$  (jenis nasabah) dan  $x_3$  (jenis pembiayaan mikro).

Berikut adalah model yang dibentuk oleh metode *random forest* menggunakan *oversampling* :



**Gambar 5.13** Decision Tree yang Dibentuk Random Forest Oversampling

Dapat dilihat pada *decision tree* pada **Gambar 5.13** yang termasuk dalam kategori saldo mencukupi jika :

1.  $x_7$  (Besarnya angsuran/bulan) yang dimiliki oleh nasabah kurang dari Rp 2.446.554,00 dan  $x_6$  (lama angsuran) kurang dari 42 bulan serta  $x_5$  (Margin) lebih dari 22.05%.
2.  $x_7$  (Besarnya angsuran/bulan) yang dimiliki oleh nasabah kurang dari Rp 2.446.554,00 dan  $x_6$  (lama angsuran) kurang dari 42 bulan lalu  $x_5$  (Margin) kurang dari 22.05% serta pekerjaan dari nasabah tersebut adalah pegawai.
3.  $x_7$  (Besarnya angsuran/bulan) yang dimiliki oleh nasabah kurang dari Rp 2.446.554,00 dan  $x_6$  (lama angsuran) kurang dari 42 bulan lalu  $x_5$  (Margin) kurang dari 22.05% serta  $x_8$  (pekerjaan) dari nasabah tersebut selain pegawai, juga  $x_2$  (jenis nasabah) bukan kerjasama KUD, dan  $x_1$  (umur) lebih dari 41 tahun serta  $x_3$  (jenis pembiayaan) nasabah bisa PM Madya dan PM Utama.
4.  $x_7$  (Besarnya angsuran/bulan) yang dimiliki oleh nasabah kurang dari Rp 2.446.554,00 dan  $x_6$  (lama angsuran) lebih dari 54 bulan.
5.  $x_7$  (Besarnya angsuran/bulan) yang dimiliki oleh nasabah kurang dari Rp 2.446.554,00 dan  $x_6$  (lama angsuran) berkisar dari 42 bulan sampai 54 bulan,  $x_1$  (umur) kurang dari 35 tahun dan  $x_8$  (pekerjaan) nasabah adalah karyawan.
6.  $x_7$  (Besarnya angsuran/bulan) yang dimiliki oleh nasabah kurang dari Rp 1.580.738,50 dan  $x_6$  (lama angsuran) berkisar dari 42 bulan sampai 54 bulan,  $x_1$  (umur) lebih dari 35 tahun,  $x_3$  (jenis pembiayaan mikro) nasabah adalah PM Utama,  $x_4$  (harga pokok pinjaman) kurang dari Rp 53.000.000,00.
7.  $x_7$  (Besarnya angsuran/bulan) yang dimiliki oleh nasabah kurang dari Rp 2.446.554,00 dan  $x_6$  (lama angsuran) berkisar dari 42 bulan sampai 54 bulan,  $x_1$  (umur) lebih dari 35 tahun,  $x_3$  (jenis pembiayaan mikro) nasabah adalah PM Utama,  $x_4$  (harga pokok pinjaman) lebih dari Rp 53.000.000,00 lalu  $x_8$  (pekerjaan) nasabah tersebut selain PNS.
8.  $x_7$  (Besarnya angsuran/bulan) yang dimiliki oleh nasabah kurang dari Rp 2.446.554,00 dan  $x_6$  (lama angsuran) berkisar dari 42 bulan sampai 54 bulan,  $x_1$  (umur) lebih dari 35 tahun,  $x_3$  (jenis pembiayaan mikro) nasabah adalah PM Utama,  $x_4$  (harga pokok pinjaman) lebih dari Rp 53.000.000,00 lalu  $x_8$  (pekerjaan) nasabah tersebut adalah PNS dan  $x_5$  (margin) lebih dari 22%.

9.  $x_7$  (Besarnya angsuran/bulan) yang dimiliki oleh nasabah lebih dari Rp 2.446.554,00 dan  $x_8$  (pekerjaan) nasabah selain karyawan dan petani lalu  $x_2$  (jenis nasabah) adalah perorangan / retail serta  $x_9$  (harga jaminan) lebih dari Rp 118.750.000,00.

## **BAB VI PENUTUP**

### **6.1 Kesimpulan**

Berdasarkan analisis terhadap hasil penelitian yang telah dilakukan dengan 232 data nasabah pembiayaan Warung Mikro KC Jambi, dapat diambil kesimpulan sebagai berikut:

1. Data yang digunakan adalah *imbalanced data* (data tidak seimbang) dimana perbandingan antara kelas saldo mencukupi dan tidak mencukupi yaitu 55 :177. Pada variabel umur, umur terkecil nasabah pembiayaan Warung Mikro BSM KC Jambi adalah 25 tahun sedangkan umur maksimal nasabah adalah 65 tahun. Selanjutnya, variabel jenis nasabah dapat dilihat bahwa lebih banyak jenis nasabah yang menjalankan usahanya perorangan atau retail dibandingkan dengan kerjasama KUD. Pada variabel jenis pembiayaan mikro, terdapat 38.36% dari 232 nasabah yang mengajukan PM Madya dan 61.64% dari 232 nasabah yang mengajukan PM Utama yang berarti bahwa lebih banyak nasabah yang melakukan pembiayaan PM Utama. Harga pokok pembiayaan terkecil sebesar Rp 15.000.000,00 dan terbesar sebesar Rp 200.000.000,00. Selanjutnya, untuk variabel margin terendah sebesar 16% dan tertinggi sebesar 36% serta jangka waktu angsuran dengan nasabah terbanyak sebesar 101 nasabah pada saat jangka waktu angsuran 36 bulan. Jumlah angsuran minimal yang harus dibayarkan nasabah tiap bulannya yaitu Rp 572.857,00 sedangkan jumlah angsuran maksimal yang harus dibayarkan nasabah tiap bulannya yaitu Rp 9.714.480,00. Jenis pekerjaan terbanyak adalah wiraswasta dengan jumlah sebanyak 152 nasabah dan harga jaminan terendah senilai Rp 16.000.000,00 sedangkan harga jaminan tertinggi Rp 250.000.000,00. Pada data yang digunakan, nasabah dengan keterangan saldo tidak mencukupi paling banyak terdapat pada jenis nasabah perorangan atau retail dan jenis pembiayaan PM Utama dan untuk nasabah dengan keterangan saldo mencukupi paling banyak juga terdapat pada jenis nasabah perorangan atau retail dan jenis pembiayaan PM Utama.

2. Dengan menggunakan data latih (*training*) sebesar 70% dan data uji (*test*) sebesar 30% serta jumlah pohon atau iterasi sebanyak 250 dan *Mtry* sebesar 1, didapatkan hasil klasifikasi dengan nilai estimasi error OOB sebesar 25.44%. Akurasi tertinggi didapatkan pada saat digunakan metode *random forest* tanpa teknik *sampling* kelas *imbalanced* yaitu sebesar 82.54%, namun nilai *sensitivity* bernilai 0 yang berarti menghasilkan akurasi yang bias.
3. Saat menggunakan teknik *sampling* kelas *imbalanced* dapat meningkatkan nilai *sensitivity*, *precision*, *G-mean*, dan AUC. Dimana teknik *oversampling* memberikan nilai *sensitivity*, *precision*, dan *G-mean* terbesar yaitu dengan nilai berturut-turut 54.55%, 35.39%, dan 65.58%. Nilai AUC merupakan nilai kinerja untuk membandingkan pengklasifikasian yang populer dalam ketidakseimbangan kelas dimana nilai AUC tertinggi didapatkan pada *random forest* teknik *oversampling* sebesar 66.69% dan terendah yaitu *random forest* teknik *undersampling* sebesar 54.46%.

## 6.2 Saran

Pada penelitian selanjutnya diharapkan agar menggunakan metode *Random Forest* lainnya, misalnya *Weighted Random Forest* ataupun *Balanced Random Forest* untuk meningkatkan hasil klasifikasi yang lebih bagus untuk data *imbalanced*. Selanjutnya untuk meningkatkan nilai akurasi dan kualitas hasil klasifikasi, sebaiknya pada penelitian selanjutnya menggunakan data yang lebih banyak lagi serta menambah variabel lainnya yang lebih mendukung penelitian ini.

## DAFTAR PUSTAKA

- Abdullah, B. 2006. Menanti Kemakmuran Negeri : Kumpulan Esai tentang Pembangunan Sosial Ekonomi Indonesia. Jakarta : PT Gramedia Pustaka Utama.
- Alpaydin, E. 2010. *Introduction to Machine Learning*. London : MIT Press.
- Alfisahrin, S.N. 2014. Komparasi Algoritma C4.5, *Naive Bayes*, dan *Neural Network* Untuk Memprediksi Penyakit Jantung. Jakarta : Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri.
- Arifin, Z. 2002. Dasar-dasar Manajemen Bank Syari'ah. Jakarta : AlfaBet.
- Attenberg, J & Ertekin, S. 2013. *Class Imbalanced and Active Learning*. In H. He & Y. Ma. *Imbalanced Learning : Foundations, Algorithms, and Applications*, New Jersey : John Wiley & Sons.
- Bank Indonesia. 2015. "Profil Bisnis Usaha Mikro, Kecil, dan Menengah (UMKM)". <http://www.bi.go.id>. Diakses pada tanggal 2 Maret 2018, pukul 12.11 WIB.
- Bank Indonesia. 2017. "Statistik Ekonomi dan Keuangan Daerah". <http://www.bi.go.id>. Diakses pada tanggal 2 Maret 2018, pukul 12.11 WIB.
- Barandela, R., Sanchez, J., Garcia, V., & Rangel, E. 2003. *Strategies for Learning in Class Imbalance Problems*. *Pattern Recognition*, 849-851.
- Basuki, A., & Syarif, I. 2003. Modul Ajar *Decision Tree*. Surabaya: PENS-ITS.
- Berry, M. J., & Linoff, G. S. 2004. *Data Mining Techniques for Marketing, Sale, Customer Relationship Management*. Wiley Publishing, Inc.
- Biau, G. 2012. Analysis of A Random Forest Model. *Journal of Machine Learning Research* 13. Paris : *Universite Pierre et Marie Curie*
- Bramer, M. 2007. *Principles of Data Mining*. London : Springer.
- Bramer, M. 2013. *Principle of Data Mining Second Edition*. London : Springer.
- Breiman, L., Friedman JH., Olshen RA, Stone CJ. 1984. *Classification and Regression Trees*. New York : Chapman & Hall.
- Breiman, L., & Cutler A. 2001. *Random Forests*. *Machine Learning* 45 : 5-32.

- Breiman L, Cutler A. 2003. Manual on Setting Up, Using, and Understanding Random Forest V4.0. [http://oz.berkeley.edu/users/breiman/Using\\_random\\_forests\\_v4.0.pdf](http://oz.berkeley.edu/users/breiman/Using_random_forests_v4.0.pdf). Diakses pada tanggal 6 April 2018, pukul 08.27 WIB.
- Chao, Chen., dkk. 2004. *Using Random Forest to Learn Imbalanced Data*.
- Chawla, N. V dkk, 2002. *SMOTE: Syntethic Minority Over-Sampling Technique*. Journal of Artificial Intelligence Research. Vol.16 Hal.321-357.
- Claudia, 2010. *Pembiayaan Murahabah Bank Syariah Mandiri Kepada Usaha Kecil [Tesis] Depok : Fakultas Hukum Universitas Indonesia*.
- Furqon. 1999. *Statistika Terapan Untuk Penelitian*. AFABETA : Bandung.
- Geevv. 2017. "Mengenal Apa Itu UMKM dan Perkembangannya di Indonesia". Geevv, 7 Agustus 2017.
- Han, K.J. 2001. *Data Mining : Concepts and Technique*. San Fransisco : John Wiley & Sons Inc.
- Han,J & Kamber, M. 2006. *Data Mining Concepts and Techniques Second Edition*. San Fransisco : Diane Cerra.
- Hasan, M. I. 2005. *Pokok-pokok Materi Statistik I (Statistik Deskriptif)*. Jakarta : Bumi Aksara.
- Hastie, T., Tibshirani, R., Friedman, J. 2008. *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*. California : Springer.
- Jatmiko, Y. A., Padmadisastra, S., & Chadidjah, A. 2017. "Perbandingan Teknik *Sampling* Dalam *Random Forest* Pada Kelas *Imbalanced*". Seminar Statistika FMIPA UNPAD 2017 (SNS VI). ISSN : 2087-2590.
- Japkowicz, N. 2000. "The Class Imbalance Problem: Significance and Strategies." In Proceedings of the 200 International Conference on Artificial Intelligence (IC-AI 2000): Special Track on Inductive Learning Las Vegas, Nevada.
- Kara, M. 2013. "Kontribusi Pembiayaan Perbankan Syariah Terhadap Pengembangan Usaha Mikro Kecil dan Menengah (UMKM) di Kota Makassar". Jurnal Ilmu Syari'ah dan Hukum. Vol.47 No.1 Hal.271.
- Kasmir. 2002. *Dasar-dasar Perbankan*. Jakarta : PT. Raja Grafindo Persada.
- Kasmir. 2003. *Manajemen Perbankan*. Jakarta : PT.Raja Grafindo Persad.

- Liaw, A. and Wiener, M. (2002) Classification and Regression by Randomforest. R News, 2, 18-22.
- Liu, X.Y & Zhou, Z.H. 2013. *Ensemble Methods for Class Imbalance Learning*. . In H. He & Y, Ma. *Imbalanced Learning : Foundations, Algorithms, and Applications*, New Jersey : John Wiley & Sons.
- Mambang & Byna, A. 2017. “Analisis Perbandingan Algoritma C4.5, *Random Forest* dengan *CHAID Decision Tree* untuk klasifikasi Tingkat Kecemasan Ibu Hamil”. Seminar Nasional Teknologi Infomrasi dan Multimedia 2017.
- Mattjik, A.A., dan Sumertajaya. 2006. *Perancangan Percobaan*. Jilid 1 Edisi ke-2. Bogor : IPB Press.
- Minarni. 2013. “Konsep Pengawasan, Kerangka Audit Syariah dan Tata Kelola Lembaga Keuangan Syariah”. *Jurnal Ekonomi Islam La\_Riba*. Vol.III No.1 Hal.32.
- Muhammad. 2005. *Manajemen Bank Syariah*. Yogyakarta : UPP AMP YKPN.
- Muzakkir. 2016. “Di Jambi Ada Ribuan UMKM”. *Tribun Jambi*, 13 Januari 2016.
- Nidhomuddin dkk. 2015. *Random Forest dan Multivariate Adaptive Regression Spline (MARS) Binary Response* Untuk Klasifikasi Penderita HIV/AIDS di Surabaya. *Jurnal Statistika*. Vol.1 No.3 Hal 49-57.
- Nikmah, C., Sukarno, H., dan Mufidah, A. 2014. “Analisis Implikasi Pembiayaan Syariah pada Pedagang Kecil di Pasar Tanjung Jember”. *Jurnal Ekonomi Bisnis dan Akuntansi*. Vol.1 No.1 Hal.8-15.
- Nugroho, Y.S & Emiliyawati, N. 2017. “Sistem Klasifikasi Varibael Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode *Random Forest*”.
- Pasaribu, A. 1975. *Pengantar Statistik*. Jakarta : Gahlia Indonesia.
- Prasetio, R.T & Pratiwi. 2015. “Penerapan Teknik Bagging Pada Algoritma Klasifikasi Untuk Mengatasi Ketidakseimbangan Kelas Dataset Medis”. *Jurnal Informatika*. Vol.II No.2 Hal.395-403.
- Republik Indonesia. 1998. Undang-undang No. 7 Tahun 1992 tentang Perbankan. Lembaran Negara RI, Direvisi dengan Undang-Undang Nomor 10 Tahun 1998. Sekretariat Negara. Jakarta.

- Republik Indonesia. 2008. Undang-undang No. 20 Tahun 2008 tentang Usaha Mikro, Kecil dan Menengah. Lembaran Negara RI Tahun 2008, No. 40. Sekretariat Negara. Jakarta.
- Republik Indonesia. 2008. Undang-undang No. 21 Tahun 2008 tentang Perbankan Syariah. Lembaran Negara RI Tahun 2008, No. 94. Sekretariat Negara. Jakarta.
- Rosyidah, N. 2017. Analisis Prosedur Pembiayaan Warung Mikro di Bank Syariah Mandiri KCP Ngaliyan Semarang Barat [skripsi]. Semarang (ID) : Universitas Islam Negeri Walisongo.
- Rudjito. 2003. *Peran Lembaga Keuangan Mikro dalam Otonomi Daerah Guna Menggerakkan Ekonomi Rakyat dan Menanggulangi Kemiskinan*. Jurnal Ekonomi Rakyat.
- Sandri, M & Zuccolotto, P. 2006. *Variable Selection Using Random Forest*. Di dalam: Zani S, Cerioli A, Riani M, Vichi M, editor. *Data Analysis, Classification and the Forward Search. Proceedings of the Meeting of the Classification and Data Analysis Group (CLADAG) of the Italian Statistical Society*; University of Parma, 6-8 Jun 2005. New York: Springer Berlin Heidelberg. Hal.263-270.
- Sembiring, K. 2007. Penerapan Teknik *Support Vector Machine* Untuk Pendeteksian Intrusi Pada Jaringan [skripsi]. Bandung (ID) : Institut Teknologi Bandung.
- Siamat, D.. 2005. *Manajemen Lembaga Keuangan “Kebijakan Moneter dan Perbankan”*. Jakarta : Fakultas Ekonomi Universitas Indonesia.
- Sutton CD. 2005. Classification and Regression Trees, Bagging, and Boosting. *Handbook of Statistics* 24:303-329
- Tedjasuksmana, B. 2014. Potret UMKM Indonesia Menghadapi Masyarakat Ekonomi Asean 2015. The 7th NCFB and Doctoral Colloquium 2014. Universitas Katolik Widya Mandala Surabaya.
- Vetihzal, R., & Veithzal, A. 2008. *Islamic Financial Management Teori , Konsep, dan Aplikasi*. Jakarta : Raja Grafindo Persada.

- Walpole, Ronald E, 1995. Pengantar Statistik Edisi Ke-4. Jakarta : PT Gramedia.
- Wang, S & Yao, X. 2013. *Using Class Imbalance Learning for Software Defect Prediction*. IEEE Transactions on Reliability, 434-443.
- Weiss, G. M. 2013. *Foundations of Imbalanced Learning*. In H. He & Y, Ma. *Imbalanced Learning : Foundations, Algorithms, and Applications*, New Jersey : John Wiley & Sons.
- Wezel, M. V & Potharst, R. 2007. *European Journal of Operational Research* 2007. Vol.181. Issue 1, Hal. 436-452.
- Wilardjo, S. B. 2005. Pengertian, Peranan Dan Perkembangan Bank Syariah Di Indonesia. Semarang: Universitas Muhammadiyah Semarang.
- Witten & Frank, E. 2005. *Data Mining : Practical Machine Learning Tools and Techniques*.
- Zhang, H & Wang, Z. 2011. *A Normal Distribution Based Over-Sampling Approach to Imbalanced Data Classification Advanced Data Mining and Application*. 7th International Conference. Beijing : Springer.
- Zhu, X. 2007. *Knowledge Discovery and Data Mining : Challenges and Realities*. IGI Global Hal.118-119.

## Paper Tugas Akhir

# KLASIFIKASI PEMBIAYAAN WARUNG MIKRO MENGGUNAKAN METODE *RANDOM FOREST* DENGAN TEKNIK *SAMPLING KELAS* *IMBALANCED*

(Studi Kasus: Data Nasabah Pembiayaan Warung Mikro Bank Syariah  
Mandiri KC Jambi)

Oleh: Julia Widiastuti

*Program Studi Statistika, Fakultas MIPA*

*Universitas Islam Indonesia*

*E-mail: [14611240@students.uii.ac.id](mailto:14611240@students.uii.ac.id)*

## INTISARI

Perekonomian rakyat merupakan prioritas utama dalam pembangunan sosial dengan upaya menciptakan lapangan kerja dan meningkatkan pendapatan masyarakat. Salah satu upaya untuk menciptakan lapangan kerja dan meningkatkan pendapatan masyarakat yaitu Usaha Mikro, Kecil dan Menengah (UMKM). UMKM telah menjadi tonggak perekonomian nasional karena mampu bertahan dan menguasai sebagian besar usaha di Indonesia. Namun, masih terdapat permasalahan pada sektor UMKM, salah satunya adalah kurangnya permodalan. Pinjaman uang dari bank merupakan solusi pada sektor UMKM dalam memenuhi kebutuhan modalnya, salah satu bank yang membantu akses pembiayaan dan pengembangan UMKM yaitu Bank Syariah Mandiri dengan produk Warung Mikro. Namun, dalam pembiayaan terdapat permasalahan seperti kredit atau pembiayaan macet, maka dari itu diperlukan analisis untuk melihat nasabah yang berpotensi untuk mengalami pembiayaan bermasalah. Random forest merupakan teknik data mining untuk melakukan klasifikasi. Namun jika dataset yang digunakan merupakan sebuah data yang imbalanced (tidak seimbang) maka diperlukan pendekatan dengan melakukan resample pada data asli menggunakan teknik sampling kelas imbalanced. Data yang digunakan yaitu data nasabah pembiayaan Warung Mikro tahun 2014-2017. Berdasarkan hasil penelitian, didapatkan bahwa dengan menggunakan random forest nilai sensitivty bernilai 0 yang berarti menghasilkan akurasi yang bias. Setelah menggunakan teknik sampling kelas imbalanced, nilai sensitivity, precision, dan G-mean meningkat dibandingkan menggunakan data asli. Random forest dengan teknik oversampling mempunyai nilai AUC tertinggi yaitu sebesar 66.69%.

**Kata Kunci** : UMKM, Warung Mikro, Klasifikasi, Random Forest, Data Imbalanced

### Pendahuluan

Prioritas utama dalam pembangunan sosial di Indonesia

adalah kesejahteraan rakyat, dimana dapat diwujudkan dengan membangun perekonomian rakyat. Hal tersebut dapat dilakukan dengan

cara menciptakan lapangan kerja yang memadai, meningkatkan pendapatan masyarakat, dan mengurangi kemiskinan [1]. Jika perekonomian Indonesia diletakkan dalam suatu kuadran dapat dilihat bahwa sebagian besar kegiatan perekonomian Indonesia berada pada perekonomian tradisional yang ditandai dengan sektor usaha kerakyatan yang bersifat *labor intensive* [2]. Pemberdayaan UMKM merupakan langkah yang strategis dalam meningkatkan dan memperkuat dasar kehidupan perekonomian dari sebagian besar rakyat Indonesia, khususnya melalui penyediaan lapangan kerja dan mengurangi kesenjangan dan tingkat kemiskinan [3].

Namun, masih terdapat permasalahan pada sektor UMKM yang sangat penting bagi pertumbuhan sektor tersebut, salah satu faktor internalnya adalah kurangnya permodalan dan terbatasnya akses pembiayaan bagi pemilik usaha [4]. Pinjaman uang dari bank merupakan solusi pada sektor Selain bank konvensional, bank syariah juga menawarkan

pinjaman sektor ekonomi khususnya sektor UMKM.

Pada Bank Syariah perkreditan dikenal dengan nama pembiayaan, pembiayaan adalah menyediakan dana guna membiayai kebutuhan nasabah yang memerlukannya dan layak memperolehnya [5]. Dalam hal tersebut, pihak bank harus melakukan tindakan agar tidak menimbulkan kerugian, misalnya memberikan keringanan berupa jangka waktu atau angsuran bagi nasabah yang terkena musibah atau penyitaan bagi penundaan pembiayaan nasabah yang sengaja lalai untuk membayar. Maka dari itu, agar menjadi acuan untuk menganalisa karakteristik nasabah pada warung mikro BSM digunakan data nasabah warung mikro Bank Syariah Mandiri KC Jambi.

Berdasarkan uraian di atas diperlukan analisis untuk melihat karakteristik nasabah yang berpotensi untuk mengalami pembiayaan bermasalah. Pada permasalahan ini dapat diterapkan teknik *data mining* untuk melakukan klasifikasi menggunakan metode *Random Forest*. *Random forest* merupakan

metode pohon gabungan yang berasal dari pengembangan metode *Classification and Regression Tree* (CART), yaitu dengan menerapkan metode *bootstrap aggregating* (*bagging*) dan *random feature selection* [6].

Akurasi merupakan ukuran kinerja ketepatan klasifikasi dari sebuah algoritma *machine learning*, namun jika dataset yang digunakan merupakan sebuah data yang *imbalanced* (tidak seimbang) maka ukuran akurasi menjadi tidak tepat [7]. Maka dari itu, peneliti menanggapi permasalahan kelas yang *imbalanced* dengan pendekatan melakukan *resample* pada data asli menggunakan tiga teknik *sampling* kelas *imbalanced* yaitu *undersampling*, *oversampling*, dan *Synthetic Minority Oversampling Technique* (SMOTE).

### **Tujuan Penelitian**

Tujuan dari penelitian ini adalah :

1. Mengetahui gambaran umum pembiayaan Warung Mikro Bank Syariah Mandiri KC Jambi.
2. Mengklasifikasikan pembiayaan Warung Mikro Bank Syariah

Mandiri KC Jambi menggunakan metode *Random Forest*.

3. Mengetahui perbandingan klasifikasi pembiayaan Warung Mikro Bank Syariah Mandiri KC Jambi setelah diterapkan teknik *sampling* kelas *imbalanced* pada metode *Random Forest*.

### **Metodologi Penelitian**

Variabel penelitian ini terdiri dari variabel dependen (terikat) dan variabel independen (bebas). Variabel dependen yang digunakan yaitu keterangan saldo mencukupi atau tidak mencukupi. Keterangan saldo mencukupi atau tidak mencukupi tersebut berpengaruh terhadap masalah pembiayaan Warung Mikro, dimana jika saldo tidak mencukupi akan mempengaruhi macet tidaknya pembiayaan tersebut. Terdapat enam variabel independen pada penelitian ini yaitu jenis kelamin, pekerjaan, jenis nasabah, jenis pembiayaan mikro, lama angsuran, jumlah pinjaman, margin, dan besarnya angsuran.

Data yang digunakan adalah data sekunder yang diperoleh dari data Bank Syariah Mandiri KC

Jambi. Populasi dalam penelitian ini adalah data nasabah pembiayaan Warung Mikro Bank Syariah Mandiri KC pada tahun 2014 – 2017. Alat pengolahan data yang digunakan yaitu Ms. Excel dan R.

. Dalam tahapan ini dilakukan pengkajian data yang telah diperoleh berdasarkan teori, yaitu analisis data secara deskriptif dan menggunakan metode *random forest*, kemudian dilakukan penerapan langkah-langkah sebagai berikut:

1. Melakukan analisis deskriptif semua variabel yang digunakan, baik variabel *dependent* (terikat) maupun variabel *independent* (tidak terikat).
2. Menentukan data latih (*training*) yang akan digunakan untuk melakukan metode *random forest*.
3. Menentukan nilai *Mtry* terbaik. Terdapat tiga cara untuk mendapatkan nilai *m* untuk mengamati *error* OOB yaitu [8]:

$$m = \frac{1}{2} |\sqrt{p}| \quad (1)$$

$$m = |\sqrt{p}| \quad (2)$$

$$m = 2 \times |\sqrt{p}| \quad (3)$$

dimana:  $p$  = total variabel .

Penggunaan *m* yang tepat akan menghasilkan *random forest* dengan korelasi antar pohon cukup kecil namun kekuatan setiap pohon cukup besar yang ditunjukkan dengan perolehan *error* OOB bernilai kecil [6].

4. Memilih nilai *n* yang menunjukkan jumlah pohon.
5. Membuat plot untuk melihat proses pelatihan data latih agar mendapatkan hasil yang lebih optimal.
6. Memilih teknik *sampling* yang akan digunakan pada metode *random forest*.
7. Mengklasifikasi data dengan metode *random forest* menggunakan data latih (*training*) yang telah ditentukan sebelumnya.
8. Melakukan pengujian akurasi menggunakan data uji (*testing*) untuk melihat seberapa besar akurasi metode *random forest* terhadap data penelitian.
9. Membuat plot ukuran tingkat kepentingan *variable importance* dari variabel *independent* (tidak terikat).

## Hasil dan Pembahasan

### **Analisis Deskriptif**

Presentase variabel keterangan saldo yang dibagi menjadi dua kategori yaitu saldo mencukupi dan saldo tidak mencukupi. Terdapat 23.71% dari 232 nasabah dengan saldo mencukupi sedangkan 76.29% dari 232 nasabah dengan saldo tidak mencukupi. analisis deskriptif dari variabel umur, dimana minimal umur data nasabah Warung Mikro BSM KC Jambi adalah 25 tahun sedangkan umur maksimal nasabah Warung Mikro BSM KC Jambi adalah 65 tahun dan rata-rata umur nasabah adalah 45 tahun.

Minimal harga pokok pembiayaan data nasabah Warung Mikro BSM KC Jambi tahun 2014-2017 sebesar Rp 15.000.000,00 sedangkan maksimal harga pokok pembiayaan nya sebesar Rp 200.000.000,00. margin terendah nasabah sebesar 16% sedangkan margin tertinggi sebesar 36%. Margin disesuaikan berdasarkan harga pokok pembiayaan, jangka waktu angsuran, dan harga jaminan.

Lebih banyak jenis nasabah yang menjalankan usahanya perorangan atau retail dibandingkan

dengan kerjasama KUD, dimana 21.12% dari 232 nasabah menjalankan usahanya perorangan atau retail dan sisanya sebesar 78.88% dari 232 nasabah menjalankan usahanya dengan kerjasama KUD. Pada variabel jangka waktu angsuran dalam bulan terdapat 101 nasabah dengan jangka waktu angsuran 36 bulan atau setara dengan 3 tahun, 82 nasabah dengan jangka waktu angsuran 48 bulan atau setara dengan 4 tahun, 29 nasabah dengan jangka waktu angsuran 60 bulan atau setara dengan 5 tahun, 14 nasabah dengan jangka waktu angsuran 24 bulan atau setara dengan 2 tahun, 4 nasabah dengan jangka waktu angsuran 12 bulan atau setara dengan 1 tahun, dan masing-masing 1 orang nasabah dengan jangka waktu angsuran 18 bulan (1.5 tahun) dan 30 bulan (2.5 tahun).

Jumlah angsuran minimal yang harus dibayarkan nasabah tiap bulannya yaitu Rp 572.857,00 sedangkan jumlah angsuran maksimal yang harus dibayarkan nasabah tiap bulannya yaitu Rp 9.714.480,00. jenis pekerjaan terbanyak pada data nasabah Warung

Mikro BSM KC Jambi adalah wiraswasta dengan jumlah sebanyak 152 nasabah, diikuti oleh petani dengan jumlah 42 nasabah, karyawan dengan jumlah 14 nasabah, Pegawai Negeri Sipil (PNS) dengan jumlah 13 nasabah, buruh harian lepas sebanyak 2 nasabah, pegawai sebanyak 2 nasabah, dan pekerjaan lainnya sebanyak 7 nasabah. Harga jaminan terendah senilai Rp 16.000.000,00 sedangkan harga jaminan tertinggi Rp 250.000.000,00.

Jenis nasabah perorangan/retail dengan saldo keterangan tidak mencukupi sebesar 135 nasabah. Ini berarti bahwa pada data yang digunakan, nasabah dengan keterangan saldo tidak mencukupi paling banyak terdapat pada jenis nasabah perorangan/retail. Lalu, nasabah terbanyak terdapat pada saat PM Utama dan keterangan saldo tidak mencukupi. Ini berarti bahwa pada data yang digunakan, terdapat 110 nasabah dengan keterangan saldo tidak mencukupi paling banyak terdapat pada jenis pembiayaan PM Utama.

Nasabah dengan keterangan saldo tidak mencukupi paling banyak

terdapat pada jenis nasabah perorangan/retail dan jenis pembiayaan PM Utama yaitu sejumlah 88 nasabah. Pada keterangan saldo mencukupi, nasabah paling banyak saat jenis nasabah perorangan/retail dan jenis pembiayaan PM Utama sejumlah 29 nasabah. Ini berarti bahwa pada data yang digunakan, nasabah dengan keterangan saldo tidak mencukupi paling banyak terdapat pada jenis nasabah perorangan/retail dan jenis pembiayaan PM Utama dan untuk nasabah dengan keterangan saldo mencukupi paling banyak juga terdapat pada jenis nasabah perorangan/retail dan jenis pembiayaan PM Utama.

### ***Random Forest***

Peneliti menggunakan 70% untuk data latih (*train*) dan 30% untuk data uji (*test*). Langkah selanjutnya adalah menentukan nilai *Mtry* atau peubah penjelas dan *Ntree* atau jumlah pohon agar mendapatkan model yang optimal dan nilai *error* OOB yang bernilai kecil. Penentuan *Mtry* agar didapatkan nilai yang optimal dapat dilakukan dengan tiga cara, yaitu:

$$4. \quad Mtry = \frac{\sqrt{\text{total variabel}}}{2}$$

$$= \frac{\sqrt{9}}{2} = 1.5 \approx 1$$

$$5. \quad Mtry = \sqrt{\text{total variabel}}$$

$$= \sqrt{9} = 3$$

$$6. \quad Mtry = \sqrt{\text{total variabel}} \times 2$$

$$= \sqrt{9} \times 2 = 3 \times 2 = 6$$

Setelah diujikan dengan masing-masing nilai *Mtry* yang didapatkan menggunakan *default* jumlah pohon (*Ntree*) sebanyak 500, maka didapatkan nilai *error* OOB sebagai berikut:

**Tabel 1.** Pengujian Nilai *Error* OOB

<i>Mtry</i>	<i>Error</i> OOB
1	26.63%
3	31.95%
6	29.59%

Kemudian, langkah selanjutnya yaitu menentukan jumlah pohon (*Ntree*) yang akan digunakan menggunakan nilai *Mtry* yang didapatkan sebelumnya. Jumlah pohon (*Ntree*) yang akan dicoba oleh peneliti dimulai dari 25, 50, 100, 250, 300, 500, dan 1000. Berikut adalah nilai *error* yang dihasilkan oleh setiap jumlah pohon dengan nilai *Mtry* sebesar 1 :

**Tabel 2.** Pengujian Nilai *Error* OOB Dengan *Ntree* Berbeda

<i>Ntree</i>	<i>Error</i> OOB
25	28.4%

50	28.99%
100	26.63%
250	25.44%
300	26.04%
500	26.63%
1000	26.63%

Nilai *error* OOB terendah pada saat jumlah pohon (*Ntree*) sebesar 250 adalah 25.44%.

Pada perhitungan 1-APER pada metode *random forest* bernilai 82.54% yang berarti terdapat 82.54% data sampel yang tepat diklasifikasikan. Namun, dikarenakan data yang digunakan merupakan dataset yang *imbalanced*, maka akurasi klasifikasi lainnya harus diperhatikan juga. *G-mean* berguna untuk menentukan keseimbangan akurasi prediksi untuk mengukur permasalahan *imbalanced*, dimana *G-mean* yang didapatkan jika menggunakan metode *random forest* adalah 0 yang menandakan ketidakseimbangan prediksi. *G-mean* diperlukan karena metode klasifikasi cenderung baik dalam memprediksi kelas dengan data sampel yang lebih banyak namun buruk dalam memprediksi kelas dengan data sampel yang sedikit. AUC merupakan cara yang baik untuk mendapatkan nilai kinerja

pengklasifikasi secara umum dimana AUC adalah ukuran kinerja yang populer dalam ketidakseimbangan kelas.

Hasil akurasi yang didapatkan menunjukkan bahwa metode *random forest* bagus dalam mengklasifikasikan kelas saldo tidak mencukupi, namun tidak dapat

mengklasifikasikan kelas saldo mencukupi dengan baik. Hal ini menandakan hasil akurasi yang bias atau disebut dengan *accuracy paradox*. Jika hal ini terjadi, maka lebih baik menggunakan ukuran ketepatan klasifikasi yang lain selain akurasi seperti *precision* dan *sensitivity*.

**Tabel 3.** Perbedaan Keempat Metode *Random Forest* yang Digunakan

	<i>Random Forest</i>	<i>Undersampling</i>	<i>Oversampling</i>	SMOTE
<i>Specificity</i>	100%	63.46%	<b>78.85%</b>	69.23%
<i>Sensitivity</i>	0%	45.45%	<b>54.55%</b>	45.46%
<i>Precision</i>	0%	20.83%	<b>35.29%</b>	23.81%
<i>APER</i>	17.46%	39.68%	25.39%	34.92%
<i>1-APER</i>	82.54%	60.32%	<b>74.60%</b>	65.08%
<i>G-mean</i>	0%	54.08%	<b>65.58%</b>	56.09%
<i>AUC</i>	50%	54.46%	<b>66.69%</b>	57.34%

Menurut **Tabel 3** nilai 1-APER (*Total Accucary Rate*) pada ketiga teknik *sampling* memberikan nilai akurasi lebih kecil dibandingkan dengan metode *random forest* tanpa teknik *sampling imbalanced data*. Walaupun nilai akurasi dari metode *random forest* paling tinggi, untuk pengklasifikasian kelas *true positive* tidak ada yang tepat. Nilai akurasi *oversampling* lebih tinggi dibandingkan dengan *undersampling* dan SMOTE. Nilai *sensitivity* tertinggi yaitu pada saat menggunakan teknik *oversampling*

yaitu sebesar 54.55% yang berarti bahwa sebesar 54.55% data nasabah dengan keterangan saldo mencukupi yang tepat diklasifikasikan dengan keterangan saldo mencukupi.

Selanjutnya adalah nilai *specificity* tertinggi saat menggunakan teknik *oversampling* yaitu sebesar 78.85% yang berarti bahwa data nasabah dengan keterangan saldo tidak mencukupi yang tepat diklasifikasikan dengan keterangan saldo tidak mencukupi adalah 78.85%. Dari ketiga teknik *sampling* kelas data *imbalanced*

memberikan nilai *precision* dan *sensitivty* yang lebih baik dibandingkan dengan menggunakan metode *random forest* tanpa teknik *sampling* data *imbalanced*.

Nilai *precision* dengan menggunakan teknik *oversampling* lebih tinggi dibandingkan kedua teknik lainnya yaitu sebesar 35.29%. Selanjutnya, yaitu nilai AUC yang merupakan cara mendapatkan nilai kinerja pengklasifikasian secara umum dan untuk membandingkannya dengan pengklasifikasian yang lain serta ukuran kinerja yang populer dalam ketidakseimbangan kelas dimana nilai AUC tertinggi yaitu jika menggunakan *random forest* dengan teknik *oversampling* yaitu sebesar 66.69% sedangkan teknik *sampling* dengan nilai AUC terendah yaitu *random forest* dengan teknik *undersampling* yaitu sebesar 54.46%.

### **Kesimpulan**

Berdasarkan analisis terhadap hasil penelitian yang telah dilakukan dengan 232 data nasabah pembiayaan Warung Mikro KC Jambi, dapat diambil kesimpulan sebagai berikut:

1. Data yang digunakan adalah *imbalanced data* (data tidak seimbang) dimana perbandingan antara kelas saldo mencukupi dan tidak mencukupi yaitu 55 :177. Pada variabel umur, umur terkecil nasabah pembiayaan Warung Mikro BSM KC Jambi adalah 25 tahun sedangkan umur maksimal nasabah adalah 65 tahun. Selanjutnya, variabel jenis nasabah dapat dilihat bahwa lebih banyak jenis nasabah yang menjalankan usahanya perorangan atau retail dibandingkan dengan kerjasama KUD. Pada variabel jenis pembiayaan mikro, terdapat 38.36% dari 232 nasabah yang mengajukan PM Madya dan 61.64% dari 232 nasabah yang mengajukan PM Utama yang berarti bahwa lebih banyak nasabah yang melakukan pembiayaan PM Utama. Harga pokok pembiayaan terkecil sebesar Rp 15.000.000,00 dan terbesar sebesar Rp 200.000.000,00. Selanjutnya, untuk variabel margin terendah sebesar 16% dan tertinggi sebesar 36% serta jangka waktu angsuran dengan nasabah terbanyak sebesar

101 nasabah pada saat jangka waktu angsuran 36 bulan. Jumlah angsuran minimal yang harus dibayarkan nasabah tiap bulannya yaitu Rp 572.857,00 sedangkan jumlah angsuran maksimal yang harus dibayarkan nasabah tiap bulannya yaitu Rp 9.714.480,00. Jenis pekerjaan terbanyak adalah wiraswasta dengan jumlah sebanyak 152 nasabah dan harga jaminan terendah senilai Rp 16.000.000,00 sedangkan harga jaminan tertinggi Rp 250.000.000,00. Pada data yang digunakan, nasabah dengan keterangan saldo tidak mencukupi paling banyak terdapat pada jenis nasabah perorangan atau retail dan jenis pembiayaan PM Utama dan untuk nasabah dengan keterangan saldo mencukupi paling banyak juga terdapat pada jenis nasabah perorangan atau retail dan jenis pembiayaan PM Utama.

2. Dengan menggunakan data latihan (*training*) sebesar 70% dan data uji (*test*) sebesar 30% serta jumlah pohon atau iterasi sebanyak 250 dan *Mtry* sebesar 1, didapatkan hasil klasifikasi dengan nilai

estimasi error OOB sebesar 25.44%. Akurasi tertinggi didapatkan pada saat digunakan metode *random forest* tanpa teknik *sampling* kelas *imbalanced* yaitu sebesar 82.54%, namun nilai *sensitivity* bernilai 0 yang berarti menghasilkan akurasi yang bias.

3. Saat menggunakan teknik *sampling* kelas *imbalanced* dapat meningkatkan nilai *sensitivity*, *precision*, *G-mean*, dan AUC. Nilai AUC tertinggi didapatkan pada *random forest* teknik *oversampling* sebesar 66.69% dan terendah yaitu *random forest* teknik *undersampling* sebesar 54.46%.

### Saran

Pada penelitian selanjutnya diharapkan agar menggunakan metode *Random Forest* lainnya, misalnya *Weighted Random Forest* ataupun *Balanced Random Forest* untuk meningkatkan hasil klasifikasi yang lebih bagus untuk data *imbalanced*. Selanjutnya untuk meningkatkan nilai akurasi dan kualitas hasil klasifikasi, sebaiknya pada penelitian selanjutnya menggunakan data yang lebih banyak lagi serta menambah variabel

lainnya yang lebih mendukung penelitian ini.

### Daftar Pustaka

- [1] Tedjasuksmana, B. 2014. Potret UMKM Indonesia Menghadapi Masyarakat Ekonomi Asean 2015. The 7th NCFB and Doctoral Colloquium 2014. Universitas Katolik Widya Mandala Surabaya.
- [2] Abdullah, B. 2006. Menanti Kemakmuran Negeri : Kumpulan Esai tentang Pembangunan Sosial Ekonomi Indonesia. Jakarta : PT Gramedia Pustaka Utama.
- [3] Nikmah, C., Sukarno, H., dan Mufidah, A. 2014. “Analisis Implikasi Pembiayaan Syariah pada Pedagang Kecil di Pasar Tanjung Jember”. Jurnal Ekonomi Bisnis dan Akuntansi. Vol.1 No.1 Hal.8-15.
- [4] Suhendar, 2010
- [5] Arifin, Z. 2002. Dasar-dasar Manajemen Bank Syari’ah. Jakarta : AlfaBet.
- [6] Breiman, L., & Cutler A. 2001. *Random Forests. Machine Learning* 45 : 5-32.
- [7] Chawla, N. V dkk, 2002. *SMOTE: Syntethic Minority Over-Sampling Technique. Journal of Artificial Intelligence Research.* Vol.16 Hal.321-357.
- [8] Breiman L, Cutler A. 2003. Manual on Setting Up, Using, and Understanding Random Forest V4.0. [http://oz.berkeley.edu/users/breiman/Using\\_random\\_forests\\_v4.0.pdf](http://oz.berkeley.edu/users/breiman/Using_random_forests_v4.0.pdf). Diakses pada tanggal 6 April 2018, pukul 08.27 WIB.



# LAMPIRAN

## LAMPIRAN

### LAMPIRAN 1 Sintaks *Random Forest*

```

#-----PERSIAPAN DATA-----
data <- read.delim ("clipboard")
str(data)
data$y <- as.factor(data$y)
table(data$y)
summary(data)
barplot(prop.table(table(data$y)),
        col = rainbow(2),
        ylim = c(0, 1),
        main = "Keterangan Sald0")

#-----TRAINING & TEST-----
set.seed(123)
ind <- sample(2, nrow(data), replace = TRUE, prob = c(0.7,
0.3))
train <- data[ind==1,]
test <- data[ind==2,]
summary(train)
summary(test)

#-----RANDOM FOREST MODEL-----
library(randomForest)
set.seed(123)
rf <- randomForest(y~., data=train,
                   ntree = 250,
                   mtry = 1,
                   importance = TRUE,
                   proximity = TRUE)

print(rf)
attributes(rf)
plot(rf)

#-----Akurasi Menggunakan Data Test-----
library(caret)
p2 <- predict(rf, test)
head(p2)
confusionMatrix(predict(rf, test), test$y, positive = 'Y')

```

## **LAMPIRAN 2 Sintaks *Random Forest* Menggunakan *Undersampling* dan *Oversampling* serta *Variable Importance***

```

#---Random Forest Menggunakan Teknik Imbalanced---
#-----Random Forest Undersampling-----
table(train$y)
prop.table(table(train$y))
summary(train)
library(ROSE)
set.seed(250)
under <- ovun.sample(y~., data=train, method = "under", N =
88)$data
table(under$y)
summary(under)
set.seed(123456)
rfunder <- randomForest(y~., data=under, ntree=250, mtry=1,
importance = TRUE,
                        proximity = TRUE)
rfunder
#--Prediksi Random Forest Undersampling dengan Data Test--
library(caret)
library(e1071)
confusionMatrix(predict(rfunder, test), test$y, positive = 'Y')

#-----Random Forest Oversampling-----
set.seed(250)
over <- ovun.sample(y~., data = train, method = "over", N =
250)$data
table(over$y)
summary(over)
set.seed(123456)
rfover <- randomForest(y~., data = over, ntree=250, mtry=1,
importance = TRUE,
                        proximity = TRUE)
rfover
#--Prediksi Random Forest Oversampling dengan Data Test--
confusionMatrix(predict(rfover, test), test$y, positive = 'Y')

#-----Random Forest SMOTE-----
Set.seed(250)
rose <- ROSE(y~., data = train, N = 500, seed=111)$data
table(rose$y)
summary(rose)
set.seed(123456)
rfrose <- randomForest(y~., data=rose, ntree=250, mtry=1,
importance = TRUE,
                        proximity = TRUE)
rfrose
#--Prediksi Random Forest SMOTE dengan Data Test--
confusionMatrix(predict(rfrose, test), test$y, positive = 'Y')

#-----variable Importance dan Pohon Model-----
varImpPlot(rfover,

```

```
        sort = T,  
        n.var = 9,  
        main = "Variable Importance")  
library(reprtree)  
set.seed(1)  
rfover1 <- randomForest(y~x1+x7+x8+x9+x4+x6+x5+x2+x3, data = over,  
ntree=250, mtry=1, importance = TRUE,  
                        proximity = TRUE)  
reprtree:::plot.getTree(rfover1)
```