

**SKRIPSI**

**DATA MINING UNTUK STRATEGI MARKETING BANK**



**Stanislaus Dendrio Evan**

**NPM: 61819010034**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS  
UNIVERSITAS KATOLIK PARAHYANGAN  
2023**



**UNDERGRADUATE THESIS**

**DATA MINING FOR BANK MARKETING STRATEGY**



**Stanislaus Dendrio Evan**

**NPM: 61819010034**

**DEPARTMENT OF INFORMATICS  
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES  
PARAHYANGAN CATHOLIC UNIVERSITY  
2023**

**LEMBAR PENGESAHAN**

**DATA MINING UNTUK STRATEGI MARKETING BANK**

**Stanislaus Dendrio Evan**

**NPM: 61819010034**

**Bandung, 6 Juli 2023**

**Menyetujui,**

**Pembimbing**

**Digitally signed  
by Veronica Sri  
Moertini**

**Dr. Veronica Sri Moertini**

**Ketua Tim Penguji  
Digitally signed  
by Raymond  
Chandra Putra**

**Raymond Chandra Putra, M.T.**

**Anggota Tim Penguji**

**Digitally signed  
by Natalia**

**Natalia, M.Si.**

**Mengetahui,**

**Ketua Program Studi  
Digitally signed  
by Mariskha Tri  
Adithia**

**Mariskha Tri Adithia, P.D.Eng**

## PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

### DATA MINING UNTUK STRATEGI MARKETING BANK

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,  
Tanggal 6 Juli 2023



Stanislaus Dendrio Evan  
NPM: 61819010034

## ABSTRAK

Kegiatan pemasaran atau *marketing* barang dan jasa telah menjadi praktik umum dalam dunia bisnis. *Marketing* bertujuan untuk menarik dan mempertahankan pelanggan baru sebanyak mungkin. Kesuksesan *marketing* tidak hanya tergantung pada jumlah produk yang terjual, tetapi juga pada kemampuan untuk terus menambah pelanggan baru sambil mempertahankan yang sudah ada. Untuk menjaga hubungan yang baik dengan pelanggan, pelaku usaha perlu memahami apa yang diinginkan oleh pelanggan. Salah satu cara untuk mengetahui kebutuhan pelanggan adalah melalui riset pasar.

Kegiatan *marketing* saat ini dapat dilakukan secara langsung maupun tidak langsung. *Marketing* langsung melibatkan penawaran langsung kepada calon pelanggan tanpa melalui perantara, sedangkan *marketing* tidak langsung melibatkan perantara, seperti ulasan dari *vlogger*. *Telemarketing* merupakan salah satu bentuk pemasaran langsung melalui telepon. Jenis *marketing* ini adalah teknik *marketing* yang telah lama ada dan masih populer. Hal ini disebabkan oleh metode yang sederhana dan mudah dilakukan.

*Telemarketing* sering digunakan dalam industri perbankan karena dianggap efektif dalam membangun hubungan yang baik dengan nasabah. Biaya telepon juga lebih rendah dibandingkan dengan media iklan lainnya. Namun, tidak semua nasabah menghargai *telemarketing*, terutama kalangan muda dan generasi milenial yang merasa privasinya terganggu. Oleh karena itu, bank perlu mencari strategi pemasaran yang sesuai dengan karakteristik nasabahnya.

*Data mining* merupakan proses untuk menemukan pola dan pengetahuan dari data dalam jumlah besar. *Data mining* melibatkan beberapa tahapan, mulai dari persiapan data, analisis data untuk mencari pola, hingga penyampaian wawasan melalui visualisasi data dan teknik presentasi. Penelitian ini akan melibatkan beberapa tahapan *data mining*, termasuk pencarian data bank dan data terkait, persiapan data, pemilihan fitur, pembuatan model dengan *data mining*, evaluasi model, pembuatan perangkat lunak untuk melakukan prediksi, dan penentuan strategi *marketing* yang tepat berdasarkan *insight* yang diperoleh dari proses segmentasi nasabah.

Hasil dari penelitian ini adalah suatu perangkat lunak yang dapat meluncurkan model klasifikasi berdasarkan *dataset* yang digunakan, yaitu *dataset* deposito berjangka dan *dataset* pinjaman pribadi bank. Model klasifikasi yang diluncurkan merupakan model dengan algoritma *decision tree* yang menggunakan fitur yang telah dipilih menggunakan beberapa teknik pemilihan fitur, seperti *information gain* dan ANOVA. Perangkat lunak ini juga dapat menampilkan hasil *clustering* untuk segmentasi nasabah bank.

**Kata-kata kunci:** pemasaran, *telemarketing*, penambangan data, pembelajaran mesin, bank, strategi pemasaran



## ABSTRACT

*Marketing activities for goods and services have become a common practice in the business world. Marketing aims to attract and retain as many new customers as possible. The success of marketing is not only determined by the quantity of products sold but also by the ability to continuously acquire new customers while retaining existing ones. To maintain a good relationship with customers, businesses need to understand their needs. One way to identify customer needs is through market research.*

*Marketing activities can be conducted directly or indirectly. Direct marketing involves offering products or services directly to potential customers without intermediaries, while indirect marketing involves intermediaries, such as vlogger reviews. Telemarketing is one form of direct marketing through telephone. This marketing technique has been around for a long time and is still popular due to its simple and easy-to-implement methods.*

*Telemarketing is often used in the banking industry because it is considered effective in building good relationships with customers. Telephone costs are also lower compared to other advertising media. However, not all customers appreciate telemarketing, especially the younger generation and millennials who feel their privacy is being invaded. Therefore, banks need to find marketing strategies that are suitable for their customers' characteristics.*

*Data mining is a process of discovering patterns and knowledge from large amounts of data. Data mining involves several stages, from data preparation, data analysis to find patterns, to conveying insights through data visualization and presentation techniques. This research will involve various data mining stages, including searching for bank-related data, data preparation, feature selection, model creation using data mining, model evaluation, software development for prediction, and determining appropriate marketing strategies based on insights obtained from customer segmentation process.*

*The result of this research is a software that can deploy classification models based on the used dataset, namely, the deposit dataset and the personal loan dataset of the bank. The launched classification model is a decision tree algorithm that employs features selected using various feature selection techniques, such as information gain and ANOVA. This software can also display clustering results for bank customer segmentation.*

**Keywords:** *marketing, telemarketing, data mining, machine learning, bank, marketing strategy*



*Dengan rasa syukur yang mendalam, skripsi ini dipersembahkan kepada:*

- 1. Orang tua dan keluarga besar penulis yang telah senantiasa membantu penulis dalam menyelesaikan skripsi.*
- 2. Kemajuan ilmu di Indonesia serta segenap civitas akademika Universitas Katolik Parahyangan.*
- 3. Rekan dan teman-teman penulis yang senantiasa memberikan dukungan moral kepada penulis.*



## KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa, atas segala rahmat dan berkat-Nya, sehingga penyusunan skripsi dapat diselesaikan dengan baik. Skripsi yang berjudul *Data Mining untuk Strategi Marketing Bank* ini diajukan sebagai syarat untuk menempuh pendidikan di Program Studi Teknik Informatika, Fakultas Teknologi Informasi dan Sains, Universitas Katolik Parahyangan, Bandung.

Penulis juga tak lupa untuk memberikan ucapan terima kasih yang sebesar-besarnya kepada pihak-pihak yang terlibat baik dalam memberi bimbingan, bantuan dan dukungan secara langsung maupun tidak langsung ke dalam proses penyusunan skripsi ini, terutama untuk:

1. Kepada Tuhan yang Maha Esa atas rahmat-Nya memperkenankan penulis untuk menyelesaikan skripsi ini.
2. Ibu Lydia Mutiara Dewi, M.A. dan Ibu Dr. Veronica Sri Moertini, selaku dosen-dosen pembimbing Skripsi *Data Mining* untuk Strategi *Marketing Bank*.
3. Bapak Raymond Chandra Putra, M.T. dan Ibu Natalia, M.Si., selaku dosen-dosen penguji.
4. Orang tua dan keluarga besar, yang senantiasa selalu memberi dukungan moral kepada penulis dalam proses penyusunan skripsi.
5. Teman-teman penulis dalam Gabut Club dan Grup yeuuuu yang senantiasa selalu memberikan dukungan dan menemani penulis selama proses penyusunan skripsi.
6. Teman-teman lainnya, seperti Ignatius Daniel Widjojo, Adella Dewianne Lipang, dan Deana Lipang yang senantiasa selalu memberikan dukungan moral kepada penulis selama proses penyusunan skripsi.
7. Indra Permana Sugianto, selaku mentor dan pembimbing yang senantiasa memberi bimbingan kepada penulis selama menyusun skripsi.
8. Seluruh pihak lainnya yang tidak dapat penulis sebutkan satu per satu yang telah memberikan dukungan baik secara langsung maupun tidak langsung dalam penyusunan. skripsi ini.

Akhir kata penulis berharap skripsi ini dapat memberikan sumbangan bagi kemajuan ilmu di Indonesia dan *civitas* akademika Universitas Katolik Parahyangan.

Bandung, Juli 2023

Penulis



# DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xxi
DAFTAR TABEL	xxiii
DAFTAR KODE PROGRAM	xxvii
<b>1 PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang . . . . .	1
1.2 Rumusan Masalah . . . . .	3
1.3 Tujuan . . . . .	3
1.4 Batasan Masalah . . . . .	3
1.5 Metodologi . . . . .	4
1.6 Sistematika Pembahasan . . . . .	4
<b>2 LANDASAN TEORI</b>	<b>5</b>
2.1 <i>Marketing</i> . . . . .	5
2.2 <i>Telemarketing</i> . . . . .	7
2.3 Bank . . . . .	8
2.3.1 Bank di Portugal . . . . .	11
2.3.2 Bank di Amerika . . . . .	11
2.4 <i>Data Mining</i> . . . . .	11
2.5 Tahapan <i>Data Mining</i> . . . . .	13
2.5.1 <i>Data Preprocessing</i> . . . . .	15
2.5.2 Analisis Data ( <i>Data Mining</i> ) . . . . .	16
2.5.3 <i>Classification</i> . . . . .	16
2.5.4 <i>Clustering</i> . . . . .	23
2.5.5 <i>Data Mining</i> dalam Perbankan . . . . .	26
2.6 Statistika . . . . .	27
2.7 <i>Correlation</i> . . . . .	33
2.8 Analisis Komparatif . . . . .	34
<b>3 PENYIAPAN DATA DAN EKSPLORASI ALGORITMA</b>	<b>37</b>
3.1 <i>Dataset Marketing</i> Deposito Berjangka di Portugal ( <i>Dataset Portugese Bank</i> ) . . . . .	37
3.2 EDA <i>Dataset</i> Deposito Berjangka . . . . .	39
3.3 Penyiapan <i>Dataset</i> Deposito Berjangka . . . . .	49
3.3.1 Pemilihan Fitur untuk <i>Dataset</i> Pinjaman Pribadi Bank . . . . .	52
3.4 Eksplorasi Algoritma dengan <i>Dataset</i> Deposito Berjangka . . . . .	54
3.4.1 Pembersihan <i>Dataset</i> . . . . .	54
3.4.2 Pembuatan <i>Train</i> dan <i>Test</i> . . . . .	55

3.4.3	Pembuatan Model dengan Semua Fitur . . . . .	55
3.4.4	Pemilihan Fitur . . . . .	57
3.4.5	Pembuatan Model dengan Fitur Terpilih . . . . .	58
3.4.6	Kesimpulan . . . . .	60
3.5	<i>Dataset Marketing</i> Pinjaman Pribadi Bank di Amerika( <i>Dataset Personal Loan</i> ) . . . . .	61
3.6	Penyiapan <i>Dataset</i> Pinjaman Pribadi Bank . . . . .	62
3.6.1	Pemilihan Fitur untuk <i>Dataset</i> Pinjaman Pribadi Bank . . . . .	64
3.7	Eksplorasi dengan <i>Dataset</i> Pinjaman Pribadi Bank . . . . .	66
3.7.1	Pembersihan <i>Dataset</i> . . . . .	66
3.7.2	Pembuatan <i>Train</i> dan <i>Test</i> . . . . .	67
3.7.3	Pembuatan Model dengan Semua Fitur . . . . .	67
3.7.4	Pemilihan Fitur . . . . .	69
3.7.5	Pembuatan Model dengan Fitur Terpilih . . . . .	70
3.7.6	Kesimpulan . . . . .	72
<b>4</b>	<b>ANALISIS DATA DAN HASIL</b>	<b>73</b>
4.1	Analisis untuk <i>Dataset</i> Deposito Berjangka di Portugis . . . . .	73
4.1.1	Model Menggunakan Fitur 'contact', 'housing', 'duration', dan 'previous' . . . . .	73
4.1.2	Model Menggunakan Fitur 'contact', 'housing', 'duration', dan 'previous' dengan <i>Undersampling</i> . . . . .	74
4.1.3	Model Menggunakan Fitur 'contact', 'housing', 'duration', dan 'poutcome' . . . . .	75
4.1.4	Model Menggunakan Fitur 'contact', 'housing', 'duration', 'previous', 'pdays', dan 'education' . . . . .	76
4.1.5	Model Menggunakan Fitur 'contact', 'housing', 'duration', 'previous', 'poutcome', dan 'month' . . . . .	77
4.1.6	Model Menggunakan Fitur 'contact', 'housing', 'duration', 'previous', 'pdays', 'education', 'campaign', dan 'month' . . . . .	78
4.1.7	Model Menggunakan Fitur 'contact', 'month', 'duration', dan 'poutcome' . . . . .	79
4.1.8	Kesimpulan . . . . .	80
4.1.9	Hasil Prediksi Model . . . . .	80
4.2	Analisis untuk <i>Dataset</i> Pinjaman Pribadi Bank di Amerika Serikat . . . . .	81
4.2.1	Model Menggunakan Fitur 'Education', 'CDAccount', 'Income', dan 'CCAvg' . . . . .	81
4.2.2	Model Menggunakan Fitur 'Education', 'CDAccount', 'Income', dan 'CCAvg' dengan <i>Undersampling</i> . . . . .	82
4.2.3	Model Menggunakan Fitur 'Mortgage', 'CDAccount', 'Income', 'CCAvg' . . . . .	83
4.2.4	Model Menggunakan 6 Fitur . . . . .	84
4.2.5	Model Menggunakan 8 Fitur . . . . .	85
4.2.6	Kesimpulan . . . . .	85
4.2.7	Hasil Prediksi Model . . . . .	86
<b>5</b>	<b>SEGMENTASI NASABAH DAN PERANCANGAN STRATEGI <i>Marketing</i> BANK</b>	<b>87</b>
5.1	Pemilihan Fitur untuk <i>Clustering</i> . . . . .	87
5.2	Penyiapan Data untuk <i>Clustering</i> . . . . .	87
5.3	Penentuan K Terbaik . . . . .	88
5.4	Pembuatan Model <i>Clustering</i> dan Hasilnya . . . . .	89
5.5	Pemilihan Fitur dan Penyiapan Data untuk <i>Clustering</i> Siklus 2 . . . . .	91
5.6	Penentuan K Terbaik Siklus 2 . . . . .	94
5.7	Pembuatan Model <i>Clustering</i> dan Hasilnya Siklus 2 . . . . .	95
<b>6</b>	<b>PERANCANGAN PERANGKAT LUNAK DAN IMPLEMENTASINYA</b>	<b>99</b>
6.1	Perancangan dan Implementasi Perangkat Lunak <i>Dataset</i> Deposito Berjangka . . . . .	99
6.1.1	Fitur Perangkat Lunak . . . . .	99

6.1.2	Diagram <i>Use Case</i> . . . . .	99
6.1.3	Implementasi Perangkat Lunak . . . . .	101
6.2	Perancangan dan Implementasi Perangkat Lunak <i>Dataset</i> Pinjaman Pribadi Bank . . . . .	106
6.2.1	Fitur Perangkat Lunak . . . . .	106
6.2.2	Diagram <i>Use Case</i> . . . . .	106
6.2.3	Implementasi Perangkat Lunak . . . . .	108
<b>7</b>	<b>KESIMPULAN DAN SARAN</b> . . . . .	<b>113</b>
7.1	Kesimpulan . . . . .	113
7.2	Saran . . . . .	113
	<b>DAFTAR REFERENSI</b> . . . . .	<b>115</b>
	<b>A KODE PROGRAM</b> . . . . .	<b>117</b>
	<b>B <i>Dataset</i></b> . . . . .	<b>159</b>





## DAFTAR GAMBAR

1.1	Ilustrasi <i>Data Mining</i> . . . . .	2
2.1	Skema <i>Data Warehouse</i> . . . . .	12
2.2	Skema <i>Tahapan Data Mining</i> . . . . .	14
2.3	Skema Ilustrasi Analisis Data . . . . .	16
2.4	Sampel dan Populasi . . . . .	28
2.5	Contoh <i>Pie Chart</i> . . . . .	29
2.6	Contoh <i>Bar Chart</i> . . . . .	30
2.7	Contoh <i>Line Chart</i> . . . . .	30
2.8	Contoh <i>Box Plot</i> . . . . .	31
2.9	Contoh <i>Scatterplot</i> . . . . .	32
2.10	Contoh kasus untuk <i>Histogram</i> . . . . .	32
2.11	Table Frekuensi untuk <i>Histogram</i> . . . . .	33
2.12	Contoh <i>Histogram</i> . . . . .	33
3.1	Distribusi dari pekerjaan nasabah dan variabel <i>y</i> untuk <i>dataset</i> deposito berjangka	39
3.2	Distribusi dari status perkawinan nasabah dan variabel <i>y</i> untuk <i>dataset</i> deposito berjangka . . . . .	40
3.3	Distribusi dari tingkat pendidikan nasabah dan variabel <i>y</i> untuk <i>dataset</i> deposito berjangka . . . . .	41
3.4	Distribusi dari jenis telepon yang digunakan dalam menghubungi nasabah dan variabel <i>y</i> untuk <i>dataset</i> deposito berjangka . . . . .	42
3.5	Distribusi dari status pinjaman bank nasabah dan variabel <i>y</i> untuk <i>dataset</i> deposito berjangka . . . . .	42
3.6	Distribusi dari status kredit rumah nasabah dan variabel <i>y</i> untuk <i>dataset</i> deposito berjangka . . . . .	43
3.7	Distribusi dari bulan nasabah dihubungi dan variabel <i>y</i> untuk <i>dataset</i> deposito berjangka . . . . .	43
3.8	Distribusi dari pekerjaan nasabah dan variabel <i>y</i> dengan rasio untuk <i>dataset</i> deposito berjangka . . . . .	44
3.9	Distribusi dari status perkawinan nasabah dan variabel <i>y</i> dengan rasio untuk <i>dataset</i> deposito berjangka . . . . .	45
3.10	Distribusi dari tingkat pendidikan nasabah dan variabel <i>y</i> dengan rasio untuk <i>dataset</i> deposito berjangka . . . . .	46
3.11	Distribusi dari status pinjaman nasabah dan variabel <i>y</i> dengan rasio untuk <i>dataset</i> deposito berjangka . . . . .	46
3.12	Distribusi dari jenis telepon yang digunakan dalam menghubungi nasabah dan variabel <i>y</i> dengan rasio untuk <i>dataset</i> deposito berjangka . . . . .	47
3.13	Distribusi dari status kredit rumah nasabah dan variabel <i>y</i> dengan rasio untuk <i>dataset</i> deposito berjangka . . . . .	48
3.14	Tabel korelasi untuk variabel yang bertipe numerik dalam <i>dataset</i> deposito berjangka	48
3.15	Portugal Bank info . . . . .	49
3.16	Distribusi umur nasabah berdasarkan target untuk <i>dataset bank-full.csv</i> . . . . .	50

3.17	Distribusi umur nasabah berdasarkan target dengan mengelompokkan pekerjaan nasabah untuk <i>dataset bank-full.csv</i> . . . . .	50
3.18	Distribusi saldo tabungan nasabah berdasarkan target untuk <i>dataset bank-full.csv</i> . . . . .	51
3.19	Distribusi durasi <i>marketing</i> nasabah berdasarkan target untuk <i>dataset bank-full.csv</i> . . . . .	52
3.20	Sampel <i>Dataset</i> Deposito Berjangka . . . . .	54
3.21	Personal Loan info . . . . .	62
3.22	Negative Experience Personnal Loan . . . . .	63
3.23	<i>Outlier</i> Pada Variabel ZIPCode . . . . .	64
3.24	Sampel <i>Dataset</i> Pinjaman Pribadi Bank . . . . .	66
5.1	Elbow 100 . . . . .	88
5.2	Elbow 20 . . . . .	89
5.3	Silhouette Plot . . . . .	91
5.4	Pairplot Fitur <i>Clustering</i> . . . . .	92
5.5	Scater Plot Age Duration <i>Clustering</i> . . . . .	93
5.6	Scater Plot Day Duration <i>Clustering</i> . . . . .	93
5.7	Elbow 100 - 3Fitur . . . . .	94
5.8	Elbow 20 - 3Fitur . . . . .	95
5.9	Scatter Plot Clustering . . . . .	96
5.10	Silhouette Plot Clustering . . . . .	97
6.1	<i>Use Case</i> Deposito Berjangka . . . . .	100
6.2	Tampilan Awal . . . . .	101
6.3	Tampilan Awal Isi . . . . .	102
6.4	Tampilan Prediksi Negatif . . . . .	102
6.5	Tampilan Prediksi Positif . . . . .	103
6.6	Tampilan Browse input . . . . .	103
6.7	Tampilan Setelah Memasukkan File . . . . .	104
6.8	Tampilan Browse Output . . . . .	104
6.9	Tampilan Setelah Memasukkan File Hasil . . . . .	105
6.10	Lokasi Hasil . . . . .	105
6.11	Hasil Clustering . . . . .	106
6.12	<i>Use Case</i> Deposito Berjangka . . . . .	107
6.13	Tampilan Awal . . . . .	108
6.14	Tampilan Awal Isi . . . . .	109
6.15	Tampilan Prediksi Negatif . . . . .	109
6.16	Tampilan Prediksi Positif . . . . .	110
6.17	Tampilan Browse input . . . . .	110
6.18	Tampilan Setelah Memasukkan File . . . . .	111
6.19	Tampilan Browse Output . . . . .	111
6.20	Tampilan Setelah Memasukkan File Hasil . . . . .	112
6.21	Lokasi Hasil . . . . .	112

## DAFTAR TABEL

2.1	Tabel Karakteristik Data yang Berkualitas . . . . .	15
2.2	<i>Confusion Matrix</i> . . . . .	17
2.3	Tabel parameter algoritma <i>decision tree</i> dalam <i>scikit-learn</i> . . . . .	18
2.4	Tabel parameter algoritma <i>random forest</i> dalam <i>scikit-learn</i> . . . . .	20
2.5	Tabel parameter algoritma <i>K-Nearest Neighbors</i> (K-NN) dalam <i>scikit-learn</i> . . . . .	22
2.6	Tabel parameter algoritma <i>Naive Bayes Classifier</i> (NBC) dalam <i>scikit-learn</i> . . . . .	23
2.7	Tabel parameter algoritma <i>k-means</i> dalam <i>library scikit-learn</i> . . . . .	26
2.8	Tabel contoh variabel kualitatif dan variabel kuantitatif . . . . .	29
3.1	Tabel deskripsi variabel <i>dataset bank-full.csv</i> . . . . .	38
3.2	Tabel hasil <i>info gain dataset</i> deposito berjangka . . . . .	53
3.3	Tabel pemilihan fitur kategorikal untuk <i>dataset</i> deposito berjangka . . . . .	53
3.4	Tabel pemilihan fitur numerik pada eksplorasi <i>dataset</i> deposito berjangka . . . . .	54
3.5	Tabel hasil prediksi pada eksplorasi <i>dataset</i> deposito berjangka dengan <i>decision tree</i> . . . . .	55
3.6	Tabel hasil metrik evaluasi pada eksplorasi <i>dataset</i> deposito berjangka dengan <i>decision tree</i> . . . . .	55
3.7	Tabel hasil prediksi pada eksplorasi <i>dataset</i> deposito berjangka dengan <i>random forest</i> . . . . .	56
3.8	Tabel hasil metrik evaluasi pada eksplorasi <i>dataset</i> deposito berjangka dengan <i>random forest</i> . . . . .	56
3.9	Tabel hasil prediksi pada eksplorasi <i>dataset</i> deposito berjangka dengan K-NN . . . . .	56
3.10	Tabel hasil metrik evaluasi pada eksplorasi <i>dataset</i> deposito berjangka dengan K-NN . . . . .	57
3.11	Tabel hasil prediksi pada eksplorasi <i>dataset</i> deposito berjangka dengan NBC . . . . .	57
3.12	Tabel hasil metrik evaluasi pada eksplorasi <i>dataset</i> deposito berjangka dengan NBC . . . . .	57
3.13	Tabel pemilihan fitur kategorikal pada eksplorasi <i>dataset</i> deposito berjangka . . . . .	57
3.14	Tabel pemilihan fitur numerik pada eksplorasi <i>dataset</i> deposito berjangka . . . . .	58
3.15	Tabel hasil prediksi pada eksplorasi <i>dataset</i> deposito berjangka dengan <i>decision tree</i> dan fitur yang terpilih . . . . .	58
3.16	Tabel hasil metrik evaluasi pada eksplorasi <i>dataset</i> deposito berjangka dengan <i>decision tree</i> dan fitur yang terpilih . . . . .	59
3.17	Tabel hasil prediksi pada eksplorasi <i>dataset</i> deposito berjangka dengan <i>random forest</i> dan fitur yang terpilih . . . . .	59
3.18	Tabel hasil metrik evaluasi pada eksplorasi <i>dataset</i> deposito berjangka dengan <i>random forest</i> dan fitur yang terpilih . . . . .	59
3.19	Tabel hasil prediksi pada eksplorasi <i>dataset</i> deposito berjangka dengan K-NN dan fitur yang terpilih . . . . .	60
3.20	Tabel hasil metrik evaluasi pada eksplorasi <i>dataset</i> deposito berjangka dengan K-NN dan fitur yang terpilih . . . . .	60
3.21	Tabel hasil prediksi pada eksplorasi <i>dataset</i> deposito berjangka dengan NBC dan fitur yang terpilih . . . . .	60
3.22	Tabel hasil metrik evaluasi pada eksplorasi <i>dataset</i> deposito berjangka dengan NBC dan fitur yang terpilih . . . . .	60
3.23	Tabel deskripsi variabel <i>dataset Personal Loan Dataset</i> . . . . .	62

3.24	Tabel hasil <i>info gain dataset</i> pinjaman pribadi bank. . . . .	65
3.25	Tabel pemilihan fitur kategorikal untuk <i>dataset</i> pinjaman pribadi bank. . . . .	65
3.26	Tabel pemilihan fitur numerik untuk <i>dataset</i> pinjaman pribadi bank. . . . .	65
3.27	Tabel hasil prediksi pada eksplorasi <i>dataset</i> pinjaman pribadi bank dengan <i>decision tree</i> . . . . .	67
3.28	Tabel hasil metrik evaluasi pada eksplorasi <i>dataset</i> pinjaman pribadi bank dengan <i>decision tree</i> . . . . .	67
3.29	Tabel hasil prediksi pada eksplorasi <i>dataset</i> pinjaman pribadi bank dengan <i>random forest</i> . . . . .	68
3.30	Tabel hasil metrik evaluasi pada eksplorasi <i>dataset</i> pinjaman pribadi bank dengan <i>random forest</i> . . . . .	68
3.31	Tabel hasil prediksi pada eksplorasi <i>dataset</i> pinjaman pribadi bank dengan K-NN . . . . .	68
3.32	Tabel hasil metrik evaluasi pada eksplorasi <i>dataset</i> pinjaman pribadi bank dengan K-NN . . . . .	68
3.33	Tabel hasil prediksi pada eksplorasi <i>dataset</i> pinjaman pribadi bank dengan NBC . . . . .	69
3.34	Tabel hasil metrik evaluasi pada eksplorasi <i>dataset</i> pinjaman pribadi bank dengan NBC . . . . .	69
3.35	Tabel pemilihan fitur kategorikal pada eksplorasi <i>dataset</i> pinjaman pribadi bank . . . . .	69
3.36	Tabel pemilihan fitur numerik pada eksplorasi <i>dataset</i> pinjaman pribadi bank . . . . .	70
3.37	Tabel hasil prediksi pada eksplorasi <i>dataset</i> pinjaman pribadi bank dengan <i>decision tree</i> dengan fitur yang terpilih . . . . .	70
3.38	Tabel hasil metrik evaluasi pada eksplorasi <i>dataset</i> pinjaman pribadi bank dengan <i>decision tree</i> dengan fitur yang terpilih . . . . .	70
3.39	Tabel hasil prediksi pada eksplorasi <i>dataset</i> pinjaman pribadi bank dengan <i>random forest</i> dengan fitur yang terpilih . . . . .	71
3.40	Tabel hasil metrik evaluasi pada eksplorasi <i>dataset</i> pinjaman pribadi bank dengan <i>random forest</i> dengan fitur yang terpilih . . . . .	71
3.41	Tabel hasil prediksi pada eksplorasi <i>dataset</i> pinjaman pribadi bank dengan K-NN dengan fitur yang terpilih . . . . .	71
3.42	Tabel hasil metrik evaluasi pada eksplorasi <i>dataset</i> pinjaman pribadi bank dengan K-NN dengan fitur yang terpilih . . . . .	72
3.43	Tabel hasil prediksi pada eksplorasi <i>dataset</i> pinjaman pribadi bank dengan NBC dengan fitur yang terpilih . . . . .	72
3.44	Tabel hasil metrik evaluasi pada eksplorasi <i>dataset</i> pinjaman pribadi bank dengan NBC dengan fitur yang terpilih . . . . .	72
4.1	Tabel hasil prediksi untuk data <i>train</i> menggunakan 4 fitur. . . . .	73
4.2	Tabel hasil metrik evaluasi untuk data <i>train</i> menggunakan 4 fitur. . . . .	73
4.3	Tabel hasil prediksi untuk data <i>test</i> menggunakan 4 fitur. . . . .	73
4.4	Tabel hasil metrik evaluasi untuk data <i>test</i> menggunakan 4 fitur. . . . .	74
4.5	Tabel hasil prediksi untuk data <i>train</i> menggunakan 4 fitur dan <i>undersampling</i> . . . . .	74
4.6	Tabel hasil metrik evaluasi untuk data <i>train</i> menggunakan 4 fitur dan <i>undersampling</i> . . . . .	74
4.7	Tabel hasil prediksi untuk data <i>test</i> menggunakan 4 fitur dan <i>undersampling</i> . . . . .	74
4.8	Tabel hasil metrik evaluasi untuk data <i>test</i> menggunakan 4 fitur dan <i>undersampling</i> . . . . .	75
4.9	Tabel hasil prediksi untuk data <i>train</i> menggunakan 4 fitur. . . . .	75
4.10	Tabel hasil metrik evaluasi untuk data <i>train</i> menggunakan 4 fitur. . . . .	75
4.11	Tabel hasil prediksi untuk data <i>test</i> menggunakan 4 fitur. . . . .	75
4.12	Tabel hasil metrik evaluasi untuk data <i>test</i> menggunakan 4 fitur. . . . .	76
4.13	Tabel hasil prediksi untuk data <i>train</i> menggunakan 6 fitur. . . . .	76
4.14	Tabel hasil metrik evaluasi untuk data <i>train</i> menggunakan 6 fitur. . . . .	76
4.15	Tabel hasil prediksi untuk data <i>test</i> menggunakan 6 fitur. . . . .	76
4.16	Tabel hasil metrik evaluasi untuk data <i>test</i> menggunakan 6 fitur. . . . .	77

4.17	Tabel hasil prediksi untuk data <i>train</i> menggunakan 6 fitur. . . . .	77
4.18	Tabel hasil metrik evaluasi untuk data <i>train</i> menggunakan 6 fitur. . . . .	77
4.19	Tabel hasil prediksi untuk data <i>test</i> menggunakan 6 fitur. . . . .	77
4.20	Tabel hasil metrik evaluasi untuk data <i>test</i> menggunakan 6 fitur. . . . .	78
4.21	Tabel hasil prediksi untuk data <i>train</i> menggunakan 8 fitur. . . . .	78
4.22	Tabel hasil metrik evaluasi untuk data <i>train</i> menggunakan 8 fitur. . . . .	78
4.23	Tabel hasil prediksi untuk data <i>test</i> menggunakan 8 fitur. . . . .	78
4.24	Tabel hasil metrik evaluasi untuk data <i>test</i> menggunakan 8 fitur. . . . .	79
4.25	Tabel hasil prediksi untuk data <i>train</i> menggunakan 4 fitur. . . . .	79
4.26	Tabel hasil metrik evaluasi untuk data <i>train</i> menggunakan 4 fitur. . . . .	79
4.27	Tabel hasil prediksi untuk data <i>test</i> menggunakan 4 fitur. . . . .	79
4.28	Tabel hasil metrik evaluasi untuk data <i>test</i> menggunakan 4 fitur. . . . .	80
4.29	Tabel hasil prediksi perangkat lunak dengan <i>dataset</i> deposito berjangka . . . . .	80
4.30	Tabel hasil prediksi untuk data <i>train</i> menggunakan 4 fitur. . . . .	81
4.31	Tabel hasil metrik evaluasi untuk data <i>train</i> menggunakan 4 fitur. . . . .	81
4.32	Tabel hasil prediksi untuk data <i>test</i> menggunakan 4 fitur. . . . .	81
4.33	Tabel hasil metrik evaluasi untuk data <i>test</i> menggunakan 4 fitur. . . . .	81
4.34	Tabel hasil prediksi untuk data <i>train</i> menggunakan 4 fitur dan <i>undersampling</i> . . .	82
4.35	Tabel hasil metrik evaluasi untuk data <i>train</i> menggunakan 4 fitur dan <i>undersampling</i> . . .	82
4.36	Tabel hasil prediksi untuk data <i>test</i> menggunakan 4 fitur dan <i>undersampling</i> . . . .	82
4.37	Tabel hasil metrik evaluasi untuk data <i>test</i> menggunakan 4 fitur dan <i>undersampling</i> . . .	82
4.38	Tabel hasil prediksi untuk data <i>train</i> menggunakan 4 fitur. . . . .	83
4.39	Tabel hasil metrik evaluasi untuk data <i>train</i> menggunakan 4 fitur. . . . .	83
4.40	Tabel hasil prediksi untuk data <i>test</i> menggunakan 4 fitur. . . . .	83
4.41	Tabel hasil metrik evaluasi untuk data <i>test</i> menggunakan 4 fitur. . . . .	83
4.42	Tabel hasil prediksi untuk data <i>train</i> dengan 6 fitur. . . . .	84
4.43	Tabel hasil metrik evaluasi untuk data <i>train</i> dengan 6 fitur. . . . .	84
4.44	Tabel hasil prediksi untuk data <i>test</i> dengan 6 fitur. . . . .	84
4.45	Tabel hasil metrik evaluasi untuk data <i>test</i> dengan 6 fitur. . . . .	84
4.46	Tabel hasil prediksi untuk data <i>train</i> dengan menggunakan 8 fitur. . . . .	85
4.47	Tabel hasil metrik evaluasi untuk data <i>train</i> dengan menggunakan 8 fitur. . . . .	85
4.48	Tabel hasil prediksi untuk data <i>test</i> dengan menggunakan 8 fitur. . . . .	85
4.49	Tabel hasil metrik evaluasi untuk data <i>test</i> dengan menggunakan 8 fitur. . . . .	85
4.50	Tabel hasil prediksi perangkat lunak dengan <i>dataset</i> pinjaman pribadi bank . . . . .	86
5.1	Tabel <i>centroid</i> hasil pembuatan model <i>clustering</i> . . . . .	89
5.2	Tabel <i>centroid</i> hasil pembuatan model <i>clustering</i> . . . . .	97
5.3	Tabel nilai minimal model <i>clustering</i> . . . . .	97
5.4	Tabel nilai maksimal model <i>clustering</i> . . . . .	98
5.5	Tabel jumlah anggota setiap <i>cluster</i> . . . . .	98
B.1	Lampiran 100 baris pertama <i>dataset</i> deposito berjangka di Portugal. . . . .	159
B.2	Lampiran 100 baris pertama <i>dataset</i> pinjaman pribadi bank di Amerika Serikat. . .	161



## DAFTAR KODE PROGRAM

3.1	Potongan kode untuk melakukan EDA fitur kategorikal pada <i>dataset</i> deposito berjangka . . . . .	39
3.2	Potongan kode untuk melakukan EDA fitur kategorikal pada <i>dataset</i> deposito berjangka dengan menggunakan rasio . . . . .	44
3.3	Potongan kode untuk melakukan transformasi variabel kategorikal pada <i>dataset</i> deposito berjangka . . . . .	49
3.4	Potongan kode untuk mencari <i>info gain dataset</i> pinjaman pribadi . . . . .	52
3.5	Potongan kode pembersihan <i>dataset</i> deposito berjangka . . . . .	54
3.6	Potongan kode untuk <i>train/test split</i> pada <i>dataset</i> deposito berjangka . . . . .	55
3.7	Potongan kode untuk melakukan transformasi pada kolom <i>Experience</i> untuk <i>dataset</i> pinjaman pribadi bank . . . . .	63
3.8	Potongan kode untuk mencari <i>info gain dataset</i> pinjaman pribadi . . . . .	64
3.9	Potongan kode pembersihan <i>dataset</i> pinjaman pribadi bank . . . . .	66
3.10	Potongan kode untuk <i>train/test split</i> pada <i>dataset</i> deposito berjangka . . . . .	67
5.1	Potongan kode untuk melakukan penyiapan data untuk <i>clustering</i> . . . . .	88
5.2	Potongan kode untuk pembuatan model <i>clustering</i> . . . . .	89
A.1	percobaan_1_skripsi.py . . . . .	117
A.2	percobaan_2_skripsi(personal_loan_selling).py . . . . .	123
A.3	skripsi_eksplorasteknologi(1).py . . . . .	128
A.4	skripsi_eksplorasteknologi(2).py . . . . .	131
A.5	skripsi_analisisbab5noeda.py . . . . .	135
A.6	model-DepositoBerjangka.py . . . . .	143
A.7	ui-DepositoBerjangka.py . . . . .	145
A.8	ui-DepositoBerjangka-4Fitur . . . . .	147
A.9	model-PinjamPribadi.py . . . . .	148
A.10	ui-PinjamanPribadi.py . . . . .	150
A.11	skripsi_clusteringcoba1.py . . . . .	152
A.12	Skripsi-Clustering2(SemuaFitur).py . . . . .	153
A.13	Skripsi-Clustering2(siklus2).py . . . . .	154

# BAB 1

## PENDAHULUAN

Pada bab ini akan dibahas tentang latar belakang penelitian yang menyebabkan munculnya permasalahan yang akan diteliti. Dilanjutkan dengan pembahasan rumusan masalah yang akan menjelaskan inti permasalahan yang terjadi, tujuan penelitian yang membahas solusi dari masalah yang dihadapi, dan batasan penelitian yang membahas asumsi sebagai batas ruang lingkup dari penelitian. Bagian terakhir dari bab ini adalah metodologi yang berisi langkah-langkah penyelesaian masalah secara terperinci dan sistematika pembahasan yang berisi penjelasan setiap bab dalam skripsi ini secara singkat.

### 1.1 Latar Belakang

Kegiatan *marketing* atau pemasaran barang dan jasa sudah menjadi hal umum yang dilakukan oleh para pelaku usaha. *Marketing* bukan sekedar kegiatan menawarkan barang dan jasa untuk mencari keuntungan sebanyak-banyaknya, melainkan, kegiatan yang bertujuan untuk mencari pelanggan baru sebanyak-banyaknya dan mempertahankannya[1]. Meskipun *marketing* erat kaitannya dengan penjualan, ukuran kesuksesan *marketing* tidak hanya dilihat dari jumlah produk yang terjual. Ukuran kesuksesan *marketing* dapat dilihat pula dari bagaimana para pelaku usaha dapat terus menambah pelanggan yang baru dengan tetap menjaganya[2]. Agar hubungan baik dapat terjaga, para pelaku usaha perlu mengetahui apa yang diinginkan oleh pelanggan. Salah satu cara untuk mengetahui apa yang diinginkan pelanggan adalah dengan melakukan riset terhadap pasar tempat pelaku usaha melakukan *marketing*. Hasil dari riset bisa menjadi dasar bagi pelaku usaha dalam melakukan kegiatan *marketing* [3].

Kegiatan *marketing* saat ini dapat dilakukan secara langsung maupun tidak langsung[1]. *Direct marketing* atau pemasaran secara langsung adalah kegiatan penawaran barang atau jasa tanpa melalui perantara kepada calon pelanggan[4]. Tanpa melalui perantara di sini bukan berarti *marketing* tersebut tidak bisa menggunakan alat seperti telepon, melainkan *marketing* yang dilakukan memang murni untuk memasarkan barang dan jasa bukan untuk tujuan lain. Sedangkan *indirect marketing* atau pemasaran secara tidak langsung adalah kegiatan penawaran barang dan jasa dengan melalui perantara kepada pelanggan, dalam artian melalui jalur lain terlebih dahulu, misalnya melalui ulasan dari *vlogger*. Kegiatan pemasaran langsung dapat dilakukan dengan berbagai teknik, seperti *telemarketing*, *online marketing*, *social media marketing*, dan *e-mail marketing* [4]. Di antara beberapa teknik pemasaran langsung, *telemarketing* merupakan salah satu bentuk *marketing* yang paling tua dan masih paling populer dikalangan para pelaku usaha. Hal ini dikarenakan *telemarketing* memiliki metode yang sederhana dan termasuk mudah dilakukan[1].

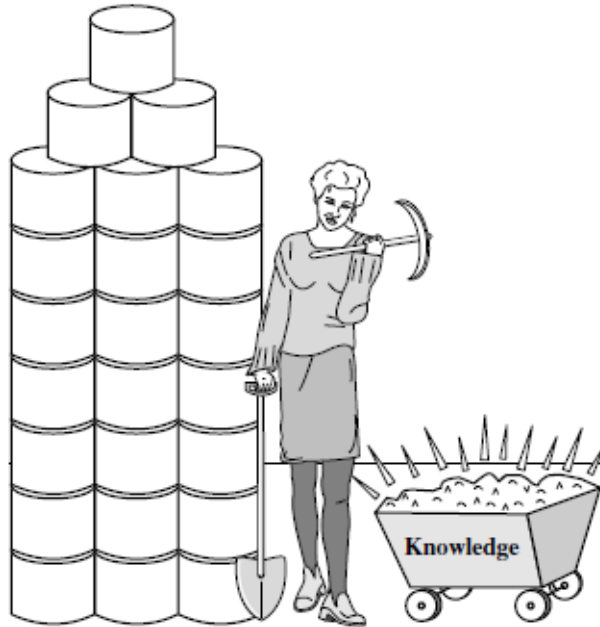
*Telemarketing* atau pemasaran melalui telepon, adalah teknik *marketing* yang menggunakan fasilitas telepon untuk menawarkan produknya[1]. Di industri perbankan, *telemarketing* merupakan salah satu teknik *marketing* yang paling sering dilakukan untuk beberapa alasan. Alasan pertama, *telemarketing* dinilai efektif dalam membantu personil bank untuk membangun hubungan yang baik dengan nasabah karena aktivitasnya yang personal dan dua arah. Nasabah cenderung akan membeli produk bank jika ada hubungan yang baik dan rasa percaya dengan personil bank[5].

Alasan lain yang membuat *telemarketing* menjadi teknik pemasaran yang populer adalah biaya.



Biaya untuk melakukan sambungan telepon dinilai jauh lebih murah jika dibandingkan dengan media lain seperti iklan[1]. Selain kedua alasan di atas, *telemarketing* juga dianggap ideal untuk menjangkau nasabah yang memiliki sedikit waktu senggang atau berada di lokasi yang jauh. Hal ini dibuktikan dengan adanya penerapan *telemarketing* yang cukup sering di Amerika Serikat oleh para pelaku usaha yang sampai memakan biaya 44 miliar *dollar* Amerika Serikat[1].

Meskipun *telemarketing* adalah salah satu teknik yang populer di industri perbankan, namun tidak semua nasabah mengapresiasi teknik ini. Tidak semua calon nasabah dapat menerima penawaran produk melalui *telemarketing*. *Telemarketing* hanya sesuai untuk beberapa golongan, seperti golongan lansia. Bagi golongan muda atau golongan milenial, teknik ini kurang sesuai karena seringkali golongan ini merasa privasinya terganggu apabila sering dihubungi oleh pihak bank melalui telepon[1]. Karena tidak ada satu teknik marketing universal yang cocok untuk semua golongan maka bank perlu mencari strategi baru yang sesuai dengan karakteristik nasabahnya. Untuk itu, pihak bank perlu memahami calon nasabahnya agar strategi *marketing* yang dilakukan pihak bank tepat sasaran.



Gambar 1.1: Ilustrasi Sederhana *Data Mining* [6]

*Data mining* dapat didefinisikan sebagai suatu proses untuk menemukan pola dan pengetahuan dari suatu data dalam jumlah banyak. Data untuk proses ini bisa berasal dari berbagai sumber, seperti basis data, gudang data, situs daring, atau data yang didapat dari penarikan secara langsung (melakukan pencarian data secara manual ke tempat kejadian perkara)[6]. Secara umum, *data mining* memiliki beberapa tahapan. Tahap pertama adalah proses penyiapan data. Proses penyiapan data terdiri atas pembersihan data, transformasi data, dan eksplorasi data. Di tahap kedua, data yang telah disiapkan akan dianalisis untuk mencari pola dari data. Tujuan dari tahap ini adalah untuk mendapatkan *insight* atau wawasan yang dapat diterapkan untuk menyelesaikan suatu masalah[6]. Terdapat 2 teknik untuk mencari pola dari suatu data, yaitu dengan menggunakan *classification* dan *clustering*. *Classification* merupakan bentuk analisis data untuk menghasilkan model yang dapat digunakan untuk mengelompokkan data ke dalam suatu kelas data, dimana data yang termasuk ke dalam suatu kelas akan memiliki label dari kelas tersebut[6]. Sedangkan *clustering* merupakan suatu bentuk data analisis dengan membuat suatu model yang dapat mengelompokkan data ke dalam suatu *cluster* atau kelompok berdasarkan kemiripan sifat. Tahapan ketiga adalah proses penyampaian *insights* yang dihasilkan melalui visualisasi data dan teknik presentasi. Visualisasi

dan presentasi diperlukan supaya *insight* yang disampaikan dapat lebih mudah dimengerti oleh orang yang memerlukan *insight* tersebut[6].

Penelitian mengenai topik menentukan strategi *marketing* dengan menggunakan *data mining* dapat dilakukan karena sudah ada penelitian yang menyinggung masalah ini. Salah satu contoh penelitian yang berhubungan dengan topik *marketing* dan *data mining* adalah penelitian mengenai penentuan strategi *marketing* yang tepat untuk bank dengan menggunakan *data mining* yang dilakukan oleh Sérgio Moro, Raul M. S. Laureano, dan Paulo Cortez. Pada penelitian tersebut, para peneliti menggunakan *data mining* untuk mempelajari dan mencari pola dari data bank. Setelah itu, peneliti memprediksi apakah strategi *direct marketing* dari bank tersebut dapat dikatakan berhasil atau tidak[7]. Di penelitian ini, pendekatan yang dilakukan adalah menggunakan *CRoss-Industry Standard Process for Data Mining* (CRISP-DM) untuk menentukan apakah *direct marketing* yang dilakukan oleh bank berhasil atau tidak[7]. CRISP-DM merupakan suatu model yang langkah kerja yang terstruktur dan sistematis dalam proses pengolahan data yang sistematis dan efektif[8]. CRISP-DM memiliki 6 tahap utama, yaitu pemahaman tentang tujuan proyek, melakukan eksplorasi terhadap data dan memahaminya, melakukan penyiapan data, menerapkan teknik analisis data, melakukan evaluasi, dan menerapkan model yang telah dibuat. Sedangkan pada penelitian yang akan dilakukan pada skripsi ini, pendekatan *data mining* yang dapat menjadi alternatif adalah penggunaan teknik *machine learning*, seperti *classification* dan *clustering* untuk mengetahui *direct marketing* yang dilakukan bank berhasil atau tidak dan menentukan strategi *marketing* yang baru apabila diperlukan.

## 1.2 Rumusan Masalah

Adapun rumusan masalah yang didapat dari latar belakang yang telah dipaparkan:

1. Apa saja faktor-faktor penentu keberhasilan bank dalam menawarkan produknya dengan menggunakan *data mining*?
2. Bagaimana cara mencari alternatif strategi *marketing* untuk bank diluar strategi yang telah dilakukan bank dengan *data mining*?
3. Bagaimana cara membuat perangkat lunak yang dapat memprediksi apakah seorang nasabah memiliki suatu produk bank atau tidak?
4. Bagaimana bentuk visualisasi yang tepat untuk menampilkan kelompok nasabah dengan karakteristik tertentu?

## 1.3 Tujuan

Adapun tujuan penelitian dari masalah yang telah dirumuskan:

1. Mencari faktor-faktor penentu keberhasilan bank dalam menawarkan produknya dengan menggunakan *data mining*.
2. Mencari alternatif strategi *marketing* untuk bank selain strategi yang telah diterapkan dengan *data mining*.
3. Membuat suatu perangkat lunak yang dapat memprediksi apakah seorang nasabah memiliki suatu produk bank atau tidak.
4. Memvisualisasikan kelompok nasabah dengan karakteristik tertentu.

## 1.4 Batasan Masalah

Adapun batasan dari penelitian yang dilakukan adalah sebagai berikut:

1. Penelitian ini menggunakan 2 *dataset*, yaitu *dataset marketing* deposito berjangka di Portugal dan *dataset marketing* pinjaman pribadi bank di Amerika Serikat.
2. Pada *dataset marketing* deposito berjangka di Portugal, akan dilakukan pembuatan model prediksi dengan klasifikasi dan segmentasi nasabah dengan *clustering*.

3. Pada *dataset marketing* pinjaman pribadi bank di Amerika Serikat, hanya akan dilakukan pembuatan model prediksi dengan klasifikasi saja karena pada *dataset* ini, tidak dijelaskan bentuk strategi *marketing* yang telah dilakukan sehingga tidak bisa untuk dicari strategi *marketing* bank alternatifnya.

## 1.5 Metodologi

Adapun metodologi yang digunakan dalam penelitian skripsi ini adalah sebagai berikut:

1. Mendefinisikan masalah.
2. Mempelajari dan melakukan studi literatur mengenai *marketing*, bank, *data mining*, dan statistika.
3. Mencari *dataset* bank yang sesuai.
4. Melakukan pembersihan *dataset* dengan menangani nilai *outlier* dan *null value*.
5. Melakukan transformasi terhadap *dataset* dan melakukan eksplorasi *dataset*.
6. Melakukan pemilihan fitur untuk pembuatan model klasifikasi.
7. Membuat model klasifikasi dengan fitur yang dipilih.
8. Melakukan evaluasi terhadap model, apabila hasil masih buruk, maka tahap 7 dan 8 akan diulang dengan menggunakan fitur yang berbeda hingga menghasilkan model yang baik.
9. Membuat perangkat lunak dengan model terbaik yang sudah dibuat.
10. Melakukan pemilihan fitur untuk pembuatan model *clustering*.
11. Membuat model *clustering* yang mengelompokkan nasabah (segmentasi nasabah).
12. Membuat suatu strategi *marketing* berdasarkan hasil *clustering*.
13. Membuat buku skripsi.

## 1.6 Sistematika Pembahasan

Adapun skripsi ini tersusun secara sistematis ke dalam 8 bab sebagai berikut:

1. Bab 1: Pendahuluan  
Pada bab ini membahas gambaran umum permasalahan yang ingin diselesaikan, rumusan masalah yang ingin diselesaikan beserta tujuan penelitian, mendefinisikan batasan penelitian, dan memaparkan tahapan-tahapan yang akan dilakukan dalam penelitian.
2. Bab 2: Landasan Teori  
Pada bab ini membahas hasil studi literatur yang dilakukan sebagai dasar teori yang digunakan untuk melakukan penelitian.
3. Bab 3: Penyiapan Data dan Eksplorasi Algoritma  
Pada bab ini membahas mengenai penjelasan *dataset* yang didapatkan, eksplorasi tahap awal pada *dataset*, pembersihan terhadap *dataset*, transformasi terhadap *dataset* pemilihan fitur untuk pembuatan model klasifikasi, dan percobaan terhadap algoritma-algoritma klasifikasi di mana algoritma-algoritma tersebut akan dibandingkan dan dipilih yang terbaik.
4. Bab 4: Analisis Data dan Hasil  
Pada bab ini membahas mengenai pembuatan model klasifikasi menggunakan algoritma terbaik yang sudah dipilih, pembuatan dilakukan beberapa kali dengan fitur yang dibedakan.
5. Bab 5: Segmentasi Nasabah dan Perancangan Strategi *Marketing* Bank  
Pada bab ini membahas pembuatan model *clustering* untuk melakukan segmentasi nasabah dan pembuatan strategi *marketing* bank berdasarkan hasil segmentasi.
6. Bab 6: Perancangan Perangkat Lunak dan Implementasinya  
Pada bab ini membahas mengenai perancangan dan implementasi perangkat lunak untuk memprediksikan nasabah berdasarkan model klasifikasi yang terbaik.
7. Bab 7: Kesimpulan dan Saran  
Pada bab ini membahas mengenai kesimpulan yang didapat dari penelitian yang telah dilakukan serta saran untuk penelitian selanjutnya.