

**PENERAPAN METODE SUPPORT VECTOR MACHINE  
UNTUK SELEKSI ATRIBUT BERBASIS PARTICLE  
SWARM OPTIMIZATION UNTUK PREDIKSI  
PEMASARAN LANGSUNG**



**TESIS**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar  
Magister Ilmu Komputer (M.Kom)**

**YUNI EKA ACHYANI**

**14001517**

**PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER  
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN  
KOMPUTER NUSA MANDIRI  
JAKARTA  
2016**

**PENERAPAN METODE SUPPORT VECTOR MACHINE  
UNTUK SELEKSI ATRIBUT BERBASIS PARTICLE  
SWARM OPTIMIZATION UNTUK PREDIKSI  
PEMASARAN LANGSUNG**



**TESIS**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar  
Magister Ilmu Komputer (M.Kom)**

**YUNI EKA ACHYANI**

**14001517**

**PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER  
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN  
KOMPUTER NUSA MANDIRI  
JAKARTA  
2016**

## SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini :

Nama : Yuni Eka Achyani  
NIM : 14001517  
Program Studi : Magister Ilmu Komputer  
Jenjang : Strata Dua (S2)  
Konsentrasi : *Management Information System*

Dengan ini menyatakan bahwa tesis yang telah saya buat dengan judul: “Penerapan Metode *Support Vector Machine* Untuk Seleksi Atribut Berbasis *Particle Swarm Optimization* Untuk Prediksi Pemasaran Langsung” adalah hasil karya sendiri, dan semua sumber baik yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar dan tesis belum pernah diterbitkan atau dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila dikemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa tesis yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri dicabut/dibatalkan.

Jakarta, 15 Agustus 2016  
Yang menyatakan,

Materai Rp. 6.000,-

Yuni Eka Achyani

## HALAMAN PENGESAHAN

Tesis ini diajukan oleh :

Nama : Yuni Eka Achyani  
NIM : 14001517  
Program Studi : Magister Ilmu Komputer  
Jenjang : Strata Dua (S2)  
Konsentrasi : *Management Information System*  
Judul Tesis : “Penerapan Metode *Support Vector Machine* Untuk Seleksi Atribut Berbasis *Particle Swarm Optimization* Untuk Prediksi Pemasaran Langsung”

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Magister Ilmu Komputer (M.Kom) pada Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri).

Jakarta, 25 Agustus 2016  
Pascasarjana Magister Ilmu Komputer  
STMIK Nusa Mandiri  
Direktur

Prof. Dr. Ir. R. Eko Indrajit, M.Sc., MBA., Mphil., MA

## DEWAN PENGUJI

Penguji I : Dr. Ing. Ir. Sukrisno Mardiyanto .....

Penguji II : Dr. Sularso Budilaksono, M.Kom .....

Penguji III / Pembimbing : Dr. Sfenrianto, M.Kom .....

|                                                                                   |                                                                         |
|-----------------------------------------------------------------------------------|-------------------------------------------------------------------------|
|  | LEMBAR KONSULTASI BIMBINGAN TESIS                                       |
|                                                                                   | <b>Pascasarjana Magister Ilmu Komputer</b><br><b>STMIK Nusa Mandiri</b> |

**NIM** : 14001517  
**Nama Lengkap** : Yuni Eka Achyani  
**Dosen Pembimbing** : Dr. Sfenrianto, M.Kom  
**Judul Tesis** : Penerapan Metode *Support Vector Machine*  
 Untuk Seleksi Atribut Berbasis *Particle Swarm*  
*Optimization* Untuk Prediksi Pemasaran Langsung

| No | Tanggal Bimbingan | Pokok Bahasan                      | Paraf Dosen Pembimbing I |
|----|-------------------|------------------------------------|--------------------------|
| 1  | 23/04/2016        | Bimbingan Perdana, Judul dan BAB I |                          |
| 2  | 07/05/2016        | Acc Judul dan BAB I dan BAB II     |                          |
| 3  | 21/05/2016        | Acc BAB II dan BAB III             |                          |
| 4  | 11/06/2016        | Acc BAB III dan BAB IV             |                          |
| 5  | 25/06/2016        | Acc BAB IV dan BAB V               |                          |
| 6  | 23/07/2016        | Acc BAB V                          |                          |
| 7  | 07/08/2016        | Acc Keseluruhan                    |                          |
| 8  | 02/09/2016        | Revisi Keseluruhan                 |                          |

Catatan untuk Dosen Pembimbing  
Bimbingan Tesis

- Dimulai pada tanggal : 23/04/2016
- Diakhiri pada tanggal : 02/09/2016
- Jumlah pertemuan bimbingan : Delapan kali Pertemuan

Disetujui oleh,  
Dosen Pembimbing

(Dr. Sfenrianto, M.Kom)

## KATA PENGANTAR

Puji syukur alhamdulillah, penulis panjatkan kehadiran Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan tesis ini dengan baik. Di mana tesis ini penulis sajikan dalam bentuk buku yang sederhana. Adapun judul penulisan tesis, yang penulis ambil adalah sebagai berikut: “Penerapan Metode *Support Vector Machine* Untuk Seleksi Atribut Berbasis *Particle Swarm Optimization* Untuk Prediksi Pemasaran Langsung”.

Tujuan penulisan tesis ini dibuat sebagai salah satu syarat untuk memperoleh gelar Magister Ilmu Komputer (M.Kom) pada Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri).

Penulis melakukan pencarian dan menganalisa berbagai macam sumber referensi, baik dalam bentuk jurnal ilmiah, buku-buku literatur, internet dan lain-lain yang terkait dengan pembahasan tesis ini.

Penulis menyadari bahwa tanpa bimbingan dan dukungan dari semua pihak dalam pembuatan tesis ini, maka penulis tidak dapat menyelesaikan tesis ini tepat pada waktunya. Untuk itu izinkanlah penulis dalam kesempatan ini untuk mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT yang selalu mencurahkan nikmat dan rahmatNya pada saya, sehingga saya dapat menyelesaikan tesis ini tepat pada waktunya.
2. Bapak Dr. Sfenrianto, M.Kom selaku pembimbing tesis yang telah menyediakan waktu, pikiran dan tenaga dalam membimbing penulis dalam menyelesaikan tesis ini.
3. Seluruh karyawan dan staf pengajar Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri, Jakarta yang telah memberikan pelajaran yang berarti bagi penulis selama menempuh studi.
4. Orang tua tercinta dan suami serta seluruh keluarga yang telah memberikan dukungan moril dan materil kepada penulis.

5. Rekan–rekan mahasiswa Program Pascasarjana Magister Ilmu Komputer, STMIK Nusa Mandiri Jakarta, atas dukungan, motivasi serta bantuan yang telah diberikan.

Serta semua pihak yang tidak dapat penulis sebutkan satu persatu, sehingga atas bantuan semua pihak tersebut penulis dapat menyelesaikan penulisan tesis ini. Penulis menyadari bahwa dalam penulisan tesis ini masih terdapat banyak kekurangan dan kekeliruan, untuk itu penulis mengharapkan kritik dan saran membangun agar penulisan karya ilmiah yang penulis hasilkan di masa yang akan datang menjadi semakin lebih baik.

Akhir kata semoga tesis ini dapat bermanfaat bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Jakarta, 15 Agustus 2016

Yuni Eka Achyani

## SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini, saya :

Nama : Yuni Eka Achyani  
NIM : 14001517  
Program Studi : Magister Ilmu Komputer  
Jenjang : Strata Dua (S2)  
Konsentrasi : *Management Information System*  
Jenis Karya : Tesis

Demi pengembangan ilmu pengetahuan, dengan ini menyetujui untuk memberikan ijin kepada pihak Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri) **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalti-Free Right*)** atas karya ilmiah kami yang berjudul :“Penerapan Metode *Support Vector Machine* Untuk Seleksi Atribut Berbasis *Particle Swarm Optimization* Untuk Prediksi Pemasaran Langsung”.

Dengan **Hak Bebas Royalti Non-Eksklusif** ini pihak STMIK Nusa Mandiri berhak menyimpan, mengalih-media atau *bentuk*-kan, mengelolanya dalam pangkalan data (*database*), mendistribusikannya dan menampilkan atau mempublikasikannya di *internet* atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari kami selama tetap mencantumkan nama kami sebagai penulis/pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak STMIK Nusa Mandiri, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 18 Agustus 2016  
Yang menyatakan,

Materai Rp. 6.000,-

Yuni Eka Achyani



## DAFTAR ISI

|                                                       | Halaman |
|-------------------------------------------------------|---------|
| HALAMAN SAMPUL .....                                  | i       |
| HALAMAN JUDUL .....                                   | ii      |
| SURAT PERNYATAAN ORISINALITAS .....                   | iii     |
| HALAMAN PENGESAHAN .....                              | iv      |
| LEMBAR KONSULTASI BIMBINGAN TESIS .....               | v       |
| KATA PENGANTAR .....                                  | iv      |
| SURAT PERNYATAAN PERSETUJUAN PUBLIKASI                |         |
| KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS .....         | viii    |
| ABSTRAK .....                                         | ix      |
| ABSTRACT .....                                        | x       |
| DAFTAR ISI .....                                      | xi      |
| DAFTAR TABEL .....                                    | xiii    |
| DAFTAR GAMBAR .....                                   | xiv     |
| DAFTAR LAMPIRAN .....                                 | xv      |
| BAB I PENDAHULUAN .....                               | 1       |
| 1.1. Latar Belakang Penulisan .....                   | 1       |
| 1.2. Masalah Penelitian .....                         | 4       |
| 1.2.1. Identifikasi Masalah .....                     | 4       |
| 1.2.2. Ruang Lingkup Penelitian .....                 | 5       |
| 1.2.3. Rumusan Masalah .....                          | 5       |
| 1.3. Tujuan dan Manfaat Penelitian .....              | 5       |
| 1.3.1. Tujuan Penelitian .....                        | 5       |
| 1.3.2. Manfaat Penelitian .....                       | 5       |
| 1.4. Sistematika Penulisan .....                      | 6       |
| BAB II LANDASAN /KERANGKA PEMIKIRAN .....             | 7       |
| 2.1. Tinjauan Pustaka .....                           | 7       |
| 2.1.1. Pemasaran Langsung .....                       | 7       |
| 2.1.2. <i>Data Mining</i> .....                       | 8       |
| 2.1.3. <i>Support Vector Machine</i> .....            | 12      |
| 2.1.4. Seleksi Atribut .....                          | 17      |
| 2.1.5. <i>Particle Swarm Optimization</i> .....       | 18      |
| 2.1.6. Pengujian K-Fold <i>Cross validation</i> ..... | 21      |
| 2.1.7. Evaluasi dan Validasi Hasil .....              | 22      |
| 2.1.7.1. <i>Confusion Matrix</i> .....                | 22      |
| 2.1.7.2. Kurva ROC .....                              | 23      |
| 2.2. Tinjauan Studi .....                             | 25      |
| 2.3. Tinjauan Objek Penelitian .....                  | 27      |
| 2.4. Kerangka Konsep Penelitian .....                 | 28      |
| 2.4.1. Kerangka Pemikiran .....                       | 28      |
| BAB III METODE PENELITIAN .....                       | 30      |
| 3.1. Desain Penelitian .....                          | 30      |
| 3.2. Pengumpulan Data .....                           | 31      |
| 3.3. Pengolahan Awal Data .....                       | 34      |
| 3.4. Model yang diusulkan .....                       | 41      |
| 3.5. Eksperimen dan Pengujian Metode .....            | 42      |

|                                                                                                               |    |
|---------------------------------------------------------------------------------------------------------------|----|
| 3.6. Evaluasi dan Validasi .                                                                                  | 43 |
| BAB IV HASIL PENELITIAN DAN PEMBAHASAN                                                                        | 45 |
| 4.1. Hasil Eksperimen dan Pengujian Metode.....                                                               | 45 |
| 4.1.1. Metode <i>Support Vector Machine</i> .....                                                             | 45 |
| 4.1.2. Metode <i>Support Vector Machine</i> berbasis <i>Particle Swarm Optimization</i> .....                 | 46 |
| 4.2. Evaluasi dan Validasi Hasil .....                                                                        | 48 |
| 4.2.1. Hasil Pengujian Metode <i>Support Vector Machine</i> .....                                             | 49 |
| 4.2.2. Hasil Pengujian Metode <i>Support Vector Machine</i> berbasis <i>Particle Swarm Optimization</i> ..... | 51 |
| 4.2.3. Analisa Evaluasi dan Validasi Model .....                                                              | 53 |
| 4.3 Pembahasan.....                                                                                           | 54 |
| 4.4. Implikasi Penelitian.....                                                                                | 55 |
| BAB V PENUTUP.....                                                                                            | 57 |
| 5.1 Kesimpulan.....                                                                                           | 57 |
| 5.2 Saran .....                                                                                               | 57 |
| DAFTAR REFERENSI .....                                                                                        | 58 |

## DAFTAR TABEL

|                                                                                                                            | Halaman |
|----------------------------------------------------------------------------------------------------------------------------|---------|
| Tabel 2.1. AND Problem .....                                                                                               | 16      |
| Tabel 2.2. <i>Confusion Matrix</i> .....                                                                                   | 22      |
| Tabel 2.3. Rangkuman Penelitian Terkait.....                                                                               | 26      |
| Tabel 2.4. Data Set Pemasaran Bank .....                                                                                   | 28      |
| Tabel 3.1. Sampel Data Set Pemasaran Bank .....                                                                            | 33      |
| Tabel 3.2. Ilustrasi <i>missing data</i> pada <i>data training</i> .....                                                   | 35      |
| Tabel 3.3. Data <i>Training</i> setelah dilakukan teknik validasi.....                                                     | 37      |
| Tabel 3.4. Atribut yang digunakan.....                                                                                     | 38      |
| Tabel 3.5. Spesifikasi Perangkat Keras dan Perangkat Lunak.....                                                            | 26      |
| Tabel 4.1. Eksperimen penentuan nilai training cycle SVM .....                                                             | 46      |
| Tabel 4.2. Eksperimen penentuan nilai training cycle SVM berbasis PSO .....                                                | 47      |
| Tabel 4.3. Hasil Seleksi Atribut .....                                                                                     | 48      |
| Tabel 4.4. Nilai <i>Accuracy</i> , <i>sensitivity</i> , <i>spesificity</i> , PPV, dan NPV metode SVM                       | 50      |
| Tabel 4.5. Nilai <i>Accuracy</i> , <i>sensitivity</i> , <i>spesificity</i> , PPV, dan NPV metode SVM<br>berbasis PSO ..... | 52      |
| Tabel 4.6. Komparasi Nilai <i>Accuracy</i> dan AUC.....                                                                    | 53      |

## DAFTAR GAMBAR

|                                                                                                                                                | Halaman |
|------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| Gambar 2.1. Tahapan data mining .....                                                                                                          | 10      |
| Gambar 2.2. Konsep SVM untuk mencari <i>hyperlane</i> terbaik.....                                                                             | 14      |
| Gambar 2.3. Struktur dasar PSO .....                                                                                                           | 21      |
| Gambar 2.4. Ilustrasi 10-Fold <i>Cross validation</i> .....                                                                                    | 22      |
| Gambar 2.5. Grafik ROC ( <i>discrete</i> dan <i>continuous</i> ).....                                                                          | 24      |
| Gambar 2.6. Kerangka Pemikiran.....                                                                                                            | 29      |
| Gambar 3.1. Uji Data <i>Training</i> dengan RapidMiner .....                                                                                   | 36      |
| Gambar 3.2. Model Desain <i>Replace Missing</i> .....                                                                                          | 38      |
| Gambar 3.3. Hasil <i>Replace Missing</i> .....                                                                                                 | 38      |
| Gambar 3.1. Model yang diusulkan .....                                                                                                         | 42      |
| Gambar 4.1. Pengujian <i>K-Fold Cross Validation</i> algoritma <i>Support Vector Machine</i> .....                                             | 45      |
| Gambar 4.2. Pengujian <i>K-Fold Cross Validation</i> algoritma <i>Support Vector Machine</i> berbasis PSO .....                                | 47      |
| Gambar 4.3. Hasil pengujian <i>Confusion Matrix</i> untuk metode SVM.....                                                                      | 49      |
| Gambar 4.4. Kurva ROC dengan metode <i>Support Vector Machine</i> .....                                                                        | 50      |
| Gambar 4.5. Hasil pengujian <i>Confusion Matrix</i> untuk metode SVM berbasis <i>Particle swarm optimization</i> .....                         | 51      |
| Gambar 4.6. Kurva ROC dengan Metode <i>Support Vector Machine</i> berbasis <i>Particle swarm optimization</i> .....                            | 52      |
| Gambar 4.7. <i>Attribute Weight</i> dengan Metode <i>Support Vector Machine</i> berbasis <i>Particle Swarm Optimization</i> .....              | 53      |
| Gambar 4.8. Kurva ROC dengan <i>support vector machine</i> dan <i>support vector machine</i> berbasis <i>particle swarm optimization</i> ..... | 54      |

## DAFTAR LAMPIRAN

|                                                        | Halaman |
|--------------------------------------------------------|---------|
| Lampiran A : Bank Portugis Data (UCI Repository) ..... | 59      |

## ABSTRAK

Nama : Yuni Eka Achyani  
NIM : 14001517  
Program Studi : Ilmu Komputer  
Jenjang : Strata Dua (S2)  
Konsentrasi : *Management Information System*  
Judul : “Penerapan Metode *Support Vector Machine* Untuk Seleksi Atribut Berbasis *Particle Swarm Optimization* Untuk Prediksi Pemasaran Langsung”

Pemasaran langsung merupakan strategi yang khas untuk meningkatkan bisnis. Perusahaan menggunakan pemasaran langsung bila menargetkan segmen pelanggan dengan menghubungi mereka untuk memenuhi tujuan tertentu. Pelaksanaan pemasaran langsung dari waktu ke waktu menghasilkan data dan informasi dalam bentuk laporan yang perlu dianalisis oleh manager dalam rangka mendukung keputusan. Namun, itu adalah tugas yang sulit bagi manusia untuk menganalisis data yang kompleks yang luas. Kesulitan ini menyebabkan perkembangan teknik intelegen bisnis, yang bertujuan mengekstraksi pengetahuan yang berguna untuk mendukung pengambilan keputusan. Peningkatan akurasi prediksi pemasaran langsung dapat dilakukan dengan cara melakukan seleksi terhadap atribut, karena seleksi atribut mengurangi dimensi dari data sehingga operasi algoritma data mining dapat berjalan lebih efektif dan lebih cepat. Salah satu metode yang paling banyak digunakan adalah metode *support vector machine*. Dalam penelitian ini akan digunakan metode *support vector machine* dan akan dilakukan seleksi atribut dengan menggunakan *particle swarm optimization* untuk prediksi pemasaran langsung. Setelah dilakukan pengujian maka hasil yang didapat adalah *support vector machine* menghasilkan nilai akurasi sebesar 88,71 %, nilai *precision* 89,47% dan nilai AUC sebesar 0,896. Kemudian dilakukan seleksi atribut dengan menggunakan *particle swarm optimization* dimana atribut yang semula berjumlah 16 variabel prediktor terpilih 12 atribut yang digunakan. Hasil menunjukkan nilai akurasi yang lebih tinggi yaitu sebesar 89,38%, nilai *precision* 89,89% dan nilai AUC sebesar 0,909 dengan nilai akurasi klasifikasi sangat baik (*excellent clasification*). Sehingga dicapai peningkatan akurasi sebesar 0,67 %, dan peningkatan AUC sebesar 0,013.

**Kata kunci :** Pemasaran Langsung, Seleksi Atribut, *Support Vector Machine*, *particle swarm optimization*.

## ***ABSTRACT***

Name : Yuni Eka Achyani  
NIM : 14001517  
Study of Program : Magister Ilmu Komputer  
Levels : Strata Dua (S2)  
Concentration : *Management Information System*  
Tittle : “Penerapan Metode *Support Vector Machine* Untuk Seleksi Atribut Berbasis *Particle Swarm Optimization* Untuk Prediksi Pemasaran Langsung”

*Direct marketing is a typical strategy to increase business. The company uses direct marketing segments when targeting customers by contacting them to meet certain goals. Implementation of direct marketing from time to time produce data and information in the form of reports that need to be analyzed by a manager in order to support the decision. However, it is a difficult task for humans to analyze complex data comprehensive. This difficulty led to the development of business intelligence techniques, aimed at extracting useful knowledge to support decision making. Improved prediction accuracy of direct marketing can be done by doing the selection of attributes, because the selection of attributes reduce the dimensionality of the data so that operation of the data mining algorithms can be run more effectively and more quickly. One of the methods most widely used method of support vector machine. In this study will be used method of support vector machine and will do the selection of attributes by using particle swarm optimization for prediction of direct marketing. After testing the results obtained is a support vector machine produces an accuracy value of 88.71%, 89.47% and a precision value AUC value of 0.896. Then the selection of attributes by using particle swarm optimization attributes which originally numbered 16 predictor variables selected 12 attributes used. The results showed higher accuracy value that is equal to 89.38%, 89.89% and a precision value AUC value of 0.909 to the value of classification accuracy was very good (excellent clasification). So as to achieve the increased accuracy of 0.67%, and an increase in AUC of 0.013.*

**Keywords :** *Direct Marketing, Selection Attributes, Support Vector Machine, particle swarm optimization.*

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang Penulisan**

Dalam menghadapi persaingan yang ketat promosi yang baik memberikan kredibilitas untuk sebuah produk yang baru. Promosi perlu mendapat perhatian yang lebih dan serius, sebab dalam kehidupan sehari-hari timbul suatu kenyataan bahwa betapapun besar keunggulan suatu produk, tetapi jika konsumen tidak mengetahuinya, kemungkinan produk yang ditawarkan kepada konsumen kurang direspon oleh pasar. Oleh karena itu perusahaan harus berusaha untuk memperkenalkan produk, meyakinkan, dan mempengaruhi konsumen agar dapat menciptakan permintaan terhadap produk tersebut langkah yang dapat dilakukan oleh perusahaan untuk melakukan hal tersebut adalah dengan promosi.

Perusahaan bisa melakukan promosi dengan iklan di media cetak seperti koran dan majalah, maupun media elektronik seperti televisi, internet dan radio hal ini menunjukkan komitmen perusahaan tersebut untuk menyampaikan pesan kepada masyarakat melalui salah satu bauran pemasaran yaitu promosi. Promosi merupakan salah satu variabel dalam bauran pemasaran yang sangat penting dilaksanakan oleh perusahaan dalam memasarkan produk atau jasa. Kegiatan promosi bukan saja berfungsi sebagai alat komunikasi antara perusahaan dengan konsumen, melainkan juga sebagai alat untuk mempengaruhi konsumen dalam kegiatan pembelian atau penggunaan jasa sesuai dengan keinginan dan kebutuhannya. Salah satu kegiatan promosi adalah melalui pemasaran langsung.

Pemasaran Langsung adalah bentuk dari pemasaran interaktif dengan menggunakan suatu media iklan atau lebih untuk mendapatkan respon atau transaksi yang dapat diukur. Pelaksanaan pemasaran langsung dari waktu ke waktu menghasilkan data dan informasi dalam bentuk laporan yang perlu dianalisis oleh manager dalam rangka mendukung keputusan. Namun, itu adalah tugas yang sulit bagi manusia untuk menganalisis data yang kompleks yang luas (Turban, 2012). Kesulitan ini menyebabkan perkembangan teknik intelegen bisnis, yang bertujuan mengekstraksi pengetahuan yang berguna untuk mendukung pengambilan keputusan.



Pemasaran langsung adalah proses mengidentifikasi potensi pembeli produk tertentu dan mempromosikan produk dengan sesuai (Sing'oei, 2013). Salah satu peran *direct marketing* adalah untuk membangun hubungan dengan konsumen, berkomunikasi secara langsung dengan konsumen. Awalnya *direct mail* menjadi alat utama dalam proses promosi, tapi seiring dengan perkembangan teknologi, database telah diperkenalkan sebagai media lain yang dapat digunakan untuk berkomunikasi secara efektif dengan konsumen (individu). Dalam pemasaran langsung, menggunakan catatan respon pelanggan yang tersedia saat ini, dapat memperkirakan jumlah tanggapan atau tingkat respon secara keseluruhan, dan penggunaan informasi tersebut dalam membuat keputusan manajerial.

Salah satu cara yang efektif untuk menganalisa laporan dari kampanye sebelumnya dan serupa dalam mencari tren dan pola adalah melalui intelegen bisnis dan teknik data mining, untuk membangun model dan kemudian ekstrak pengetahuan (Written, 2011). Intelegen bisnis adalah sebuah konsep luas yang mencakup data mining yang terdiri dalam ekstraksi pengetahuan dari data mentah. Selain itu, marketing memiliki sedikit pengetahuan tentang data mining, maka ada kebutuhan untuk mengembangkan suatu kerangka kerja yang disederhanakan untuk membantu marketing dalam memanfaatkan metode data mining untuk pemasaran langsung. Beberapa studi yang dilakukan untuk memprediksi pemasaran dengan metode komputasi antara lain: *support vector machine* (SVM) (Moro & Laureano, 2012) dan *multi layer perceptron* (MLP) (Elsalamony & Elsayad, 2013).

*Multi layer perceptron* (MLP) diketahui berfungsi untuk memprediksi dan mengklasifikasi sebuah masalah yang rumit, yang memungkinkan pengakuan terhadap data yang besar (Elsalamony & Elsayad, 2013). Tetapi kinerja MLP tergantung pada parameter, bobot dan fungsi pengalihan, banyak variable dan *overfitting* (Kahrizi & Hashemi, 2014). *Support vector machine* (SVM) bekerja lebih baik daripada MLP, dengan standar pembelajaran backpropagation, kinerja SVM lebih unggul, hal ini disebabkan karena kemampuan generalisasi support vector machine berdimensi tinggi dalam ruang high dimensional (Martinez, Sanchez, & Velez, 2010). *Support vector machine* digunakan untuk klasifikasi

pola, pemetaan dalam ruang input dengan nonlinear merubah ke ruang berdimensi tinggi, di mana masalah linear klasifikasi menjadi optimal (Ren, 2012).

Klasifikasi merupakan topik yang penting dalam penelitian *data mining*. Dimana terdiri dari sekumpulan data yang masing-masing sudah dikelompokkan kedalam kelas tertentu, masalah klasifikasi yang diperhatikan yaitu dengan penentuan aturan yang memungkinkan adanya klasifikasi walaupun data yang ada belum menunjukkan hal tersebut. Banyak algoritma yang dibangun untuk mengolah dataset yang jumlahnya besar khususnya model klasifikasi dan menghasilkan hasil yang efektif. Algoritma *data mining* seperti pohon keputusan berdasar algoritma *SVM*, *BOAT*, *C4.5*, *PUBLIC*, *Rain-Forest*, *SLIQ*, *SPRINT* dapat digunakan untuk klasifikasi. Walaupun demikian ketika sebuah pohon keputusan yang berbasis algoritma dikembangkan untuk mendeterminasi sekelompok kemungkinan yang diasosiasikan dari beberapa klasifikasi. Ini mungkin terdapat beberapa yang keluar dari pohon keputusan yang mempunyai kesamaan kelas probabilitas.

Dewasa ini SVM telah berhasil diaplikasikan dalam problema dunia nyata (real-world problems), dan secara umum memberikan solusi yang lebih baik dibandingkan metode konvensional seperti misalnya artificial neural network. SVM dapat mengatasi masalah klasifikasi dan regresi dengan linier ataupun nonlinier kernel yang dapat menjadi satu kemampuan algoritma pembelajaran untuk klasifikasi serta regresi. SVM juga memiliki akurasi tinggi dan tingkat kesalahan yang relative kecil, kemampuan untuk mengatasi overfitting tidak membutuhkan data yang terlalu besar dan dapat digunakan untuk melakukan prediksi.

Kinerja SVM sangat tergantung pada pilihan yang memadai dari nilai-nilai parameter, termasuk misalnya, kernel dan parameter regularisasi. Pemilihan parameter SVM umumnya sebagai masalah optimasi di mana teknik pencarian digunakan untuk menemukan konfigurasi parameter yang memaksimalkan kinerja SVM (Rossi & Soares, 2012). Ada banyak teknik optimasi yang telah digunakan untuk mengoptimasi parameter pada machine learning, seperti *algoritma genetika* (GA) (Ilhan & Tezel, 2013) dan *particle swarm optimization* (PSO) (Khoshahval, Minuchehr, & Zolfaghari, 2011).

Salah satu algoritma optimasi yang cukup populer adalah PSO (*Particle Swarm Optimization*). *Particle Swarm Optimization* merupakan teknik komputasi evolusioner yang mampu menghasilkan solusi secara global optimal dalam ruang pencarian melalui interaksi individu dalam segerombolan partikel. Setiap partikel menyampaikan informasi berupa posisi terbaiknya kepada partikel yang lain dan menyesuaikan posisi dan kecepatan masing-masing berdasarkan informasi yang diterima mengenai posisi yang terbaik tersebut.

Karena konsep sederhana, implementasi mudah, dan konvergensi cepat, *particle swarm optimization* (PSO) dapat diterapkan untuk berbagai aplikasi di berbagai bidang untuk memecahkan masalah optimasi (Liu, Tian, Chen, & Li, 2013). Implementasi PSO sangat mudah dan hanya di butuhkan beberapa baris kode pemrograman, dan juga tidak dibutuhkannya operator matematika yang rumit, maka dari itu dapat mengefesienkan perhitungan baik dari segi memori yang di butuhkan dan juga dari segi kecepatan. PSO, sebagai alat optimasi, yang dapat membantu menentukan parameter optimum. Tetapi PSO memiliki ketergantungan yang sensitif pada parameter yang digunakan (Yusup, N., Zain, A. M. and Hashim, S. Z. M, 2012).

Dari uraian diatas, pada penelitian ini akan digunakan klasifikasi *Support Vector Machine* (SVM) untuk membentuk model prediksi pemasaran langsung, untuk membantu mengolah data numerik serta memilih atribut yang baik, algoritma *Particle Swarm Optimization* (PSO) akan diterapkan untuk seleksi atribut dalam penentuan model prediksi pemasaran langsung sehingga akan diperoleh peningkatan akurasi.

## **1.2. Masalah Penelitian**

### **1.2.1. Identifikasi Masalah**

Berdasarkan latar belakang yang telah diuraikan, maka dapat diketahui bahwa *Support Vector Machine* (SVM) memiliki kemampuan generalisasi yang sangat baik untuk memecahkan masalah walaupun dengan sampel yang terbatas. Akan tetapi, jumlah atribut yang besar dalam data akan mengurangi akurasi dan menambah kompleksitas dari algoritma *Support Vector Machine* (SVM) tersebut. Untuk itu diperlukan seleksi terhadap atribut yang ada.

### 1.2.2. Ruang Lingkup Penelitian

Ruang lingkup pembahasan dalam penelitian ini dibatasi pada penerapan algoritma *Support Vector Machine* (SVM) dengan *Particle Swarm Optimization* (PSO) yang digunakan untuk seleksi atribut dalam penentuan model pemasaran langsung. Data yang digunakan adalah data kampanye pemasaran langsung (panggilan telepon) dari lembaga perbankan Portugis pada tahun 2008 sampai 2010 yang bersumber dari UCI machine learning repository. Sedangkan *software* yang digunakan adalah Rapid Miner dimana *software* tersebut memiliki sistem yang komprehensif untuk analisa data serta banyak digunakan karena kemampuan, fleksibilitas dan kemudahan dalam penggunaannya.

### 1.2.3. Rumusan Masalah

Berdasarkan identifikasi masalah dan ruang lingkup penelitian, maka rumusan masalah yang diangkat pada penelitian ini adalah seberapa besar akurasi metode *Support Vector Machine* (SVM) yang ditingkatkan dengan *Particle Swarm Optimization* (PSO) dengan cara melakukan seleksi atribut untuk prediksi model pemasaran langsung?

## 1.3. Tujuan dan Manfaat Penelitian

### 1.3.1. Tujuan Penelitian

Tujuan dari penelitian ini adalah menerapkan *Particle Swarm Optimization* (PSO) dengan melakukan seleksi atribut pada *Support Vector Machine* (SVM) untuk meningkatkan akurasi prediksi dalam memprediksi model pemasaran langsung.

### 1.3.2. Manfaat Penelitian

Manfaat dari penelitian ini adalah:

- a. Manfaat praktis dari hasil penelitian ini adalah dimaksudkan agar dapat digunakan perusahaan sebagai alat bantu pendukung pengambilan keputusan dalam prediksi pelaksanaan pemasaran langsung secara akurat.
- b. Manfaat Teoritis dari penelitian ini yaitu diharapkan dapat memberikan sumbangan pengembangan teori yang berkaitan dengan penerapan *Particle Swarm Optimization* (PSO) pada *Support Vector Machine* (SVM) untuk meningkatkan akurasi prediksi pemasaran langsung.

#### **1.4. Sistematika Penulisan**

Dalam penulisan tesis ini terdapat beberapa bab yang dibahas, diantaranya adalah sebagai berikut:

##### **Bab I      Pendahuluan**

Bab ini berisi tentang latar belakang permasalahan, identifikasi masalah, tujuan penelitian, manfaat penelitian, ruang lingkup penelitian, dan sistematika penulisan.

##### **Bab II     Landasan Teori**

Bab ini membahas tentang landasan teori yang melandasi penelitian, yang mencakup tinjauan studi dan tinjauan pustaka serta teori-teori lain yang mendukung penelitian ini.

##### **Bab III    Metode Penelitian**

Bab ini berisi tentang metode penelitian yang membahas tentang perancangan metode penelitian.

##### **Bab IV    Hasil dan Pembahasan**

Bab ini berisi tentang hasil dan pembahasan yang menguraikan tentang implementasi, pengukuran serta implikasi penelitian.

##### **Bab V     Penutup**

Bab ini berisi kesimpulan dari hasil penelitian dan saran-saran yang dibutuhkan untuk pengembangan sistem lebih lanjut.

## **BAB II**

### **LANDASAN TEORI/KERANGKA PEMIKIRAN**

#### **2.1. Tinjauan Pustaka**

Dalam penulisan tesis ini, digunakan beberapa media diantaranya buku, jurnal baik jurnal nasional maupun internasional, prosiding serta artikel yang didapatkan internet sebagai referensi untuk menjelaskan tentang penerapan algoritma *Particle Swarm Optimization* (PSO) untuk seleksi atribut pada *Support Vector Machine* (SVM) untuk meningkatkan prediksi pemasaran langsung.

##### **2.1.1. Pemasaran Langsung**

Pemasaran produk merupakan salah satu kunci untuk memaksimalkan tujuan perusahaan dalam menghasilkan laba. Tanpa adanya pemasaran maka usaha/bisnis yang kita jalankan tidak akan berkembang pesat. Pemasaran setiap perusahaan pun berbeda-beda. Ada yang memakai sistem tradisional yang mana yang dilakukan adalah promosi dari mulut ke mulut, dalam artian bahwa pemasarannya tidak efisien dan strategik. Pemasaran Langsung merupakan salah satu kegiatan promosi yang jika dilaksanakan secara efektif dapat mempengaruhi keputusan konsumen untuk membeli produk atau menggunakan jasa yang ditawarkan.

Pemasaran langsung adalah bentuk pemasaran yang dikirim secara langsung kepada konsumen yang di desain sedemikian rupa untuk menghasilkan respon sesegera mungkin. Respon ini seringkali dalam bentuk menciptakan penjualan secara langsung, tetapi juga respon bisa berupa komitmen untuk menghadiri sebuah acara peluncuran atau untuk memperkenalkan informasi personal dalam rangka untuk mendapatkan tempat dalam sebuah kompetisi. *Direct marketing* merupakan proses interaktif, sesuatu yang melibatkan konsumen, hal tersebut lebih dari hanya sekedar mendapatkan informasi. (Chitty, Barker & Shimp, 2008:301).

Pemasaran langsung merupakan strategi yang khas untuk meningkatkan bisnis. Perusahaan menggunakan pemasaran langsung bila menargetkan segmen pelanggan dengan menghubungi mereka untuk memenuhi tujuan tertentu.

Sentralisasi interaksi pelanggan kontak center memudahkan manajemen dalam operasional pemasaran. Pusat tersebut memungkinkan berkomunikasi dengan pelanggan melalui berbagai saluran, telepon (*fixed-line* atau *mobile*) menjadi salah satu yang paling banyak digunakan (Brinson, Lee & Rountree, 2011).

Panggilan dapat dibagi menjadi masuk dan keluar, tergantung pada sisi hubungi (klien atau *contact center*), dengan setiap kasus tantangan yang berbeda (misalnya, panggilan keluar yang sering dianggap lebih mengganggu). Teknologi memungkinkan memikirkan kembali pemasaran dengan berfokus pada memaksimalkan nilai pelanggan yang bertahan lama melalui evaluasi informasi yang tersedia dan *customer metrics*, sehingga memungkinkan kita untuk membangun hubungan yang lebih panjang dan ketat sejalan dengan tuntutan bisnis (Rust, 2010).

Perlu ditekankan juga bahwa tugas memilih kumpulan klien terbaik, yaitu memilih klien yang lebih mungkin untuk berlangganan produk dianggap lebih baik untuk menjadi referensi (Talla Nobibon, Leus & Spieksma, 2011). Banyak perusahaan yang mengadopsi pemasaran langsung sebagai cara komunikasi langsung dengan konsumen yang sudah ditargetkan secara cermat. Konsumen tertentu yang dipilih sebagai target promosi dipilih untuk memperoleh tanggapan langsung (Liao, Chen & Hsieh, 2011).

### **2.1.2. Data Mining**

*Data Mining* adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi didalam database. *Data Mining* didefinisikan sebagai proses penemuan pola dalam data. Beberapa aplikasi *data mining* fokus pada prediksi, mereka meramalkan apa yang akan terjadi dalam situasi baru dari data yang menggambarkan apa yang terjadi di masa lalu (Witten, Frank, & Hall, 2011). Secara khusus, koleksi metode yang dikenal sebagai *data mining* menawarkan metodologi dan solusi teknis untuk mengatasi analisis data medis dan konstruksi prediksi model (Bellazzi & Zupand, 2008).

Menurut Zafra dan Ventura (2012) data mining adalah proses penggalian pengetahuan yang berguna dan informasi dari kumpulan data. Saat ini data mining telah digunakan di banyak domain aplikasi seperti industri biomedis, ritel dan

pemasaran, telekomunikasi, pertambangan web, audit komputer, industri keuangan, obat-obatan dan seterusnya.

Menurut Gorunescu (2011) data mining juga didefinisikan suatu tindakan melakukan ekstraksi untuk mendapatkan informasi penting yang sifatnya implisit dan sebelumnya tidak diketahui, dari suatu data.

Menurut Pandey dan Pal (2011) data mining merupakan sebuah teknik yang digunakan untuk beroperasi pada data yang memiliki volume besar untuk menemukan pola-pola yang tersembunyi dan mencari hubungan agar dapat digunakan untuk membantu dalam pengambilan keputusan.

Menurut Abdous , He dan Yen (2012) data mining adalah multidisiplin yang terdiri dari beberapa berbagai disiplin ilmu termasuk ilmu komputer, kecerdasan buatan, statistik, dan biometrik yang menggunakan berbagai pendekatan (seperti klasifikasi, clustering, aturan asosiasi, dan visualisasi).

Data mining sering juga disebut analisis data eksploratif. Data dalam jumlah besar yang diperoleh dari mesin kasir, pemindaian *barcode* dan dari berbagai basis data dalam perusahaan, kemudian ditelaah, dianalisis, dihapus dan dipakai ulang. Pencarian dilakukan pada model yang berbeda untuk memprediksi penjualan, respon pasar, keuntungan dan lain-lain (Olson & Shi, 2008).

Berdasarkan pengertian-pengertian di atas dapat di simpulkan bahwa data mining merupakan disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data.

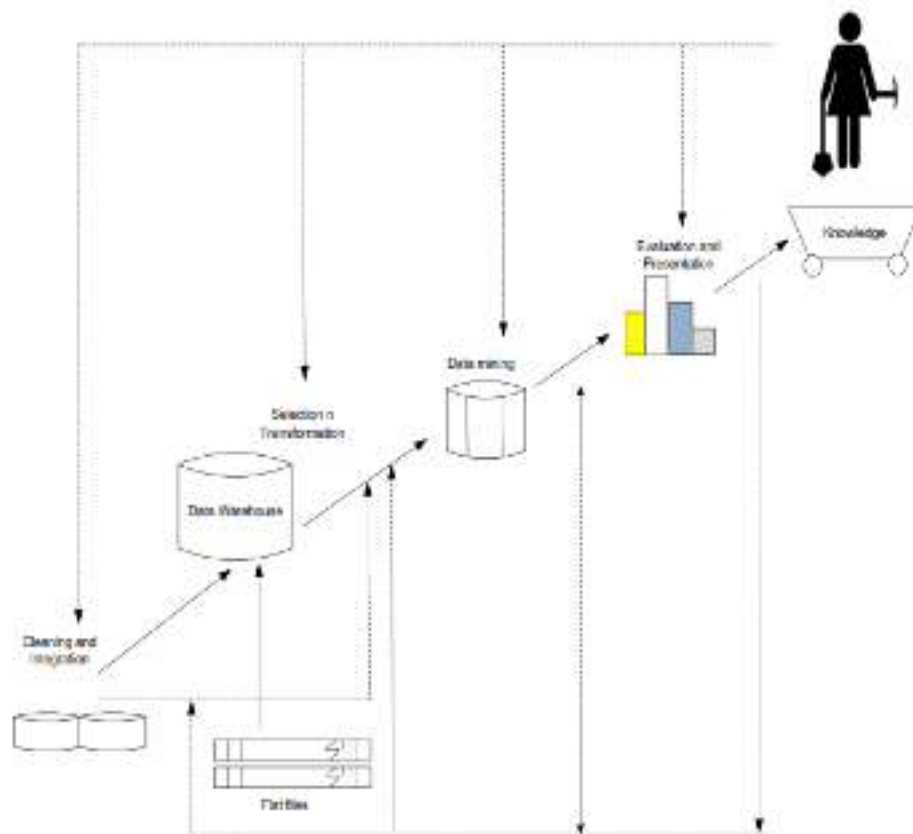
Data mining merupakan bagian dari proses *Knowledge Discovery from Data* (KDD). proses penjelajahan pengetahuan dimulai dari beberapa database dilakukan proses *cleaning* dan *integration* sehingga menghasilkan data warehouse. Dilakukan proses *selection* dan *transformation* yang kemudian disebut sebagai data mining hingga menemukan pola dan memperoleh pengetahuan dari data (*knowledge*).

Tahapan data mining dalam proses penemuan pengetahuan (Han & Kamber, 2007):

1. Pembersihan data (untuk menghilangkan noise dan data tidak konsisten)
2. Integrasi data (di mana beberapa sumber data dapat dikombinasikan)



3. Data seleksi (di mana data yang relevan dengan tugas analisis basis data yang akan diambil)
4. Data transformasi (dimana data diubah atau dikonsolidasikan ke dalam bentuk yang sesuai untuk pertambangan dengan melakukan operasi ringkasan atau agregasi)
5. Data mining (proses esensial dimana metode cerdas diaplikasikan untuk mengekstrak pola data)
6. Pola evaluasi (untuk mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan didasarkan pada beberapa langkah-langkah interestingness)
7. Pengetahuan presentasi (dimana visualisasi dan teknik representasi pengetahuan digunakan untuk menyajikan pengetahuan ditambang kepada pengguna)



**Gambar 2.1 Tahapan data mining (Han & Kamber, 2007)**

Terdapat empat pelompokan dalam data mining yaitu klasifikasi, asosiasi, clustering dan prediksi (Witten, Frank, & Hall, 2011):

## 1. Klasifikasi

Klasifikasi adalah proses menempatkan obyek atau konsep tertentu kedalam satu set kategori, berdasarkan sifat obyek atau konsep yang bersangkutan (Gorunescu, 2011). Metode klasifikasi ditujukan untuk pembelajaran fungsi-fungsi berbeda yang memetakan masing-masing data terpilih kedalam salah satu dari kelompok kelas yang telah ditetapkan sebelumnya. Proses klasifikasi didasarkan pada komponen (Gorunescu, 2011):

### a. Kelas (*class*)

Variabel dependen dari model yang merupakan kategori variabel yang mewakili label-label yang diletakkan pada obyek setelah pengklasifikasian. Contoh: kelas bintang dan kelas gempa bumi.

### b. Prediktor (*predictors*)

Variabel independen dari model yang diwakili oleh karakteristik atau atribut dari data yang diklasifikasikan berdasarkan klasifikasi yang dibuat. Contoh: tekanan darah, status perkawinan dan musim.

### c. Dataset pelatihan (*training dataset*)

Merupakan dataset yang berisi dua komponen nilai yang digunakan untuk pelatihan untuk mengenali model yang sesuai dengan kelasnya, berdasarkan prediktor yang ada. Contoh: database penelitian gempa, database badai dan database pelanggan supermarket.

### d. Dataset pengujian (*testing dataset*)

Merupakan dataset baru yang akan diklasifikasikan oleh model yang dibangun sehingga dapat dievaluasi hasil akurasi klasifikasi tersebut.

## 2. Asosiasi

Setiap asosiasi antara fitur-fitur yang dicari, bukan hanya satu yang memprediksi nilai kelas tertentu (Witten, Frank, & Hall, 2011). Pada prinsipnya, penemuan aturan asosiasi/asosiasi mempelajari aturan bagaimana kita memahami proses mengidentifikasi aturan antara ketergantungan yang berbeda dari fenomena kelompok. Dengan demikian, mari kita perkirakan kumpulan set yang kita punya masing-masing berisi sejumlah objek/benda.

Jadi tujuan kita untuk mencari peraturan yang menghubungkan (asosiasi), obyek ini berdasarkan peraturan ini, untuk dapat memprediksi terjadinya objek/item, berdasarkan kejadian lain (Gorunescu, 2011).

### 3. *Clustering*

Cluster adalah menemukan kelompok (kelompok) objek, berdasarkan kemiripan (semacam kemiripan), sehingga dalam setiap kelompok ada kemiripan yang besar, sementara kelompok cukup berbeda dari satu sama lain (Gorunescu, 2011).

### 4. Prediksi

Prediksi/ perkiraan model yang berkaitan dengan kemampuan untuk memprediksi tanggapan terbaik (*output*), yang paling dekat dengan kenyataan, berdasarkan *input* data.

Dengan demikian, semakin kecil perbedaan antara apa yang diharapkan terjadi (hasil yang diharapkan) dari apa yang sebenarnya terjadi (diamati), semakin baik prediksi, contohnya prediksi ramalan cuaca (misalnya, untuk 24 atau 48 jam) atau diagnosis untuk penyakit tertentu yang diberikan kepada pasien tertentu, yang didasarkan pada data medis (Gorunescu, 2011).

Konsep data mining, menemukan pola berharga dalam data, adalah respon yang jelas untuk pengumpulan dan penyimpanan volume data yang besar (Weiss, Indurkha, & Zhang, 2010). Secara khusus, koleksi metode yang dikenal sebagai 'data mining' menawarkan metodologi dan solusi teknis untuk mengatasi analisis data medis dan konstruksi dari prediksi (Bellazzi & Zupanb, 2008). Untuk semua aplikasi data mining, akurasi prediksi tergantung pada kualitas prediksi atribut (Weiss, Indurkha, & Zhang, 2010).

#### **2.1.3. *Support Vector Machine (SVM)***

*Support Vector Machine* adalah metode learning machine yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan hyperplane terbaik yang memisahkan dua buah *class* pada *input space* (T. Bellotti and J. Crook, 2009, pp. 3302-3308). Hyperplane terbaik adalah hyperplane yang terletak ditengah-tengah antara dua set obyek dari dua class. Hyperplane pemisah terbaik antara kedua class dapat ditemukan dengan mengukur margin hyperplane

tersebut dan mencari titik maksimalnya. Margin adalah jarak antara hyperplane tersebut dengan *pattern* terdekat dari masing-masing *class*. *Pattern* yang paling dekat ini disebut sebagai support vector (I. Aydin, M. Karakose, and E. Akin, 2011, pp. 120-129).

Menurut Y. Yin, Han, & Cai, (2011) *Support Vector Machine* (SVM) didefinisikan sebagai seperangkat metode pembelajaran terkait yang menganalisis data dan mengenali pola, yang kemudian digunakan untuk klasifikasi dan analisis regresi. SVM mengambil satu set *data input* dan memprediksi untuk setiap masukan yang diberikan, yang berasal dari dua kelas yang kemudian di klasifikasikan dengan mencari nilai *hyperplane* terbaik.

Menurut H. Yin et al., (2015) *Support Vector Machine* (SVM) adalah suatu metode klasifikasi untuk mencari nilai *hyperplane* terbaik yang mampu menemukan solusi global optimal. Sehingga nilai akurasi tidak mudah berubah-ubah.

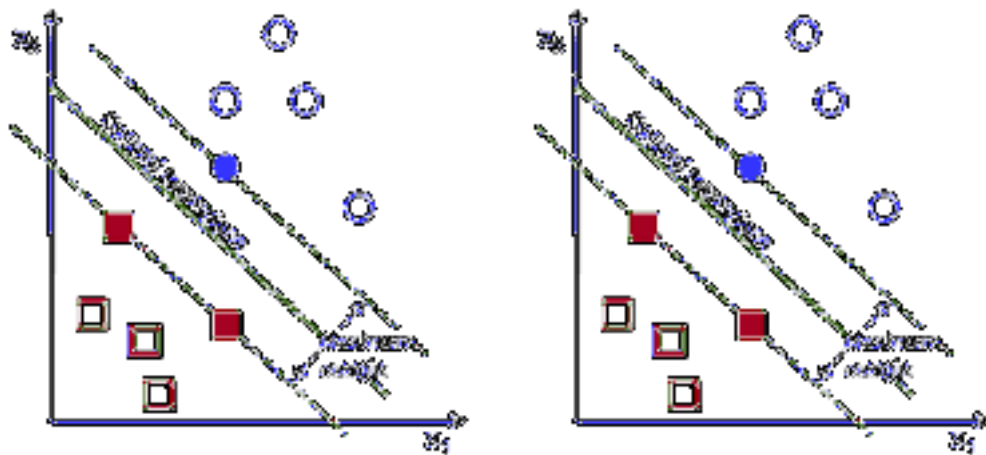
Menurut Li, You, & Liu (2015) *Support Vector Machine* (SVM) merupakan pembelajaran yang mengarah ke pemrograman kuadratik dengan kendala linear. Berdasarkan minimalisasi risiko prinsip terstruktur, SVM berusaha untuk meminimalkan batas atas kesalahan generalisasi bukan kesalahan empiris, sehingga model prediksi baru efektif menghindari *over-pas* masalah. Selain itu, model SVM bekerja di ruang fitur berdimensi tinggi yang dibentuk oleh pemetaan nonlinear dari N-dimensi vektor input  $x$  ke dalam ruang fitur K-dimensi ( $K > N$ ) melalui penggunaan fungsi  $\phi$  nonlinear ( $x$ ).

Berdasarkan pengertian di atas dapat disimpulkan *Support Vector Machine* (SVM) merupakan suatu metode klasifikasi yang memaksimalkan batas *hyperplane* (*maximal margin hyperplane*). Jika dalam ANN semua data latih akan dipelajari selama proses pelatihan, sedangkan pada SVM berbeda karena hanya sejumlah data terpilih saja yang berkontribusi untuk membentuk model yang digunakan dalam klasifikasi yang akan dipelajari (Prasetuo, 2012). Hal ini menjadi kelebihan SVM karena tidak semua data latih akan di pandang untuk dilibatkan dalam setiap iterasi pelatihnnya. Data-data yang berkontribusi tersebut disebut sebagai support vector sehingga metodenya disebut *Support Vector Machine*.

Yang menjadi karakteristik dari *Support Vector Machine* (SVM) adalah sebagai berikut:

1. Secara prinsip SVM adalah *linear classifier*.
2. Pattern recognition dilakukan dengan mentransformasikan data pada input space ke ruang yang berdimensi lebih tinggi, dan optimisasi dilakukan pada ruang vector yang baru tersebut. Hal ini membedakan SVM dari solusi pattern recognition pada umumnya, yang melakukan optimisasi parameter pada ruang hasil transformasi yang berdimensi lebih rendah daripada dimensi input space.
3. Menerapkan strategi *Structural Risk Minimization* (SRM).
4. Prinsip kerja SVM pada dasarnya hanya mampu menangani klasifikasi dua kelas.

Secara sederhana konsep SVM adalah sebagai usaha mencari *hyperlane* terbaik yang berfungsi sebagai pemisah dua buah *class* pada *input space*, dimana dapat dilihat pada gambar dibawah ini:



Sumber : Prasetyo ( 2012)

**Gambar 2.2 Konsep SVM untuk mencari *hyperlane* terbaik**

Gambar diatas memperlihatkan beberapa pola yang merupakan anggota dari dua buah kelas data : +1 dan -1. Data yang tergabung pada kelas -1 disimbolkan dengan bentuk lingkaran, sedangkan data pada kelas +1 disimbolkan dengan bentuk bujur sangkar.

*Hyperplane* (batas keputusan) pemisah terbaik antara kedua kelas dapat ditemukan dengan mengukur *margin hyperplane* tersebut dan mencari titik maksimalnya. *Margin* adalah jarak antara *hyperplane* tersebut dengan data terdekat dari masing-masing kelas. Data yang paling dekat ini disebut *support vector*. Garis solid pada gambar di atas menunjukkan *hyperplane* yang terbaik, yaitu yang terletak tepat pada tengah-tengah kedua kelas, sedangkan data lingkaran dan bujur sangkar yang dilewati garis batas margin (garis putus-putus) adalah *support vector*. Usaha untuk mencari lokasi *hyperplane* ini merupakan inti dari proses pelatihan *Support Vector Machine* (SVM) (Prasetyo, 2012).

Data yang tersedia dinotasikan sebagai  $x \in \mathbb{R}^d$  sedangkan label masing-masing dinotasikan  $y_i \in \{-1, +1\}$  untuk  $i = 1, 2, \dots, l$  yang mana  $l$  adalah banyaknya data. Diasumsikan kedua class  $-1$  dan  $+1$  dapat terpisah secara sempurna oleh *hyperplane* berdimensi  $d$ , yang didefinisikan:

$$- \quad w \cdot x + b = 0 \quad (2.1)$$

Sebuah pattern  $x_i$  yang termasuk class  $-1$  (sampel negatif) dapat dirumuskan sebagai pattern yang memenuhi pertidaksamaan:

$$- \quad w \cdot x + b = -1 \quad (2.2)$$

sedangkan pattern yang termasuk class  $+1$  (sampel positif):

$$- \quad w \cdot x + b = +1 \quad (2.3)$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperplane* dan titik terdekatnya, yaitu  $1/\|w\|$ . Hal ini dapat dirumuskan sebagai *Quadratic Programming* (QP) problem, yaitu mencari titik minimal persamaan 2.4, dengan memperhatikan constraint persamaan 2.5.

$$\frac{\min}{w} t(w) = \frac{1}{2} \|w\|^2 \quad (2.4)$$

$$y_i(x_i \cdot w + b) - 1 \geq 0, \forall_i \quad (2.5)$$

Problem ini dapat dipecahkan dengan berbagai teknik komputasi, diantaranya *Lagrange Multiplier* sebagaimana ditunjukkan pada persamaan 2.6:

$$L(w, b, \alpha) = \frac{1}{2} ||w||^2 w \sum_{i=1} (\alpha_i) \left( y_i \left( (x_i \cdot w + b) - 1 \right) \right) \quad (i = 1, 2, \dots, 1) \quad (2.6)$$

$\alpha_i$  adalah Lagrange multipliers, yang bernilai nol atau positif ( $\alpha_i \geq 0$ ). Nilai optimal dari persamaan (2.6) dapat dihitung dengan meminimalkan  $L$  terhadap  $w$  dan  $b$ , dan memaksimalkan  $L$  terhadap  $\alpha_i$ . Dengan memperhatikan sifat bahwa pada titik optimal gradient  $L = 0$ , persamaan langkah 2.6 dapat dimodifikasi sebagai maksimisasi problem yang hanya mengandung saja  $\alpha_i$ , sebagaimana persamaan 2.7.

$$\text{Maximize : } \sum_{i=1} \alpha_i - \frac{1}{2} \sum_{i,j=1}^1 \alpha_i \alpha_j y_i y_j x_i \cdot x_j \quad (2.7)$$

Subject to:

$$\alpha_i \geq 0 \quad (i = 1, 2, \dots, 1) \quad \sum_{i=1}^1 \alpha_i y_i = 0 \quad (2.8)$$

Dari hasil dari perhitungan ini diperoleh  $\alpha_i$  yang kebanyakan bernilai positif. Data yang berkorelasi dengan  $\alpha_i$  yang positif inilah yang disebut sebagai *support vector*. Sebagai contoh digunakan *problem AND*. *Problem AND* adalah klasifikasi dua kelas dengan empat data (lihat Tabel 2.1). Karena ini *problem linier*, kernelisasi tidak diperlukan.

**Table 2.1 AND Problem**

| X1 | X2 | Y  |
|----|----|----|
| 1  | 1  | 1  |
| -1 | 1  | -1 |
| 1  | -1 | -1 |
| -1 | -1 | 1  |

Sumber : Prasetyo (2012)

dapatkan formulasi masalah optimisasi sebagai berikut:

$$\min \frac{1}{2} (w_1^2 + w_2^2) + C(t_1 + t_2 + t_3 + t_4)$$

Subject to :

$$w_1 + w_2 + b + t_1 \geq 1$$

$$w_1 - w_2 - b + t_2 \geq 1$$

$$-w_1 + w_2 - b + t_3 \geq 1$$

$$w_1 + w_2 - b + t_4 \geq 1$$

$$t_1, t_2, t_3, t_4 \geq 0$$

Karena fungsi AND adalah kasus klasifikasi linier, maka bisa dipastikan nilai variable slack  $t_i=0$ . Jadi Kita bisa masukkan nilai  $C=0$ . Setelah menyelesaikan problem optimasi di atas didapat solusi

$$w_1 = 1, w_2 = 1, b = -1$$

Persamaan fungsi pemisahnya adalah

$$f(x) = x_1 + x_2 - 1.$$

Untuk menentukan *output* atau label dari setiap titik data/obyek kita gunakan fungsi  $g(x) = \text{sign}(x)$ . Dengan fungsi sign ini semua nilai  $f(x) < 0$  diberi label  $-1$  dan lainnya diberi label  $+1$ .

#### 2.1.4. Seleksi Atribut

Proses data mining membutuhkan biaya komputasi yang tinggi ketika berhadapan dengan kumpulan data dalam jumlah besar. Mengurangi dimensi yaitu jumlah atribut set data atau kelompok atribut, secara efektif dapat memotong biaya tersebut. Pengurangan dimensi tersebut dilakukan dengan menekan seminimal mungkin kerugian yang dapat terjadi akibat kehilangan sebagian informasi. Tujuan pengurangan dimensi dalam domain data mining adalah untuk mengidentifikasi biaya terkecil di mana algoritma data mining dapat menjaga tingkat kesalahan di bawah perbatasan garis efisiensi. Yang dimaksud dengan biaya adalah fungsi dari kompleksitas teoritis dari algoritma data mining yang berasal dari model, dan berkorelasi dengan waktu yang dibutuhkan algoritma tersebut dalam menjalankan model, serta ukuran dari kumpulan data (Maimon & Rokach, 2010).

Tujuan seleksi atribut adalah untuk pengurangan atribut dari dataset untuk menghilangkan variabel yang dianggap tidak relevan. Metode seleksi fitur dapat diklasifikasikan ke dalam tiga kategori utama (Vercellis, 2009):

##### 1. Metode *filter*

Metode Filter adalah memilih atribut yang relevan sebelum pindah ke tahap pembelajaran berikutnya, atribut yang dianggap paling penting yang dipilih untuk pembelajar, sedangkan sisanya dikecualikan.



## 2. Metode *wrapper*

Metode *wrapper* menilai sekelompok variabel dengan menggunakan klasifikasi yang sama atau algoritma regresi digunakan untuk memprediksi nilai dari variabel target.

## 3. Metode *embedded*

Untuk metode *embedded*, proses seleksi atribut terletak di dalam algoritma pembelajaran, sehingga pemilihan set optimal atribut secara langsung dibuat selama fase generasi model.

Di sisi lain, seleksi atribut adalah proses yang mahal, dan juga bertentangan dengan asumsi awal yaitu bahwa semua informasi atau atribut diperlukan dalam rangka mencapai akurasi maksimal. Ada empat alasan utama untuk melakukan pengurangan dimensi yaitu (Maimon & Rokach, 2010):

1. Penurunan biaya model pembelajaran
2. Meningkatkan kinerja model pembelajaran
3. Mengurangi dimensi yang tidak relevan
4. Mengurangi dimensi yang berlebihan

### 2.1.5. *Particle Swarm optimization (PSO)*

*Particle Swarm Optimization (PSO)* merupakan teknik optimasi *heuristic* global yang diperkenalkan oleh Dokter Kennedy dan Eberhart pada tahun 1995 (Ratna Wati, 2011) yang terinspirasi oleh perilaku sosial kawanan burung yang mencoba mencapai tujuan yang tidak di ketahui (Jung Hsieh, Fen Hsiao, Chang Yeh, 2012).

Menurut Yun, Qiu-yan & Hua, (2011) *Particle swarm optimization (PSO)* adalah jenis algoritma kecerdasan yang mampu melakukan optimasi terhadap sebuah variable terkait yang paling efektif.

Menurut Huang & Dun, (2008) *Particle swarm optimization (PSO)* adalah perhitungan evolusi Teknik. Serupa dengan algoritma genetika, PSO adalah sebuah alat optimasi. Hal ini terinspirasi oleh perilaku sosial di antara individu. Partikel (individu) yang mewakili potensi solusi masalah bergerak melalui ruang pencarian n-dimensi. Setiap partikel  $i$  memelihara catatan posisi kinerja terbaik dalam vektor yang disebut  $pbest$ .

Menurut Y. Yin et al., (2011) *Particle Swarm Optimization* (PSO) adalah metode komputasi yang mengoptimalkan masalah dengan iteratif untuk meningkatkan solusi kandidat berkaitan dengan ukuran tertentu kualitas. Gerakan setiap partikel dipengaruhi oleh posisi lokal yang dipandu menuju posisi paling dikenal dalam pencarian ruang, yang diperbarui sebagai posisi yang lebih baik dari partikel lain.

Menurut Zhao, Liu, Zhang, & Wang, (2015) *Particle Swarm Optimization* (PSO) merupakan teknik komputasi evolusioner yang mampu menghasilkan solusi secara global optimal dalam ruang pencarian melalui interaksi individu dalam segerombolan partikel. Setiap partikel menyampaikan informasi berupa posisi terbaiknya kepada partikel yang lain dan menyesuaikan posisi dan kecepatan masing-masing berdasarkan informasi yang diterima mengenai posisi yang terbaik tersebut.

*Particle Swarm Optimization*(PSO) adalah alat untuk menangani masalah optimasi(Yuan Shu, L., Jinn Horng, S., He, M., Fan, P., Wann Kao, T., Khan, M. K., Shine Run, R., Lin Lai, J., Jian Chen, R., 2011). Walaupun relatif baru, sudah banyak yang mengaplikasikan algoritma PSO, karena cukup sederhana serta memiliki kecepatan komputasi yang lebih cepat di bandingkan dengan algoritma optimasi lain nya seperti Genetic Algorithm(GA). Setiap partikel dalam PSO juga dikaitkan dengan kecepatan partikel terbang melalui ruang pencarian dengan kecepatan yang dinamis disesuaikan untuk perilaku histori mereka. Oleh karena itu, partikel memiliki kecenderungan untuk terbang menuju daerah pencarian yang lebih baik selama proses pencarian.

Berdasarkan pengertian di atas dapat disimpulkan bahwa *Particle Swarm Optimization* (PSO) merupakan metode optimasi yang mampu mengoptimalkan variable terdekat untuk mencapai akurasi yang maksimal.

*Swarm Intelligence (SI)* adalah sebuah paradigma cerdas terdistribusi inovatif untuk memecahkan masalah optimasi yang awalnya mengambil inspirasi dari contoh biologis dengan berkerumun, berkelompok dan menggiring fenomena dalam hewan vertebrata. *Particle Swarm Optimization (PSO)* menggabungkan perilaku berkerumun yang dilakukan hewan contoh dalam kawanan burung,

sekelompok ikan, atau kawanan lebah, dan perilaku sosial pada manusia (Abraham, Grosan, & Ramos, 2006)

Untuk menemukan solusi yang optimal, maka setiap partikel akan bergerak ke arah posisi yang terbaik sebelumnya ( $pbest$ ) dan posisi terbaik secara global ( $gbest$ ). Sebagai contoh, partikel ke-  $i$  dinyatakan sebagai :  $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,d})$  dalam ruang  $d$ -dimensi. Posisi terbaik sebelumnya dari partikel ke-  $i$  disimpan dan dinyatakan sebagai  $pbest_i = (pbest_{i,1}, pbest_{i,2}, \dots, pbest_{i,d})$ . Modifikasi kecepatan dan posisi tiap partikel dapat dihitung menggunakan kecepatan saat ini dan jarak  $pbest_{i,d}$  ke  $pbest_d$  seperti ditunjukkan oleh persamaan berikut:

$$v_{i,m} = w \cdot v_{i,m} + c_1 * R * (pbest_{i,m} - x_{i,m}) + c_2 * R * (gbest_m - x_{i,m}) \quad (2.9)$$

$$x_{id} = x_{i,m} + v_{i,m} \quad (2.10)$$

Dimana:

$n$  : jumlah partikel dalam kelompok

$d$  : dimensi

$v_{i,m}$  : kecepatan partikel ke- $i$  pada iterasi ke-  $i$

$w$  : faktor bobot inersia

$c_1, c_2$  : konstanta akselerasi (learning rate)

$R$  : bilangan random (0-1)

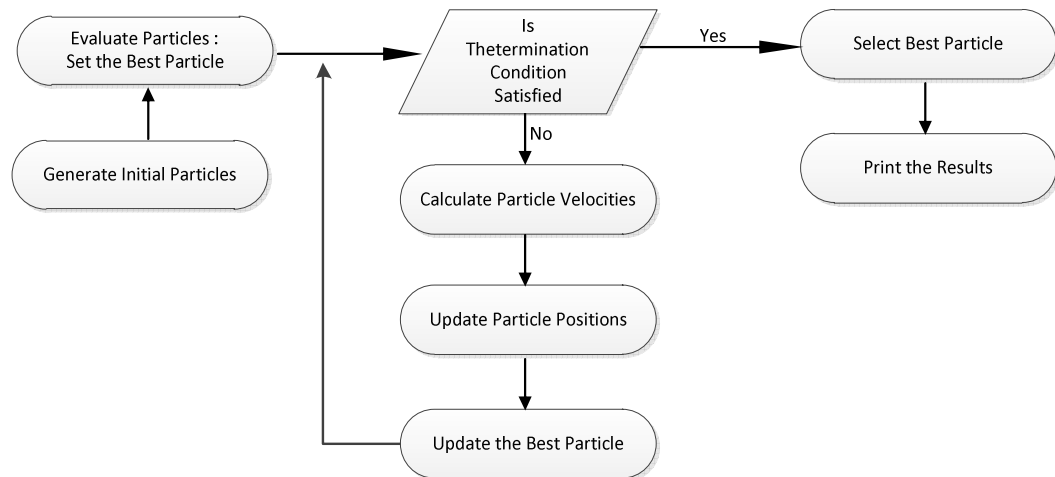
$x_{i,d}$  : posisi saat ini dari partikel ke-  $i$  pada iterasi ke-  $i$

$pbest_i$  : posisi terbaik sebelumnya dari partikel ke-  $i$

$gbest$  : partikel terbaik diantara semua partikel dalam satu kelompok atau populasi

Persamaan (2.9) menghitung kecepatan baru untuk tiap partikel (solusi potensial) berdasarkan pada kecepatan sebelumnya ( $v_{i,m}$ ), lokasi partikel dimana nilai fitness terbaik telah dicapai ( $pbest$ ), dan lokasi populasi global ( $gbest$  untuk versi global,  $lbest$  untuk versi local) atau *local neighborhood* pada algoritma versi local dimana nilai fitness terbaik telah dicapai.

Persamaan (2.10) memperbaharui posisi tiap partikel pada ruang solusi. Dua bilangan acak  $c_1$  dan  $c_2$  dibangkitkan sendiri. Penggunaan berat inersia  $w$  telah memberikan performa yang meningkat pada sejumlah aplikasi. Secara garis besar, struktur dasar dari PSO dapat digambarkan dalam bagan dibawah ini:



Sumber: Abraham, Grosan & Ramos (2006)

**Gambar 2.3 Struktur dasar PSO**

#### 2.1.6. Pengujian K-Fold Cross Validation

Salah satu pendekatan alternatif untuk “train dan test” yang sering di adopsi dalam beberapa kasus (dan beberapa lainnya terlepas dari ukurannya) yang di sebut dengan k-fold cross validation (Bramer, 2007), dengan cara menguji besarnya error pada data test (Santosa, 2007).

*Cross validation* adalah teknik validasi dengan membagi data secara acak kedalam k bagian dan masing- masing bagian akan dilakukan proses klasifikasi (Han & Kamber, 2006). Dengan menggunakan *Cross validation* akan dilakukan percobaan sebanyak k. Data yang digunakan dalam percobaan ini adalah data training untuk mencari nilai error rate secara keseluruhan. Secara umum pengujian nilai k dilakukan sebanyak 10 kali untuk memperkirakan akurasi estimasi. Dalam penelitian ini nilai k yang digunakan berjumlah 10 atau 10-fold *Cross validation*. tiap percobaan akan menggunakan satu data testing dan k-1 bagian akan menjadi data training, kemudian *data testing* itu akan ditukar dengan satu buah data training sehingga untuk tiap percobaan akan didapatkan data testing yang berbeda-beda.

Data *training* adalah data yang akan dipakai dalam melakukan pembelajaran sedangkan data testing adalah data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian kebenaran atau keakurasian hasil pembelajaran (Witten & Frank, 2011).

| Split 1  | Split 2 | Split 3 | Split 4 | Split 5 | Split 6 | Split 7 | Split 8 | Split 9 | Split 10 |
|----------|---------|---------|---------|---------|---------|---------|---------|---------|----------|
| Training |         |         |         |         |         |         |         |         | Test     |
|          |         |         |         |         |         |         |         |         |          |
| Training |         |         |         |         |         |         |         | Test    |          |
|          |         |         |         |         |         |         |         |         |          |
| Training |         |         |         |         |         |         | Test    |         |          |
|          |         |         |         |         |         |         |         |         |          |
| Training |         |         |         |         |         | Test    |         |         |          |
|          |         |         |         |         |         |         |         |         |          |
| Training |         |         |         |         | Test    |         |         |         |          |
|          |         |         |         |         |         |         |         |         |          |
| Training |         |         |         | Test    |         |         |         |         |          |
|          |         |         |         |         |         |         |         |         |          |
| Training |         |         | Test    |         |         |         |         |         |          |
|          |         |         |         |         |         |         |         |         |          |
| Training |         | Test    |         |         |         |         |         |         |          |
|          |         |         |         |         |         |         |         |         |          |
| Test     |         |         |         |         |         |         |         |         |          |

Sumber: (Witten & Frank, 2011)

**Gambar 2.4 Ilustrasi 10-Fold *Cross validation***

### 2.1.7. Evaluasi dan Validasi Hasil

#### 2.1.7.1. *Confision Matrix*

*Confusion matrix* memberikan keputusan yang diperoleh dalam traning dan testing, *Confusion matrix* memberikan penilaian performance klasifikasi berdasarkan objek dengan benar atau salah (Gorunescu, 2011). *Confusion matrix* berisi informasi aktual (*actual*) dan prediksi (*predicted*) pada sistem klasifikasi.

**Tabel 2.2 Confusion Matrix**

| Clarification  | Predicted Class |                           |                           |
|----------------|-----------------|---------------------------|---------------------------|
| Observed Class |                 | Class = Yes               | Class = No                |
|                | Class = Yes     | a<br>(True Positive – TP) | b<br>(Fals Negative – FN) |
|                | Class = No      | c<br>(Fals Positive – FP) | d<br>(True Negative – TN) |

Sumber: Gorunescu (2011)

Keterangan:

*True Positive* (TP) = proporsi positif dalam data set yang diklasifikasikan positif

*True Negative* (TN) = proporsi negative dalam data set yang diklasifikasikan negative

*False Positive* (FP) = proporsi negatif dalam data set yang diklasifikasikan positif

*False Negative* (FN) = proporsi negative dalam data set yang diklasifikasikan negatif

Berikut adalah persamaan model *Confusion matrix* :

- a. Nilai *Accuracy* adalah proporsi jumlah prediksi yang benar. Dapat dihitung dengan menggunakan persamaan:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- b. *Sensitivity* digunakan untuk membandingkan proporsi TP terhadap tupel yang positif, yang dihitung dengan menggunakan persamaan:

$$Sensitivity = \frac{TP}{TP + FN}$$

- c. *Specificity* digunakan untuk membandingkan proporsi TN terhadap tupel yang negatif, yang dihitung dengan menggunakan persamaan:

$$Specificity = \frac{TN}{TN + FP}$$

- d. PPV (*positive predictive value*) adalah proporsi kasus dengan hasil diagnosa positif, yang dihitung dengan menggunakan persamaan:

$$PPV = \frac{TP}{TP + FP}$$

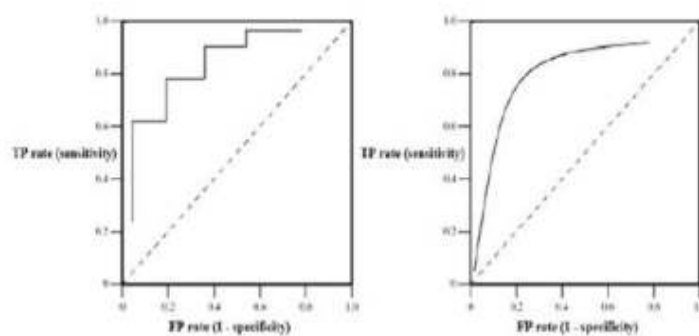
- e. NPV (*negative predictive value*) adalah proporsi kasus dengan hasil diagnosa negatif, yang dihitung dengan menggunakan persamaan:

$$NPV = \frac{TN}{TN + FN}$$

### 2.1.7.2 Kurva ROC

Kurva ROC (*Receiver Operating Characteristic*) adalah alat visual yang berguna untuk membandingkan dua model klasifikasi. ROC mengekspresikan *Confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horisontal dan *true positives* sebagai garis vertikal (Vecellis, 2009). Dengan kurva ROC, kita dapat melihat *trade off* antara tingkat dimana suatu model dapat

mengenali tuple positif secara akurat dan tingkat dimana model tersebut salah mengenali tuple negatif sebagai tuple positif. Sebuah grafik ROC adalah plot dua dimensi dengan proporsi positif salah (fp) pada sumbu X dan proporsi positif benar (tp) pada sumbu Y. Titik (0,1) merupakan klasifikasi yang sempurna terhadap semua kasus positif dan kasus negatif. Nilai positif salah adalah tidak ada ( $fp = 0$ ) dan nilai positif benar adalah tinggi ( $tp = 1$ ). Titik (0,0) adalah klasifikasi yang memprediksi setiap kasus menjadi negatif  $\{-1\}$ , dan titik (1,1) adalah klasifikasi yang memprediksi setiap kasus menjadi positif  $\{1\}$ . Grafik ROC menggambarkan *trade-off* antara manfaat (*true positive*) dan biaya (*false positives*). Berikut tampilan dua jenis kurva ROC (*discrete dan continuous*).



Sumber: Gorunescu (2011)

**Gambar 2.5 Grafik ROC (*discrete dan continuous*)**

Poin diatas garis diagonal merupakan hasil klasifikasi yang baik, sedangkan point dibawah garis diagonal merupakan hasil klasifikasi yang buruk. Dapat disimpulkan bahwa, satu point pada kurva ROC adalah lebih baik dari pada yang lainnya jika arah garis melintang dari kiri bawah ke kanan atas didalam grafik. Untuk tingkat akuransi nilai AUC dalam klasifikasi data mining dibagi menjadi lima kelompok (Gorunescu, 2011), yaitu:

- a. 0.90 - 1.00 = klasifikasi sangat baik (*excellent classification*)
- b. 0.80 - 0.90 = klasifikasi baik (*good classification*)
- c. 0.70 - 0.80 = klasifikasi cukup (*fair classification*)
- d. 0.60 - 0.70 = klasifikasi buruk (*poor classification*)
- e. 0.50 - 0.60 = klasifikasi salah (*failure*)

## 2.2. Tinjauan Studi

Berikut adalah beberapa penelitian terdahulu yang terkait dengan pembahasan prediksi menggunakan *support vector machine* yang telah banyak dilakukan oleh peneliti sebelumnya. Literatur mengenai pembahasan analisis prediksi pemasaran langsung telah dilakukan dengan beberapa metode. Penelitian yang dilakukan adalah menggunakan data set yang bersifat *public*, dimana data set diambil dari repositori publik yang disepakati oleh para peneliti data mining yaitu UCI Repository. Berikut metode-metode yang pernah digunakan untuk menyelesaikan masalah prediksi pemasaran langsung.

### 2.2.1. Model Penelitian Hany A. Elsalamony dan Alaa M. Elyasad

Penelitian yang dilakukan oleh Hany A. Elsalamony dan Alaa M. Elyasad yang berjudul *Bank Direct Marketing Based on Neural Network and C5.0 Models*, menggunakan model *Multilayer Perception Neural Network* (MLPNN) dan *Ross Quinlan New Decision Tree Model* (C5.0). Tujuan dari penelitian ini adalah untuk memeriksa kinerja MLPNN dan C5.0 pada data real-dunia deposito bank. Tujuannya untuk meningkatkan efektivitas kampanye dengan mengidentifikasi karakteristik utama yang mempengaruhi keberhasilan (deposito berlangganan oleh klien). Hasil eksperimen telah menunjukkan efektivitas model C5.0 mencapai kinerja yang lebih baik dibandingkan dengan MLPNN. Analisis pentingnya telah menunjukkan bahwa atribut “*Duration*” pada model C5.0 merupakan atribut yang paling penting.

### 2.2.2. Model Penelitian Hany A. Elsalamony

Penelitian berikutnya dilakukan kembali oleh Hany A. Elsalamony dengan judul *Bank Direct Marketing Analysis Of Data Mining Techniques*, pada penelitian ini penulis memperkenalkan model data mining untuk membantu kinerja pemasaran serta menganalisa dan memperkenalkan aplikasi yang penting didalam teknik data mining. Dalam penelitian sebelumnya penulis hanya menggunakan dua model, dan sekarang menggunakan dua model tambahan yaitu *Tree Augmented Naive Bayes* (TAN) atau dikenal juga sebagai *Bayesian Networks* dan *Nominal Regression* atau *Logistic Regression* (LR). Tujuannya pun untuk



meningkatkan efektivitas kampanye dengan mengidentifikasi karakteristik utama yang mempengaruhi keberhasilan (deposito berlangganan oleh klien) berdasarkan model MLPNN, TAN, LR dan C5.0. Klasifikasi dari empat model tersebut menggunakan tiga ukuran statistik yaitu klasifikasi akurasi, sensitivitas dan spesifisitas.

Kumpulan data ini telah dipartisi menjadi data training dan test dengan rasio masing-masing 70% dan 30%. Hasil eksperimen telah menunjukkan efektivitas model C5.0 tetap menjadi yang lebih baik dibandingkan model lain, serta analisis pentingnya telah menunjukkan bahwa atribut “*Duration*” di C5.0, LR dan model MLPNN telah mencapai atribut yang paling penting, namun atribut “*age*” juga dinilai sebagai atribut penting pada model TAN.

### **2.2.3. Model Penelitian Pin Lio, Xin Zhang, Kunli, Yang Fu, Mingyan Wang dan Sensen Wang**

Penelitian yang dilakukan oleh Pin Lio, Xin Zhang, dan Kun Li, Yang Fu, Mingyan Wang dan Sensen Wang, yang berjudul *Parameter Optimization for Support Vector Machine Based on Nested Genetic Algorithms*, support vector machine (SVM) adalah metode klasifikasi populer dan landmark berdasarkan ide dari minimalisasi risiko struktural, yang telah memperoleh adopsi luas diberbagai bidang seperti pengenalan pola, regresi, peringkat dan lain-lain. Pada penelitian ini menyajikan sebuah metode asli berdasarkan dua algoritma *nested real-valued genetic* (NRGA), yang dapat mengoptimalkan parameter SVM secara efisien dan mempercepat parameter optimasi dengan metode tradisional yang mengoptimalkan semua parameter secara bersamaan. Hasil eksperimen menunjukkan dengan menggunakan metode optimasi parameter yang diusulkan yaitu NRGA dapat mengembangkan klasifikasi SVM dengan akurasi klasifikasi unggul karena efisiensi yang luar biasa dan kekuatan pencarian yang konsekuen.

### **2.2.4. Model Penelitian Fei-Long Chena dan Feng-Chia Li**

Penelitian yang dilakukan oleh Fei-Long Chena dan Feng-Chia Li tahun 2010 yang berjudul *Combination of feature selection approaches with SVM in credit scoring* membahas tentang credit scoring dengan metode hybrid-SVM serta

mengusulkan empat pendekatan untuk seleksi atribut yaitu LDA, Decision Tree, Rough Sets dan F-score dengan menggunakan dua dataset UCI. Dari pendekatan yang diusulkan kemudian dievaluasi dan hasilnya dibandingkan dengan pengujian menggunakan nonparametrik Wilcoxon signed rank test untuk menunjukan apakah ada perbedaan yang signifikan. Hasilnya menunjukkan bahwa pendekatan hybrid SVM + F-core menunjukkan hasil yang besar, efektif dan optimal dengan Nilai akurasi 76,70% (F.-L Chen and F-C, Li, 2010, pp. 4402-4909).

**Tabel 2.3.**  
**Rangkuman Penelitian Terkait**

| No | Peneliti                                                              | Tahun | Masalah Penelitian                                                                                                                            | Metode                                                                                       | Hasil                                                                                                                                                                                     |
|----|-----------------------------------------------------------------------|-------|-----------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | Hany A. Elsalamony dan Alaa M. Elyasad                                | 2013  | Menentukan Model Prediksi Terbaik                                                                                                             | Multilayer Perception Neural Network (MLPNN) dan Ross Quinlan New Decision Tree Model (C5.0) | efektivitas model C5.0 mencapai kinerja yang lebih baik dibandingkan dengan MLPNN                                                                                                         |
| 2  | Hany A. Elsalamony                                                    | 2014  | menganalisa pengaruh keberhasilan efektifitas kampanye pemasaran langsung dengan mengidentifikasi karakteristik utama dari model yg digunakan | MLPNN, TAN, LR dan C5.0                                                                      | Hasil eksperimen telah menunjukkan efektifitas model C5.0 tetap menjadi yang lebih baik dibandingkan model lain                                                                           |
| 3  | Pin Lio, Xin Zhang, dan Kun Li, Yang Fu, Mingyan Wang dan Sensen Wang | 2015  | menentukan model optimasi parameter terbaik                                                                                                   | algoritma nested real-valued genetic (NRGA) dan SVM                                          | optimasi parameter yang diusulkan yaitu NRGA dapat mengembangkan klasifikasi SVM dengan akurasi klasifikasi unggul karena efisiensi yang luar biasa dan kekuatan pencarian yang konsekuen |

|   |                                 |      |                                                                                                                               |                                                        |                                                                                                                                         |
|---|---------------------------------|------|-------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|
| 4 | Fei-Long Chena dan Feng-Chia Li | 2010 | Menentukan model terbaik untuk seleksi atribut                                                                                | hybrid-SVM, LDA, Decision Tree, Rough Sets dan F-score | Hasilnya menunjukkan bahwa pendekatan hybrid SVM + F-core menunjukkan hasil yang besar, efektif dan optimal dengan Nilai akurasi 76,70% |
| 5 | Yuni Eka Achyani                | 2016 | Menentukan model terbaik untuk seleksi atribut untuk meningkatkan akurasi prediksi dalam memprediksi model pemasaran langsung | SVM-PSO                                                | ?                                                                                                                                       |

Pada penelitian ini akan digunakan metode *Support Vector Machine* berbasis *Particle Swarm Optimization* untuk menentukan fitur terbaik pada bobot atribut yang sesuai dengan optimal sehingga hasil prediksi lebih akurat.

### 2.3. Tinjauan Objek Penelitian

Data yang digunakan pada penelitian ini menggunakan data sekunder, dimana data sekunder adalah data yang diperoleh tidak secara langsung dari objek penelitian, namun berasal dari data yang telah dikumpulkan sebelumnya oleh pihak lain.

Data yang digunakan dalam penelitian ini adalah data yang terkait dengan kampanye pemasaran langsung (panggilan telepon) dari lembaga perbankan Portugis. Tujuan klasifikasi adalah untuk memprediksi jika klien berlangganan deposito berjangka. Data tersebut diperoleh dari <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing#>. Data set berdasarkan dari data pemasaran bank dari lembaga perbankan portugis dari bulan May 2008 sampai bulan November 2010. Data pemasaran bank digambarkan pada tabel 2.2.

**Tabel 2.4. Data Set Pemasaran Bank**

| age | job           | marital  | education | default | balance | housing | loan | contact   | day | month | duration | campaign | pdays | previous | outcome | y   |
|-----|---------------|----------|-----------|---------|---------|---------|------|-----------|-----|-------|----------|----------|-------|----------|---------|-----|
| 78  | retired       | divorced | primary   | no      | 229     | no      | no   | telephone | 22  | oct   | 97       | 1        | -1    | 0        | unknown | yes |
| 32  | blue-collar   | married  | secondary | no      | 2089    | yes     | no   | cellular  | 14  | nov   | 132      | 1        | -1    | 0        | unknown | yes |
| 33  | management    | married  | secondary | no      | 3935    | yes     | no   | cellular  | 6   | may   | 765      | 1        | 342   | 2        | failure | yes |
| 30  | unemployed    | married  | primary   | no      | 1787    | no      | no   | cellular  | 19  | oct   | 79       | 1        | -1    | 0        | unknown | no  |
| 33  | services      | married  | secondary | no      | 4789    | yes     | yes  | cellular  | 11  | may   | 220      | 1        | 339   | 4        | failure | no  |
| 35  | management    | single   | tertiary  | no      | 1350    | yes     | no   | cellular  | 16  | apr   | 185      | 1        | 330   | 1        | failure | no  |
| 30  | management    | married  | tertiary  | no      | 1476    | yes     | yes  | unknown   | 3   | jun   | 199      | 4        | -1    | 0        | unknown | no  |
| 20  | student       | single   | secondary | no      | 502     | no      | no   | cellular  | 30  | apr   | 261      | 1        | -1    | 0        | unknown | yes |
| 59  | blue-collar   | married  | secondary | no      | 0       | yes     | no   | unknown   | 5   | may   | 226      | 1        | -1    | 0        | unknown | no  |
| 35  | management    | single   | tertiary  | no      | 747     | no      | no   | cellular  | 23  | feb   | 141      | 2        | 176   | 3        | failure | no  |
| 36  | self-employed | married  | tertiary  | no      | 307     | yes     | no   | cellular  | 14  | may   | 341      | 1        | 330   | 2        | other   | no  |
| 68  | retired       | divorced | secondary | no      | 4189    | no      | no   | telephone | 14  | jul   | 897      | 2        | -1    | 0        | unknown | yes |

Sumber: UCI Repository (2012)

University of California Irvine Machine Learning Repository adalah kumpulan database, teori dan data generator yang digunakan oleh komunitas machine learning untuk analisis empiris dari algoritma machine learning. Arsip yang ada telah dibuat pada tahun 1987 oleh David Aha dan mahasiswa pascasarjana universitas California Irvine. Arsip-arsip tersebut telah banyak digunakan sebagai sumber data set primer dan telah dijadikan sebagai rujukan lebih dari 1000 kali, sehingga menjadikannya sebagai salah satu dari 100 paper yang paling banyak dikutip dalam bidang ilmu komputer.

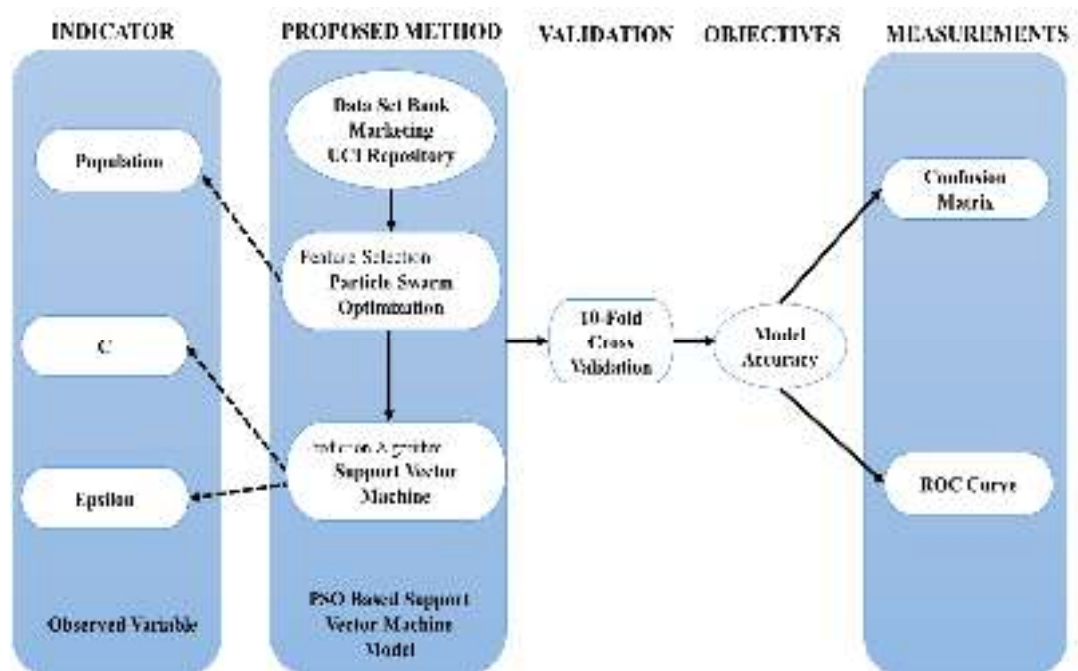
## **2.4. Kerangka Konsep Penelitian**

### **2.4.1. Kerangka Pemikiran**

Berdasarkan penjelasan di atas, penelitian ini bertujuan untuk meneliti apakah dengan menggunakan PSO untuk seleksi atribut di dalam SVM dapat meningkatkan akurasi dari prediksi pemasaran langsung, dimana atribut yang digunakan berdasarkan data set pemasaran bank.

Kerangka pemikiran pada penelitian ini terdiri dari beberapa tahap seperti terlihat pada Gambar 2.7. Permasalahan pada penelitian ini adalah banyaknya atribut sehingga akan mengurangi akurasi dan menambah kompleksitas dari algoritma *Support Vector Machine*. Untuk itu *Particle swarm optimization* untuk memecahkan permasalahan tersebut. Untuk mengembangkan aplikasi berdasarkan model yang dibuat, digunakan RapidMiner 5.3. Kemudian dilakukan pengujian dengan *10-Fold Cross validation* terhadap kinerja dari kedua metode tersebut, hasil ditunjukkan oleh *Confusion matrix* dan Kurva ROC. Dibawah ini adalah kerangka pemikiran dalam bentuk bagan:

kerangka pemikiran dalam bentuk bagan:



**Gambar 2.7 Kerangka Pemikiran**

## **BAB III**

### **METODE PENELITIAN**

#### **3.1. Desain Penelitian**

Penelitian merupakan kegiatan pemecahan masalah yang sistematis, yang dilakukan dengan perhatian dan kepedulian dalam konteks situasi yang dihadapi, penelitian dalam akademik digunakan untuk mengacu pada aktivitas yang rajin dan penyelidikan sistematis atau investigasi di suatu daerah, dengan tujuan menemukan atau merevisi fakta, teori, aplikasi dan tujuannya adalah untuk menemukan dan menyebarkan pengetahuan baru (Berndtsson, Olsson & Lundell, 2008). Sedangkan menurut (Dawson, 2009) penelitian adalah mencari melalui proses yang metodis untuk menambahkan pengetahuan itu sendiri dan dengan yang lainnya, oleh penemuan fakta dan wawasan yang tidak biasa.

Dalam konteks penelitian, metode yang dilakukan mengacu kepada pemecahan masalah yang meliputi mengumpulkan data, merumuskan hipotesis atau proposisi, pengujian hipotesis, menafsirkan hasil dan kesimpulan (Berndtsson, Hansson, Olsson & Lundell, 2008). Terdapat empat metode penelitian yang umumnya digunakan, diantaranya: *Action Research*, *Experiment*, *Case Study* dan *Survei* (Dawson, 2009).

Penelitian ini menggunakan metode penelitian eksperimen, yakni melibatkan penyelidikan perlakuan pada paran memeter atau variabel tergantung dari peneliti itu sendiri, dan menggunakan tes yang dikendalikan oleh si peneliti itu sendiri. Langkah-langkah yang dilakukan pada proses penelitian adalah sebagai berikut:

##### **1. Pengumpulan Data**

Pada bagian ini dijelaskan tentang bagaimana dan darimana data dalam penelitian ini didapatkan. Ada dua tipe dalam pengumpulan data, yaitu pengumpulan data primer dan pengumpulan data sekunder. Data primer adalah data yang dikumpulkan pertama kali untuk melihat apa yang sesungguhnya terjadi. Sedangkan data sekunder adalah data yang sebelumnya pernah dibuat oleh seseorang baik diterbitkan atau tidak. Pengumpulan data primer bisa didapat dari model interview terhadap responden, maupun dengan

model observasi terhadap suatu badan yang sedang diteliti. Pada tahap ini ditentukan data yang akan diproses. Mencari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua data kedalam data set, termasuk variabel yang diperlukan dalam proses.

## 2. Pengolahan awal data

Pada bagian ini dijelaskan tentang tahapan awal data mining. Pengolahan awal data meliputi proses input data ke format yang dibutuhkan, pengelompokan dan penentuan atribut data, serta pemecahan data (*split*) untuk digunakan dalam proses pembelajaran (*training*) dan pengujian (*testing*).

## 3. Model yang diusulkan

Pada tahap ini data dianalisis, kemudian dikelompokkan variabel mana yang berhubungan dengan satu sama lainnya. Setelah data dianalisis lalu diterapkan model-model yang sesuai dengan jenis data. Pembagian data kedalam data latihan (*training data*) dan data uji (*testing data*) juga diperlukan untuk pembuatan model.

## 4. Eksperimen dan pengujian model

Pada bagian ini dijelaskan tentang langkah-langkah eksperimen meliputi cara pemilihan arsitektur yang tepat dari model atau metode yang diusulkan sehingga didapatkan hasil yang dapat membuktikan bahwa metode yang digunakan adalah tepat.

## 5. Evaluasi dan validasi hasil

Pada bagian ini dilakukan evaluasi dan validasi hasil penerapan terhadap model penelitian yang dilakukan untuk mengetahui tingkat keakurasian model.

### 3.2. Pengumpulan Data

Data yang digunakan pada penelitian ini menggunakan data sekunder, dimana data sekunder adalah data yang diperoleh tidak secara langsung dari objek penelitian, namun berasal dari data yang telah dikumpulkan sebelumnya oleh pihak lain.

Data yang digunakan dalam penelitian ini adalah data yang terkait dengan kampanye pemasaran langsung (panggilan telepon) dari lembaga perbankan



Portugis. Tujuan klasifikasi adalah untuk memprediksi jika klien berlangganan deposito berjangka. Data tersebut diperoleh dari <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing#>. Data set berdasarkan dari data pemasaran bank dari lembaga perbankan portugis dari bulan May 2008 sampai bulan November 2010.

Parameter nasabah yang digunakan termasuk dalam faktor penentuan nasabah yang akan berlangganan deposito dan analisi data pribadi nasabah. Data terdiri dari 17 atribut. Variabel tersebut terbagi menjadi dua, yaitu variabel prediktor atau pemrediksi yaitu variabel yang dijadikan penentu hasil penilaian pemasaran langsung, dan variabel tujuan yaitu variabel yang dijadikan sebagai hasil penilaian pemasaran langsung. Adapun variabel prediktornya yaitu: usia (*age*), pekerjaan (*job*), perkawinan (*marital*), pendidikan (*education*), memiliki kredit (*default*), saldo rata-rata tahunan (*balance*), perumahan (*housing*), pinjaman (*loan*), kontak (*contact*), hari (*day*), bulan (*month*), durasi panggilan (*duration*), kampanye (*campaign*), jumlah hari yang berlalu setelah klien terakhir dihubungi dari panggilan sebelumnya (*pdays*), jumlah kontak yang dilakukan sebelum kampanye ini dan untuk klien ini (*previous*), hasil dari kampanye pemasaran sebelumnya (*poutcome*). Sedangkan variabel tujuannya yaitu klien berlangganan deposito berjangka (*y*). Berikut adalah data pemasaran bank dari lembaga perbankan portugis pada tabel 3.1:

**Tabel 3.1. Sampel Data Set Pemasaran Bank (*UCI Repository*)**

| age | job          | marital  | education           | default | housing | Loan    | contact   | month | day_of_week | duration | campaign | pdays | previous | outcome     | emp.var.rate | cons.price.idx | cons.conf.idx | euribor3m | nr.employed | y  |
|-----|--------------|----------|---------------------|---------|---------|---------|-----------|-------|-------------|----------|----------|-------|----------|-------------|--------------|----------------|---------------|-----------|-------------|----|
| 30  | blue-collar  | married  | basic.9y            | no      | yes     | No      | cellular  | may   | fri         | 487      | 2        | 999   | 0        | nonexistent | -1.8         | 92.893         | -46.2         | 1.313     | 5099.1      | no |
| 39  | services     | single   | high.school         | no      | no      | No      | telephone | may   | fri         | 346      | 4        | 999   | 0        | nonexistent | 1.1          | 93.994         | -36.4         | 4.855     | 5191        | no |
| 25  | services     | married  | high.school         | no      | yes     | No      | telephone | jun   | wed         | 227      | 1        | 999   | 0        | nonexistent | 1.4          | 94.465         | -41.8         | 4.962     | 5228.1      | no |
| 38  | services     | married  | basic.9y            | no      | unknown | Unknown | telephone | jun   | fri         | 17       | 3        | 999   | 0        | nonexistent | 1.4          | 94.465         | -41.8         | 4.959     | 5228.1      | no |
| 47  | admin.       | married  | university.degree   | no      | yes     | No      | cellular  | nov   | mon         | 58       | 1        | 999   | 0        | nonexistent | -0.1         | 93.2           | -42           | 4.191     | 5195.8      | no |
| 32  | services     | Single   | university.degree   | no      | no      | No      | cellular  | sep   | thu         | 128      | 3        | 999   | 2        | failure     | -1.1         | 94.199         | -37.5         | 0.884     | 4963.6      | no |
| 32  | admin.       | Single   | university.degree   | no      | yes     | No      | cellular  | sep   | mon         | 290      | 4        | 999   | 0        | nonexistent | -1.1         | 94.199         | -37.5         | 0.879     | 4963.6      | no |
| 41  | entrepreneur | Married  | university.degree   | unknown | yes     | No      | cellular  | nov   | mon         | 44       | 2        | 999   | 0        | nonexistent | -0.1         | 93.2           | -42           | 4.191     | 5195.8      | no |
| 31  | services     | Divorced | professional.course | no      | no      | No      | cellular  | nov   | tue         | 68       | 1        | 999   | 1        | failure     | -0.1         | 93.2           | -42           | 4.153     | 5195.8      | no |

Sumber: Uci Repository (2012)

### 3.3. Pengolahan Awal Data

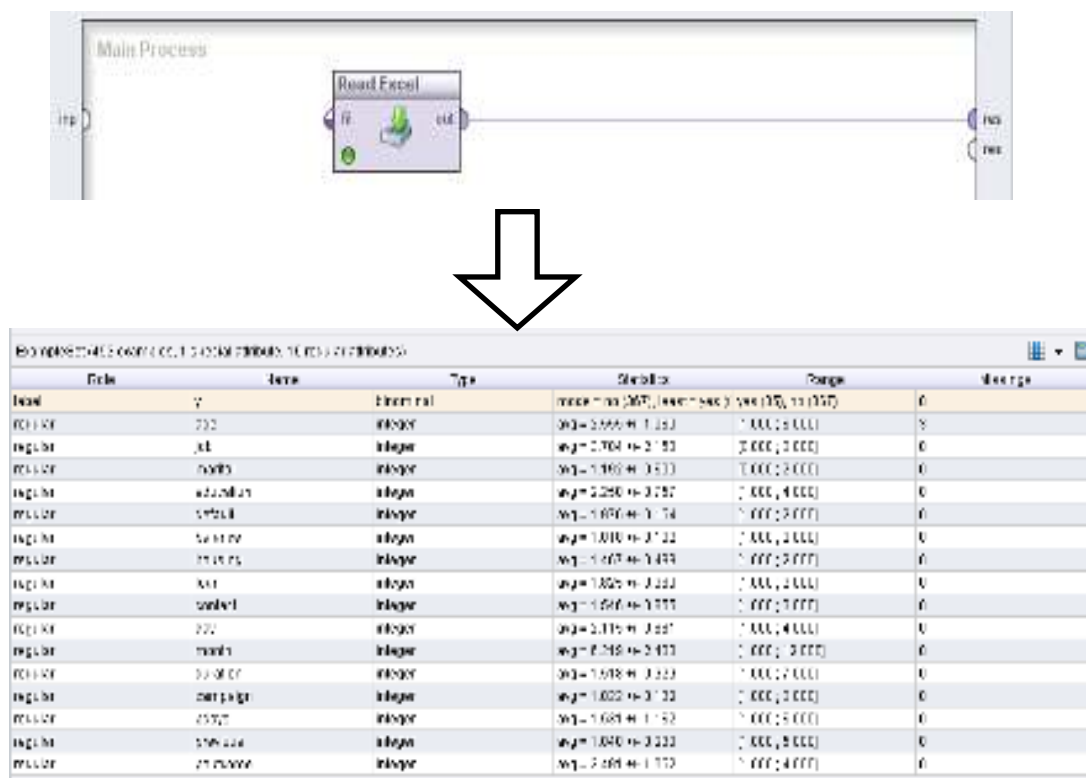
Jumlah data awal yang diperoleh dari pengumpulan data yaitu sebanyak 4.522 data. Pada dataset karena jumlah *record* aslinya terlalu besar, maka dataset hanya akan diambil *record* sebanyak 10%, yakni 452 data secara random. Untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan adalah sebagai berikut (Vercellis, 2009):

1. Data validation, untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*). Untuk mendapatkan data yang valid dan berkualitas maka dilakukan uji data validation untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*). Missing data terlihat pada tabel 3.2 dan hasilnya terlihat pada tabel 3.3.

Tabel 3.2. Ilustrasi *missing data* pada data *training*

| age | job | marital | education | default | balance | housing | loan | contact | day | month | duration | campaign | pdays | previous | outcome | y   |
|-----|-----|---------|-----------|---------|---------|---------|------|---------|-----|-------|----------|----------|-------|----------|---------|-----|
| 1   | 5   | 2       | 2         | 2       | 1       | 2       | 2    | 1       | 4   | 4     | 1        | 1        | 1     | 1        | 4       | yes |
| 5   | 4   | 0       | 2         | 2       | 1       | 2       | 2    | 2       | 2   | 7     | 3        | 1        | 1     | 1        | 4       | yes |
| 1   | 8   | 1       | 1         | 2       | 1       | 2       | 2    | 1       | 2   | 10    | 1        | 1        | 1     | 1        | 4       | no  |
|     | 6   | 1       | 2         | 2       | 1       | 1       | 1    | 1       | 2   | 5     | 1        | 1        | 5     | 1        | 1       | no  |
| 2   | 3   | 2       | 3         | 2       | 1       | 1       | 2    | 1       | 2   | 4     | 1        | 1        | 5     | 1        | 1       | no  |
| 2   | 3   | 2       | 3         | 2       | 1       | 1       | 2    | 1       | 3   | 8     | 4        | 1        | 1     | 1        | 4       | yes |
| 3   | 7   | 1       | 3         | 2       | 1       | 2       | 2    | 1       | 2   | 8     | 2        | 1        | 1     | 1        | 4       | yes |
|     | 4   | 0       | 1         | 2       | 1       | 2       | 2    | 2       | 3   | 10    | 1        | 1        | 1     | 1        | 4       | yes |
| 1   | 3   | 1       | 3         | 2       | 1       | 1       | 1    | 3       | 1   | 6     | 1        | 1        | 1     | 1        | 4       | no  |
| 2   | 2   | 1       | 2         | 2       | 1       | 1       | 2    | 1       | 2   | 11    | 1        | 1        | 1     | 1        | 4       | yes |
| 2   | 3   | 1       | 2         | 2       | 1       | 1       | 2    | 1       | 1   | 5     | 3        | 1        | 5     | 1        | 1       | yes |
| 4   | 2   | 1       | 2         | 2       | 1       | 1       | 2    | 3       | 1   | 5     | 1        | 1        | 1     | 1        | 4       | no  |
| 2   | 3   | 2       | 3         | 2       | 1       | 2       | 2    | 1       | 3   | 2     | 1        | 1        | 3     | 1        | 1       | no  |
|     | 3   | 1       | 3         | 2       | 1       | 1       | 2    | 1       | 2   | 5     | 2        | 1        | 5     | 1        | 2       | no  |
| 2   | 7   | 1       | 2         | 2       | 1       | 1       | 2    | 1       | 1   | 5     | 1        | 1        | 1     | 1        | 4       | no  |
| 3   | 3   | 1       | 3         | 2       | 1       | 1       | 2    | 3       | 2   | 5     | 1        | 1        | 1     | 1        | 4       | no  |
| 3   | 6   | 1       | 1         | 2       | 1       | 1       | 1    | 1       | 2   | 4     | 2        | 1        | 3     | 1        | 1       | no  |
| 2   | 6   | 1       | 2         | 2       | 1       | 1       | 2    | 3       | 3   | 5     | 1        | 1        | 1     | 1        | 4       | no  |

Sumber: UCI Repository (2012)



Sumber: Hasil Penelitian (2016)

### Gambar 3.1. Uji Data *Training* dengan RapidMiner

Data pemasaran bank sebanyak 452 record ternyata terdapat data *missing value* atau *noise*. Data yang tidak ada nilainya atau bernilai tidak konsisten tersebut dihilangkan sehingga dari 452 record diperoleh 449 record data valid. Dari 449 data, jumlah data pemasaran bank dengan output “no” yaitu sebanyak 365 data dan jumlah data pemasaran bank dengan output “yes” sebanyak 84 data. Sampel data training yang digunakan seperti terlihat pada tabel 3.3.

Tabel 3.3. Data *Training* setelah dilakukan teknik validasi

| age | job | marital | education | default | balance | housing | loan | contact | day | month | duration | campaign | pdays | previous | outcome | y   |
|-----|-----|---------|-----------|---------|---------|---------|------|---------|-----|-------|----------|----------|-------|----------|---------|-----|
| 1   | 5   | 2       | 2         | 2       | 1       | 2       | 2    | 1       | 4   | 4     | 1        | 1        | 1     | 1        | 4       | yes |
| 5   | 4   | 0       | 2         | 2       | 1       | 2       | 2    | 2       | 2   | 7     | 3        | 1        | 1     | 1        | 4       | yes |
| 1   | 8   | 1       | 1         | 2       | 1       | 2       | 2    | 1       | 2   | 10    | 1        | 1        | 1     | 1        | 4       | no  |
| 2   | 6   | 1       | 2         | 2       | 1       | 1       | 1    | 1       | 2   | 5     | 1        | 1        | 5     | 1        | 1       | no  |
| 2   | 3   | 2       | 3         | 2       | 1       | 1       | 2    | 1       | 2   | 4     | 1        | 1        | 5     | 1        | 1       | no  |
| 2   | 3   | 2       | 3         | 2       | 1       | 1       | 2    | 1       | 3   | 8     | 4        | 1        | 1     | 1        | 4       | yes |
| 3   | 7   | 1       | 3         | 2       | 1       | 2       | 2    | 1       | 2   | 8     | 2        | 1        | 1     | 1        | 4       | yes |
| 6   | 4   | 0       | 1         | 2       | 1       | 2       | 2    | 2       | 3   | 10    | 1        | 1        | 1     | 1        | 4       | yes |
| 1   | 3   | 1       | 3         | 2       | 1       | 1       | 1    | 3       | 1   | 6     | 1        | 1        | 1     | 1        | 4       | no  |
| 2   | 2   | 1       | 2         | 2       | 1       | 1       | 2    | 1       | 2   | 11    | 1        | 1        | 1     | 1        | 4       | yes |
| 2   | 3   | 1       | 2         | 2       | 1       | 1       | 2    | 1       | 1   | 5     | 3        | 1        | 5     | 1        | 1       | yes |
| 4   | 2   | 1       | 2         | 2       | 1       | 1       | 2    | 3       | 1   | 5     | 1        | 1        | 1     | 1        | 4       | no  |
| 2   | 3   | 2       | 3         | 2       | 1       | 2       | 2    | 1       | 3   | 2     | 1        | 1        | 3     | 1        | 1       | no  |
| 2   | 3   | 1       | 3         | 2       | 1       | 1       | 2    | 1       | 2   | 5     | 2        | 1        | 5     | 1        | 2       | no  |
| 2   | 7   | 1       | 2         | 2       | 1       | 1       | 2    | 1       | 1   | 5     | 1        | 1        | 1     | 1        | 4       | no  |
| 3   | 3   | 1       | 3         | 2       | 1       | 1       | 2    | 3       | 2   | 5     | 1        | 1        | 1     | 1        | 4       | no  |
| 3   | 6   | 1       | 1         | 2       | 1       | 1       | 1    | 1       | 2   | 4     | 2        | 1        | 3     | 1        | 1       | no  |
| 2   | 6   | 1       | 2         | 2       | 1       | 1       | 2    | 3       | 3   | 5     | 1        | 1        | 1     | 1        | 4       | no  |
| 3   | 1   | 1       | 2         | 2       | 1       | 1       | 2    | 1       | 2   | 4     | 1        | 1        | 1     | 1        | 4       | no  |

Sumber: UCI Repository (2012)

Jika data yang ada masih terdapat duplikasi atau anomaly maka terlebih dahulu dilakukan proses *replace missing* seperti gambar berikut:



Sumber: Hasil Penelitian (2016)

**Gambar 3.2. Model Desain *Replace Missing***

| Case | Age | Sex    | Marital | Occupation | Income | Education   |
|------|-----|--------|---------|------------|--------|-------------|
| 1    | 18  | Male   | Married | Unemployed | 10000  | High School |
| 2    | 25  | Female | Single  | Unemployed | 10000  | High School |
| 3    | 30  | Male   | Married | Unemployed | 10000  | High School |
| 4    | 35  | Female | Single  | Unemployed | 10000  | High School |
| 5    | 40  | Male   | Married | Unemployed | 10000  | High School |
| 6    | 45  | Female | Single  | Unemployed | 10000  | High School |
| 7    | 50  | Male   | Married | Unemployed | 10000  | High School |
| 8    | 55  | Female | Single  | Unemployed | 10000  | High School |
| 9    | 60  | Male   | Married | Unemployed | 10000  | High School |
| 10   | 65  | Female | Single  | Unemployed | 10000  | High School |
| 11   | 70  | Male   | Married | Unemployed | 10000  | High School |
| 12   | 75  | Female | Single  | Unemployed | 10000  | High School |
| 13   | 80  | Male   | Married | Unemployed | 10000  | High School |
| 14   | 85  | Female | Single  | Unemployed | 10000  | High School |
| 15   | 90  | Male   | Married | Unemployed | 10000  | High School |

Sumber: Hasil Penelitian (2016)

**Gambar 3.3. Hasil *Replace Missing***

Dari hasil diatas diketahui bahwa tidak terdapat *missing attribute* yang terjadi.

2. Data *integration and transformation*, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal. Data ditransformasikan kedalam *software Rapidminer*. Tabel kategorikal atribut terlihat pada table 3.4.

**Tabel 3.4 Atribut yang digunakan**

| No | Atribut | Nilai | Keterangan     |
|----|---------|-------|----------------|
| 1  | Age     | 1     | 18 <= ... < 30 |
|    |         | 2     | 30 <= ... < 40 |
|    |         | 3     | 40 <= ... < 50 |
|    |         | 4     | 50 <= ... < 60 |
|    |         | 5     | 60 <= ... < 70 |
|    |         | 6     | 70 <= ... < 80 |
|    |         | 7     | 80 <= ... < 90 |

| No | Atribut   | Nilai | Keterangan                              |
|----|-----------|-------|-----------------------------------------|
| 2  | Job       | 0     | unknown                                 |
|    |           | 1     | admin                                   |
|    |           | 2     | housemaid,blue-collar                   |
|    |           | 3     | management, entrepreneur, self-employed |
|    |           | 4     | retired                                 |
|    |           | 5     | pelajar                                 |
|    |           | 6     | services                                |
|    |           | 7     | technician                              |
|    |           | 8     | unemployed                              |
| 3  | Marital   | 0     | divorced                                |
|    |           | 1     | married                                 |
|    |           | 2     | single                                  |
| 4  | Education | 1     | primary                                 |
|    |           | 2     | secondary                               |
|    |           | 3     | tertiary                                |
|    |           | 4     | unknown                                 |
| 5  | Default   | 1     | yes                                     |
|    |           | 2     | no                                      |
| 6  | Balance   | 1     | -3000 <= ... < 10.000                   |
|    |           | 2     | 10.000 <= ... < 20.000                  |
|    |           | 3     | 20.000 <= ... < 30.000                  |
|    |           | 4     | 30.000 <= ... < 40.000                  |
|    |           | 5     | 40.000 <= ... < 50.000                  |
|    |           | 6     | 50.000 <= ... < 60.000                  |
|    |           | 7     | 60.000 <= ... < 70.000                  |
|    |           | 8     | 70.000 <= ... < 80.000                  |
| 7  | Housing   | 1     | yes                                     |
|    |           | 2     | no                                      |
| 8  | Loan      | 1     | yes                                     |
|    |           | 2     | no                                      |
| 9  | Contact   | 1     | cellular                                |
|    |           | 2     | telephone                               |
|    |           | 3     | unknown                                 |
| 10 | Day       | 1     | 1 <= ... <10                            |
|    |           | 2     | 10 <= ... <20                           |
|    |           | 3     | 20 <= ... <30                           |
|    |           | 4     | >=30                                    |
| 11 | Month     | 1     | Januari                                 |
|    |           | 2     | Februari                                |
|    |           | 3     | Maret                                   |
|    |           | 4     | April                                   |



| No | Atribut  | Nilai | Keterangan        |
|----|----------|-------|-------------------|
|    |          | 5     | Mei               |
|    |          | 6     | Juni              |
|    |          | 7     | Juli              |
|    |          | 8     | Agustus           |
|    |          | 9     | September         |
|    |          | 10    | Oktober           |
|    |          | 11    | November          |
|    |          | 12    | Desember          |
| 12 | Duration | 1     | ... < 300         |
|    |          | 2     | 300 <= .. < 600   |
|    |          | 3     | 600 <= .. < 900   |
|    |          | 4     | 900 <= .. < 1200  |
|    |          | 5     | 1200 <= .. < 1500 |
|    |          | 6     | 1500 <= .. < 1800 |
|    |          | 7     | 1800 <= .. < 2100 |
|    |          | 8     | 2100 <= .. < 2400 |
|    |          | 9     | 2400 <= .. < 2700 |
|    |          | 10    | 2700 <= .. < 3000 |
|    |          | 11    | >= 3000           |
| 13 | Campaign | 1     | .. < 10           |
|    |          | 2     | 10 <= .. < 20     |
|    |          | 3     | 20 <= .. < 30     |
|    |          | 4     | 30 <= .. < 40     |
|    |          | 5     | >= 40             |
| 14 | Pdays    | 1     | -1                |
|    |          | 2     | .. < 100          |
|    |          | 3     | 100 <= .. < 200   |
|    |          | 4     | 200 <= .. < 300   |
|    |          | 5     | 300 <= .. < 400   |
|    |          | 6     | 400 <= .. < 500   |
|    |          | 7     | 500 <= .. < 600   |
|    |          | 8     | 600 <= .. < 700   |
|    |          | 9     | 700 <= .. < 800   |
|    |          | 10    | >= 800            |
| 15 | Previous | 1     | .. < 5            |
|    |          | 2     | 5 <= .. < 10      |
|    |          | 3     | 10 <= .. < 15     |
|    |          | 4     | 15 <= .. < 20     |
|    |          | 5     | >= 20             |

| No | Atribut  | Nilai | Keterangan |
|----|----------|-------|------------|
| 16 | Poutcome | 1     | failure    |
|    |          | 2     | other      |
|    |          | 3     | success    |
|    |          | 4     | unknown    |
| 17 | Y        | 1     | yes        |
|    |          | 2     | no         |

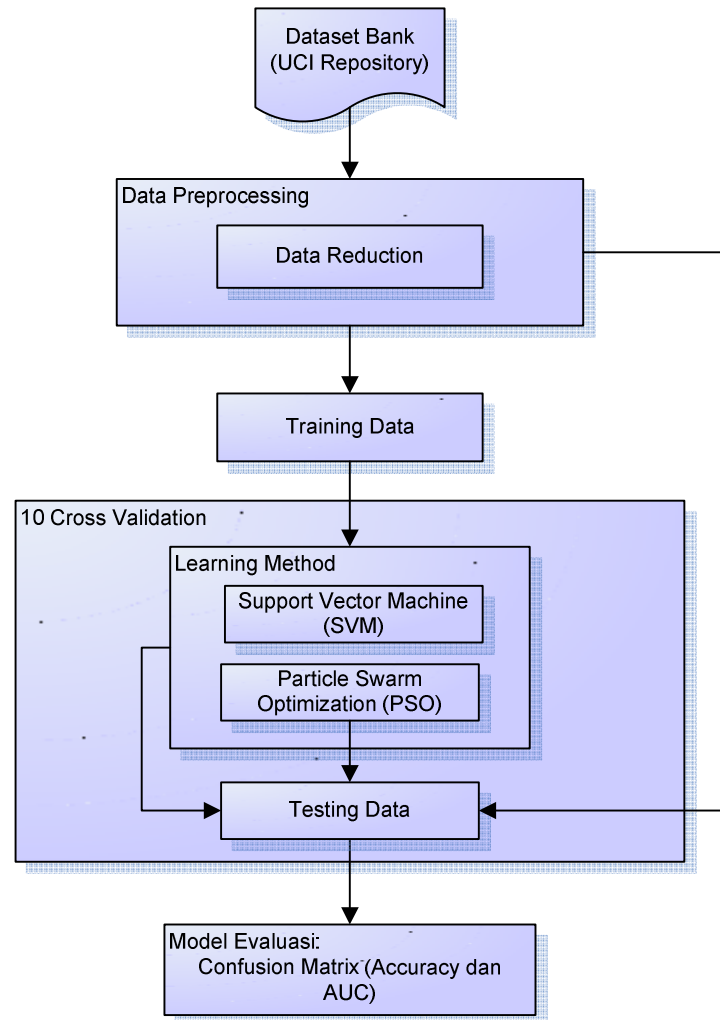
Sumber: Hasil Penelitian (2016)

3. *Data size reduction and discrization*, teknik ini digunakan untuk mereduksi ukuran data jika dataset yang digunakan cukup besar yang berguna untuk membuat model pembelajaran algoritma menjadi lebih efisien tanpa mengurangi kualitas dari hasil yang diperoleh, dengan kata lain untuk memperoleh data set dengan jumlah atribut dan *record* yang lebih sedikit tetapi bersifat informative.

### 3.4. Model yang Diusulkan

Model yang diusulkan pada penelitian ini menggunakan model pada data *training* yang sudah diseleksi variabelnya. Kemudian, data training tersebut diuji dengan model yang digunakan yaitu Algoritma *Support Vector Machine* dan *Support Vector Machine* berbasis *Particle swarm optimization* akan menghasilkan akurasi dalam proses prediksi pemasaran langsung.

Pengujian yang dilakukan terhadap data *training* dengan metode tersebut dilakukan dengan teknik *10 folds cross validation*. Hal ini dilakukan dengan tujuan menghasilkan akurasi yang paling tinggi pada hasil uji-t metode tersebut. Hasil akurasi tersebut dapat digunakan sebagai nilai yang paling baik setelah teknik *10 folds cross validation*. Semua *dataset* diuji dengan metode/model yang diusulkan pada aplikasi RapidMiner. Eksperimen dilakukan satu persatu terhadap *dataset* dan model yang digunakan.



Sumber: Hasil Penelitian (2016)

**Gambar 3.4. Model yang Diusulkan**

### 3.5. Eksperimen dan Pengujian Metode

Tahap modeling untuk menyelesaikan prediksi pemasaran langsung dengan menggunakan dua metode yaitu algoritma *support vector machine* dan algoritma *particle swarm optimization*.

1. *Support vector machine* yaitu model untuk menemukan *hyperplane* terbaik yang memisahkan dua buah class.
2. *Particle Swarm Optimization* yaitu pencarian solusi optimal secara global dalam ruang pencarian melalui interaksi individu dalam segerombolan partikel dengan cara melakukan seleksi terhadap atribut yang ada.

Dalam penelitian eksperimen ini digunakan spesifikasi *software* dan *hardware* sebagai alat bantu dalam penelitian ini, dapat dilihat pada Tabel 3.5 sebagai berikut:

**Tabel 3.5 Spesifikasi Perangkat Keras dan Perangkat Lunak**

| Spesifikasi        | Tipe                                      |
|--------------------|-------------------------------------------|
| Processor          | Intel (R) Celeron (R) CPU 1007U @ 1.50GHz |
| Memori             | 4 GB                                      |
| Monitor            | Generic PnP Monitor                       |
| Sistem Operasi     | Windows 7                                 |
| Aplikasi olah data | RapidMiner 5.3                            |

### 3.6. Evaluasi dan Validasi

Setelah tahap modelling dilakukan, hasil dari model Algoritma *Support Vector Machine* dan *Support Vector Machine* berbasis *Particle Swarm Optimization* ini kemudian akan dilihat tingkat akurasi menggunakan *Confusion matrix* dan nilai AUC (*Area Under Curve*) untuk mengetahui model/metode yang memiliki tingkat akurasi yang paling tinggi, sehingga tujuan penerapan model Algoritma *Support Vector Machines* dan *Support Vector Machine* berbasis *Particle Swarm Optimization* untuk klasifikasi pada prediksi pemasaran langsung akan tercapai.

Dari hasil akurasi yang paling tinggi menandakan bahwa model tersebut yang diuji menghasilkan model baru yang paling baik. Model tersebut yang dianggap sementara menjadi model yang paling baik untuk menghasilkan klasifikasi dalam penentuan prediksi pemasaran langsung. Proses selanjutnya adalah proses pengujian model tersebut terhadap data *training* yang menggunakan variabel yang sudah diseleksi. Hasil dari proses ini akan menghasilkan sebuah nilai AUC (*Area Under Curve*) dan angka yang berisi *False Positif*, *False Negatif*, *True Positif*, dan *True Negatif*. Hasil tersebut dikenal dengan *confusion matrix* dimana akan terdapat tabel dengan keadaan klasifikasi prediksi pemasaran langsung tersebut untuk mendapatkan persentase ketepatan prediksi/akurasi.

Setelah diperoleh nilai AUC, nilai tersebut digunakan sebagai penentuan metode yang menghasilkan akurasi yang paling tinggi. Semua nilai AUC pada *dataset* akan dibandingkan. Hasil AUC akan dimasukkan ke dalam sebuah tabel dan kemudian akan disimpulkan metode mana yang memiliki akurasi paling tinggi dalam klasifikasi prediksi pemasaran langsung.

Sedangkan nilai confusion matrix akan digunakan untuk menghitung performa prediksi. Performa prediksi pemasaran langsung menggunakan perhitungan nilai *precision*. *Precision* digunakan untuk mendapatkan persentase nilai *True* terhadap nilai *True* dan nilai yang prediksi *True* tapi kenyataannya *False*.

Pada dataset yang memiliki kesalahan prediksi yang sedikit akan menghasilkan nilai *precision* yang besar. Setelah diperoleh semua nilai performa prediksi pada semua dataset, kemudian nilai-nilai tersebut dibandingkan. Nilai performa prediksi yang paling tinggi dianggap sebagai yang paling baik. Kesimpulan dari nilai performa prediksi yang paling tinggi menandakan bahwa model tersebut merupakan model yang paling cocok untuk klasifikasi penilaian prediksi pemasaran langsung.

## BAB IV

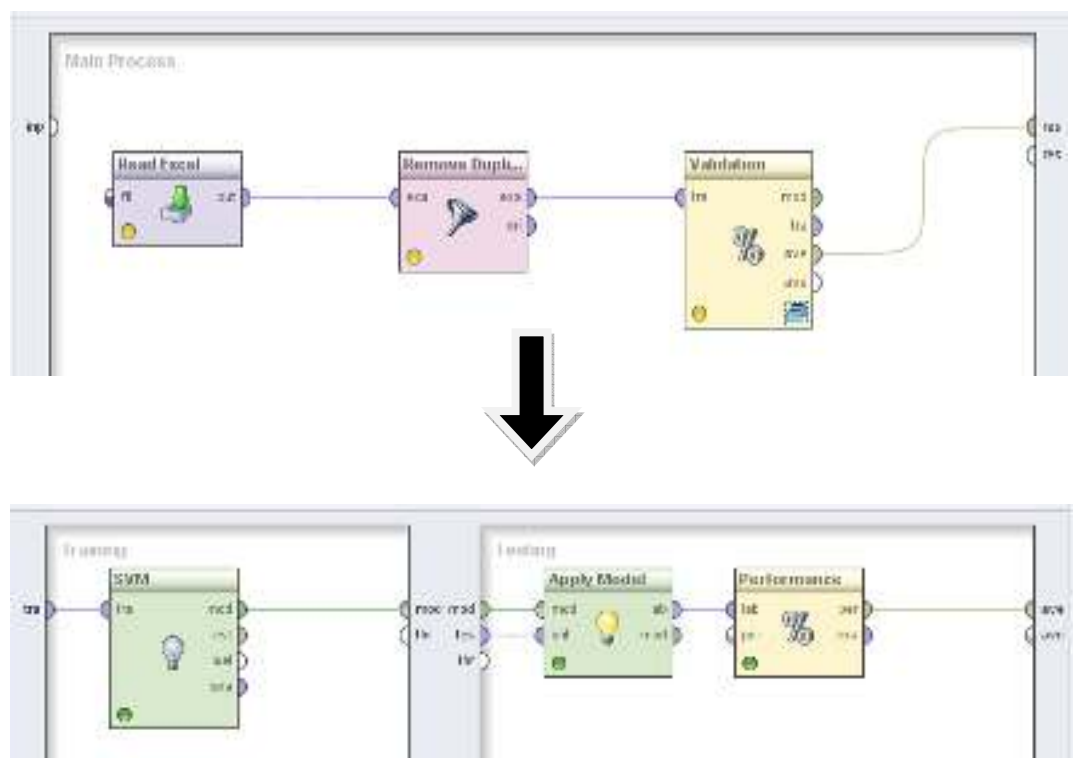
### HASIL PENELITIAN DAN PEMBAHASAN

#### 4.1. Hasil Eksperimen dan Pengujian Metode

Tujuan dari penelitian ini adalah mengembangkan model yang sudah terbentuk dengan klasifikasi algoritma *Support Vector Machine* (SVM) dan menerapkan *Particle Swarm Optimization* (PSO) dengan melakukan seleksi atribut pada *Support Vector Machine* (SVM) untuk meningkatkan akurasi keakuratan model prediksi pemasaran langsung.

##### 4.1.1. Metode *Support Vector Machine*

Berikut adalah gambar pengujian algoritma *Support Vector Machine* (SVM) menggunakan metode *K-Fold Cross Validation* dengan menggunakan RapidMiner:



Sumber: Hasil Penelitian (2016)

**Gambar 4.1** Pengujian *K-Fold Cross Validation* algoritma *Support Vector Machine*

Nilai *training cycles* dalam penelitian ini ditentukan dengan cara melakukan uji coba memasukkan C dan epsilon. Berikut ini adalah hasil dari percobaan yang telah dilakukan untuk penentuan nilai *training cycles*:

**Tabel 4.1 Eksperimen penentuan nilai training cycle SVM**

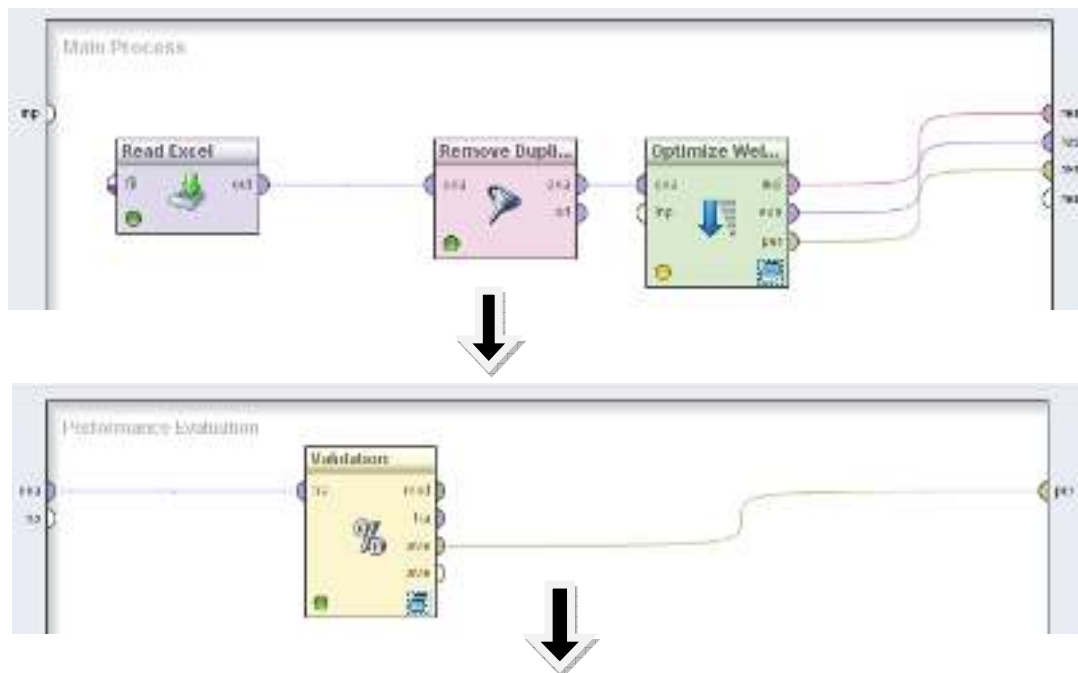
| C   | Epsilon | SVM      |       |
|-----|---------|----------|-------|
|     |         | Accuracy | AUC   |
| 0.0 | 0.0     | 88.50%   | 0.895 |
| 1.0 | 1.0     | 81.20%   | 0.500 |
| 1.0 | 0.0     | 88.71%   | 0.896 |

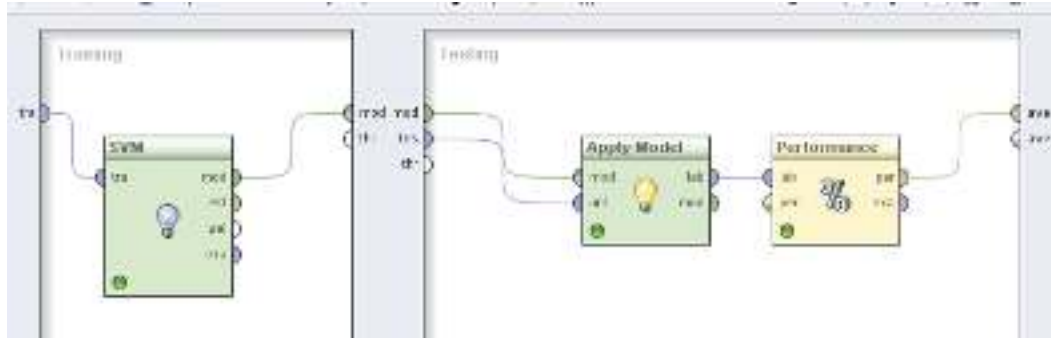
Sumber: Hasil Penelitian (2016)

Hasil terbaik pada eksperimen SVM diatas adalah C=1.0 dan Epsilon=0.0 dihasilkan *accuracy* 88.71% dan AUC 0.896. Lalu diikuti SVM dengan C=0.0 dan Epsilon=0.0 dihasilkan *accuracy* 88.50% dan AUC 0.895, kemudian SVM dengan C=1.0 dan Epsilon=1.0 dihasilkan *accuracy* 81.20% dan AUC 0.500.

#### 4.1.2. Metode *Support Vector Machine* berbasis *Particle Swarm Optimization*

Berikut adalah gambar pengujian algoritma *Support Vector Machine* berbasis *Particle swarm optimization* menggunakan metode *K-Fold Cross Validation* dengan menggunakan RapidMiner:





Sumber: Hasil Penelitian (2016)

**Gambar 4.2. Pengujian *K-Fold Cross Validation* algoritma *Support Vector Machine* berbasis PSO**

Nilai *training cycles* dalam penelitian ini ditentukan dengan cara melakukan uji coba memasukkan C dan epsilon. Selanjutnya dilakukan observasi terhadap variabel C dan  $\epsilon$  dari *Support Vector Machine* dan *Particle swarm optimization*. Berikut ini adalah hasil dari percobaan yang telah dilakukan untuk penentuan nilai *training cycles*:

**Tabel 4.2 Eksperimen penentuan nilai training cycle SVM berbasis PSO**

| C   | Epsilon | SVM-PSO  |       |
|-----|---------|----------|-------|
|     |         | Accuracy | AUC   |
| 0.0 | 0.0     | 88.53%   | 0.897 |
| 1.0 | 1.0     | 81.20%   | 0.500 |
| 1.0 | 0.0     | 89.38%   | 0.909 |

Sumber: Hasil Penelitian (2016)

Hasil terbaik pada eksperimen SVM berbasis PSO diatas adalah dengan C= 1.0 dan Epsilon = 0.0 yang diperoleh dengan hasil accuracy 89.38% dan AUCnya 0.909. maka nilai-nilai tersebut yang akan digunakan dalam penelitian ini.

Langkah selanjutnya adalah menyeleksi atribut yang digunakan yaitu *age*, *job*, *marital*, *education*, *default*, *balance*, *housing*, *loan*, *contact*, *day*, *month*, *duration*, *campaign*, *pdays*, *previous*, *poutcome* dan 1 atribut sebagai label yaitu Y. Dari hasil eksperimen dengan menggunakan algoritma *support vector machine* berbasis *particle swarm optimization* diperoleh hasil seperti dalam Tabel 4.3 dibawah ini.



**Tabel 4.3 Hasil Seleksi Atribut**

| No | Attribute | weight |
|----|-----------|--------|
| 1  | age       | 0.415  |
| 2  | job       | 1      |
| 3  | marital   | 1      |
| 4  | education | 0      |
| 5  | default   | 0.387  |
| 6  | balance   | 0.386  |
| 7  | housing   | 0      |
| 8  | loan      | 1      |
| 9  | contact   | 0.781  |
| 10 | day       | 1      |
| 11 | month     | 0      |
| 12 | duration  | 1      |
| 13 | campaign  | 0.020  |
| 14 | pdays     | 0.559  |
| 15 | previous  | 0.159  |
| 16 | poutcome  | 0      |

Sumber: Hasil Penelitian (2016)

Dari 16 variabel prediktor dilakukan seleksi atribut sehingga menghasilkan terpilihnya 12 atribut yang digunakan, yaitu: *age*, *job*, *marital*, *default*, *balance*, *loan*, *contact*, *day*, *duration*, *campaign*, *pdays* dan *previous*. Sedangkan atribut lainnya seperti: *education*, *housing*, *month*, dan *poutcome*, tidak berpengaruh terhadap bobot atribut.

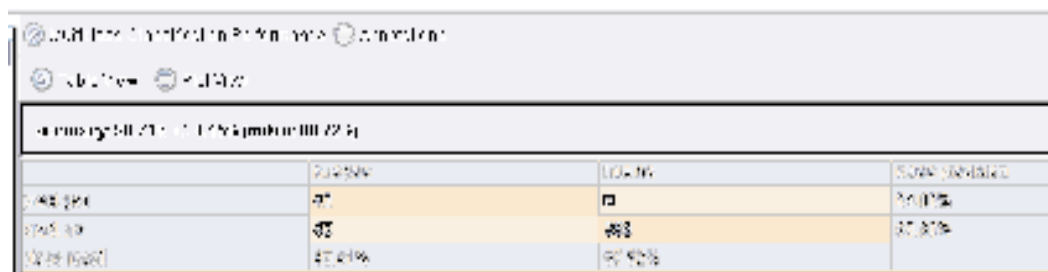
#### 4.2. Evaluasi dan Validasi Hasil

Hasil dari pengujian model yang dilakukan adalah memprediksi pemasaran langsung dengan *support vector machine* berbasis *particle swarm optimization* untuk menentukan nilai *accuracy* dan *AUC (Area Under Curve)*. Model klasifikasi bisa dievaluasi berdasarkan kriteria seperti tingkat akurasi, kecepatan, kehandalan, skalabilitas dan interpretabilitas (Vecellis, 2009). Setelah data diolah maka dapat diuji tingkat akurasinya untuk melihat kinerja dari masing-masing model.

#### 4.2.1. Hasil Pengujian Metode *Support Vector Machine*

##### 1. *Confusion Matrix*

Gambar 4.3 menunjukkan hasil dari *confusion matrix* metode *support vector machine*. Berdasarkan Gambar 4.3 dapat diketahui bahwa dari 452 data, 42 data diprediksikan *yes* sesuai dengan prediksi yang dilakukan dengan metode SVM, kemudian 8 data diprediksi *yes* tetapi ternyata hasilnya prediksi *no*. Kemudian 359 data *class no* diprediksi sesuai dengan prediksi yang dilakukan dengan metode SVM, dan 43 data diprediksi *no* tetapi ternyata hasil prediksinya *yes*.



|               | Actual Yes | Actual No | Sum |
|---------------|------------|-----------|-----|
| Predicted Yes | 42         | 8         | 50  |
| Predicted No  | 8          | 359       | 367 |
| Sum           | 50         | 367       | 417 |

Sumber: Hasil Penelitian (2016)

#### Gambar 4.3 Hasil Pengujian *Confusion Matrix* untuk Metode SVM

Tingkat akurasi yang dihasilkan dengan algoritma *Support Vector Machine* adalah sebesar 88.71% dan dapat dihitung untuk mencari nilai *accuracy*, *sensitivity*, *specificity*, *ppv* dan *npv* pada persamaan dibawah ini:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{42 + 359}{42 + 359 + 43 + 8} = 0.8871 = 88.71\%$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{42}{42 + 8} = 0.84 = 84.00\%$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{359}{359 + 43} = 0.8930 = 89.30\%$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{42}{42 + 43} = 0.4941 = 49.41\%$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} = \frac{359}{359 + 8} = 0.9782 = 97.82\%$$

Hasil perhitungan terlihat pada tabel 4.4 dibawah ini:

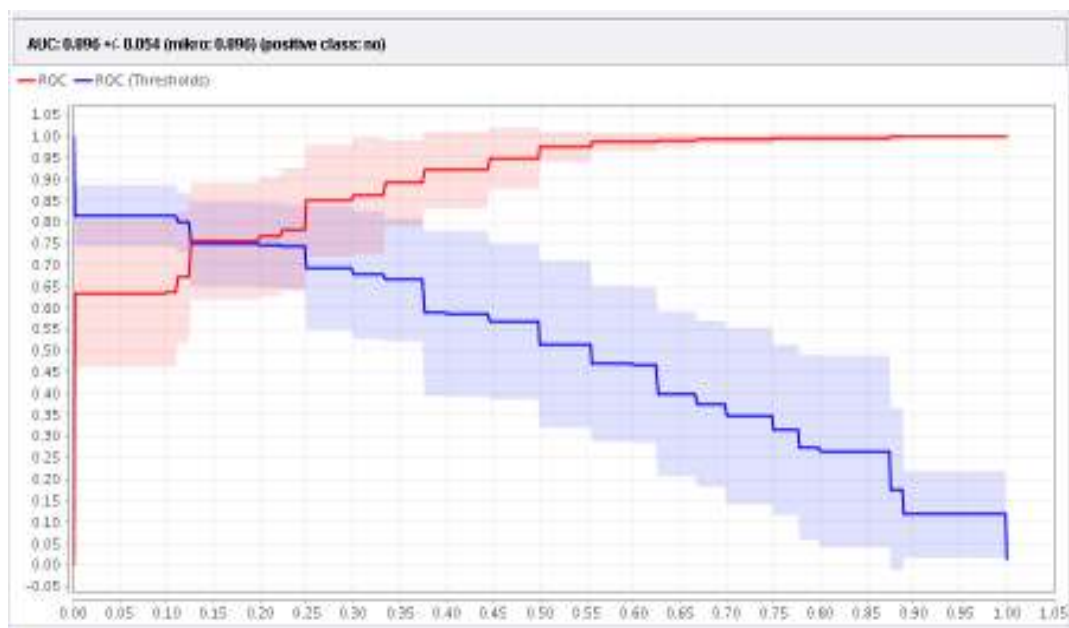
**Tabel 4.4 Nilai *Accuracy*, *sensitivity*, *specificity*, PPV dan NPV metode SVM**

|                    | Nilai (%) |
|--------------------|-----------|
| <i>Accuracy</i>    | 88.71     |
| <i>Sensitivity</i> | 84.00     |
| <i>Specificity</i> | 89.30     |
| PPV                | 49.41     |
| NPV                | 97.82     |

Sumber: Hasil Penelitian (2016)

## 2. Kurva ROC

Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan kedua class bisa dilihat pada Gambar 4.5 yang merupakan kurva ROC untuk algoritma *Support Vector Machine*. Kurva ROC pada gambar 4.4 mengekspresikan *confusion matrix* dari gambar 4.3. Garis horizontal adalah *false positives* dan garis vertikal *true positives*. Menghasilkan nilai AUC (*Area Under Curve*) sebesar 0.896 dengan nilai akurasi klasifikasi baik (*good classification*).



Sumber: Hasil Penelitian (2016)

**Gambar 4.4 Kurva ROC dengan Metode *Support Vector Machine***

#### 4.2.2. Hasil Pengujian Metode *Support Vector Machine* berbasis *Particle Swarm Optimization*

##### 1. *Confusion Matrix*

Gambar 4.5 menunjukkan hasil dari *confusion matrix* metode *support vector machine* berbasis *particle swarm optimization*.

|               | Actual Yes | Actual No |     |
|---------------|------------|-----------|-----|
| Predicted Yes | 44         | 41        | 85  |
| Predicted No  | 7          | 360       | 367 |
|               | 51         | 401       |     |

Sumber: Hasil Penelitian (2016)

**Gambar 4.5 Hasil Pengujian *Confusion Matrix* untuk Metode SVM berbasis PSO**

Berdasarkan Gambar 4.5 dapat diketahui bahwa dari 452 data, 44 diklasifikasikan *yes* sesuai dengan prediksi yang dilakukan dengan metode SVM berbasis *Particle Swarm Optimization* (PSO), kemudian 7 data di prediksi *yes* tetapi ternyata hasil prediksi *no*, 360 data *class no* diprediksi sesuai, dan 41 data diprediksi *no* tetapi ternyata hasil prediksinya *yes*.

Tingkat akurasi yang dihasilkan dengan algoritma *Support Vector Machine* adalah sebesar 89.38% dan dapat dihitung untuk mencari nilai *accuracy*, *sensitifity*, *specificity*, *ppv* dan *npv* pada persamaan dibawah ini:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{44 + 360}{44 + 360 + 41 + 7} = 0.8938 = 89.38\%$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{44}{44 + 7} = 0.8627 = 86.27\%$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{360}{360 + 41} = 0.8978 = 89.78\%$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{44}{44 + 41} = 0.5176 = 51.76\%$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} = \frac{360}{360 + 7} = 0.9809 = 98.09\%$$

Hasil perhitungan terlihat pada tabel 4.5 dibawah ini:

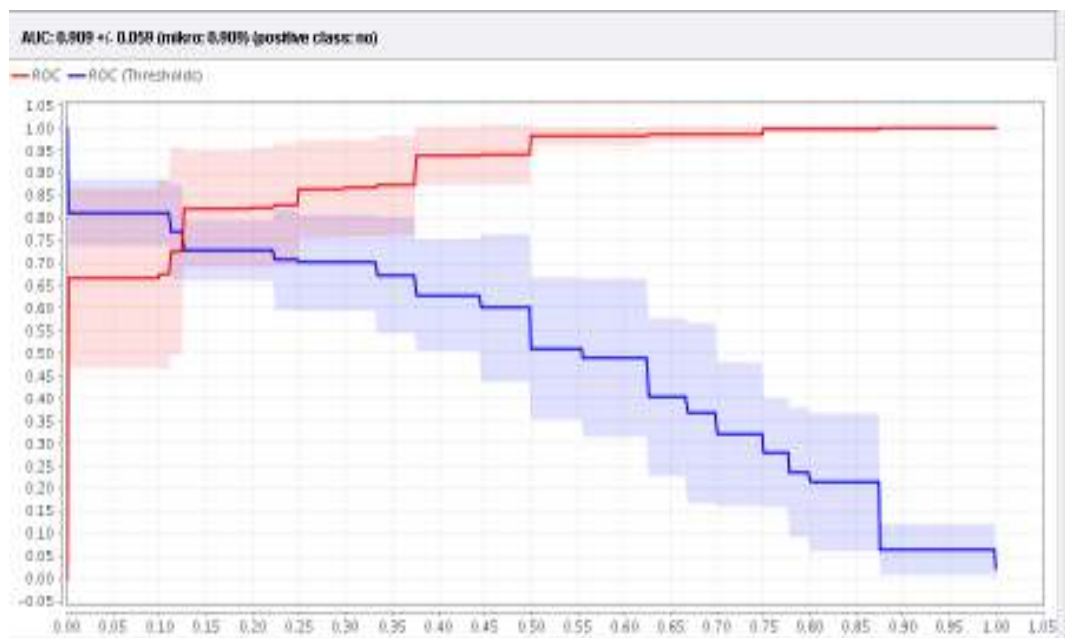
**Tabel 4.5 Nilai *Accuracy*, *sensitivity*, *specificity*, PPV dan NPV metode SVM berbasis PSO**

|                    | Nilai (%) |
|--------------------|-----------|
| <i>Accuracy</i>    | 89.38     |
| <i>Sensitivity</i> | 86.27     |
| <i>Specificity</i> | 89.78     |
| PPV                | 51.76     |
| NPV                | 98.09     |

Sumber: Hasil Penelitian (2016)

## 2. Kurva ROC

Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan kedua class bisa dilihat pada Gambar 4.6 yang merupakan kurva ROC untuk algoritma *Support Vector Machine* berbasis PSO. Kurva ROC pada gambar 4.6 mengekspresikan *confusion matrix* dari tabel 4.6. Garis horizontal adalah *false positives* dan garis vertikal *true positives*. Menghasilkan nilai AUC (*Area Under Curve*) sebesar 0.909 dengan nilai akurasi klasifikasi sangat baik (*excellent classification*).



Sumber: Hasil Penelitian (2016)

**Gambar 4.6 Kurva ROC dengan Metode *Support Vector Machine* berbasis PSO**

### 3. *Attribute Weight*

Hasil *Attribute Weight* yang didapat dari penelitian ini adalah:

| atribut        | weight |
|----------------|--------|
| sex            | 0.1    |
| age            | 0.1    |
| education      | 0.1    |
| marital-status | 0.1    |
| occupation     | 0.1    |
| relationship   | 0.1    |
| income         | 0.1    |
| workclass      | 0.1    |
| hours-per-week | 0.1    |
| native-born    | 0.1    |
| foreign-born   | 0.1    |
| race           | 0.1    |
| ethnicity      | 0.1    |
| religion       | 0.1    |
| politics       | 0.1    |
| party          | 0.1    |
| ideology       | 0.1    |

Sumber: Hasil Penelitian (2016)

**Gambar 4.7 *Attribute Weight* dengan Metode *Support Vector Machine* berbasis *Particle Swarm Optimization***

#### 4.2.3. Analisa Evaluasi dan Validasi Model

Hasil analisis dari metode algoritma *Support Vector Machine* dan *Support Vector Machine* berbasis *Particle Swarm Optimization* dirangkumkan dalam tabel dibawah ini.

**Tabel 4.6 Komparasi Nilai *Accuracy* dan AUC**

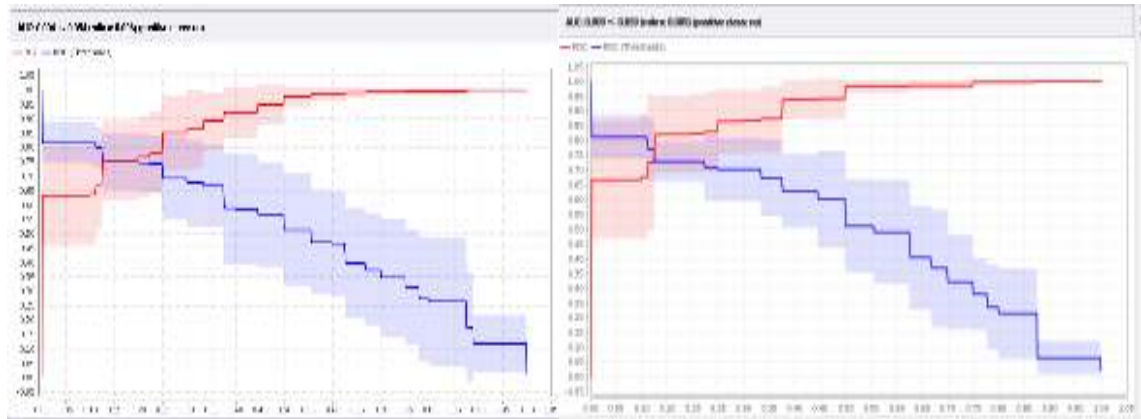
|                 | <i>Support Vector Machine</i> | <i>Support Vector Machine</i> Berbasis PSO | Peningkatan |
|-----------------|-------------------------------|--------------------------------------------|-------------|
| <i>Accuracy</i> | 88.71%                        | 89.38 %                                    | 0.67%       |
| AUC             | 0.896                         | 0.909                                      | 0.013       |

Sumber: Hasil Penelitian (2016)

Tabel 4.6 membandingkan *Accuracy* dan AUC dari tiap metode. Terlihat bahwa nilai *accuracy* dan AUC *Support Vector Machine* berbasis *Particle Swarm Optimization* lebih tinggi dibandingkan *Support Vector Machine* tunggal. Penerapan *Particle Swarm Optimization* untuk seleksi atribut menghasilkan peningkatan akurasi sebesar 0.67%.

Untuk evaluasi menggunakan ROC *curve* sehingga menghasilkan nilai AUC (*Area Under Curve*) untuk model algoritma SVM menghasilkan nilai 0.896 dengan nilai akurasi *Good Classification*, sedangkan untuk algoritma SVM

berbasis PSO (*Particle Swarm Optimization*) menghasilkan nilai 0.909 dengan nilai akurasi *Excellent Classification*, dan selisih nilai keduanya sebesar 0.013. Dapat dilihat pada Gambar 4.7 dibawah ini.



Sumber: Hasil Penelitian (2016)

**Gambar 4.8 Kurva ROC dengan *support vector machine* dan *support vector machine* berbasis *particle swarm optimization***

#### 4.3. Pembahasan

Berdasarkan hasil eksperimen dan analisis data dalam penelitian ini, maka dapat diperoleh perbandingan nilai akurasi tertinggi pada pengujian model SVM sebelum dan sesudah dilakukan optimasi parameter dengan algoritma PSO pada data testing yang diterapkan pada prediksi pemasaran langsung.

Pada penelitian ini *Support Vector Machine* (SVM) digunakan karena diketahui dari hasil penelitian sebelumnya bahwa SVM memiliki kemampuan generalisasi yang sangat baik untuk memecahkan masalah walaupun dengan sampel terbatas. Eksperimen menggunakan metode *support vector machine* menghasilkan tingkat akurasi sebesar 88.71% dan mempunyai nilai AUC sebesar 0.896. Dari hasil tersebut diketahui bahwa keberhasilan *support vector machine* (SVM) sangat dipengaruhi oleh pemilihan atribut yang tepat. Semakin banyak atribut dan informasi yang digunakan akan mengakibatkan banyaknya waktu dan biaya yang dikorbankan bahkan akan mengurangi tingkat akurasi dan kompleksitas yang lebih tinggi.

Mengingat pentingnya seleksi atribut dalam *Support Vector Machine* (SVM) maka diterapkan *Particle swarm optimization* (PSO) untuk melakukan

tugas tersebut. *Particle Swarm Optimization* (PSO) diketahui dapat digunakan sebagai teknik optimasi untuk mengoptimalkan subset fitur. Algoritma PSO sederhana dan memiliki kompleksitas yang lebih rendah, sehingga dapat memastikan solusi optimal dengan menyesuaikan pencarian global dan lokal, sehingga kinerja klasifikasi *Support Vector Machine* (SVM) dapat ditingkatkan. Eksperiment dilakukan kembali dengan menerapkan *Particle swarm optimization* (PSO) untuk seleksi atribut dalam *Support Vector Machine* (SVM) dan dilakukan penyesuaian pada parameter C dan  $\epsilon$ . Dari 16 variabel prediktor dilakukan seleksi atribut sehingga menghasilkan terpilihnya 12 atribut yang digunakan. Hasil eksperimen menunjukkan akurasi sebesar 89.38% dan nilai AUC sebesar 0.909. Hasil dari *eksperiment* tersebut menunjukkan pengujian data pemasaran langsung data set menggunakan *Support Vector Machine* (SVM) dan penerapan *Particle Swarm Optimization* dalam pemilihan atribut menunjukkan hasil yang lebih akurat dalam penentuan prediksi pemasaran langsung ditandai dengan peningkatan nilai akurasi sebesar 0.67 % dan nilai AUC sebesar 0.013.

#### 4.4. Implikasi Penelitian

Dari hasil pengujian diatas, dengan dilakukan evaluasi baik secara *confusion matrix* maupun *ROC curve* ternyata terbukti bahwa pengujian yang dilakukan algoritma *Support Vector Machine* berbasis *Particle Swarm Optimization* (PSO) memiliki nilai akurasi yang lebih tinggi dibandingkan hanya dengan menggunakan algoritma *Support Vector Machine* (SVM).

Hasil dari penelitian ini mempunyai implikasi terhadap pemasaran langsung yaitu meliputi implikasi terhadap aspek sistem pendukung keputusan pemasaran langsung, aspek manajerial, dan aspek terhadap penelitian-penelitian selanjutnya yang akan diuraikan dibawah ini:

##### 1. Implikasi terhadap aspek sistem pendukung keputusan pemasaran langsung

Hasil evaluasi menunjukkan penerapan *Particle swarm optimization* (PSO) untuk seleksi atribut dalam *Support Vector Machine* (SVM) dan dilakukan penyesuaian pada parameter C dan  $\epsilon$  mampu menunjukkan aspek berupa atribut yang berpengaruh terhadap hasil dari prediksi pemasaran langsung. Dari hasil ini menunjukkan bahwa metode *Support Vector Machine* berbasis *Particle swarm*



*optimization* merupakan metode yang cukup baik dalam pengklasifikasian data. Dengan demikian metode *Support Vector Machine* berbasis *Particle swarm optimization* dapat memberikan pemecahan permasalahan analisa penilaian prediksi pemasaran langsung dan dapat mendukung pengambilan keputusan serta pengembangan sistem informasi manajemen pada lembaga keuangan dan perbankan dengan menggunakan bantuan *software* RapidMiner.

## 2. Implikasi terhadap aspek manajerial

Dari hasil penelitian ini diketahui bahwa metode *Support Vector Machine* berbasis *Particle swarm optimization* dapat mendukung pengambilan keputusan dan pengembangan sistem informasi manajemen pada lembaga keuangan dan perbankan dengan menggunakan bantuan *software* RapidMiner, untuk itu diperlukan peningkatan kemampuan manajerial dalam hal pemasaran langsung agar mampu membuat perencanaan marketing dengan perencanaan yang baik.

## 3. Implikasi terhadap penelitian-penelitian selanjutnya

Dalam penelitian ini menggunakan metode *Support Vector Machine* (SVM) berbasis *Particle swarm optimization*, untuk mengetahui kehandalan metode, maka pada penelitian selanjutnya dapat dilakukan penggunaan data set lebih dari satu. Penelitian seperti ini dapat dikembangkan pada unit bisnis serupa atau yang lain. Pengembangan dapat dilakukan dengan pemilihan parameter menggunakan metode *Genetic Algorithm* untuk meningkatkan hasil optimasi. Penggunaan algoritma yang lain juga dapat digunakan misalkan dengan metode *Neural Network*, *C4.5*, *K-Nearest Neighbor* dan sebagainya. Karena lembaga keuangan semakin berkembang maka kajian semacam ini dapat dilakukan secara periodik.

## **BAB V**

### **PENUTUP**

#### **5.1. Kesimpulan**

Dalam penelitian ini dilakukan pengujian model dengan menggunakan *Support Vector Machine* dan *Support Vector Machine* berbasis *Particle Swarm Optimization* dengan menggunakan data yang terkait kampanye pemasaran langsung dari lembaga perbankan Portugis yang diperoleh dari UCI Repository tahun 2012. Model yang dihasilkan diuji untuk mendapatkan nilai *accuracy* dan AUC dari setiap algoritma sehingga didapat pengujian dengan menggunakan *Support Vector Machine* didapat nilai *accuracy* adalah 88.71% dan nilai AUC adalah 0.896. Sedangkan pengujian dengan menggunakan *Support Vector Machine* berbasis *Particle Swarm Optimization* dilakukan seleksi atribut dan penyesuaian pada parameter C dan  $\epsilon$ . Dari 16 variabel prediktor dilakukan seleksi atribut sehingga menghasilkan terpilihnya 12 atribut yang digunakan. didapatkan nilai *accuracy* 89.38% dan nilai AUC adalah 0.909 Maka dapat disimpulkan pengujian dataset bank (pemasaran langsung) menggunakan *Support Vector Machine* dan penerapan *Particle Swarm Optimization* dalam pemilihan atribut didapat bahwa metode tersebut lebih akurat dalam penentuan prediksi pemasaran langsung dibandingkan dengan metode *Support Vector Machine* tunggal, ditandai dengan peningkatan nilai akurasi sebesar 0.67% dan nilai AUC sebesar 0.013, dengan nilai tersebut masuk kedalam klasifikasi akurasi sangat baik (*excellent classification*).

#### **5.2. Saran**

Berdasarkan proses pengujian dan kesimpulan yang telah dilakukan, maka ada beberapa saran dalam penelitian ini adalah:

1. Menambahkan jumlah data yang lebih besar dan atribut yang lebih banyak, sehingga hasil pengukuran yang akan didapatkan lebih baik lagi.
2. Mengembangkan metode SVM yang dipadukan dengan lebih dari satu kernel (*multi kernel*) dengan tujuan meningkatkan tingkat akurasi klasifikasi data dengan metode SVM.

3. Menggunakan metode optimasi lain seperti *Adaboost*, *Ant Colony Optimization* (ACO), *Genetik Algorithm* (GA), dan lainnya.
4. Melakukan pengembangan dengan menggunakan metode seleksi atribut yang lain seperti *chi-square*, *information index* dan sebagainya untuk ketepatan penyeleksian atribut.
5. Meningkatkan lagi faktor sistem prediksi pemasaran langsung untuk penentuan nasabah yang akan berlangganan deposito dan analisi data pribadi nasabah.

## DAFTAR REFERENSI

- Abdous, M., He, W., & Yen, C. J. (2012). Using data mining for predicting relationships between online question theme and final grade. *Educational Technology and Society*, 15(3), 77–88.
- Abraham, A., Grosan, C., Ramos, V., (2006). *Swarm Intelligence in Data Mining. Springer-Verlag Berlin Heidelberg.*
- Aydin, I., Karakose, M., & Akin, E. (2011). A multi-objective artificial immune algorithm for parameter optimization in support vector machine. *Journal Applied Soft Computing*, 11, 120-129.
- Berndtsson, M., Hansson, J., Olsson, B., & Lundell, B. (2008). *Thesis Guide – A Guide for Students in Computer Science and Information Systems.*
- Bellazzi, R., & Zupanb, B. (2008). Predictive Data Mining In Clinical Medicine: Current Issues And And Guidelines. *International Journal Of Medical Informatics* 77, 81–97.
- Bellotti, T., & Crook, J. (2007) Support vector machines for credit scoring and discovery of significant features. *Expert System with Application: An International Journal*, 36, 3302-3308.
- Bramer, M. (2007). *Principles Of Data Mining.* Verlag London: Springer.
- Brinson, A., Lee, M.-Y., & Rountree, B. (2011). Direct marketing strategies: The rise of community supported fishery programs. *Marine Policy*, 35(4), 542–548.
- Chitty, William., Nigel Barker., & Terence A. Shimp. (2008). *Integrated Marketing Communication.* Australasian ed of 2nd revised ed edition.
- Dawson, C. W. (2009). *Project in Computing and Information System A Student's Guide. England: Addison-Wesley.*
- Elsalamony, H. A. (2014). Bank Direct Marketing Analysis of Data Mining Techniques. *International Journal of Computer Applications* (0975 – 8887) Volume 85 – No 7.
- Elsalamony, H. A., & Elsayad, A. M. (2013). Bank Direct Marketing Based on Neural Network and C5 . 0 Models. *International Journal of Engineering and Advanced Technology (IJEAT)*, 2(6).

- Fei-Long, Chena., & Feng-Chia Li. (2010). Combination of Feature Selection Approaches with SVM in Credit Scoring. *Expert Systems with Applications: An International Journal archive*. Volume 37 Issue 7. Pages 4902-4909.
- Frias-Martinez, E., Sanchez, A., & Velez, J. (2010). Support vector machines versus multi-layer perceptrons for efficient off-line signature recognition. *Engineering Applications of Artificial Intelligence*, 19(6), 693–704.
- Gorunescu, Florin (2011). *Data Mining: Concepts, Models, and Techniques*. *Verlag Berlin Heidelberg: Springer*.
- Huang, C. L., & Dun, J. F. (2008). A distributed PSO-SVM hybrid system with feature selection and parameter optimization. *Applied Soft Computing Journal*, 8(4), 1381–1391.
- Han, J., & Kamber, M. (2006). *Data Mining Concepts and technique*. *San Francisco: Diane Cerra*
- Ilhan, I., & Tezel, G. (2013). A genetic algorithm-support vector machine method with parameter optimization for selecting the tag SNPs. *Journal of Biomedical Informatics*, 46(2), 328–40.
- Khoshahval, F., Minuchehr, H., & Zolfaghari, a. (2011). Performance evaluation of PSO and GA in PWR core loading pattern optimization. *Nuclear Engineering and Design*, 241(3), 799–808.
- Kahrizi, A., & Hashemi, H. (2014). Neuron curve as a tool for performance evaluation of MLP and RBF architecture in first break picking of seismic data. *Journal of Applied Geophysics*, 108, 159–166.
- Li, G., You, J., & Liu, X. (2015). Support Vector Machine (SVM) based prestack AVO inversion and its applications. *Journal of Applied Geophysics*, 120, 60–68.
- Liu, H., Tian, H., Chen, C., & Li, Y. (2013). *Electrical Power and Energy Systems* An experimental investigation of two Wavelet-MLP hybrid frameworks for wind speed prediction using GA and PSO optimization, 52, 161–173.
- Liao, S., Chen, Y., & Hsieh, H. (2011). Mining customer knowledge for direct selling and marketing. *Expert Systems with Applications*, 38(5), 6059–6069.
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook (2nd ed)*. *New York: Springer Dordrecht Heidelberg London*.
- Moro, S., & Laureano, R. M. S. (2012). Using Data Mining for Bank Direct Marketing: An application of the CRISP-DM methodology. *European Simulation and Modelling Conference*, (Figure 1), 117–121.

- Olson, D., & Shi, Y. (2008). *Pengantar Ilmu Penggalian Data Bisnis*. Jakarta: Penerbit Salemba Empat.
- Pandey, U. K., & Pal, S. (2011). Data Mining: A prediction of performer or underperformer using classification, 2(2), 5.
- Pin Lio, Xin Zhang, & Kun Li, Yang Fu, Mingyan Wang dan Sensen Wang. (2015). *Parameter Optimization for Support Vector Machine Based on Nested Genetic Algorithms*. *Journal of Automation and Control Engineering*. Vol. 3, No. 6.
- Prasetyo, E. (2012). *Data Mining : Konsep dan Aplikasi Menggunakan Matlab*. Indonesia: Andi Yogyakarta.
- Ren, J. (2012). ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging. *Knowledge-Based Systems*, 26, 144–153.
- Rossi, L. D., & Soares, C. (2012). Neurocomputing Combining meta-learning and search techniques to select parameters for support vector machines. *Neurocomputing*, 75, 3–13.
- Rust, R. T. (2010). Rethinking marketing. *Harvard Business Review*, 1, 1–8.
- Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Sing'oei, Lilian. & Jiayang Wang. (2013). Data Mining Framework for Direct Marketing: A Case Study of Bank Marketing. *International Journal of Computer Science and Issue*, 10(2), 198-203.
- Talla, F., Leus, R. &, & Spieksma, F. C. R. (2011). Optimization models for targeted offers in direct marketing: Exact and heuristic algorithms. *European Journal of Operational Research*, 210(3), 670–683.
- Turban, E. (2012). *Information Technology for Management*. (B. L. Golub, Ed.) (8th ed.). United State of America: John Wiley & Sons, Inc.
- UCI (2010, November). Bank Marketing Data Set. Februari 14, 2012. <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing#>
- Vercellis, Carlo (2009). *Business Intelligent: Data Mining and Optimization for Decision Making*. *Southern Gate, Chichester, West Sussex*: John Willey & Sons, Ltd.
- Vieira, S. M., & Mendonc, L. F. (2013). Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients, 13, 3494–3504.

- Weiss, S. M., Indurkha, N., & Zhang, T. (2010). *Fundamentals Of Predictive Text Mining*. London: Springer.
- Written, I. H. (2011). *Data Mining Practical Machine Learning Tools and Techniques* (3rd ed.). USA: Elsevier.
- Written, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning and Tools*. Burlington: Morgan Koufmann Publisher.
- Yin, Y., Han, D., & Cai, Z. (2011). Explore Data Classification Algorithm Based on SVM and PSO for Education Decision. *Journal of Convergence Information Technology*, 6(10), 122–128.
- Yin, H., Jiao, X., Chai, Y., & Fang, B. (2015). Scene classification based on single-layer SAE and SVM. *Expert Systems with Applications*, 42(7), 3368–3380.
- Yin, Y., Han, D., & Cai, Z. (2011). Explore Data Classification Algorithm Based on SVM and PSO for Education Decision. *Journal of Convergence Information Technology*, 6(10), 122–128.
- Yusup, N., Zain, A. M., Zaiton, S., & Hashim, M. (2012). *Procedia Engineering Overview of PSO for Optimizing Process Parameters of Machining*.
- Zafra, A., & Ventura, S. (2012). Multi-instance genetic programming for predicting student performance in web based educational environments. *Applied Soft Computing Journal*, 12(8), 2693–2706.
- Zhao, F., Liu, Y., Zhang, C., & Wang, J. (2015). A self-adaptive harmony PSO search algorithm and its performance analysis. *Expert Systems with Applications*, 42(21), 7436–7455.

## DAFTAR RIWAYAT HIDUP

### I. Biodata Mahasiswa

NIM : 11110787  
Nama Lengkap : Yuni Eka Achyani  
Tempat & Tanggal Lahir : Jakarta, 04 Juni 1990  
Alamat Lengkap : Kp. Pulo Rt.001 Rw.35 No.135  
Sumber Jaya Tambun Selatan 17510

### II. Pendidikan

1. SDN Sumber Jaya 02 Tambun Selatan, lulus tahun 2002
2. SMP Negeri 3 Tambun Selatan, lulus tahun 2005
3. SMA Negeri 1 Tambun Selatan, lulus tahun 2008
4. D.III AMIK Bina Sarana Informatika, Bekasi, lulus tahun 2011
5. S1 STMIK Nusa Mandiri, Sistem Informasi, Lulus Tahun 2014



Jakarta, 15 Agustus 2016

Yuni Eka Achyani



## LAMPIRAN

| age | job           | marital  | education | default | balance | housing | loan | contact   | day | month | duration | campaign | pdays | previous | poutcome | y   |
|-----|---------------|----------|-----------|---------|---------|---------|------|-----------|-----|-------|----------|----------|-------|----------|----------|-----|
| 20  | student       | single   | secondary | no      | 502     | no      | no   | cellular  | 30  | apr   | 261      | 1        | -1    | 0        | unknown  | yes |
| 68  | retired       | divorced | secondary | no      | 4189    | no      | no   | telephone | 14  | jul   | 897      | 2        | -1    | 0        | unknown  | yes |
| 32  | management    | single   | tertiary  | no      | 2536    | yes     | no   | cellular  | 26  | aug   | 958      | 6        | -1    | 0        | unknown  | yes |
| 30  | unemployed    | married  | primary   | no      | 1787    | no      | no   | cellular  | 19  | oct   | 79       | 1        | -1    | 0        | unknown  | no  |
| 33  | services      | married  | secondary | no      | 4789    | yes     | yes  | cellular  | 11  | may   | 220      | 1        | 339   | 4        | failure  | no  |
| 35  | management    | single   | tertiary  | no      | 1350    | yes     | no   | cellular  | 16  | apr   | 185      | 1        | 330   | 1        | failure  | no  |
| 30  | management    | married  | tertiary  | no      | 1476    | yes     | yes  | unknown   | 3   | jun   | 199      | 4        | -1    | 0        | unknown  | no  |
| 49  | technician    | married  | tertiary  | no      | 1235    | no      | no   | cellular  | 13  | aug   | 354      | 3        | -1    | 0        | unknown  | yes |
| 59  | blue-collar   | married  | secondary | no      | 0       | yes     | no   | unknown   | 5   | may   | 226      | 1        | -1    | 0        | unknown  | no  |
| 35  | management    | single   | tertiary  | no      | 747     | no      | no   | cellular  | 23  | feb   | 141      | 2        | 176   | 3        | failure  | no  |
| 36  | self-employed | married  | tertiary  | no      | 307     | yes     | no   | cellular  | 14  | may   | 341      | 1        | 330   | 2        | other    | no  |
| 78  | retired       | divorced | primary   | no      | 229     | no      | no   | telephone | 22  | oct   | 97       | 1        | -1    | 0        | unknown  | yes |
| 39  | technician    | married  | secondary | no      | 147     | yes     | no   | cellular  | 6   | may   | 151      | 2        | -1    | 0        | unknown  | no  |
| 41  | entrepreneur  | married  | tertiary  | no      | 221     | yes     | no   | unknown   | 14  | may   | 57       | 2        | -1    | 0        | unknown  | no  |
| 43  | services      | married  | primary   | no      | -88     | yes     | yes  | cellular  | 17  | apr   | 313      | 1        | 147   | 2        | failure  | no  |
| 32  | blue-collar   | married  | secondary | no      | 2089    | yes     | no   | cellular  | 14  | nov   | 132      | 1        | -1    | 0        | unknown  | yes |
| 39  | services      | married  | secondary | no      | 9374    | yes     | no   | unknown   | 20  | may   | 273      | 1        | -1    | 0        | unknown  | no  |
| 43  | admin.        | married  | secondary | no      | 264     | yes     | no   | cellular  | 17  | apr   | 113      | 2        | -1    | 0        | unknown  | no  |
| 36  | technician    | married  | tertiary  | no      | 1109    | no      | no   | cellular  | 13  | aug   | 328      | 2        | -1    | 0        | unknown  | no  |
| 31  | blue-collar   | married  | secondary | no      | 360     | yes     | yes  | cellular  | 29  | jan   | 89       | 1        | 241   | 1        | failure  | no  |
| 40  | management    | married  | tertiary  | no      | 194     | no      | yes  | cellular  | 29  | aug   | 189      | 2        | -1    | 0        | unknown  | no  |
| 33  | management    | married  | secondary | no      | 3935    | yes     | no   | cellular  | 6   | may   | 765      | 1        | 342   | 2        | failure  | yes |
| 56  | technician    | married  | secondary | no      | 4073    | no      | no   | cellular  | 27  | aug   | 239      | 5        | -1    | 0        | unknown  | no  |
| 37  | admin.        | single   | tertiary  | no      | 2317    | yes     | no   | cellular  | 20  | apr   | 114      | 1        | 152   | 2        | failure  | no  |
| 25  | blue-collar   | single   | primary   | no      | -221    | yes     | no   | unknown   | 23  | may   | 250      | 1        | -1    | 0        | unknown  | no  |
| 31  | services      | married  | secondary | no      | 132     | no      | no   | cellular  | 7   | jul   | 148      | 1        | 152   | 1        | other    | no  |
| 38  | management    | divorced | unknown   | no      | 0       | yes     | no   | cellular  | 18  | nov   | 96       | 2        | -1    | 0        | unknown  | no  |
| 61  | admin.        | married  | unknown   | no      | 4629    | yes     | no   | cellular  | 27  | jan   | 181      | 1        | 92    | 1        | success  | yes |
| 42  | management    | divorced | tertiary  | no      | 16      | no      | no   | cellular  | 19  | nov   | 140      | 3        | -1    | 0        | unknown  | no  |
| 44  | services      | single   | secondary | no      | 106     | no      | no   | unknown   | 12  | jun   | 109      | 2        | -1    | 0        | unknown  | no  |

| age | job           | marital  | education | default | balance | housing | loan | contact   | day | month | duration | campaign | pdays | previous | poutcome | y   |
|-----|---------------|----------|-----------|---------|---------|---------|------|-----------|-----|-------|----------|----------|-------|----------|----------|-----|
| 44  | entrepreneur  | married  | secondary | no      | 93      | no      | no   | cellular  | 7   | jul   | 125      | 2        | -1    | 0        | unknown  | no  |
| 26  | housemaid     | married  | tertiary  | no      | 543     | no      | no   | cellular  | 30  | jan   | 169      | 3        | -1    | 0        | unknown  | no  |
| 41  | management    | married  | tertiary  | no      | 5883    | no      | no   | cellular  | 20  | nov   | 182      | 2        | -1    | 0        | unknown  | no  |
| 45  | blue-collar   | divorced | primary   | no      | 844     | no      | no   | unknown   | 5   | jun   | 1018     | 3        | -1    | 0        | unknown  | yes |
| 34  | technician    | married  | tertiary  | no      | 1539    | yes     | no   | cellular  | 15  | jun   | 441      | 1        | 56    | 1        | other    | yes |
| 37  | management    | married  | tertiary  | no      | 0       | no      | no   | cellular  | 16  | jul   | 268      | 2        | 182   | 3        | success  | yes |
| 55  | blue-collar   | married  | primary   | no      | 627     | yes     | no   | unknown   | 5   | may   | 247      | 1        | -1    | 0        | unknown  | no  |
| 67  | retired       | married  | unknown   | no      | 696     | no      | no   | telephone | 17  | aug   | 119      | 1        | 105   | 2        | failure  | no  |
| 56  | self-employed | married  | secondary | no      | 784     | no      | yes  | cellular  | 30  | jul   | 149      | 2        | -1    | 0        | unknown  | no  |
| 53  | admin.        | married  | secondary | no      | 105     | no      | yes  | cellular  | 21  | aug   | 74       | 2        | -1    | 0        | unknown  | no  |
| 31  | technician    | married  | secondary | no      | 171     | no      | no   | cellular  | 27  | aug   | 81       | 3        | -1    | 0        | unknown  | no  |
| 59  | management    | married  | secondary | no      | 42      | no      | no   | cellular  | 21  | nov   | 40       | 1        | -1    | 0        | unknown  | no  |
| 42  | admin.        | divorced | secondary | no      | 1811    | yes     | no   | unknown   | 14  | may   | 150      | 1        | -1    | 0        | unknown  | no  |
| 23  | services      | single   | tertiary  | no      | 363     | yes     | no   | unknown   | 30  | may   | 16       | 18       | -1    | 0        | unknown  | no  |
| 38  | management    | single   | tertiary  | no      | 11971   | yes     | no   | unknown   | 17  | nov   | 609      | 2        | 101   | 3        | failure  | no  |
| 36  | management    | single   | tertiary  | no      | 553     | no      | no   | cellular  | 11  | aug   | 106      | 2        | -1    | 0        | unknown  | no  |
| 27  | admin.        | divorced | secondary | no      | 451     | yes     | no   | cellular  | 16  | jul   | 652      | 1        | -1    | 0        | unknown  | yes |
| 52  | services      | married  | secondary | no      | 657     | no      | no   | telephone | 7   | jul   | 398      | 2        | 460   | 2        | failure  | yes |
| 37  | management    | married  | tertiary  | no      | 1315    | no      | no   | cellular  | 30  | jun   | 224      | 1        | -1    | 0        | unknown  | yes |
| 52  | blue-collar   | married  | secondary | no      | 1117    | yes     | no   | cellular  | 13  | may   | 365      | 1        | -1    | 0        | unknown  | no  |
| 32  | technician    | married  | tertiary  | no      | 396     | yes     | no   | cellular  | 13  | may   | 205      | 3        | -1    | 0        | unknown  | no  |
| 32  | technician    | single   | tertiary  | no      | 2204    | yes     | no   | cellular  | 21  | nov   | 11       | 4        | -1    | 0        | unknown  | no  |
| 34  | management    | single   | tertiary  | no      | 872     | yes     | no   | unknown   | 7   | may   | 105      | 2        | -1    | 0        | unknown  | no  |
| 55  | blue-collar   | married  | primary   | no      | 145     | no      | no   | telephone | 2   | feb   | 59       | 3        | 5     | 2        | other    | no  |
| 26  | blue-collar   | married  | primary   | no      | 0       | yes     | no   | unknown   | 21  | may   | 425      | 1        | -1    | 0        | unknown  | no  |
| 32  | entrepreneur  | single   | primary   | yes     | -849    | yes     | yes  | cellular  | 4   | feb   | 204      | 1        | -1    | 0        | unknown  | no  |
| 37  | technician    | single   | secondary | no      | 228     | yes     | no   | cellular  | 20  | aug   | 1740     | 2        | -1    | 0        | unknown  | no  |
| 38  | technician    | single   | secondary | no      | 50      | no      | no   | cellular  | 29  | aug   | 98       | 2        | -1    | 0        | unknown  | no  |
| 53  | blue-collar   | married  | secondary | no      | 2231    | yes     | no   | cellular  | 14  | jul   | 272      | 3        | -1    | 0        | unknown  | no  |
| 27  | services      | single   | secondary | no      | -195    | yes     | no   | cellular  | 18  | may   | 391      | 1        | -1    | 0        | unknown  | yes |
| 31  | unemployed    | single   | primary   | no      | 406     | no      | no   | cellular  | 4   | feb   | 736      | 1        | -1    | 0        | unknown  | yes |
| 56  | unemployed    | married  | primary   | no      | 3391    | no      | no   | cellular  | 21  | apr   | 243      | 1        | -1    | 0        | unknown  | yes |
| 48  | admin.        | married  | secondary | no      | 3064    | yes     | no   | telephone | 15  | jul   | 159      | 2        | -1    | 0        | unknown  | no  |

| age | job           | marital  | education | default | balance | housing | loan | contact   | day | month | duration | campaign | pdays | previous | poutcome | y   |
|-----|---------------|----------|-----------|---------|---------|---------|------|-----------|-----|-------|----------|----------|-------|----------|----------|-----|
| 57  | management    | married  | secondary | no      | 82      | no      | yes  | telephone | 4   | feb   | 140      | 1        | -1    | 0        | unknown  | no  |
| 33  | self-employed | single   | tertiary  | no      | 2155    | no      | no   | cellular  | 17  | nov   | 295      | 1        | -1    | 0        | unknown  | no  |
| 36  | retired       | single   | secondary | no      | 101     | yes     | yes  | unknown   | 26  | may   | 314      | 1        | -1    | 0        | unknown  | no  |
| 54  | technician    | divorced | secondary | no      | 784     | yes     | yes  | unknown   | 15  | may   | 579      | 1        | -1    | 0        | unknown  | no  |
| 41  | blue-collar   | married  | primary   | no      | -516    | no      | yes  | telephone | 8   | jul   | 554      | 3        | -1    | 0        | unknown  | no  |
| 63  | retired       | married  | secondary | no      | 415     | yes     | no   | cellular  | 7   | oct   | 323      | 1        | -1    | 0        | unknown  | no  |
| 48  | management    | married  | tertiary  | no      | 5887    | no      | no   | cellular  | 13  | aug   | 227      | 2        | -1    | 0        | unknown  | no  |
| 48  | services      | married  | secondary | no      | 1355    | no      | no   | cellular  | 4   | jul   | 134      | 2        | -1    | 0        | unknown  | no  |
| 56  | admin.        | married  | secondary | no      | 16873   | no      | no   | cellular  | 7   | oct   | 223      | 1        | -1    | 0        | unknown  | no  |
| 21  | student       | single   | secondary | no      | 2488    | no      | no   | cellular  | 30  | jun   | 258      | 6        | 169   | 3        | success  | yes |
| 25  | student       | single   | secondary | no      | 331     | no      | no   | telephone | 26  | aug   | 170      | 4        | -1    | 0        | unknown  | yes |
| 25  | technician    | single   | secondary | no      | 505     | no      | yes  | cellular  | 17  | nov   | 386      | 2        | -1    | 0        | unknown  | yes |
| 51  | blue-collar   | divorced | secondary | no      | 203     | yes     | no   | cellular  | 8   | may   | 134      | 1        | 170   | 5        | failure  | no  |
| 31  | services      | married  | secondary | no      | 338     | yes     | no   | cellular  | 28  | jan   | 155      | 1        | -1    | 0        | unknown  | no  |
| 29  | technician    | married  | secondary | no      | 444     | yes     | no   | cellular  | 21  | jul   | 130      | 4        | -1    | 0        | unknown  | no  |
| 41  | management    | married  | secondary | no      | 0       | no      | yes  | cellular  | 7   | jul   | 630      | 3        | -1    | 0        | unknown  | no  |
| 32  | technician    | single   | tertiary  | no      | 360     | no      | no   | cellular  | 19  | nov   | 164      | 2        | -1    | 0        | unknown  | no  |
| 42  | entrepreneur  | divorced | tertiary  | yes     | 2       | yes     | no   | unknown   | 5   | may   | 380      | 1        | -1    | 0        | unknown  | no  |
| 31  | management    | single   | tertiary  | no      | 6248    | yes     | no   | unknown   | 9   | may   | 154      | 2        | -1    | 0        | unknown  | no  |
| 41  | services      | married  | tertiary  | no      | 412     | yes     | no   | unknown   | 3   | jun   | 154      | 5        | -1    | 0        | unknown  | no  |
| 56  | retired       | married  | secondary | no      | 344     | yes     | no   | unknown   | 15  | may   | 221      | 1        | -1    | 0        | unknown  | no  |
| 54  | management    | divorced | tertiary  | no      | 3222    | no      | no   | cellular  | 14  | aug   | 67       | 2        | -1    | 0        | unknown  | no  |
| 34  | self-employed | single   | tertiary  | no      | 462     | no      | no   | cellular  | 21  | aug   | 1877     | 3        | -1    | 0        | unknown  | yes |
| 41  | blue-collar   | divorced | secondary | no      | 174     | yes     | no   | cellular  | 14  | may   | 367      | 2        | 297   | 1        | other    | no  |
| 40  | technician    | single   | secondary | no      | 591     | no      | yes  | cellular  | 14  | aug   | 87       | 6        | -1    | 0        | unknown  | no  |
| 50  | blue-collar   | divorced | primary   | no      | 388     | no      | no   | cellular  | 5   | feb   | 701      | 1        | -1    | 0        | unknown  | no  |
| 40  | unemployed    | married  | secondary | no      | 219     | yes     | no   | cellular  | 17  | nov   | 204      | 2        | 196   | 1        | failure  | no  |
| 60  | retired       | married  | primary   | no      | 5       | no      | no   | cellular  | 26  | aug   | 63       | 2        | -1    | 0        | unknown  | no  |
| 28  | admin.        | single   | secondary | no      | 177     | yes     | yes  | cellular  | 6   | may   | 150      | 1        | -1    | 0        | unknown  | no  |
| 51  | blue-collar   | married  | primary   | no      | 1466    | yes     | no   | unknown   | 7   | may   | 406      | 2        | -1    | 0        | unknown  | no  |
| 41  | technician    | single   | secondary | no      | 879     | no      | yes  | cellular  | 7   | jul   | 60       | 4        | -1    | 0        | unknown  | no  |
| 35  | blue-collar   | single   | primary   | no      | 293     | yes     | no   | unknown   | 30  | may   | 521      | 2        | -1    | 0        | unknown  | no  |

| age | job           | marital  | education | default | balance | housing | loan | contact   | day | month | duration | campaign | pdays | previous | poutcome | y   |
|-----|---------------|----------|-----------|---------|---------|---------|------|-----------|-----|-------|----------|----------|-------|----------|----------|-----|
| 38  | admin.        | married  | secondary | no      | 424     | yes     | no   | cellular  | 17  | apr   | 279      | 1        | -1    | 0        | unknown  | no  |
| 34  | blue-collar   | married  | secondary | no      | 1831    | yes     | no   | unknown   | 20  | may   | 203      | 2        | -1    | 0        | unknown  | no  |
| 39  | blue-collar   | married  | primary   | no      | 111     | no      | no   | cellular  | 18  | nov   | 201      | 2        | -1    | 0        | unknown  | no  |
| 34  | blue-collar   | married  | primary   | no      | 455     | yes     | no   | unknown   | 20  | jun   | 372      | 3        | -1    | 0        | unknown  | no  |
| 32  | management    | single   | tertiary  | no      | 3616    | no      | no   | cellular  | 14  | aug   | 165      | 2        | -1    | 0        | unknown  | no  |
| 57  | blue-collar   | married  | secondary | no      | 14093   | no      | no   | cellular  | 11  | aug   | 231      | 5        | -1    | 0        | unknown  | no  |
| 77  | retired       | divorced | tertiary  | no      | 4659    | no      | no   | cellular  | 14  | apr   | 161      | 1        | -1    | 0        | unknown  | yes |
| 52  | management    | married  | tertiary  | no      | -970    | yes     | no   | cellular  | 4   | feb   | 489      | 1        | -1    | 0        | unknown  | yes |
| 55  | admin.        | married  | secondary | no      | 602     | no      | no   | cellular  | 1   | jun   | 324      | 3        | 90    | 2        | success  | yes |
| 28  | management    | single   | tertiary  | no      | 80      | no      | no   | cellular  | 20  | oct   | 676      | 2        | 146   | 2        | failure  | yes |
| 56  | retired       | married  | secondary | no      | -1206   | yes     | no   | cellular  | 15  | jun   | 382      | 2        | -1    | 0        | unknown  | yes |
| 41  | unknown       | single   | tertiary  | no      | 1567    | no      | no   | cellular  | 1   | jul   | 291      | 2        | -1    | 0        | unknown  | no  |
| 41  | admin.        | married  | tertiary  | no      | 5426    | yes     | no   | unknown   | 14  | may   | 149      | 2        | -1    | 0        | unknown  | no  |
| 30  | admin.        | single   | tertiary  | no      | 261     | no      | no   | cellular  | 19  | oct   | 233      | 1        | 137   | 20       | failure  | no  |
| 36  | blue-collar   | divorced | secondary | no      | 2843    | no      | no   | cellular  | 12  | feb   | 473      | 1        | 182   | 1        | success  | no  |
| 36  | blue-collar   | married  | secondary | no      | 0       | yes     | no   | unknown   | 6   | may   | 337      | 1        | -1    | 0        | unknown  | no  |
| 38  | management    | single   | tertiary  | no      | 493     | yes     | no   | cellular  | 11  | may   | 553      | 1        | 367   | 7        | failure  | no  |
| 49  | technician    | married  | secondary | no      | 5996    | no      | no   | telephone | 21  | nov   | 345      | 2        | -1    | 0        | unknown  | no  |
| 57  | services      | single   | primary   | no      | 3777    | yes     | no   | telephone | 13  | may   | 65       | 2        | -1    | 0        | unknown  | no  |
| 34  | management    | single   | secondary | no      | 524     | no      | no   | unknown   | 13  | aug   | 9        | 1        | -1    | 0        | unknown  | no  |
| 32  | management    | single   | tertiary  | no      | 574     | yes     | no   | cellular  | 14  | apr   | 259      | 2        | 145   | 3        | failure  | no  |
| 75  | retired       | married  | secondary | no      | 3771    | no      | no   | telephone | 15  | apr   | 185      | 1        | 181   | 2        | success  | yes |
| 35  | technician    | single   | tertiary  | no      | 2658    | yes     | no   | cellular  | 19  | aug   | 768      | 2        | 141   | 3        | failure  | yes |
| 29  | blue-collar   | single   | tertiary  | no      | 455     | no      | no   | telephone | 21  | apr   | 403      | 1        | -1    | 0        | unknown  | yes |
| 37  | blue-collar   | married  | secondary | no      | 427     | yes     | yes  | unknown   | 9   | jun   | 371      | 3        | -1    | 0        | unknown  | no  |
| 33  | self-employed | married  | tertiary  | no      | 483     | yes     | no   | cellular  | 3   | feb   | 280      | 2        | -1    | 0        | unknown  | no  |
| 27  | housemaid     | married  | primary   | no      | 0       | yes     | no   | cellular  | 23  | jul   | 435      | 3        | -1    | 0        | unknown  | no  |
| 58  | management    | married  | tertiary  | no      | 1517    | yes     | no   | telephone | 29  | jul   | 7        | 5        | -1    | 0        | unknown  | no  |
| 32  | self-employed | married  | secondary | no      | 217     | yes     | yes  | cellular  | 15  | jul   | 317      | 5        | -1    | 0        | unknown  | no  |
| 42  | management    | married  | tertiary  | no      | 372     | yes     | no   | cellular  | 3   | aug   | 654      | 2        | -1    | 0        | unknown  | yes |
| 29  | blue-collar   | single   | secondary | no      | 482     | no      | yes  | cellular  | 10  | jul   | 1097     | 3        | -1    | 0        | unknown  | yes |
| 24  | blue-collar   | single   | secondary | no      | 174     | yes     | yes  | cellular  | 18  | may   | 487      | 1        | -1    | 0        | unknown  | yes |

| age | job          | marital  | education | default | balance | housing | loan | contact   | day | month | duration | campaign | pdays | previous | poutcome | y   |
|-----|--------------|----------|-----------|---------|---------|---------|------|-----------|-----|-------|----------|----------|-------|----------|----------|-----|
| 33  | blue-collar  | single   | secondary | no      | 22      | no      | no   | cellular  | 30  | jan   | 76       | 2        | 207   | 1        | failure  | no  |
| 30  | services     | married  | secondary | no      | 1808    | yes     | no   | cellular  | 18  | may   | 83       | 2        | -1    | 0        | unknown  | no  |
| 22  | admin.       | single   | secondary | no      | 4111    | no      | yes  | cellular  | 19  | aug   | 65       | 1        | -1    | 0        | unknown  | no  |
| 39  | blue-collar  | married  | primary   | no      | 52      | no      | no   | cellular  | 13  | aug   | 69       | 1        | -1    | 0        | unknown  | no  |
| 46  | management   | divorced | tertiary  | no      | 25      | no      | no   | unknown   | 17  | jun   | 564      | 2        | -1    | 0        | unknown  | no  |
| 34  | admin.       | married  | tertiary  | no      | 69      | no      | yes  | cellular  | 4   | aug   | 120      | 3        | 1     | 6        | success  | yes |
| 31  | admin.       | divorced | secondary | no      | 1890    | yes     | no   | cellular  | 21  | jul   | 588      | 1        | -1    | 0        | unknown  | no  |
| 45  | blue-collar  | married  | primary   | no      | 419     | no      | no   | cellular  | 18  | aug   | 113      | 4        | -1    | 0        | unknown  | no  |
| 32  | technician   | married  | secondary | no      | 2693    | yes     | no   | unknown   | 21  | may   | 16       | 1        | -1    | 0        | unknown  | no  |
| 36  | blue-collar  | married  | secondary | no      | -231    | no      | yes  | cellular  | 15  | jul   | 779      | 2        | -1    | 0        | unknown  | no  |
| 51  | management   | single   | tertiary  | yes     | -55     | yes     | no   | cellular  | 11  | may   | 281      | 2        | 266   | 6        | failure  | no  |
| 24  | unemployed   | single   | secondary | no      | 119     | no      | no   | unknown   | 15  | sep   | 51       | 1        | -1    | 0        | unknown  | no  |
| 45  | technician   | single   | secondary | no      | 586     | no      | yes  | cellular  | 28  | jul   | 32       | 3        | -1    | 0        | unknown  | no  |
| 32  | blue-collar  | single   | secondary | no      | 228     | no      | no   | telephone | 5   | feb   | 176      | 1        | 288   | 3        | failure  | no  |
| 41  | blue-collar  | married  | primary   | no      | 168     | yes     | no   | cellular  | 18  | nov   | 187      | 3        | -1    | 0        | unknown  | no  |
| 37  | blue-collar  | married  | primary   | no      | 0       | yes     | no   | unknown   | 26  | may   | 24       | 1        | -1    | 0        | unknown  | no  |
| 43  | blue-collar  | married  | unknown   | yes     | -715    | yes     | no   | unknown   | 30  | may   | 67       | 3        | -1    | 0        | unknown  | no  |
| 48  | admin.       | married  | unknown   | no      | 0       | yes     | no   | cellular  | 8   | may   | 85       | 1        | 168   | 2        | failure  | no  |
| 50  | blue-collar  | married  | primary   | no      | 4108    | yes     | no   | cellular  | 7   | aug   | 526      | 2        | -1    | 0        | unknown  | yes |
| 49  | entrepreneur | divorced | unknown   | yes     | -701    | yes     | no   | cellular  | 30  | jul   | 988      | 2        | -1    | 0        | unknown  | yes |
| 45  | blue-collar  | divorced | primary   | no      | -311    | yes     | no   | unknown   | 23  | may   | 1030     | 1        | -1    | 0        | unknown  | yes |
| 27  | technician   | single   | secondary | no      | 769     | yes     | no   | unknown   | 9   | may   | 317      | 2        | -1    | 0        | unknown  | no  |
| 33  | management   | single   | tertiary  | no      | 0       | no      | no   | cellular  | 28  | aug   | 236      | 10       | -1    | 0        | unknown  | no  |
| 29  | admin.       | single   | secondary | no      | 428     | yes     | yes  | cellular  | 7   | may   | 54       | 1        | 345   | 2        | failure  | no  |
| 34  | management   | single   | tertiary  | no      | 0       | no      | no   | cellular  | 5   | feb   | 71       | 1        | -1    | 0        | unknown  | no  |
| 60  | management   | married  | unknown   | no      | 6313    | no      | no   | unknown   | 19  | jun   | 39       | 2        | -1    | 0        | unknown  | no  |
| 32  | management   | divorced | tertiary  | no      | 948     | no      | no   | cellular  | 12  | aug   | 204      | 2        | -1    | 0        | unknown  | no  |
| 32  | management   | single   | tertiary  | no      | 0       | yes     | no   | cellular  | 22  | jul   | 455      | 2        | -1    | 0        | unknown  | no  |
| 56  | services     | divorced | unknown   | no      | 1877    | yes     | no   | unknown   | 30  | may   | 86       | 1        | -1    | 0        | unknown  | no  |
| 31  | management   | single   | tertiary  | no      | 874     | yes     | no   | cellular  | 18  | nov   | 190      | 2        | -1    | 0        | unknown  | no  |
| 42  | admin.       | married  | secondary | no      | 323     | yes     | yes  | unknown   | 8   | may   | 280      | 2        | -1    | 0        | unknown  | no  |
| 38  | entrepreneur | married  | secondary | no      | 593     | yes     | yes  | cellular  | 24  | jul   | 1484     | 24       | -1    | 0        | unknown  | yes |
| 32  | blue-collar  | married  | secondary | no      | 759     | yes     | no   | cellular  | 9   | apr   | 317      | 5        | -1    | 0        | unknown  | yes |
| 49  | entrepreneur | married  | secondary | no      | 687     | yes     | no   | cellular  | 17  | apr   | 772      | 1        | -1    | 0        | unknown  | yes |

| age | job           | marital  | education | default | balance | housing | loan | contact   | day | month | duration | campaign | pdays | previous | poutcome | y   |
|-----|---------------|----------|-----------|---------|---------|---------|------|-----------|-----|-------|----------|----------|-------|----------|----------|-----|
| 35  | management    | married  | tertiary  | no      | 106     | no      | yes  | cellular  | 11  | aug   | 588      | 2        | -1    | 0        | unknown  | no  |
| 56  | retired       | married  | primary   | no      | 1906    | no      | yes  | unknown   | 19  | jun   | 45       | 9        | -1    | 0        | unknown  | no  |
| 48  | technician    | married  | secondary | no      | 1152    | yes     | no   | cellular  | 21  | jul   | 168      | 5        | -1    | 0        | unknown  | no  |
| 46  | admin.        | married  | secondary | no      | 179     | yes     | no   | cellular  | 29  | jan   | 194      | 4        | 436   | 1        | other    | no  |
| 37  | technician    | single   | unknown   | no      | 391     | yes     | no   | unknown   | 8   | may   | 103      | 3        | -1    | 0        | unknown  | no  |
| 41  | entrepreneur  | married  | unknown   | no      | 89      | yes     | no   | unknown   | 6   | may   | 333      | 2        | -1    | 0        | unknown  | no  |
| 36  | management    | single   | tertiary  | no      | 442     | yes     | yes  | cellular  | 29  | jul   | 102      | 3        | -1    | 0        | unknown  | no  |
| 45  | unemployed    | divorced | primary   | yes     | -249    | yes     | yes  | unknown   | 1   | jul   | 92       | 1        | -1    | 0        | unknown  | no  |
| 55  | entrepreneur  | married  | secondary | no      | 8104    | no      | no   | cellular  | 6   | feb   | 213      | 2        | -1    | 0        | unknown  | no  |
| 39  | technician    | married  | secondary | no      | 501     | no      | no   | unknown   | 26  | may   | 289      | 1        | -1    | 0        | unknown  | no  |
| 35  | technician    | divorced | secondary | no      | 308     | no      | no   | cellular  | 22  | aug   | 77       | 2        | -1    | 0        | unknown  | no  |
| 55  | technician    | married  | secondary | no      | 273     | yes     | no   | cellular  | 29  | jan   | 84       | 3        | 183   | 3        | failure  | no  |
| 27  | self-employed | single   | secondary | no      | 50      | no      | no   | cellular  | 29  | jan   | 168      | 1        | -1    | 0        | unknown  | no  |
| 34  | technician    | married  | secondary | no      | 1641    | yes     | no   | cellular  | 17  | apr   | 380      | 1        | -1    | 0        | unknown  | no  |
| 34  | blue-collar   | married  | tertiary  | no      | 20      | no      | no   | unknown   | 4   | oct   | 10       | 1        | -1    | 0        | unknown  | no  |
| 25  | admin.        | single   | secondary | no      | 4590    | yes     | no   | unknown   | 28  | may   | 199      | 2        | -1    | 0        | unknown  | no  |
| 41  | admin.        | single   | secondary | no      | -256    | no      | yes  | unknown   | 9   | may   | 35       | 3        | -1    | 0        | unknown  | no  |
| 43  | services      | married  | secondary | no      | 978     | yes     | yes  | unknown   | 26  | may   | 82       | 2        | -1    | 0        | unknown  | no  |
| 31  | admin.        | single   | secondary | no      | 2104    | yes     | no   | unknown   | 3   | jun   | 455      | 1        | -1    | 0        | unknown  | no  |
| 78  | housemaid     | married  | secondary | no      | 499     | no      | no   | telephone | 16  | mar   | 80       | 4        | -1    | 0        | unknown  | no  |
| 25  | admin.        | single   | tertiary  | no      | 760     | yes     | no   | cellular  | 27  | may   | 223      | 3        | 89    | 2        | failure  | yes |
| 35  | entrepreneur  | married  | secondary | no      | 2971    | no      | no   | cellular  | 4   | may   | 429      | 4        | 164   | 2        | failure  | yes |
| 38  | management    | married  | tertiary  | no      | 1199    | no      | no   | cellular  | 4   | aug   | 178      | 1        | 1     | 1        | success  | yes |
| 45  | blue-collar   | divorced | secondary | no      | 1269    | yes     | no   | cellular  | 29  | jan   | 549      | 2        | -1    | 0        | unknown  | no  |
| 48  | technician    | married  | tertiary  | no      | 310     | no      | no   | unknown   | 6   | may   | 135      | 1        | -1    | 0        | unknown  | no  |
| 32  | blue-collar   | single   | secondary | no      | 1143    | yes     | yes  | cellular  | 10  | jul   | 412      | 1        | -1    | 0        | unknown  | no  |
| 42  | technician    | married  | secondary | no      | 2980    | no      | no   | unknown   | 20  | jun   | 101      | 3        | -1    | 0        | unknown  | no  |
| 37  | management    | single   | tertiary  | no      | 37      | no      | no   | cellular  | 28  | jan   | 253      | 2        | -1    | 0        | unknown  | no  |
| 40  | admin.        | single   | secondary | no      | 462     | yes     | yes  | cellular  | 6   | apr   | 272      | 1        | 335   | 4        | other    | no  |
| 32  | admin.        | married  | secondary | no      | 1031    | yes     | no   | cellular  | 17  | jul   | 84       | 1        | -1    | 0        | unknown  | no  |
| 42  | management    | married  | tertiary  | no      | 9009    | no      | no   | cellular  | 20  | nov   | 135      | 2        | -1    | 0        | unknown  | no  |
| 43  | admin.        | married  | secondary | no      | -465    | yes     | yes  | cellular  | 23  | jul   | 166      | 1        | -1    | 0        | unknown  | no  |
| 44  | admin.        | married  | secondary | no      | 5181    | yes     | yes  | cellular  | 31  | jul   | 18       | 7        | -1    | 0        | unknown  | no  |

| age | job           | marital  | education | default | balance | housing | loan | contact   | day | month | duration | campaign | pdays | previous | outcome | y   |
|-----|---------------|----------|-----------|---------|---------|---------|------|-----------|-----|-------|----------|----------|-------|----------|---------|-----|
| 37  | admin.        | married  | secondary | no      | 176     | yes     | no   | cellular  | 5   | may   | 147      | 2        | 347   | 1        | failure | no  |
| 37  | unknown       | married  | unknown   | no      | 1699    | no      | no   | unknown   | 9   | jun   | 63       | 1        | -1    | 0        | unknown | no  |
| 37  | admin.        | married  | tertiary  | no      | 0       | yes     | no   | cellular  | 17  | nov   | 65       | 2        | 119   | 1        | failure | no  |
| 28  | technician    | single   | tertiary  | no      | 172     | yes     | yes  | cellular  | 2   | feb   | 194      | 2        | -1    | 0        | unknown | no  |
| 32  | services      | single   | secondary | no      | 0       | no      | yes  | cellular  | 11  | jul   | 333      | 2        | -1    | 0        | unknown | no  |
| 34  | admin.        | single   | secondary | no      | 6979    | yes     | no   | cellular  | 15  | may   | 14       | 5        | -1    | 0        | unknown | no  |
| 35  | entrepreneur  | single   | tertiary  | no      | 145     | yes     | no   | cellular  | 18  | jul   | 799      | 2        | -1    | 0        | unknown | yes |
| 35  | unemployed    | married  | secondary | no      | 978     | no      | no   | cellular  | 29  | dec   | 1370     | 3        | 209   | 3        | success | yes |
| 40  | admin.        | divorced | secondary | no      | 6       | no      | no   | unknown   | 11  | jun   | 140      | 1        | -1    | 0        | unknown | no  |
| 34  | services      | married  | secondary | no      | -62     | yes     | no   | unknown   | 16  | jun   | 109      | 3        | -1    | 0        | unknown | no  |
| 34  | self-employed | married  | tertiary  | no      | 1972    | yes     | no   | cellular  | 23  | jul   | 61       | 2        | -1    | 0        | unknown | no  |
| 34  | admin.        | married  | tertiary  | no      | 899     | yes     | no   | telephone | 19  | nov   | 377      | 3        | 7     | 4        | success | no  |
| 37  | services      | single   | secondary | no      | 21      | yes     | no   | cellular  | 2   | feb   | 96       | 1        | 271   | 2        | other   | no  |
| 36  | technician    | married  | secondary | no      | 23      | yes     | no   | cellular  | 8   | may   | 152      | 2        | 347   | 1        | failure | no  |
| 49  | blue-collar   | married  | primary   | no      | 0       | yes     | yes  | telephone | 23  | jul   | 97       | 6        | -1    | 0        | unknown | no  |
| 33  | self-employed | married  | secondary | no      | 215     | yes     | no   | cellular  | 29  | jul   | 543      | 12       | -1    | 0        | unknown | no  |
| 30  | management    | single   | tertiary  | no      | 1372    | no      | no   | cellular  | 20  | nov   | 240      | 2        | -1    | 0        | unknown | no  |
| 36  | services      | married  | secondary | no      | 1981    | no      | no   | unknown   | 17  | jun   | 48       | 2        | -1    | 0        | unknown | no  |
| 23  | student       | single   | secondary | no      | 9216    | no      | no   | cellular  | 5   | jun   | 471      | 2        | -1    | 0        | unknown | no  |
| 57  | retired       | divorced | secondary | no      | 39      | yes     | no   | cellular  | 30  | jul   | 354      | 3        | -1    | 0        | unknown | no  |
| 27  | blue-collar   | single   | secondary | no      | 5563    | yes     | no   | unknown   | 30  | may   | 435      | 2        | -1    | 0        | unknown | no  |
| 38  | services      | married  | secondary | no      | 1       | no      | yes  | cellular  | 21  | nov   | 152      | 2        | -1    | 0        | unknown | no  |
| 42  | technician    | married  | tertiary  | no      | 782     | no      | no   | cellular  | 3   | apr   | 285      | 1        | -1    | 0        | unknown | no  |
| 34  | technician    | single   | tertiary  | no      | 992     | yes     | no   | cellular  | 4   | may   | 301      | 1        | 88    | 2        | success | no  |
| 48  | technician    | married  | secondary | no      | 1526    | no      | no   | cellular  | 6   | aug   | 273      | 2        | -1    | 0        | unknown | no  |
| 39  | management    | divorced | tertiary  | no      | 1315    | yes     | no   | cellular  | 19  | nov   | 1337     | 4        | 126   | 1        | failure | no  |
| 45  | technician    | single   | tertiary  | no      | 0       | yes     | no   | cellular  | 18  | nov   | 139      | 2        | -1    | 0        | unknown | no  |
| 42  | technician    | married  | secondary | no      | 2030    | yes     | yes  | cellular  | 9   | jul   | 196      | 1        | -1    | 0        | unknown | no  |
| 29  | technician    | married  | secondary | no      | 1599    | yes     | no   | cellular  | 8   | jul   | 115      | 1        | 61    | 1        | failure | no  |
| 38  | services      | single   | secondary | no      | -196    | yes     | no   | cellular  | 2   | feb   | 60       | 1        | 271   | 1        | failure | no  |
| 56  | retired       | married  | secondary | no      | 1333    | no      | no   | cellular  | 28  | aug   | 17       | 14       | -1    | 0        | unknown | no  |
| 36  | technician    | single   | secondary | no      | 9149    | yes     | no   | cellular  | 13  | aug   | 141      | 3        | -1    | 0        | unknown | no  |

| age | job           | marital  | education | default | balance | housing | loan | contact   | day | month | duration | campaign | pdays | previous | poutcome | y   |
|-----|---------------|----------|-----------|---------|---------|---------|------|-----------|-----|-------|----------|----------|-------|----------|----------|-----|
| 25  | blue-collar   | single   | secondary | no      | -350    | yes     | no   | unknown   | 23  | jun   | 95       | 3        | -1    | 0        | unknown  | no  |
| 38  | technician    | married  | tertiary  | no      | 432     | yes     | no   | cellular  | 15  | may   | 10       | 2        | 373   | 1        | other    | no  |
| 57  | self-employed | married  | secondary | no      | 11494   | no      | no   | cellular  | 19  | nov   | 198      | 1        | -1    | 0        | unknown  | no  |
| 31  | management    | married  | tertiary  | yes     | -253    | no      | no   | cellular  | 12  | aug   | 220      | 4        | -1    | 0        | unknown  | no  |
| 45  | technician    | married  | secondary | no      | 1410    | no      | no   | cellular  | 11  | aug   | 71       | 2        | -1    | 0        | unknown  | no  |
| 30  | blue-collar   | married  | secondary | no      | 975     | yes     | no   | cellular  | 13  | may   | 256      | 2        | 351   | 2        | failure  | no  |
| 32  | services      | married  | primary   | no      | 228     | yes     | no   | cellular  | 21  | jul   | 1017     | 6        | -1    | 0        | unknown  | yes |
| 49  | blue-collar   | married  | primary   | no      | 305     | yes     | yes  | telephone | 10  | jul   | 834      | 10       | -1    | 0        | unknown  | no  |
| 49  | services      | married  | secondary | no      | 400     | no      | no   | cellular  | 21  | aug   | 151      | 3        | -1    | 0        | unknown  | no  |
| 53  | retired       | married  | tertiary  | no      | -22     | yes     | no   | cellular  | 31  | jul   | 20       | 3        | -1    | 0        | unknown  | no  |
| 37  | technician    | married  | secondary | no      | 4030    | yes     | no   | unknown   | 26  | may   | 178      | 5        | -1    | 0        | unknown  | no  |
| 49  | blue-collar   | married  | primary   | no      | 365     | no      | no   | unknown   | 26  | may   | 59       | 2        | -1    | 0        | unknown  | no  |
| 41  | self-employed | married  | primary   | no      | 204     | no      | no   | cellular  | 13  | aug   | 111      | 7        | -1    | 0        | unknown  | no  |
| 54  | management    | married  | primary   | no      | 1691    | yes     | no   | cellular  | 20  | nov   | 186      | 1        | -1    | 0        | unknown  | no  |
| 32  | services      | single   | secondary | no      | 406     | yes     | no   | unknown   | 6   | may   | 227      | 1        | -1    | 0        | unknown  | no  |
| 33  | technician    | married  | secondary | no      | -204    | no      | no   | cellular  | 15  | may   | 297      | 3        | -1    | 0        | unknown  | no  |
| 33  | technician    | married  | secondary | no      | 359     | yes     | no   | cellular  | 20  | apr   | 210      | 1        | 242   | 1        | other    | no  |
| 36  | blue-collar   | married  | primary   | no      | 123     | yes     | no   | cellular  | 18  | may   | 9        | 1        | -1    | 0        | unknown  | no  |
| 35  | technician    | married  | tertiary  | no      | 0       | yes     | yes  | cellular  | 23  | sep   | 112      | 1        | 62    | 6        | other    | no  |
| 52  | unknown       | married  | secondary | no      | 142     | no      | no   | cellular  | 25  | aug   | 112      | 6        | -1    | 0        | unknown  | no  |
| 40  | entrepreneur  | married  | tertiary  | no      | 3571    | no      | no   | unknown   | 9   | jun   | 222      | 3        | -1    | 0        | unknown  | no  |
| 44  | technician    | single   | secondary | no      | 33      | yes     | no   | cellular  | 13  | may   | 195      | 3        | 351   | 10       | failure  | no  |
| 45  | blue-collar   | single   | secondary | no      | 0       | yes     | no   | cellular  | 26  | jul   | 768      | 2        | 91    | 5        | success  | no  |
| 30  | student       | single   | secondary | no      | 3096    | no      | no   | cellular  | 26  | jan   | 123      | 1        | -1    | 0        | unknown  | no  |
| 48  | technician    | divorced | secondary | yes     | 20      | no      | no   | cellular  | 27  | aug   | 145      | 3        | -1    | 0        | unknown  | no  |
| 33  | blue-collar   | married  | secondary | no      | 3576    | yes     | no   | cellular  | 2   | apr   | 124      | 1        | 308   | 2        | other    | no  |
| 30  | management    | single   | tertiary  | no      | 424     | no      | no   | cellular  | 16  | apr   | 298      | 4        | -1    | 0        | unknown  | yes |
| 55  | management    | married  | primary   | no      | 2587    | no      | no   | cellular  | 4   | sep   | 328      | 2        | 94    | 2        | success  | yes |
| 57  | technician    | married  | unknown   | no      | 206     | yes     | no   | unknown   | 5   | may   | 216      | 3        | -1    | 0        | unknown  | no  |
| 57  | housemaid     | married  | primary   | no      | 7082    | no      | no   | unknown   | 6   | jun   | 69       | 3        | -1    | 0        | unknown  | no  |
| 51  | unemployed    | married  | tertiary  | no      | 1634    | yes     | no   | cellular  | 22  | jul   | 168      | 4        | -1    | 0        | unknown  | no  |
| 34  | admin.        | single   | secondary | no      | 1       | yes     | no   | cellular  | 22  | jul   | 483      | 7        | -1    | 0        | unknown  | no  |



| age | job           | marital  | education | default | balance | housing | loan | contact   | day | month | duration | campaign | pdays | previous | poutcome | y  |
|-----|---------------|----------|-----------|---------|---------|---------|------|-----------|-----|-------|----------|----------|-------|----------|----------|----|
| 50  | self-employed | married  | tertiary  | no      | 4012    | no      | no   | cellular  | 19  | nov   | 690      | 3        | -1    | 0        | unknown  | no |
| 44  | management    | married  | tertiary  | no      | 3762    | no      | no   | cellular  | 28  | aug   | 125      | 12       | -1    | 0        | unknown  | no |
| 27  | admin.        | married  | secondary | no      | -247    | yes     | yes  | unknown   | 4   | jun   | 344      | 2        | -1    | 0        | unknown  | no |
| 50  | blue-collar   | married  | secondary | no      | 606     | yes     | no   | unknown   | 14  | may   | 673      | 2        | -1    | 0        | unknown  | no |
| 36  | management    | married  | tertiary  | no      | 0       | yes     | no   | cellular  | 28  | jan   | 60       | 1        | 250   | 1        | other    | no |
| 42  | management    | married  | secondary | no      | -189    | yes     | no   | unknown   | 12  | may   | 247      | 2        | -1    | 0        | unknown  | no |
| 32  | unemployed    | single   | secondary | no      | 0       | no      | no   | cellular  | 27  | may   | 123      | 2        | -1    | 0        | unknown  | no |
| 42  | management    | married  | tertiary  | no      | 484     | yes     | no   | cellular  | 11  | jul   | 144      | 2        | -1    | 0        | unknown  | no |
| 43  | technician    | married  | secondary | no      | 0       | no      | yes  | cellular  | 8   | may   | 9        | 2        | 172   | 5        | failure  | no |
| 31  | management    | married  | tertiary  | no      | 395     | yes     | no   | cellular  | 9   | feb   | 155      | 3        | -1    | 0        | unknown  | no |
| 32  | management    | married  | tertiary  | no      | 2349    | no      | no   | cellular  | 14  | aug   | 134      | 5        | -1    | 0        | unknown  | no |
| 55  | blue-collar   | divorced | primary   | no      | 410     | no      | yes  | cellular  | 11  | jul   | 221      | 2        | -1    | 0        | unknown  | no |
| 58  | retired       | married  | primary   | no      | 3401    | yes     | no   | cellular  | 3   | feb   | 84       | 3        | 265   | 2        | failure  | no |
| 46  | blue-collar   | married  | primary   | no      | 222     | no      | no   | cellular  | 8   | aug   | 71       | 2        | -1    | 0        | unknown  | no |
| 55  | blue-collar   | married  | secondary | no      | 989     | yes     | yes  | unknown   | 23  | may   | 246      | 4        | -1    | 0        | unknown  | no |
| 34  | management    | married  | tertiary  | no      | 415     | no      | yes  | cellular  | 23  | jul   | 361      | 2        | -1    | 0        | unknown  | no |
| 28  | unemployed    | single   | secondary | no      | 16      | no      | no   | cellular  | 12  | aug   | 119      | 4        | -1    | 0        | unknown  | no |
| 42  | technician    | divorced | tertiary  | no      | 7190    | yes     | no   | cellular  | 29  | jul   | 375      | 2        | -1    | 0        | unknown  | no |
| 54  | housemaid     | married  | secondary | no      | 209     | yes     | yes  | cellular  | 25  | jul   | 97       | 1        | -1    | 0        | unknown  | no |
| 32  | technician    | single   | secondary | no      | 0       | no      | no   | cellular  | 26  | aug   | 181      | 2        | -1    | 0        | unknown  | no |
| 39  | technician    | single   | tertiary  | no      | 5803    | no      | no   | cellular  | 2   | feb   | 83       | 3        | -1    | 0        | unknown  | no |
| 53  | entrepreneur  | married  | tertiary  | no      | 624     | no      | yes  | cellular  | 21  | jul   | 180      | 4        | -1    | 0        | unknown  | no |
| 58  | technician    | married  | tertiary  | no      | 883     | no      | no   | cellular  | 18  | aug   | 543      | 1        | 78    | 2        | failure  | no |
| 40  | self-employed | married  | secondary | no      | 1516    | yes     | yes  | unknown   | 23  | may   | 373      | 1        | -1    | 0        | unknown  | no |
| 29  | student       | single   | secondary | no      | 872     | yes     | no   | unknown   | 7   | may   | 230      | 1        | -1    | 0        | unknown  | no |
| 33  | technician    | married  | tertiary  | no      | -988    | yes     | yes  | cellular  | 15  | jul   | 83       | 3        | -1    | 0        | unknown  | no |
| 38  | admin.        | single   | secondary | no      | 902     | yes     | no   | telephone | 31  | jul   | 58       | 13       | -1    | 0        | unknown  | no |
| 56  | services      | divorced | secondary | no      | 1007    | no      | no   | cellular  | 29  | jul   | 88       | 6        | -1    | 0        | unknown  | no |
| 40  | services      | married  | secondary | no      | 0       | yes     | no   | cellular  | 2   | jun   | 297      | 1        | 28    | 3        | failure  | no |
| 46  | management    | married  | secondary | no      | 12186   | no      | no   | unknown   | 20  | jun   | 29       | 3        | -1    | 0        | unknown  | no |
| 35  | self-employed | married  | secondary | no      | 524     | yes     | no   | cellular  | 21  | nov   | 484      | 4        | -1    | 0        | unknown  | no |

| age | job           | marital  | education | default | balance | housing | loan | contact   | day | month | duration | campaign | pdays | previous | poutcome | y  |
|-----|---------------|----------|-----------|---------|---------|---------|------|-----------|-----|-------|----------|----------|-------|----------|----------|----|
| 35  | blue-collar   | single   | tertiary  | no      | 0       | yes     | no   | cellular  | 21  | jul   | 262      | 4        | -1    | 0        | unknown  | no |
| 34  | technician    | married  | secondary | no      | 4       | yes     | no   | unknown   | 15  | may   | 644      | 1        | -1    | 0        | unknown  | no |
| 35  | housemaid     | married  | tertiary  | no      | 11219   | no      | no   | cellular  | 12  | aug   | 699      | 2        | 79    | 1        | failure  | no |
| 38  | technician    | single   | tertiary  | no      | 0       | no      | no   | cellular  | 28  | aug   | 247      | 2        | -1    | 0        | unknown  | no |
| 50  | blue-collar   | married  | primary   | no      | 5776    | no      | no   | cellular  | 18  | nov   | 49       | 3        | -1    | 0        | unknown  | no |
| 51  | blue-collar   | divorced | primary   | no      | 37      | yes     | no   | cellular  | 13  | may   | 341      | 1        | -1    | 0        | unknown  | no |
| 59  | self-employed | married  | tertiary  | no      | 593     | no      | no   | unknown   | 13  | may   | 64       | 2        | -1    | 0        | unknown  | no |
| 32  | blue-collar   | married  | secondary | no      | 1148    | yes     | no   | cellular  | 20  | apr   | 121      | 1        | -1    | 0        | unknown  | no |
| 37  | blue-collar   | single   | secondary | no      | 3561    | yes     | no   | unknown   | 2   | jun   | 98       | 1        | -1    | 0        | unknown  | no |
| 42  | admin.        | divorced | secondary | no      | 63      | no      | no   | unknown   | 16  | may   | 88       | 6        | -1    | 0        | unknown  | no |
| 40  | student       | married  | secondary | no      | 260     | no      | no   | unknown   | 20  | jun   | 197      | 6        | -1    | 0        | unknown  | no |
| 28  | technician    | single   | tertiary  | no      | 2269    | yes     | no   | unknown   | 9   | may   | 331      | 1        | -1    | 0        | unknown  | no |
| 47  | admin.        | married  | secondary | no      | 1060    | yes     | no   | unknown   | 6   | jun   | 83       | 3        | -1    | 0        | unknown  | no |
| 27  | self-employed | married  | secondary | no      | -220    | yes     | no   | unknown   | 26  | may   | 138      | 2        | -1    | 0        | unknown  | no |
| 41  | management    | married  | tertiary  | no      | 49      | yes     | no   | cellular  | 17  | jul   | 312      | 5        | -1    | 0        | unknown  | no |
| 47  | blue-collar   | married  | secondary | no      | 480     | yes     | no   | unknown   | 28  | may   | 211      | 2        | -1    | 0        | unknown  | no |
| 53  | services      | married  | secondary | no      | 4994    | no      | no   | cellular  | 18  | aug   | 62       | 6        | -1    | 0        | unknown  | no |
| 53  | technician    | divorced | secondary | no      | 0       | yes     | yes  | cellular  | 15  | jul   | 135      | 5        | -1    | 0        | unknown  | no |
| 47  | management    | married  | tertiary  | no      | 3681    | no      | no   | cellular  | 14  | aug   | 101      | 2        | -1    | 0        | unknown  | no |
| 29  | management    | married  | tertiary  | no      | 451     | no      | no   | cellular  | 28  | jan   | 451      | 1        | -1    | 0        | unknown  | no |
| 70  | retired       | divorced | primary   | no      | 4531    | no      | no   | cellular  | 18  | may   | 445      | 1        | -1    | 0        | unknown  | no |
| 65  | retired       | divorced | primary   | no      | 1840    | no      | no   | cellular  | 1   | jun   | 383      | 2        | 188   | 5        | failure  | no |
| 56  | blue-collar   | married  | secondary | no      | 345     | yes     | no   | telephone | 22  | jul   | 605      | 1        | -1    | 0        | unknown  | no |
| 35  | blue-collar   | married  | secondary | no      | 1287    | yes     | no   | cellular  | 14  | jul   | 330      | 2        | -1    | 0        | unknown  | no |
| 56  | management    | married  | unknown   | no      | 353     | no      | yes  | cellular  | 25  | jul   | 171      | 2        | -1    | 0        | unknown  | no |
| 56  | technician    | married  | secondary | no      | 205     | no      | yes  | cellular  | 23  | jul   | 442      | 2        | -1    | 0        | unknown  | no |
| 38  | unemployed    | married  | primary   | no      | 1147    | yes     | yes  | unknown   | 8   | may   | 249      | 5        | -1    | 0        | unknown  | no |
| 35  | management    | married  | tertiary  | no      | 4286    | yes     | no   | unknown   | 9   | may   | 357      | 1        | -1    | 0        | unknown  | no |
| 38  | admin.        | married  | tertiary  | no      | 605     | no      | no   | cellular  | 18  | nov   | 147      | 2        | -1    | 0        | unknown  | no |
| 49  | blue-collar   | married  | secondary | no      | 8545    | yes     | yes  | cellular  | 6   | may   | 199      | 7        | 167   | 1        | failure  | no |
| 55  | management    | married  | unknown   | no      | 559     | no      | no   | cellular  | 28  | jan   | 271      | 2        | -1    | 0        | unknown  | no |
| 43  | blue-collar   | single   | primary   | no      | 2       | yes     | no   | cellular  | 30  | jul   | 783      | 3        | -1    | 0        | unknown  | no |

| age | job           | marital  | education | default | balance | housing | loan | contact   | day | month | duration | campaign | pdays | previous | poutcome | y   |
|-----|---------------|----------|-----------|---------|---------|---------|------|-----------|-----|-------|----------|----------|-------|----------|----------|-----|
| 43  | management    | married  | tertiary  | no      | 2       | no      | yes  | cellular  | 20  | aug   | 472      | 2        | -1    | 0        | unknown  | no  |
| 30  | blue-collar   | married  | primary   | no      | 1317    | yes     | no   | unknown   | 21  | may   | 395      | 2        | -1    | 0        | unknown  | no  |
| 46  | blue-collar   | married  | secondary | no      | 182     | no      | no   | cellular  | 4   | feb   | 56       | 1        | -1    | 0        | unknown  | no  |
| 50  | admin.        | married  | primary   | no      | 276     | no      | yes  | unknown   | 17  | jun   | 641      | 6        | -1    | 0        | unknown  | no  |
| 26  | unemployed    | single   | secondary | no      | 1064    | no      | no   | unknown   | 31  | may   | 113      | 1        | -1    | 0        | unknown  | no  |
| 45  | admin.        | married  | unknown   | no      | 1783    | no      | no   | unknown   | 20  | jun   | 11       | 11       | -1    | 0        | unknown  | no  |
| 37  | blue-collar   | married  | secondary | no      | -67     | no      | no   | cellular  | 2   | feb   | 157      | 1        | -1    | 0        | unknown  | no  |
| 60  | self-employed | single   | tertiary  | no      | 179     | no      | yes  | cellular  | 29  | jul   | 79       | 4        | -1    | 0        | unknown  | no  |
| 47  | blue-collar   | married  | primary   | no      | 214     | yes     | yes  | unknown   | 9   | jun   | 168      | 1        | -1    | 0        | unknown  | no  |
| 33  | technician    | married  | secondary | no      | 6699    | no      | no   | cellular  | 9   | nov   | 425      | 1        | 462   | 1        | failure  | no  |
| 51  | unemployed    | married  | secondary | no      | 1760    | no      | yes  | cellular  | 19  | nov   | 162      | 1        | -1    | 0        | unknown  | no  |
| 48  | admin.        | single   | secondary | no      | 1150    | yes     | no   | unknown   | 28  | may   | 357      | 1        | -1    | 0        | unknown  | no  |
| 37  | admin.        | married  | secondary | no      | 121     | no      | no   | cellular  | 22  | jul   | 22       | 6        | -1    | 0        | unknown  | no  |
| 29  | technician    | single   | tertiary  | no      | 828     | yes     | no   | cellular  | 16  | apr   | 215      | 3        | 321   | 3        | other    | no  |
| 34  | admin.        | married  | secondary | no      | 244     | yes     | no   | cellular  | 31  | jul   | 56       | 4        | -1    | 0        | unknown  | no  |
| 47  | management    | divorced | tertiary  | no      | 447     | no      | yes  | unknown   | 7   | may   | 126      | 1        | -1    | 0        | unknown  | no  |
| 33  | blue-collar   | married  | primary   | no      | -195    | no      | no   | unknown   | 20  | sep   | 9        | 1        | -1    | 0        | unknown  | no  |
| 33  | management    | single   | tertiary  | no      | 1410    | yes     | no   | cellular  | 5   | feb   | 8        | 1        | 254   | 5        | failure  | no  |
| 63  | retired       | married  | primary   | no      | 1084    | no      | no   | cellular  | 14  | oct   | 201      | 1        | 183   | 1        | success  | no  |
| 51  | self-employed | divorced | primary   | no      | 10924   | no      | no   | cellular  | 6   | may   | 106      | 2        | -1    | 0        | unknown  | no  |
| 41  | management    | married  | tertiary  | no      | 832     | yes     | yes  | cellular  | 21  | nov   | 164      | 1        | 176   | 1        | failure  | no  |
| 56  | retired       | married  | tertiary  | no      | 478     | no      | no   | cellular  | 5   | feb   | 97       | 1        | -1    | 0        | unknown  | no  |
| 60  | management    | single   | tertiary  | no      | 0       | no      | yes  | cellular  | 23  | jul   | 151      | 1        | -1    | 0        | unknown  | no  |
| 41  | self-employed | single   | tertiary  | no      | 3950    | yes     | no   | unknown   | 30  | may   | 96       | 2        | -1    | 0        | unknown  | no  |
| 48  | admin.        | divorced | secondary | no      | 4099    | no      | no   | cellular  | 2   | apr   | 397      | 2        | -1    | 0        | unknown  | yes |
| 60  | technician    | single   | primary   | no      | 11262   | no      | no   | cellular  | 26  | aug   | 323      | 1        | -1    | 0        | unknown  | yes |
| 37  | management    | married  | tertiary  | no      | 393     | yes     | no   | cellular  | 14  | may   | 567      | 3        | 275   | 4        | failure  | yes |
| 59  | retired       | married  | tertiary  | no      | 9676    | no      | no   | telephone | 27  | mar   | 234      | 2        | -1    | 0        | unknown  | yes |
| 37  | management    | single   | unknown   | no      | 504     | yes     | no   | cellular  | 16  | apr   | 555      | 1        | -1    | 0        | unknown  | no  |
| 36  | blue-collar   | single   | secondary | no      | 219     | yes     | yes  | cellular  | 11  | may   | 270      | 2        | 347   | 1        | failure  | no  |
| 57  | management    | married  | tertiary  | no      | 1602    | no      | no   | cellular  | 29  | jul   | 339      | 3        | -1    | 0        | unknown  | no  |

| age | job           | marital  | education | default | balance | housing | loan | contact   | day | month | duration | campaign | pdays | previous | poutcome | y   |
|-----|---------------|----------|-----------|---------|---------|---------|------|-----------|-----|-------|----------|----------|-------|----------|----------|-----|
| 44  | admin.        | married  | secondary | no      | 0       | yes     | yes  | unknown   | 16  | may   | 159      | 1        | -1    | 0        | unknown  | no  |
| 47  | self-employed | married  | tertiary  | no      | 2303    | yes     | no   | cellular  | 6   | may   | 323      | 2        | 364   | 1        | failure  | no  |
| 31  | blue-collar   | married  | secondary | no      | 1660    | yes     | no   | cellular  | 20  | apr   | 342      | 1        | -1    | 0        | unknown  | no  |
| 35  | management    | married  | tertiary  | yes     | 32      | yes     | no   | cellular  | 21  | nov   | 180      | 1        | -1    | 0        | unknown  | no  |
| 39  | management    | married  | tertiary  | no      | 278     | no      | yes  | cellular  | 28  | jul   | 60       | 3        | -1    | 0        | unknown  | no  |
| 50  | management    | married  | tertiary  | no      | 4       | no      | no   | cellular  | 17  | nov   | 223      | 1        | 96    | 2        | failure  | no  |
| 29  | management    | married  | tertiary  | no      | 318     | yes     | no   | cellular  | 18  | may   | 1434     | 2        | 356   | 4        | failure  | no  |
| 28  | blue-collar   | single   | secondary | no      | 517     | no      | yes  | cellular  | 8   | jul   | 681      | 1        | -1    | 0        | unknown  | yes |
| 54  | admin.        | married  | secondary | no      | 0       | no      | no   | cellular  | 28  | jan   | 161      | 1        | 98    | 2        | failure  | yes |
| 75  | retired       | divorced | tertiary  | no      | 3810    | yes     | no   | cellular  | 16  | nov   | 262      | 1        | 183   | 1        | failure  | yes |
| 49  | management    | divorced | tertiary  | no      | 755     | no      | no   | cellular  | 31  | jul   | 1212     | 4        | -1    | 0        | unknown  | yes |
| 30  | management    | married  | tertiary  | no      | 1221    | yes     | no   | telephone | 25  | jul   | 279      | 4        | -1    | 0        | unknown  | yes |
| 35  | technician    | married  | secondary | no      | 4508    | yes     | no   | cellular  | 18  | nov   | 854      | 1        | -1    | 0        | unknown  | yes |
| 34  | self-employed | single   | tertiary  | no      | 5462    | no      | no   | cellular  | 30  | apr   | 430      | 1        | -1    | 0        | unknown  | yes |
| 32  | blue-collar   | married  | secondary | no      | 889     | yes     | no   | cellular  | 10  | sep   | 185      | 1        | 99    | 5        | success  | yes |
| 61  | admin.        | married  | secondary | no      | 4608    | no      | no   | cellular  | 12  | mar   | 212      | 1        | -1    | 0        | unknown  | yes |
| 77  | retired       | married  | tertiary  | no      | 7802    | no      | no   | telephone | 4   | may   | 421      | 1        | 92    | 2        | success  | yes |
| 53  | admin.        | married  | secondary | no      | 1796    | no      | no   | telephone | 16  | mar   | 295      | 1        | 91    | 2        | success  | yes |
| 30  | services      | single   | secondary | no      | 209     | yes     | no   | unknown   | 6   | may   | 30       | 2        | -1    | 0        | unknown  | no  |
| 46  | services      | married  | secondary | no      | 271     | yes     | no   | unknown   | 16  | may   | 134      | 4        | -1    | 0        | unknown  | no  |
| 27  | management    | single   | unknown   | no      | 3196    | no      | no   | cellular  | 9   | feb   | 10       | 2        | -1    | 0        | unknown  | no  |
| 26  | student       | single   | secondary | no      | 2572    | yes     | no   | cellular  | 5   | may   | 620      | 1        | -1    | 0        | unknown  | no  |
| 36  | management    | single   | tertiary  | no      | 339     | no      | no   | cellular  | 21  | nov   | 6        | 6        | -1    | 0        | unknown  | no  |
| 42  | blue-collar   | married  | primary   | no      | 0       | yes     | yes  | cellular  | 7   | may   | 209      | 2        | -1    | 0        | unknown  | no  |
| 50  | housemaid     | married  | primary   | no      | 395     | yes     | no   | unknown   | 15  | may   | 419      | 2        | -1    | 0        | unknown  | no  |
| 49  | blue-collar   | married  | secondary | no      | 2039    | yes     | no   | cellular  | 18  | nov   | 283      | 2        | -1    | 0        | unknown  | no  |
| 45  | management    | married  | tertiary  | no      | 5450    | no      | no   | cellular  | 8   | aug   | 188      | 2        | -1    | 0        | unknown  | no  |
| 23  | student       | single   | secondary | no      | 780     | no      | no   | cellular  | 16  | sep   | 267      | 1        | -1    | 0        | unknown  | no  |
| 46  | admin.        | married  | secondary | no      | 22      | no      | no   | cellular  | 26  | aug   | 388      | 6        | -1    | 0        | unknown  | yes |
| 44  | technician    | single   | secondary | no      | 244     | yes     | no   | cellular  | 12  | aug   | 1735     | 4        | -1    | 0        | unknown  | yes |
| 19  | student       | single   | primary   | no      | 103     | no      | no   | cellular  | 10  | jul   | 104      | 2        | -1    | 0        | unknown  | yes |
| 30  | blue-collar   | single   | secondary | no      | 315     | no      | no   | cellular  | 23  | jun   | 1407     | 1        | -1    | 0        | unknown  | yes |

| age | job           | marital  | education | default | balance | housing | loan | contact   | day | month | duration | campaign | pdays | previous | poutcome | y   |
|-----|---------------|----------|-----------|---------|---------|---------|------|-----------|-----|-------|----------|----------|-------|----------|----------|-----|
| 52  | technician    | married  | secondary | no      | 195     | yes     | no   | cellular  | 18  | feb   | 220      | 1        | 63    | 5        | failure  | yes |
| 36  | blue-collar   | married  | primary   | no      | 1925    | yes     | no   | cellular  | 20  | apr   | 904      | 2        | -1    | 0        | unknown  | yes |
| 47  | admin.        | single   | secondary | no      | 3696    | no      | no   | cellular  | 12  | jul   | 250      | 2        | 181   | 4        | success  | yes |
| 38  | technician    | single   | secondary | no      | 0       | yes     | no   | cellular  | 7   | apr   | 196      | 2        | -1    | 0        | unknown  | no  |
| 32  | management    | single   | tertiary  | no      | 8436    | yes     | no   | cellular  | 17  | apr   | 245      | 1        | 149   | 2        | failure  | no  |
| 30  | entrepreneur  | married  | secondary | no      | 946     | yes     | no   | cellular  | 19  | nov   | 1065     | 1        | 183   | 1        | other    | no  |
| 58  | housemaid     | married  | tertiary  | no      | 497     | no      | no   | unknown   | 11  | jun   | 138      | 5        | -1    | 0        | unknown  | no  |
| 60  | retired       | married  | primary   | no      | 55      | no      | no   | cellular  | 4   | jun   | 207      | 1        | -1    | 0        | unknown  | no  |
| 38  | services      | married  | secondary | no      | -140    | no      | no   | unknown   | 25  | jun   | 456      | 3        | -1    | 0        | unknown  | no  |
| 51  | technician    | married  | secondary | no      | -379    | yes     | no   | cellular  | 5   | aug   | 261      | 1        | -1    | 0        | unknown  | no  |
| 42  | admin.        | married  | secondary | no      | -396    | yes     | no   | cellular  | 11  | may   | 80       | 1        | 308   | 4        | failure  | no  |
| 30  | services      | single   | secondary | no      | 0       | yes     | no   | cellular  | 18  | may   | 131      | 6        | 363   | 9        | failure  | no  |
| 40  | management    | married  | primary   | no      | 733     | yes     | no   | unknown   | 6   | jun   | 94       | 7        | -1    | 0        | unknown  | no  |
| 53  | technician    | divorced | primary   | no      | 7290    | yes     | no   | unknown   | 30  | may   | 153      | 1        | -1    | 0        | unknown  | no  |
| 27  | services      | single   | secondary | no      | 468     | yes     | no   | cellular  | 13  | may   | 109      | 1        | -1    | 0        | unknown  | no  |
| 39  | blue-collar   | married  | secondary | no      | 0       | yes     | no   | cellular  | 30  | apr   | 53       | 1        | 325   | 1        | failure  | no  |
| 39  | blue-collar   | married  | primary   | no      | 3705    | yes     | no   | unknown   | 16  | jun   | 77       | 1        | -1    | 0        | unknown  | no  |
| 37  | services      | married  | secondary | no      | 1328    | yes     | no   | telephone | 13  | may   | 121      | 2        | 341   | 4        | failure  | no  |
| 41  | blue-collar   | married  | primary   | no      | 8       | yes     | no   | cellular  | 15  | jul   | 155      | 2        | -1    | 0        | unknown  | no  |
| 45  | blue-collar   | married  | primary   | no      | 61      | no      | no   | cellular  | 17  | apr   | 108      | 1        | -1    | 0        | unknown  | no  |
| 48  | management    | married  | tertiary  | no      | 6613    | no      | no   | unknown   | 17  | jun   | 159      | 3        | -1    | 0        | unknown  | no  |
| 48  | technician    | married  | secondary | no      | 302     | yes     | no   | unknown   | 26  | may   | 208      | 3        | -1    | 0        | unknown  | no  |
| 32  | blue-collar   | married  | secondary | no      | 1005    | no      | no   | cellular  | 12  | may   | 215      | 1        | -1    | 0        | unknown  | no  |
| 35  | services      | married  | secondary | no      | 109     | yes     | no   | unknown   | 6   | may   | 597      | 3        | -1    | 0        | unknown  | no  |
| 33  | management    | single   | tertiary  | no      | 1240    | yes     | no   | cellular  | 4   | feb   | 505      | 2        | 260   | 1        | other    | no  |
| 28  | blue-collar   | single   | primary   | no      | 0       | yes     | yes  | unknown   | 3   | jun   | 380      | 4        | -1    | 0        | unknown  | no  |
| 64  | retired       | married  | unknown   | no      | 2923    | no      | no   | cellular  | 12  | mar   | 120      | 1        | -1    | 0        | unknown  | no  |
| 35  | technician    | married  | secondary | no      | 0       | no      | no   | cellular  | 27  | aug   | 106      | 4        | -1    | 0        | unknown  | no  |
| 43  | management    | single   | tertiary  | no      | 1304    | no      | no   | cellular  | 6   | aug   | 501      | 2        | -1    | 0        | unknown  | yes |
| 41  | blue-collar   | married  | secondary | yes     | 720     | no      | yes  | cellular  | 24  | jul   | 651      | 1        | -1    | 0        | unknown  | yes |
| 28  | technician    | single   | secondary | no      | 227     | no      | no   | cellular  | 30  | apr   | 278      | 1        | -1    | 0        | unknown  | yes |
| 32  | services      | single   | secondary | no      | 2881    | no      | no   | cellular  | 2   | jul   | 356      | 1        | -1    | 0        | unknown  | yes |
| 41  | self-employed | married  | tertiary  | no      | 231     | no      | no   | cellular  | 7   | aug   | 352      | 2        | -1    | 0        | unknown  | yes |

| age | job           | marital  | education | default | balance | housing | loan | contact   | day | month | duration | campaign | pdays | previous | poutcome | y   |
|-----|---------------|----------|-----------|---------|---------|---------|------|-----------|-----|-------|----------|----------|-------|----------|----------|-----|
| 35  | technician    | married  | tertiary  | no      | 7050    | no      | no   | cellular  | 11  | aug   | 222      | 3        | 182   | 4        | success  | yes |
| 30  | management    | single   | tertiary  | no      | 2       | yes     | yes  | cellular  | 25  | jul   | 1031     | 8        | -1    | 0        | unknown  | yes |
| 58  | self-employed | divorced | tertiary  | no      | 1382    | no      | no   | unknown   | 9   | jun   | 700      | 1        | -1    | 0        | unknown  | yes |
| 27  | management    | single   | tertiary  | no      | 703     | yes     | no   | unknown   | 5   | jun   | 590      | 3        | -1    | 0        | unknown  | yes |
| 39  | management    | married  | tertiary  | no      | 2763    | yes     | no   | unknown   | 2   | jun   | 526      | 2        | -1    | 0        | unknown  | yes |
| 46  | self-employed | married  | secondary | no      | 548     | no      | no   | cellular  | 12  | aug   | 1181     | 4        | -1    | 0        | unknown  | yes |
| 28  | management    | single   | tertiary  | no      | 0       | yes     | no   | cellular  | 16  | jul   | 767      | 5        | -1    | 0        | unknown  | yes |
| 38  | technician    | single   | tertiary  | no      | 25      | yes     | no   | cellular  | 1   | jun   | 232      | 2        | -1    | 0        | unknown  | yes |
| 28  | services      | single   | secondary | no      | 168     | no      | no   | cellular  | 1   | jun   | 347      | 3        | 140   | 2        | success  | yes |
| 32  | management    | divorced | tertiary  | no      | 874     | no      | no   | cellular  | 12  | aug   | 1282     | 6        | -1    | 0        | unknown  | yes |
| 59  | retired       | married  | unknown   | no      | 1033    | no      | no   | unknown   | 11  | jun   | 1199     | 1        | -1    | 0        | unknown  | yes |
| 33  | blue-collar   | married  | secondary | no      | 43      | yes     | no   | cellular  | 14  | may   | 332      | 2        | 358   | 2        | failure  | no  |
| 41  | technician    | married  | secondary | no      | 288     | yes     | no   | unknown   | 15  | may   | 221      | 4        | -1    | 0        | unknown  | no  |
| 34  | admin.        | married  | secondary | no      | 613     | yes     | no   | cellular  | 4   | jun   | 80       | 1        | -1    | 0        | unknown  | no  |
| 37  | technician    | single   | tertiary  | no      | 6101    | no      | no   | cellular  | 4   | may   | 60       | 5        | 87    | 1        | success  | no  |
| 38  | blue-collar   | married  | primary   | no      | 168     | yes     | yes  | telephone | 13  | may   | 212      | 3        | -1    | 0        | unknown  | no  |
| 39  | blue-collar   | married  | primary   | no      | 7685    | yes     | no   | cellular  | 14  | apr   | 493      | 3        | -1    | 0        | unknown  | no  |
| 35  | technician    | divorced | tertiary  | no      | 4515    | yes     | no   | cellular  | 19  | nov   | 98       | 2        | -1    | 0        | unknown  | no  |
| 27  | blue-collar   | single   | secondary | no      | 148     | no      | no   | cellular  | 22  | jul   | 203      | 1        | -1    | 0        | unknown  | no  |
| 48  | retired       | married  | primary   | yes     | -285    | no      | yes  | cellular  | 19  | nov   | 157      | 1        | -1    | 0        | unknown  | no  |
| 50  | blue-collar   | married  | secondary | no      | 887     | no      | no   | unknown   | 16  | jun   | 287      | 1        | -1    | 0        | unknown  | no  |
| 45  | self-employed | married  | secondary | no      | 3098    | yes     | no   | cellular  | 7   | aug   | 159      | 2        | -1    | 0        | unknown  | no  |
| 41  | technician    | married  | tertiary  | no      | 744     | yes     | no   | unknown   | 14  | may   | 125      | 1        | -1    | 0        | unknown  | no  |
| 39  | technician    | married  | secondary | no      | 130     | no      | no   | cellular  | 8   | jul   | 202      | 2        | -1    | 0        | unknown  | no  |
| 44  | management    | married  | secondary | no      | 438     | yes     | no   | telephone | 8   | may   | 37       | 3        | 303   | 1        | failure  | no  |
| 49  | management    | married  | tertiary  | no      | 322     | no      | no   | cellular  | 6   | aug   | 72       | 4        | -1    | 0        | unknown  | no  |
| 34  | blue-collar   | single   | secondary | no      | 186     | no      | no   | cellular  | 2   | feb   | 342      | 1        | -1    | 0        | unknown  | no  |
| 42  | management    | married  | tertiary  | no      | 0       | no      | no   | cellular  | 13  | aug   | 325      | 2        | -1    | 0        | unknown  | no  |
| 53  | blue-collar   | married  | secondary | no      | 4380    | no      | no   | telephone | 19  | aug   | 223      | 7        | -1    | 0        | unknown  | no  |
| 31  | services      | single   | secondary | no      | 2070    | no      | no   | cellular  | 29  | aug   | 319      | 2        | -1    | 0        | unknown  | no  |
| 45  | blue-collar   | single   | primary   | no      | 999     | no      | no   | cellular  | 29  | jan   | 182      | 2        | -1    | 0        | unknown  | no  |

| age | job           | marital  | education | default | balance | housing | loan | contact  | day | month | duration | campaign | pdays | previous | poutcome | y  |
|-----|---------------|----------|-----------|---------|---------|---------|------|----------|-----|-------|----------|----------|-------|----------|----------|----|
| 39  | self-employed | married  | secondary | no      | 167     | yes     | yes  | cellular | 17  | jul   | 330      | 1        | -1    | 0        | unknown  | no |
| 46  | technician    | married  | secondary | no      | 1322    | yes     | yes  | cellular | 19  | nov   | 514      | 1        | -1    | 0        | unknown  | no |
| 42  | admin.        | divorced | secondary | no      | 1076    | yes     | no   | unknown  | 14  | may   | 213      | 1        | -1    | 0        | unknown  | no |
| 37  | management    | married  | tertiary  | no      | 502     | yes     | no   | cellular | 7   | may   | 551      | 1        | 351   | 1        | failure  | no |