

# IMPLEMENTASI MODEL CRISP-DM UNTUK MENENTUKAN SALES PIPELINE PADA PT X

Dina Feblian<sup>1)</sup>, Dadan Umar Daihani<sup>2)</sup>

<sup>1)</sup>Mahasiswa Magister Teknik Industri, Fakultas Teknologi Industri, Universitas Trisakti

<sup>2)</sup>Dosen Magister Teknik Industri, Fakultas Teknologi Industri, Universitas Trisakti

dina.feblian@gmail.com, dadan@trisakti.ac.id

## ABSTRACT

*PT X is a joint venture subsidiary of a bank in Indonesia, which is a company, engaged in general insurance business. PT X is required to increase its share of the parent company by the shareholders in developing its business potential. In developing the potential of the business, companies need enough information to be analyzed further. Sources of information that can be used is the business data of the company, both of PT X as a subsidiary and its holding company. By utilizing business data in large numbers, the company will find a variety of information needed. For the business data processing required a technique of data mining. In a study conducted, the implementation of the model refers to the six stages of Cross-Industry Standard Process for Data Mining or commonly known by the acronym CRISP-DM. In doing data mining, the six phases of the CRISP-DM is business understanding, understanding the data, the data preparation, modeling, evaluation, and deployment. At this stage of deployment, the report will be made by the customer segmentation by region. The expected result is the information required in the form of a new customer pipeline.*

**Keywords:** Data Mining, CRISP-DM, Customer Segmentation, Customer Pipeline

## 1. PENDAHULUAN

### 1.1 Latar Belakang

Data merupakan sebuah aset yang berharga. Seiring dengan pesatnya perkembangan teknologi pada pengumpulan dan penyimpanan data, menghasilkan basis data yang terlampau besar. Sehingga, data yang dikumpulkan jarang dilihat karena terlalu panjang, tidak menarik dan memakan waktu yang cukup lama untuk mengolah data tersebut. Keputusan yang seharusnya dibuat berdasarkan data seringkali dibuat tidak lagi berdasarkan data, tetapi berdasar intuisi pembuat keputusan, sehingga terciptalah cabang ilmu *data mining*.

*Data mining* merupakan suatu proses pencarian data secara otomatis yang bersumber dari *database* dalam skala besar untuk mendapatkan informasi yang berguna [12]. Perkembangan *data mining* tidak dapat lepas dari teknologi informasi yang memungkinkan data dalam jumlah besar terakumulasi. Dari keterkaitan tersebut, menjadikan *data mining* telah berkembang pada berbagai sektor bisnis.

Dalam dunia bisnis yang penuh dengan persaingan dan bersifat dinamis, membuat para

pelakunya untuk selalu memikirkan strategi agar dapat terus bertahan dan jika mungkin mengembangkan skala bisnis mereka. Untuk mencapai hal tersebut dapat dilakukan upaya seperti peningkatan kapasitas produk, efisiensi biaya operasional perusahaan dan peningkatan efektivitas pemasaran. Salah satu cara untuk memenuhi kebutuhan diatas yaitu dengan melakukan analisis data perusahaan.

Seperti PT X yang didirikan di Jakarta pada tahun 2011 merupakan salah satu anak perusahaan *joint venture* dari sebuah bank di Indonesia, yaitu perusahaan yang bergerak dalam bidang asuransi kerugian. Tidak hanya PT X yang bergerak dalam bidang usaha tersebut, masih cukup banyak perusahaan lain yang bergerak dibidang serupa. Hal ini tentu saja menimbulkan persaingan bisnis antar perusahaan. Persaingan yang terjadi dalam dunia bisnis asuransi kerugian memaksa para pelakunya untuk selalu memikirkan strategi yang dapat menjamin kelangsungan dari bisnis asuransi yang mereka jalankan serta meningkatkan pendapatan perusahaan. Sebagai anak perusahaan, PT X juga dituntut untuk meningkatkan pangsa pasarnya terhadap

perusahaan induk oleh para pemegang saham dalam mengembangkan potensi bisnisnya.

Untuk dapat melaksanakan hal itu, perusahaan memerlukan informasi yang cukup untuk dapat dianalisa lebih lanjut. Sumber informasi yang dapat digunakan yaitu data bisnis perusahaan, baik dari PT X sebagai perusahaan anak dan perusahaan induknya. Data bisnis dalam jumlah yang besar merupakan salah satu aset berharga yang dimiliki oleh perusahaan, namun apabila data tersebut tidak terlihat lagi nilainya karena selain berupa tumpukan data, data tersebut adalah data yang tidak terpakai lagi. Akan tetapi, dengan memanfaatkan data bisnis dalam jumlah yang besar, perusahaan akan menemukan beragam informasi yang dibutuhkan. Untuk proses pengolahan data bisnis tersebut diperlukan suatu teknik dari *data mining*.

## 1.2 Tujuan Penelitian

Tujuan dari pelaksanaan penelitian ini yaitu:

1. Mengembangkan pemodelan melalui teknik dari *data mining* untuk mengendalikan data bisnis dengan segmentasi pasar.
2. Menemukan informasi yang dibutuhkan berupa *sales pipeline* dari pemodelan yang dilakukan pada data bisnis yang bersumber dari kedua buah perusahaan yaitu, PT X dan perusahaan induknya.
3. Mengetahui ukuran pangsa pasar pada segmen *Commercial* yang telah dicapai oleh PT X sebagai anak perusahaan terhadap perusahaan induknya.

## 2. TINJAUAN PUSTAKA

### 2.1 Sales Pipeline

Menurut Cefkin (2007), manajemen *sales pipeline* digunakan pada perusahaan komersial sebagai sarana untuk mengarahkan fokus dan perhatian personil serta untuk mengantisipasi kinerja pada masa yang akan datang. Sistem akuntansi pada *sales pipeline* digunakan untuk memantau potensi pendapatan, mengelola keputusan investasi, dan untuk menetapkan proyeksi yang akan datang. Karena itu, *sales pipeline* dapat dijadikan sebagai petunjuk oleh pemegang saham dan analis pasar.

*Sales Pipeline* menurut Kimla (2013) dibuat secara wajib dan digunakan untuk penjualan serta *forecasting* dan analisis. Secara berkala dilakukan evaluasi untuk efektivitas dan perbaikan. *Pipeline* yang fleksibel merupakan metode yang paling sukses digunakan oleh beberapa perusahaan. Umpan balik secara konsisten dari tenaga penjual dibutuhkan untuk akurasi *pipeline* tersebut dan sebagai refleksi dari kondisi nyata yang ada. *Pipeline* disesuaikan setiap kali akan kebutuhannya untuk mencerminkan perubahan yang dibutuhkan dalam proses penjualan [5].

Model operasi *pipeline* secara umum dijelaskan oleh Yan, et al., (2015), yaitu seperti mengidentifikasi prospek penjualan baru, kemudian penjual memasukkan peluang prospek penjualan pada sistem manajemen *pipeline*. Prospek yang telah teridentifikasi ini selanjutnya dievaluasi dan ditetapkan beberapa yang berpotensi menjadi peluang penjualan. Sebuah peluang penjualan terdiri dari satu set dari satu atau lebih produk atau jasa yang penjual berusaha untuk mengkonversi kedalam pembelian klien yang sebenarnya [14].

Berdasarkan penjelasan diatas, dapat disimpulkan bahwa *sales pipeline* menggambarkan pendekatan untuk penjualan, dibuat berdasarkan prinsip yang mendasari proses penjualan dari pihak yang melakukan penjualan. Menggambarkan langkah *sales* dari pelanggan potensial atau prospek untuk mengkualifikasi prospek tersebut menjadi petunjuk dan memvalidasinya lebih lanjut yang mengarah kepada peluang penjualan. *Sales pipeline* digunakan agar angka penjualan dapat dikendalikan dan pekerjaan dalam *sales* terorganisir dengan baik.

### 2.2 Segmentasi Pasar

Menurut Lamb, Hair, dan McDaniel (2010), segmen pasar adalah subkelompok orang atau organisasi yang berbagi satu atau lebih karakteristik yang menyebabkan memiliki kebutuhan produk serupa [7]. Dari perspektif pemasaran, segmen pasar dapat digambarkan sebagai suatu tempat antara dua ekstrem. Segmentasi Pasar atau *market segmentation* menurut Kotler dan Keller (2012), adalah pasar yang telah dibagi terdiri dari sekelompok pelanggan yang memiliki sekumpulan kebutuhan dan keinginan yang serupa [6]. Proses membagi pasar menjadi segmen-segmen yang bermakna, relatif sama, dan atau

kelompok yang dapat diidentifikasi disebut sebagai segmentasi pasar. Tujuan dari segmentasi pasar adalah untuk memungkinkan pemasar menyesuaikan bauran pemasaran untuk memenuhi kebutuhan suatu segmen secara spesifik [7].

Segmentasi dapat dilakukan berdasar atas variabel utama seperti dibawah ini, yaitu: [6]

1. Segmentasi geografis, segmentasi yang membagi pasar menjadi unit geografis yang berbeda-beda seperti negara, daerah otonomi, kota, iklim, atau kawasan pemukiman.
2. Segmentasi demografis, segmentasi yang membagi pasar menjadi berbagai kelompok berdasarkan variabel seperti usia, ukuran keluarga, siklus kehidupan keluarga, jenis kelamin, penghasilan, pekerjaan, agama, ras, generasi kewarganegaraan, dan kelas sosial.
3. Segmentasi psikografis, segmen pasar ini dilakukan dengan mengelompokkan konsumen atau pembeli menjadi bagian pasar menurut variabel-variabel pola atau gaya hidup dan kepribadian.
4. Segmentasi perilaku, pada segmentasi ini pasar diklasifikasi dalam kelompok-kelompok yang dibedakan berdasarkan pengetahuan, sikap, penggunaan atau respon terhadap suatu produk.

Pada penelitian ini, untuk menentukan segmentasi dan mengetahui ukuran pangsa pasar PT X terhadap perusahaan induknya yang difokuskan pada satu segmen, yaitu segmen *Commercial* dengan menggunakan data bisnis dari kedua belah pihak perusahaan. Dalam proses pengolahan data bisnis tersebut menggunakan penerapan dari *data mining*.

### 2.3 Data Mining

Menurut Gartner Group, *data mining* merupakan suatu proses pengidentifikasian sekumpulan data dalam jumlah yang besar serta tersimpan dalam tempat penyimpanan, melalui teknik-teknik pengenalan pola seperti matematika dan teknik statistik [8]. *Data mining* adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi didalam *database*. *Data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi

pengetahuan potensial dan berguna yang bermanfaat yang tersimpan didalam *database* besar [13].

*Data mining* menurut David Hand, Heikki Mannila dan Padhraic Smyth dari MIT adalah analisa terhadap data yang biasanya berukuran besar untuk menemukan hubungan yang jelas serta menyimpulkan hasil dari analisa data yang belum diketahui sebelumnya dengan cara terkini untuk dipahami dan dapat berguna bagi pemilik data tersebut [8].

Selain definisi diatas, *data mining* adalah analisis data untuk membangun hubungan dan mengidentifikasi pola, berfokus pada identifikasi hubungan dalam data. *Data mining* merupakan sebuah kegiatan ekstraksi informasi yang tujuannya adalah untuk menemukan fakta-fakta tersembunyi yang terkandung dalam *database*. Menggunakan kombinasi dari pembelajaran mesin, analisis statistik, teknik pemodelan dan teknologi *database*. *Data mining* menemukan pola dan hubungan dalam data dan menyimpulkan aturan yang memungkinkan sebagai prediksi untuk masa yang akan datang [9].

*Data mining* merupakan suatu proses pencarian data secara otomatis yang bersumber dari *database* dalam skala besar untuk mendapatkan informasi yang berguna [12]. *Data mining* merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. *Data mining* merupakan bidang dari beberapa keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, *database* dan visualisasi untuk penanganan permasalahan dalam pengambilan informasi dari *database* yang besar [8].

Beberapa faktor pendorong yang mendukung kemajuan yang luar biasa dalam bidang *data mining*, antara lain: [8]

1. Pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam *data warehouse* sehingga seluruh perusahaan memiliki akses yang baik ke dalam *database* yang ada.
3. Adanya peningkatan akses data melalui navigasi web dan intranet.
4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.

5. Perkembangan teknologi perangkat lunak untuk *data mining*.
6. Perkembangan yang pesat dalam kemajuan komputasi dan pengembangan kapasitas media penyimpanan.

Berdasarkan pada definisi yang telah disampaikan, hal penting yang berhubungan dengan *data mining*, yaitu:

1. *Data mining* merupakan suatu proses otomatis terhadap data yang sudah ada sebelumnya.
2. Data yang akan diproses berupa data dalam skala besar.
3. Tujuan dari *data mining* adalah mendapatkan hubungan atau pola yang dapat berupa suatu informasi yang mungkin memberikan indikasi yang bermanfaat.

Pencarian suatu hubungan atau pola didalam *data mining* dapat berupa hubungan antara dua atau lebih dalam satu dimensi. Misalnya dalam dimensi produk, dapat dilihat keterkaitan pembelian suatu produk dengan produk yang lain. Selain itu, hubungannya juga dapat dilihat antara dua atribut atau lebih dan dua objek ataupun lebih [10].

Sementara itu, penemuan suatu pola merupakan *output* lain dari *data mining*. Misalnya, sebuah perusahaan perbankan yang akan meningkatkan fasilitas pinjaman untuk nasabah. Perusahaan tersebut akan mencari pola dari daftar nasabah yang ada untuk mengetahui nasabah yang memiliki potensial dan nasabah yang tidak memiliki potensial, sehingga perusahaan juga dapat melakukan analisa perihal nasabah yang dapat dinaikkan fasilitas pinjamannya [10].

Beberapa penjelasan mengenai definisi awal dari *data mining* menyertakan fokus pada proses otomatisasi. Berry dan Linoff, (2004) dalam buku *Data Mining Technique for Marketing, Sales, and Customer Support* mendefinisikan *data mining* sebagai suatu proses eksplorasi dan analisis secara otomatis maupun semi otomatis terhadap data dalam jumlah besar dengan tujuan menemukan pola atau aturan yang berarti. Tiga tahun kemudian dalam buku *Mastering Data Mining*, mereka memberikan definisi ulang terhadap pengertian dari *data mining* dan memberikan pernyataan bahwa “jika ada yang kami sesalkan adalah frasa secara otomatis maupun semi otomatis, karena kami merasa hal tersebut memberikan fokus berlebih pada teknik otomatis dan kurang pada eksplorasi dan analisis”. Hal tersebut

memberikan pemahaman yang salah bahwa *data mining* merupakan produk yang dapat dibeli dibandingkan keilmuan yang harus dikuasai [8].

Pernyataan tersebut menegaskan bahwa dalam *data mining* otomatisasi tidak menggantikan campur tangan manusia. Manusia harus ikut aktif dalam setiap fase dalam proses *data mining*. Kehebatan kemampuan algoritma *data mining* yang terdapat dalam perangkat lunak analisis yang terdapat saat ini memungkinkan terjadinya kesalahan penggunaan yang dapat berakibat fatal. Pengguna mungkin menerapkan analisis yang tidak tepat terhadap kumpulan data dengan menggunakan pendekatan yang berbeda. Oleh karenanya, dibutuhkan pemahaman tentang statistik dan struktur model matematika yang mendasari kinerja dari perangkat lunak [8].

*Data Mining* bukanlah suatu bidang ilmu yang sama sekali baru [1][4] dan teorinya pun sudah banyak dibahas dalam berbagai literatur. Teori-teori tersebut antara lain yaitu, Naive-Bayes dan Nearest Neighbour, Pohon Keputusan, Aturan Asosiasi, *K-Means Clustering*, dan *Text Mining* [1]. Sedangkan perkembangan terkini menghadirkan algoritma-algoritma yang dikembangkan antara lain, Jaringan Syaraf Tiruan (JST), Algoritma Genetik, *Fuzzy C-Means*, *Support Vector Machine* (SVM), dan lain-lain [8]. Sejak perkembangannya dari tahun 1990-an, *data mining* yang sebelumnya dikuasai oleh para pakar statistik, ternyata pada tahun 2000-an banyak dikembangkan oleh para pakar ilmu komputer yang berusaha untuk memperbaiki kinerja dari teori-teori statistik yang ada. Menurut Daryl Pregibon disebutkan bahwa *data mining* adalah perpaduan dari Statistik, *Artificial Intelligence* dan *Database* [3].

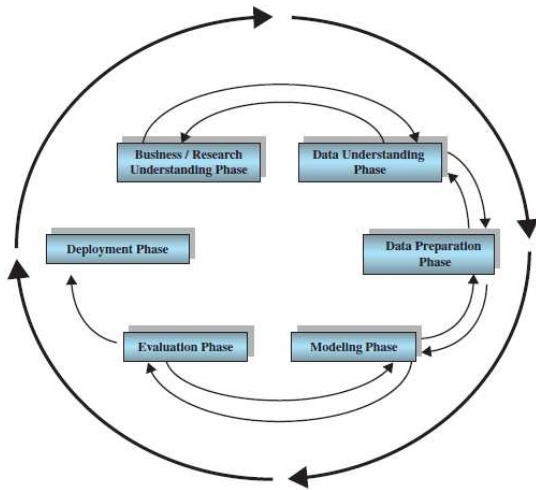
#### **2.4 Cross-Industry Standard Process untuk Data Mining**

*Cross-Industry Standard Process* untuk *Data Mining* atau CRISP-DM dikembangkan tahun 1996 oleh analis dari beberapa industri seperti Daimler Chrysler, SPSS dan NCR [8][11]. CRISP-DM merupakan standarisasi proses *data mining* sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian [8].

Dalam CRISP-DM sebuah proyek *data mining* memiliki siklus hidup yang terbagi dalam enam fase [8][11]. Berikut merupakan



gambar 2.3 yang akan memberikan gambaran secara umum mengenai siklus hidup dalam CRISP-DM.



**Gambar 2.1** Siklus Hidup dalam CRISP-DM (Larose, 2014)

Keseluruhan fase berurutan yang ada tersebut bersifat adaptif. Fase berikutnya dalam urutan bergantung kepada keluaran dari fase sebelumnya. Hubungan penting antar fase digambarkan dengan panah. Sebagai contoh, jika proses berada pada fase *modeling*. Berdasar pada perilaku dan karakteristik model, proses mungkin kembali kepada fase *data preparation* untuk perbaikan lebih lanjut terhadap data atau berpindah maju kepada fase *evaluation* [8].

Dari gambar 2.3 dapat dilihat bahwa siklus hidup dari CRISP-DM terdiri dari enam fase, yaitu: [8]

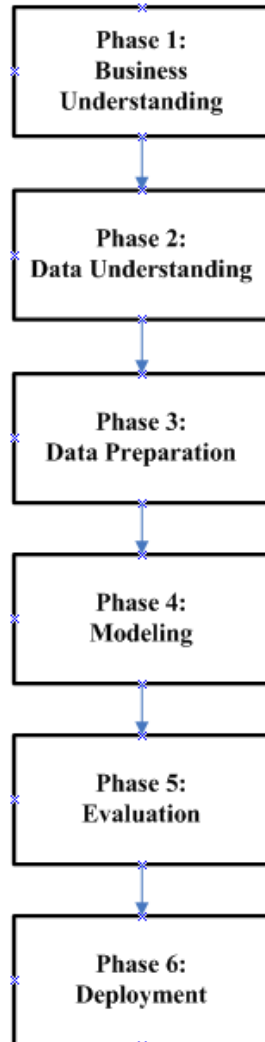
1. *Business Understanding Phase* atau Fase Pemahaman Bisnis
  - a. Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan.
  - b. Menerjemahkan tujuan dan batasan menjadi formula dan definisi dari permasalahan *data mining*.
  - c. Menyiapkan strategi awal untuk mencapai tujuan.
2. *Data Understanding Phase* atau Fase Pemahaman Data
  - a. Mengumpulkan data.
  - b. Menggunakan analisis penyelidikan data untuk mengenali lebih lanjut mengenai data dan pencarian awal pengetahuan yang ada didalamnya.
  - c. Mengevaluasi kualitas data.

- d. Jika diinginkan, pilih sebagian kecil kelompok data yang mungkin mengandung pola dari permasalahan.
3. *Data Preparation Phase* atau Fase Pengolahan Data
  - a. Menyiapkan data yang tersedia dari awal, karena merupakan kumpulan data yang akan digunakan untuk keseluruhan fase berikutnya. Fase ini merupakan pekerjaan berat yang perlu dilaksanakan secara intensif.
  - b. Memilih kasus dan variabel yang ingin dianalisis dan yang sesuai dengan analisis yang akan dilakukan.
  - c. Melakukan perubahan pada beberapa variabel jika dibutuhkan.
  - d. Menyiapkan data awal sehingga siap untuk perangkat pemodelan.
4. *Modeling Phase* atau Fase Pemodelan
  - a. Memilih dan mengaplikasikan teknik pemodelan yang sesuai.
  - b. Kalibrasi aturan model untuk mengoptimalkan hasil.
  - c. Yang perlu diperhatikan bahwa beberapa teknik mungkin untuk digunakan pada permasalahan *data mining* yang sama.
  - d. Jika diperlukan, proses dapat kembali kepada fase pengolahan data untuk menjadikan data kedalam bentuk yang sesuai dengan spesifikasi kebutuhan dari teknik *data mining* tertentu.
5. *Evaluation Phase* atau Fase Evaluasi
  - a. Mengevaluasi satu atau lebih model yang digunakan dalam fase pemodelan untuk mendapatkan kualitas dan efektivitas sebelum disebarkan untuk digunakan.
  - b. Menetapkan apakah terdapat model yang memenuhi tujuan pada fase awal.
  - c. Menentukan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik.
  - d. Mengambil keputusan berkaitan dengan penggunaan hasil dari *data mining*.
6. *Deployment Phase* atau Fase Penyebaran
  - a. Menggunakan model yang dihasilkan. Terbentuknya model tidak menandakan telah terselesaikannya proyek.
  - b. Contoh sederhana dari penyebaran yaitu pembuatan laporan.

- c. Contoh kompleks dari penyebaran yaitu penerapan proses *data mining* secara paralel pada departemen lain.

### 3. METODOLOGI PENELITIAN

Model referensi khusus metodologi CRISP-DM pada penelitian ini telah dikembangkan. Gambar 3.1 merupakan model referensinya secara umum.



**Gambar 3.1** Model Referensi Umum CRISP-DM

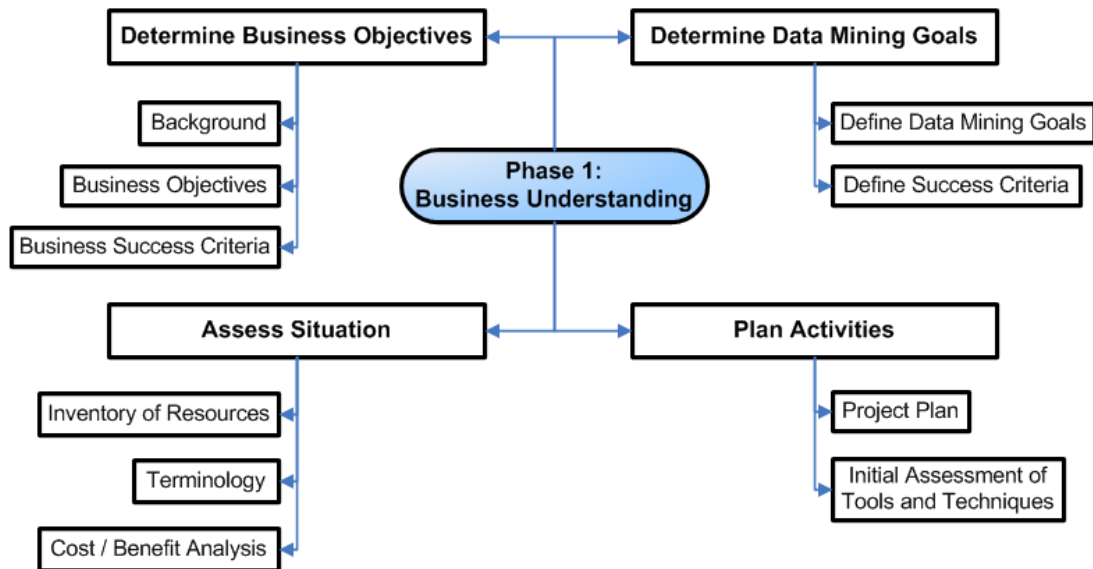
Berdasar pada model referensi yang telah ditampilkan pada gambar 3.1, berikut merupakan penjelasannya:

1. *Business Understanding*, adalah pemahaman tentang substansi dari kegiatan *data mining* yang akan dilakukan, kebutuhan dari perspektif bisnis. Kejadiannya antara lain menentukan sasaran atau tujuan bisnis, memahami situasi bisnis, menerjemahkan tujuan bisnis kedalam tujuan *data mining*.
2. *Data Understanding*, adalah pengumpulan data, mempelajari data untuk dapat memahami data yang akan digunakan dalam penelitian, mengidentifikasi masalah yang berkaitan dengan data.
3. *Data Preparation*, pada tahap ini struktur basis data akan dipersiapkan sehingga mempermudah proses *mining*.
4. *Modeling Phase*, adalah fase menentukan teknik *data mining* yang digunakan, menentukan *tools data mining*, algoritma *data mining*, menentukan parameter dengan nilai yang optimal.
5. *Evaluation Phase*, adalah fase interpretasi terhadap hasil *data mining* yang ditunjukkan dalam proses pemodelan yang terdapat pada fase sebelumnya. Evaluasi dilakukan secara mendalam dengan tujuan menyesuaikan model yang didapat agar sesuai dengan sasaran yang ingin dicapai dalam fase pertama.

6. *Deployment Phase* atau fase penyebaran adalah fase penyusunan laporan atau

presentasi dari pengetahuan yang didapat dari evaluasi pada proses *data mining*.

### 3.1 Phase 1



**Gambar 3.2** Fase Pertama Model Referensi CRISP-DM

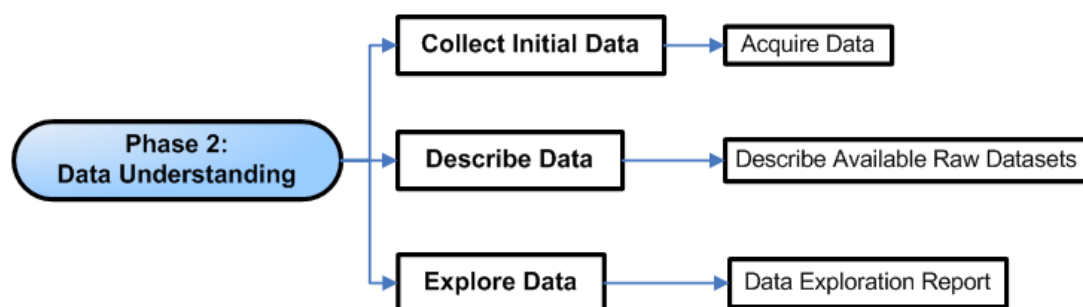
Pada fase ini terdapat empat proses yaitu:

1. *Determine business objectives* dilakukan untuk memahami tujuan yang ingin dicapai oleh PT X dari perspektif bisnis kepada perusahaan induknya secara menyeluruh.
2. *Asses situation* dilakukan untuk melibatkan fakta yang ada pada PT X tentang sumber daya, kendala, asumsi dan faktor-faktor lain yang harus dipertimbangkan dalam menentukan tujuan analisis data dan

rencana yang akan dilakukan dalam penelitian.

3. *Determine data mining goals* yaitu menyatakan tujuan *data mining* secara teknis terhadap penelitian yang dilakukan.
4. *Plan activities* yaitu menjelaskan rencana yang ditujukan untuk mencapai tujuan *data mining* dan dengan demikian dapat mencapai tujuan bisnis, kemudian menentukan *tools* yang akan dipergunakan.

### 3.2 Phase 2



**Gambar 3.3** Fase Kedua Model Referensi CRISP-DM

Pada fase ini terdapat tiga proses yaitu:

1. *Collect initial data* yaitu memperoleh data atau akses untuk membuka data yang akan digunakan dalam penelitian. Pengumpulan

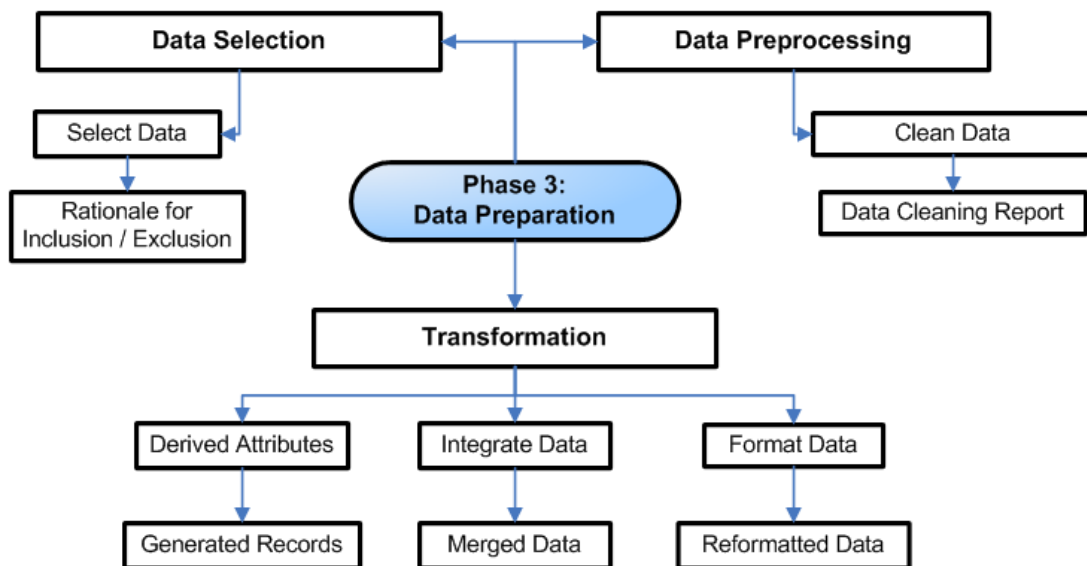
awal ini termasuk *loading data* jika diperlukan untuk memahami data tersebut.

2. *Describe data* yaitu memeriksa data yang diperoleh dan melaporkan hasilnya.

3. *Explore data* yaitu melakukan eksplorasi terhadap isi data yang diperoleh, kemudian

memberikan deskripsi mengenai isi data tersebut.

### 3.3 Phase 3



**Gambar 3.4** Fase Ketiga Model Referensi CRISP-DM

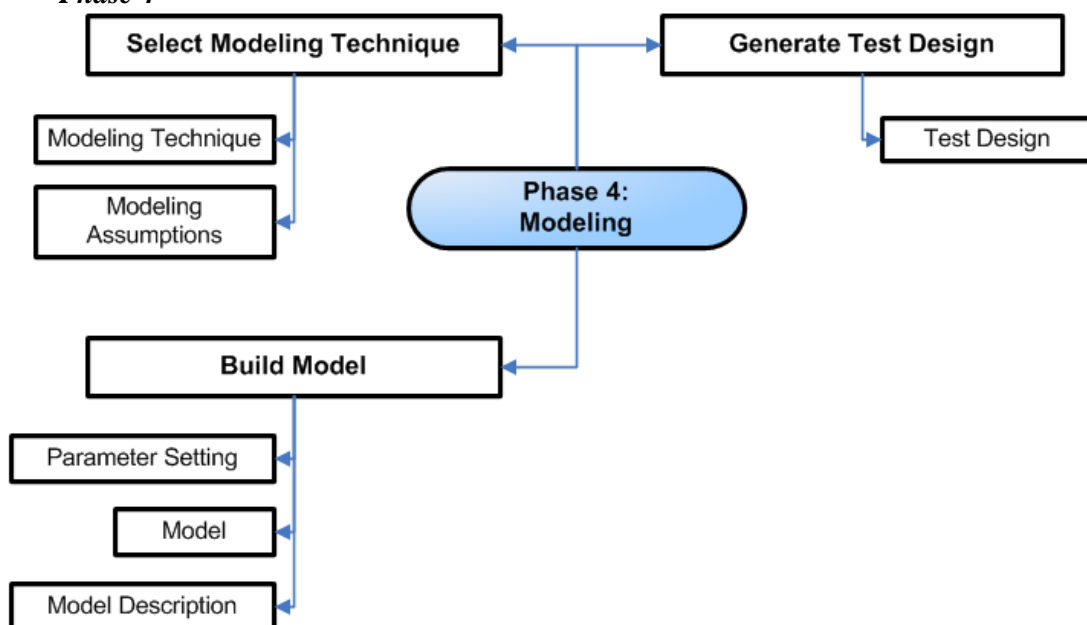
Pada fase ini terdapat tiga proses yaitu:

1. *Data selection* yaitu memilih data yang akan digunakan dalam proses *data mining*. Dalam proses ini dilakukan juga pemilihan atribut-atribut yang disesuaikan dengan proses *data mining*.
2. *Data preprocessing* yaitu memastikan kualitas data yang telah dipilih pada tahap *data selection*, pada tahap ini masalah yang harus dihadapi adalah apabila terdapat

*noisy data* dan *missing values*. Proses pembersihan data atau *cleansing* dilakukan untuk menemukan anomali data yang mungkin saja masih terdapat pada data tersebut.

3. *Transformation* yaitu pengelompokkan atribut-atribut atau *field* yang telah terpilih kedalam basis data baru untuk bahan *data mining*.

### 3.4 Phase 4



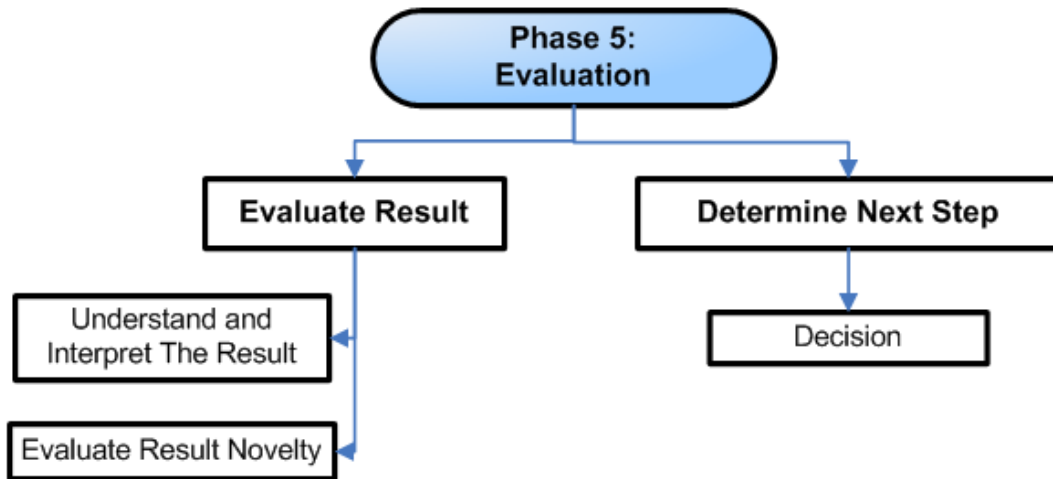
**Gambar 3.5** Fase Keempat Model Referensi CRISP-DM



- Pada fase ini terdapat tiga proses yaitu:
1. *Select modeling technique* merupakan langkah pertama dalam pemodelan, yaitu dengan menggunakan teknik pemodelan yang telah ditetapkan pada fase *business understanding*.

2. *Build model* yaitu menjalankan alat pemodelan sesuai dengan prosedur.
3. *Generate test design* yaitu menguji kualitas model dan validitas dengan menggunakan dataset yang telah dipersiapkan.

### 3.5 Phase 5

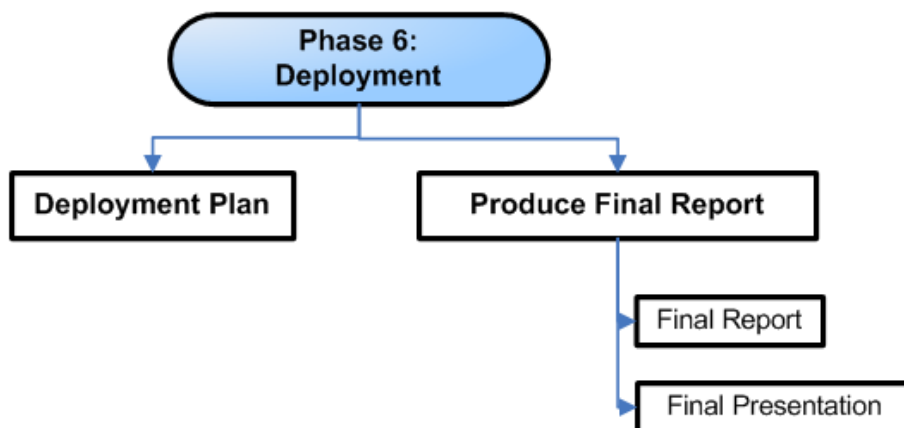


**Gambar 3.6** Fase Kelima Model Referensi CRISP-DM pada Penelitian

- Pada fase ini terdapat dua proses yaitu:
1. *Evaluate result* yaitu merangkum hasil penilaian dari segi kriteria keberhasilan bisnis, termasuk pernyataan akhir mengenai apakah penelitian sudah memenuhi tujuan bisnis.

2. *Determine next step* yaitu memberikan keputusan apakah teknik pemodelan yang dipergunakan dapat dijadikan standar dalam menentukan tujuan penelitian

### 3.6 Phase 6



**Gambar 3.7** Fase Keenam Model Referensi CRISP-DM pada Penelitian

Pada fase ini terdapat dua proses yaitu:

1. *Deployment plan* yaitu menjelaskan gambaran mengenai rencana terhadap pembuatan laporan yang akan dibuat.

2. *Produce final report* yaitu memberikan visualisasi dari laporan yang telah dibuat berdasarkan pada *deployment plan*.

#### 4. HASIL DAN PEMBAHASAN

Data yang digunakan dalam proses *modeling* sebanyak 34 *file* yang terdiri dari 17 *file* data primer dan 17 *file* data sekunder. Kedua buah data *file* tersebut kemudian masing-masing digabung menjadi satu sesuai

dengan urutan nama *file* yang telah disesuaikan menurut isi dan penamaan *file*. Dari hasil proses *fuzzy lookup* yang dilakukan dalam tahap *modeling*, maka terdapat 17 *file* yang telah disimpan. Tabel berikut ini akan menampilkan ringkasan yang menerangkan seberapa banyak jumlah *customer* yang termasuk dalam *pipeline customer* baru yang telah dikelompokkan berdasarkan bulan dan disegmentasi berdasarkan wilayah pembagian kerja dari segmen *Commercial* pada PT X.

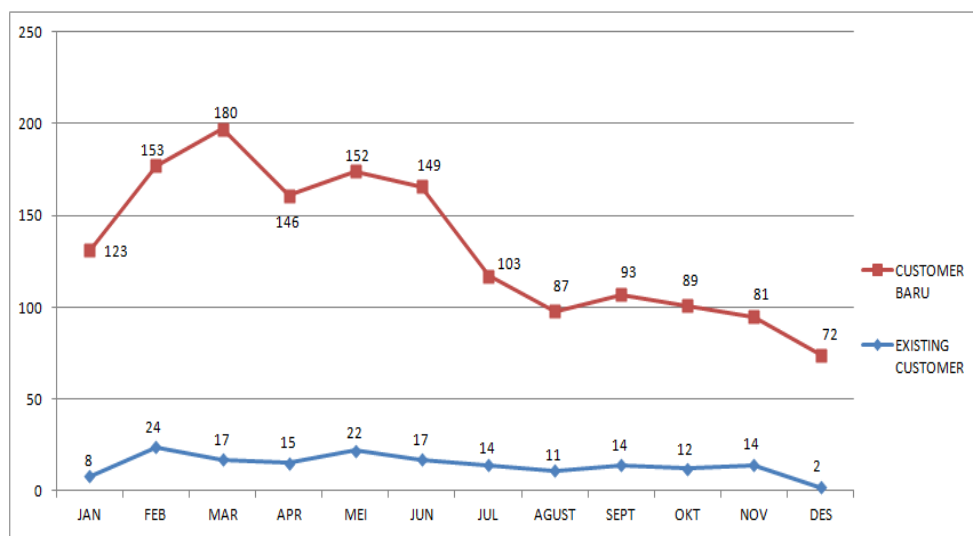
NO	WILAYAH	JUMLAH CUSTOMER												TOTAL
		JAN	FEB	MAR	APR	MEI	JUN	JUL	AGT	SEPT	OKT	NOV	DES	
1	MEDAN	17	26	16	23	19	15	10	17	18	9	15	12	197
2	PALEMBANG	19	11	13	4	7	18	4	1	7	7	2	1	94
3	JAKARTA 1	11	17	22	14	27	20	16	11	10	10	6	2	166
4	JAKARTA 2	10	9	6	12	17	14	9	16	9	19	9	8	138
5	JAKARTA 3	7	10	13	11	6	8	6	6	3	2	5	8	85
6	BANDUNG	7	14	16	17	11	14	3	7	8	6	8	5	116
7	BANJARMASIN	26	16	23	7	13	15	8	2	11	6	7	3	137
8	SEMARANG	10	14	14	12	17	6	8	2	2	11	3	5	104
9	SURABAYA	10	23	34	21	18	24	20	20	11	9	15	16	221
10	DENPASAR	1	4	8	10	5	3	12	5	5	5	5	4	67
11	MAKASSAR	5	9	15	15	12	12	7	7	9	5	6	8	110
GRAND TOTAL		123	153	180	146	152	149	103	94	93	89	81	72	1435

**Tabel 4.1** Jumlah *Pipeline Customer* Baru

Berdasarkan pada tabel 4.1 yang telah ditampilkan dapat dilihat bahwa jumlah *pipeline customer* baru secara total untuk wilayah Medan berjumlah 197 *customer*, wilayah Palembang berjumlah 94 *customer*, wilayah Jakarta 1 berjumlah 166 *customer*, wilayah Jakarta 2 berjumlah 138 *customer*, wilayah Jakarta 3 berjumlah 85 *customer*, wilayah Bandung berjumlah 116 *customer*, wilayah Banjarmasin berjumlah 137 *customer*, wilayah Semarang berjumlah 104 *customer*,

wilayah Surabaya berjumlah 221 *customer*, wilayah Denpasar berjumlah 67 *customer*, dan wilayah Makassar berjumlah 110 *customer*. Sehingga jumlah *pipeline customer* baru secara keseluruhan berjumlah 1.435 *customer*.

Gambar berikut ini merupakan grafik yang akan menampilkan jumlah *customer* dari PT Bank X yang belum dan sudah menjadi *customer* PT X sebagai anak perusahaan PT Bank X.



**Gambar 4.1** Grafik Perbandingan Jumlah *Customer* Baru dan *Existing*

Berdasarkan pada grafik yang telah ditampilkan, dapat dilihat bahwa pada bulan Januari terdapat 8 jumlah *customer* yang sudah menjadi *customer* PT X dan sebanyak 123 *customer* yang belum menjadi *customer* PT X dari total 131 *customer*. Pada bulan Februari terdapat 24 jumlah *customer* yang sudah menjadi *customer* PT X dan sebanyak 153 *customer* yang belum menjadi *customer* PT X dari total 177 *customer*. Pada bulan Maret terdapat 17 jumlah *customer* yang sudah menjadi *customer* PT X dan sebanyak 180 *customer* yang belum menjadi *customer* PT X dari total 197 *customer*. Pada bulan April terdapat 15 jumlah *customer* yang sudah menjadi *customer* PT X dan sebanyak 146 *customer* yang belum menjadi *customer* PT X dari total 161 *customer*. Pada bulan Mei terdapat 22 jumlah *customer* yang sudah menjadi *customer* PT X dan sebanyak 152 *customer* yang belum menjadi *customer* PT X dari total 174 *customer*. Pada bulan Juni terdapat 17 jumlah *customer* yang sudah menjadi *customer* PT X dan sebanyak 149 *customer* yang belum menjadi *customer* PT X dari total 166 *customer*. Pada bulan Juli terdapat 14 jumlah *customer* yang sudah menjadi *customer* PT X dan sebanyak 103 *customer* yang belum menjadi *customer* PT X dari total 117 *customer*. Pada bulan Agustus terdapat 11 jumlah *customer* yang sudah menjadi *customer* PT X dan sebanyak 87 *customer* yang belum menjadi *customer* PT X dari total 98 *customer*. Pada bulan September terdapat 14 jumlah *customer* yang sudah menjadi *customer* PT X dan sebanyak 93

*customer* yang belum menjadi *customer* PT X dari total 107 *customer*. Pada bulan Oktober terdapat 12 jumlah *customer* yang sudah menjadi *customer* PT X dan sebanyak 89 *customer* yang belum menjadi *customer* PT X dari total 101 *customer*. Pada bulan November terdapat 14 jumlah *customer* yang sudah menjadi *customer* PT X dan sebanyak 81 *customer* yang belum menjadi *customer* PT X dari total 95 *customer*. Pada bulan Desember terdapat 2 jumlah *customer* yang sudah menjadi *customer* PT X dan sebanyak 72 *customer* yang belum menjadi *customer* PT X dari total 74 *customer*.

## 5. KESIMPULAN

Panduan langkah-langkah CRISP-DM tidak secara eksplisit menyebutkan dataset sebagai kiriman untuk tugas persiapan data pada setiap fase, tetapi dataset tersebut sebaiknya diarsipkan dan didokumentasikan. Dataset tidak akan sesuai tepat satu dengan tugas yang ada pada setiap fase, tetapi informasi tentang data yang digunakan harus dimasukkan dalam setiap laporan yang harus disampaikan.

*Data mining* dapat menciptakan lingkungan bisnis yang intelligence, yaitu efisiensi dalam mendapatkan informasi yang diperlukan, menyediakan tool khusus untuk kemampuan analitis yang jauh lebih baik, mengurangi beban berlebih pada sistem operasional untuk membuat laporan dan kegiatan analitik, serta memberikan kemampuan bagi pengguna dalam ruang lingkup bisnis proses.

## DAFTAR PUSTAKA

- Bramer, M. (2007). *Principles of Data Mining*. London: Springer-Verlag.
- Cefkin, M. (2007). *Ethnographic Praxis in Industry Conference. Numbers may speak louder than words, but is anyone listening? The rhythmscape and sales pipeline management*, 187–199.
- Gorunescu, F. (2011). *Data Mining - Concepts, Models, and Techniques*. Berlin: Springer-Verlag Berlin Heidelberg.
- Kantardzic, M. (2011). *Data Mining : Concepts, Models, Methods, and Algorithms*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Kimla, N. (2013). *Pipeline Management: Success with CRM*. Los Angeles: Pipelinersales Inc.
- Kotler, P., & Keller, K. (2012). *Marketing Management*. Upper Saddle River, New Jersey: Prentice Hall.
- Lamb, C. W., Hair, J. F., & McDaniel, C. (2010). *Marketing*. Boston, MA: South-Western College.

- Larose, D. T. (2014). *Discovering Knowledge in Data : An Introduction to Data Mining*. New Jersey: John Wiley & Sons, Inc.
- Pal, N. R., & Jain, L. C. (2005). *Advanced Techniques in Knowledge Discovery and Data Mining*. New York: Springer.
- Ponniah, P. (2011). *Data Warehousing Fundamentals for IT Professionals*. New Jersey: John Wiley & Sons, Inc.
- Shafique, U., & Qaiser, H. (2014). International Journal of Innovation and Scientific Research. *A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA)*, 12 (1), 217-222.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2005). *Introduction To Data Mining*. New York: Pearson.
- Turban, E., Aronson, J. E., & Liang, T.-P. (2005). *Decision Support Systems and Intelligent Systems*. New Jersey: Prentice-Hall, Inc.
- Yan, J., Zhang, C., Zha, H., Gong, M., Sun, C., Huang, J., et al. (2015). Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. *On Machine Learning towards Predictive Sales Pipeline Analytics*, 1945-1951.