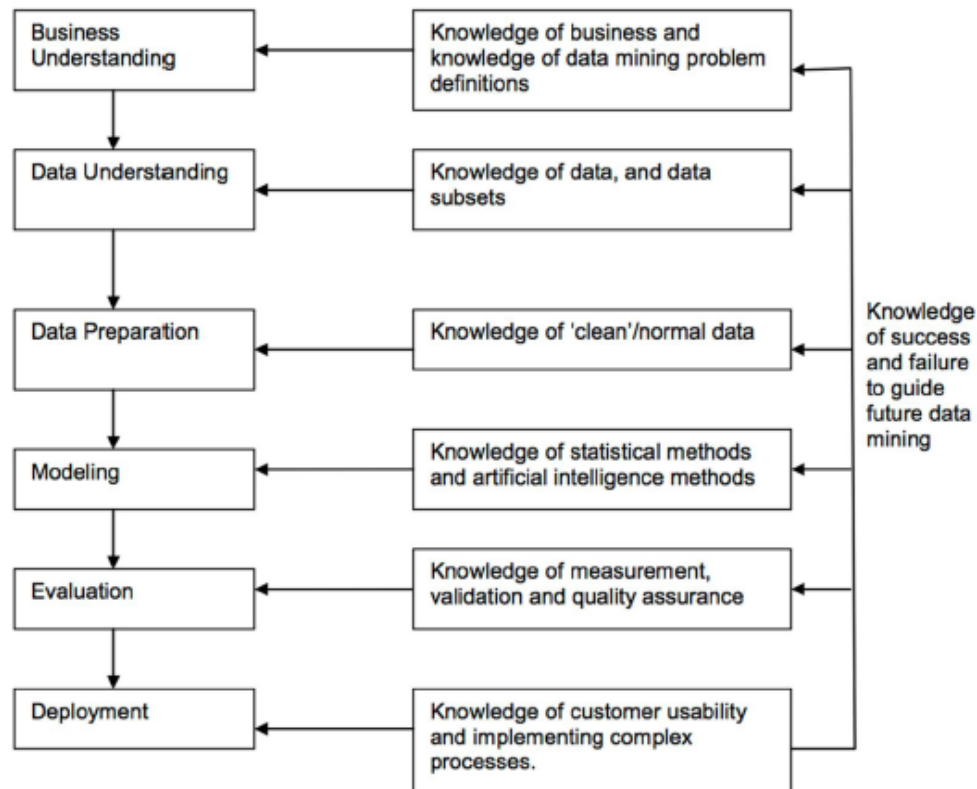


## Prediksi Kreditor Potensial menggunakan Model *Decision Tree*

Sejak tahun 1990an seluruh konsep perbankan telah bergeser ke *Centralized Databases*, transaksi online dan ATM di seluruh dunia. Pertumbuhan data transaksi semakin lama semakin cepat bertambah besar, namun sebagian besar bank hanya dapat manfaat yang sedikit sekali dari *database* yang sangat besar tersebut, padahal dengan teknologi *Data Mining* akan dapat ditemukan berbagai *knowledge* yang sangat berguna untuk kepentingan bisnis. Saat ini ada tren peningkatan minat dalam data mining, termasuk karena biaya penyimpanan data yang semakin murah dan kemudahan untuk mendapatkan teknologi pengumpulan data, pengembangan *Machine Learning Algorithms* yang kuat dan efisien untuk mengolah data. Pada industri perbankan, area paling banyak dapat menggunakan *Data Mining* adalah *marketing*. Tim marketing bank dapat menggunakan *Data Mining* untuk menganalisis database pelanggan dan mengembangkan statistik profil suara preferensi pelanggan secara individual untuk produk dan jasa.

Proses Knowledge Discovery adalah proses untuk menemukan *Knowledge* dengan menggunakan Data Mining, proses ini menggabungkan teknik dari berbagai bidang, termasuk *Database*, *Artificial Intelligence*, statistik dan visualisasi.(Scott, 2007). Proses ini juga biasa disebut sebagai proses *Knowledge Discovery and Data Mining* (KDDM) model yang terdiri dari berbagai tahapan proses interaktif. Proses KDDM model yang sudah dikenal luas sebagai standar industri adalah proses CRISP-DM model, singkatan dari *Cross Industry Standard Process (CRISP) for Data Mining (DM)*, Model ini terdiri dari enam tahap proses yaitu (Sharma, 2010):



(vinekar,2009)

Desain *framework* ini bersifat *generic* artinya bisa digunakan dengan multiple DMs, setiap DMs akan dibangun dengan menggunakan model CRISP DM, yang terdiri dari enam tahap. Metodologi yang digunakan untuk membangun suatu DMs akan menggunakan CRISP DM, metodologi ini terdiri dari enam langkah sebagai berikut:

### ***Step-1. Business Understanding***

Bank menargetkan laba sebelum pajak (EBITDA) perseroan naik dari Rp 77 miliar di 2012 menjadi Rp 87 miliar di 2013. Keyakinan itu dikontribusikan dari peningkatan fasilitas kredit. Tim *Sales* dan *Marketing* akan menargetkan spesifik nasabah untuk melakukan Telemarketing. Mula-mula harus ditemukan pola atribut nasabah yang telah diketahui menggunakan fasilitas kredit dengan angsuran yang lancar. Kemudian dengan pola tersebut, Bank berharap akan menemukan nasabah lain yang potensial, yang kemungkinan besar akan menggunakan fasilitas kredit.

### ***Step-2. Data Preparation***

Data akan diambil dari aplikasi Core-banking yaitu Alphabits. Data yang diperlukan adalah tabel master individual nasabah yaitu M4CUI dan tabel master Loan yaitu LLOAN. Walaupun data yang dibutuhkan tidak semua *field*, namun pada proses EOD (End of Day) telah disiapkan semua tabel dengan semua *field*. Struktur tabel untuk *field* yang dibutuhkan saja adalah sebagai berikut:

Struktur Tabel MACUI

NO.	FIELD	LENGTH	DESCRIPTION
1	CUCODE	8A	Customer code
2	CUTITL	10A	Titel
3	CUSHOR	10A	Short name
4	CUJEKL	1A	Jenis kelamin
5	CUAGAM	1A	Agama
6	CUDTLH	5P	Birth date
7	CUMRST	1A	Marital Status
8	CUNCHL	2N	Numbers Child
9	CUEDUC	2A	Education Code
10	CUPSTN	25A	Position Job
11	CUINCM	2A	Income per Month
12	CUZIPP	7A	Zip Code
13	CUPPA1	4A	Kode Area
14	CUPPN1	9A	Phone #1
15	CUEXT1	5A	Ext #1
16	CUPPA2	4A	Kode Area
17	CUPPN2	9A	Phone #2
18	CUEXT2	5A	Ext #2
19	CUPFXA	4A	Fax Area
20	CUPFXN	9A	Facsimile

Struktur Tabel LLOAN

NO.	FIELD	LENGTH	DESCRIPTION
1	L0STAT	1A	Status record
2	L0STAD	1A	Status Data
3	L0CSNO	8A	Customer code
4	L0FCTY	3A	Facility Type
5	L0FCSQ	2N	Fac. Seq. No
6	L0LNTY	5A	Loan Type
7	L0CYCD	3A	Loan Ccy Code
8	L0ECON	4A	Economical Sector Code
9	L0TYUS	2A	Type of Use
10	L0COLS	1A	Collectibility System
11	L0STDT	5P	Start Date
12	L0MTDT	5P	End Date

### Step-3. Data Understanding

Data akan diekstrak dari DB2/400 menjadi text file lalu di download di PC dan import ke *SQL Server* dan untuk sementara disimpan pada *Database Staging Area*. Di *Staging Area* ini data ditingkatkan kualitasnya dengan proses:

- *Data Cleansing*
- *Data Enrichment*
- *Data Standarization*

Untuk keperluan proses Data Mining akan dibuat beberapa View sebagai berikut:

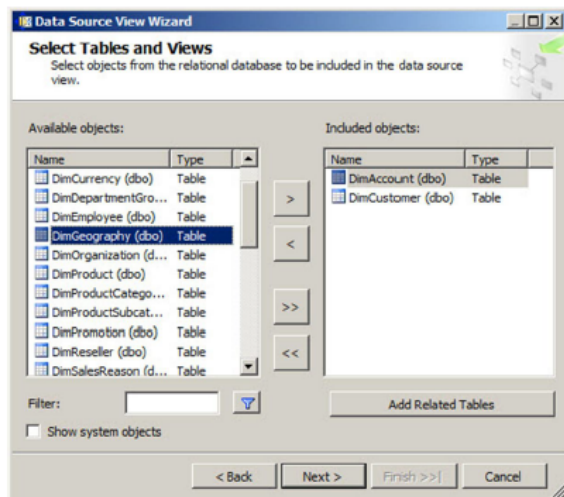
- TargetCall – sebagai *Training* dan *Testing* data set, data ini diambil dari tabel nasabah tahun lalu dengan flag yang pernah mengambil kredit atau yang belum pernah.
- ProspectiveCreditor – sebagai *Prediction* data set, data ini diambil dari tabel nasabah tahun berjalan yang belum pernah menggunakan fasilitas kredit yang dipromosikan.

### Step-4. Modeling

#### a. Preparing the Analysis Services Database

Untuk membangun aplikasi BI pada SQL Server dapat menggunakan *BI Development Studio* dengan *project template* pada *SQL Server Analysis Services*. Setelah membuat Analysis Services Project maka perlu ditentukan data source nya baik berupa tabel atau view yaitu:

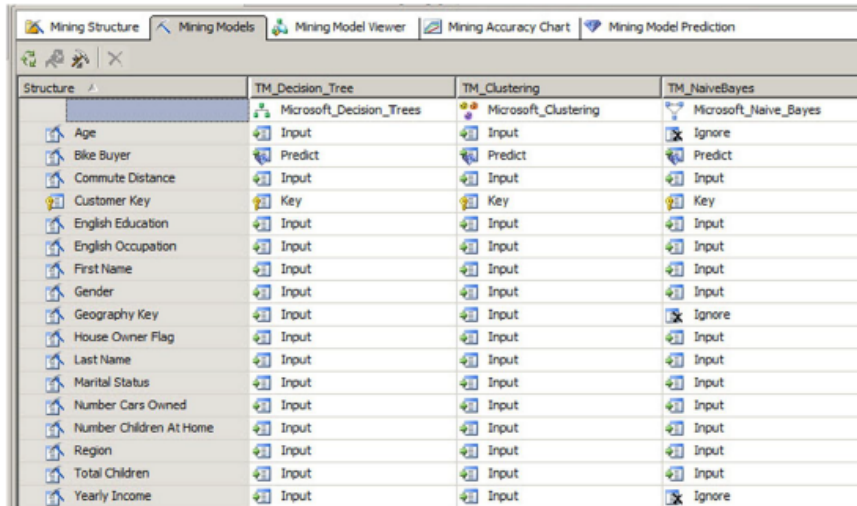
- vTargetCall
- vProspectiveCreditor



#### b. Membangun Targeted Marketing Scenario

Pada langkah ini akan dibuat Mining Model Structure baru, kemudian ditentukan Data Mining Algorithm yang akan dipakai yaitu Microsoft *Decision Trees*, *Training* Data yang akan dipakai yaitu vTargetCall lalu ditentukan juga data demografi nasabah yang akan digunakan sebagai atribut yang akan diproses *Data Mining* yaitu:

- Kode Nasabah
- Jenis Kelamin
- Status Pernikahan
- Jumlah anak
- Pendidikan Terakhir
- Pendapatan
- Kode Pos



Structure	TM_Decision_Tree	TM_Clustering	TM_NaiveBayes
	Microsoft_Decision_Trees	Microsoft_Clustering	Microsoft_Naive_Bayes
Age	Input	Input	Ignore
Bike Buyer	Predict	Predict	Predict
Commute Distance	Input	Input	Input
Customer Key	Key	Key	Key
English Education	Input	Input	Input
English Occupation	Input	Input	Input
First Name	Input	Input	Input
Gender	Input	Input	Input
Geography Key	Input	Input	Ignore
House Owner Flag	Input	Input	Input
Last Name	Input	Input	Input
Marital Status	Input	Input	Input
Number Cars Owned	Input	Input	Input
Number Children At Home	Input	Input	Input
Region	Input	Input	Input
Total Children	Input	Input	Input
Yearly Income	Input	Input	Ignore

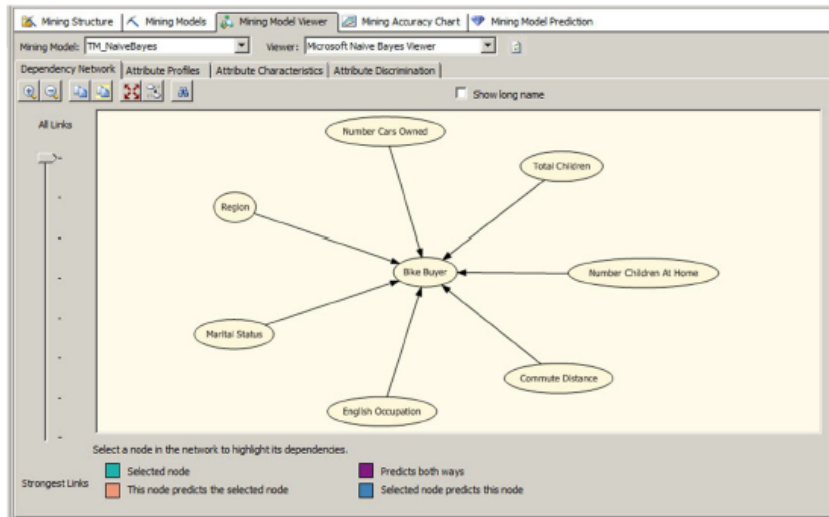
### c. Evaluation

Setelah model diproses, maka model dapat diperiksa dengan menggunakan tab Mining Model Viewer dalam *Data Mining Designer*. Masing-masing Mining model dalam mining structure dapat dipilih untuk di evaluasi.

#### Microsoft Naive Bayes Model

Viewer untuk model ini, berisi empat tab yaitu:

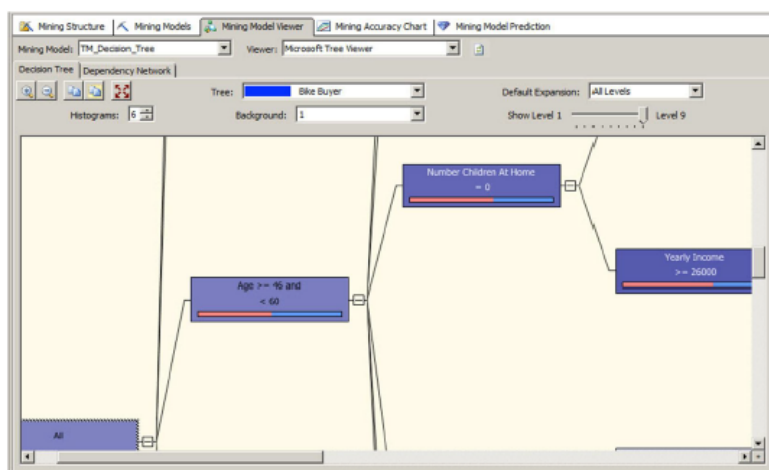
- *Dependency Network* - Setiap node dalam viewer merepresentasikan suatu atribut, dan garis antara node merepresentasikan hubungan antara node tersebut. Dalam viewer, semua atribut yang mempengaruhi keadaan atribut yang sedang diprediksi yaitu Creditor dapat dilihat.
- *Attribute Profiles* - Tab ini menggambarkan bagaimana State yang berbeda dari atribut input yang mempengaruhi hasil dari atribut yang sedang diprediksi.
- *Attribute Characteristics* - Dengan tab ini dapat dipilih atribut dan nilai untuk melihat seberapa sering nilai untuk atribut yang lain muncul dalam kasus nilai yang dipilih.
- *Attribute Discrimination* - Dengan tab ini dapat diselidiki hubungan antara dua nilai diskrit dari atribut pilihan yang diprediksi dengan nilai atribut lainnya. Karena model TM\_NaiveBayes hanya memiliki dua keadaan yaitu 0 dan 1.



### Microsoft Decision Tree Model

Viewer untuk model ini, berisi dua tab yaitu:

- *Tab Decision Tree* - pada tab ini, semua tree models yang membentuk sebuah mining model dapat diperiksa. Karena pada model Targeted Marketing ini yang ditargetkan hanya berisi atribut tunggal yaitu Creditor, maka hanya akan ada satu tree yang dapat dilihat.
- *Tab Dependency Network* - tab ini menampilkan hubungan antara atribut yang berkontribusi terhadap kemampuan prediksi mining model tersebut. Pusat node pada Dependency Network yaitu Creditor, merepresentasikan atribut yang diprediksi dalam mining model tersebut. Setiap node di sekitarnya merupakan atribut yang mempengaruhi hasil dari atribut yang diprediksi.



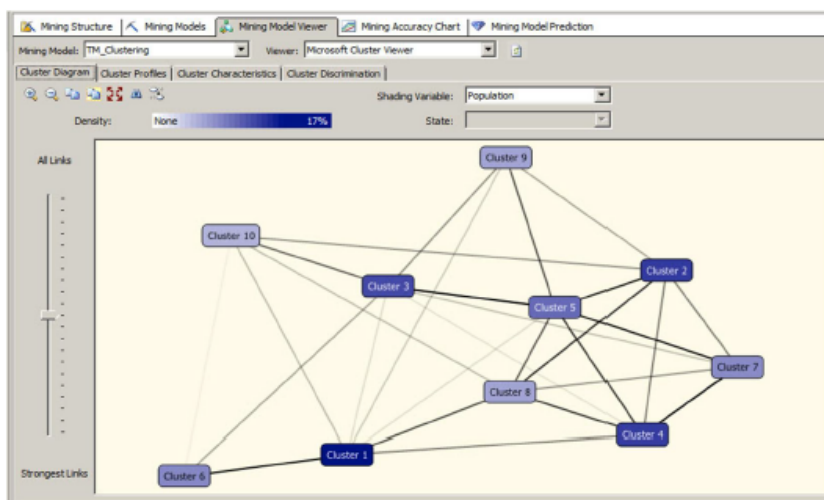
### Microsoft Clustering Model

Viewer untuk model ini memiliki empat tab yaitu:

- *Tab Cluster Diagram* – Tab ini dapat digunakan untuk mengeksplorasi hubungan antar cluster yang ditemukan algoritma. Garis antar cluster merupakan "kedekatan"

dan di-shaded berdasarkan pada seberapa mirip cluster tersebut. Warna dari setiap cluster mewakili frekuensi variabel dan state di cluster tersebut.

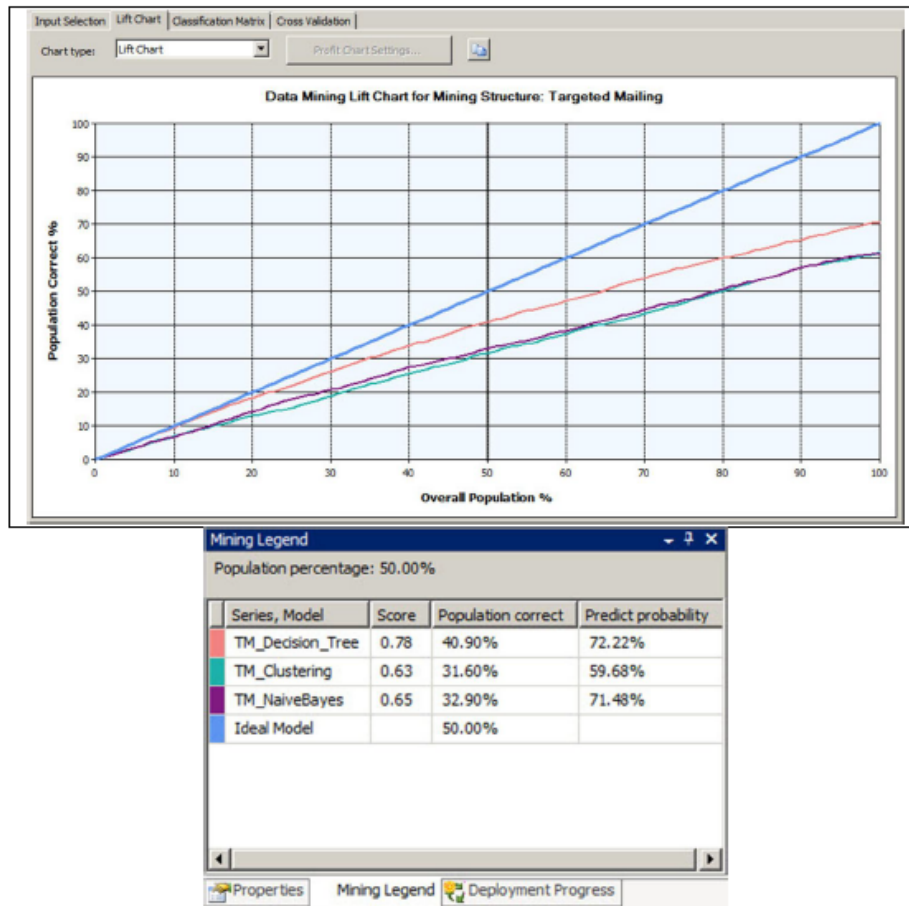
- *Tab Cluster Profiles* - Tab ini memberikan gambaran menyeluruh dari model TM\_Clustering. Tab Cluster Profiles berisi kolom untuk setiap cluster dalam model. Kolom pertama berisi daftar atribut yang berkaitan dengan setidaknya satu cluster. Sisa tampilan berisi distribusi atribut state untuk setiap cluster. Distribusi dari variabel diskrit ditampilkan sebagai bar berwarna dengan jumlah maksimum bar ditampilkan dalam daftar bar Histogram. Continuous atribut ditampilkan dengan diamond chart, yang merupakan rata-rata dan standard deviation di setiap cluster.
- *Tab Cluster Characteristics* - Dengan tab ini dapat digunakan untuk memeriksa secara lebih rinci karakteristik yang membentuk cluster. Misalnya, jika menggunakan daftar cluster untuk menampilkan Creditor yang bernilai tinggi dalam skenario Targeted Marketing, Akan dapat dilihat karakteristik orang di cluster ini misalnya mereka biasanya pernah menggunakan fasilitas credit di masa lalu, sudah menikah, pendidikan terakhir S1 dan berpenghasilan tetap.
- *Tab Cluster Discrimination* - Dengan tab ini dapat digunakan untuk mengeksplorasi karakteristik yang membedakan satu cluster dari cluster yang lain. Setelah Anda memilih dua clusters, satu dari daftar cluster-1, dan satu dari daftar Cluster-2, maka Viewer akan menghitung perbedaan antara cluster dan menampilkan daftar atribut yang paling membedakan kedua cluster tersebut.



### Testing the Accuracy of the Mining Models

Pada tab Mining Accuracy Chart pada Data Mining Designer, dapat dihitung seberapa baik masing-masing model dalam membuat prediksi, dan masing-masing model secara langsung dapat dibandingkan hasilnya terhadap hasil model-model lain. Metode perbandingan dikenal sebagai Lift chart. Tab Mining Accuracy membandingkan prediksi terhadap hasil diketahui. Hasil perbandingan ini kemudian diurutkan dan diplot pada grafik. Model ideal yaitu model teoritis yang memprediksi hasil dengan 100 persen benar, juga akan diplot juga pada grafik.

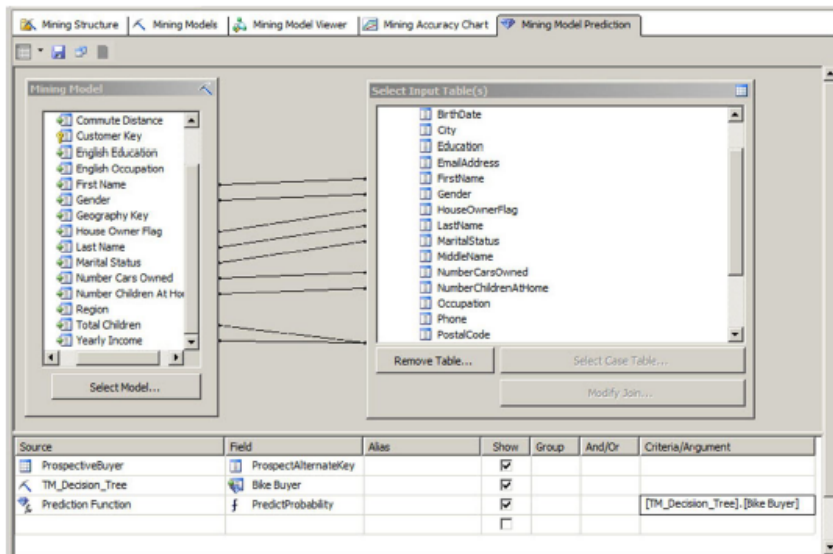




### Step-5. Deployment

- *Creating the Query*

Setelah melakukan uji akurasi pada mining model, langkah berikutnya dapat dibuat prediksi dengan membuat Data Mining Extensions (DMX) query prediksi dengan menggunakan *Prediction Query Builder* yang tersedia pada tab Mining Model Prediction di *Data Mining Designer*. Pada langkah ini akan dipilih Mining Model TM\_Decision\_Tree dan Input Table yang akan diprediksi adalah viewvProspectiveCreditor.



- Viewing the Results

Pada panel query result dapat ditampilkan kolom kolom ProspectAlternateKey sebagai identitas nasabah, Creditor sebagai indikator bahwa nasabah yang bersangkutan adalah Creditor yang potensial, dan Expression yang menunjukkan probabilitas dari prediksi yang akan benar. Perusahaan dapat menggunakan hasil ini untuk menentukan mana nasabah yang potensial yang harus dikelompokkan sebagai Targeted Marketing. Hasil Query Data Mining ini dapat disimpan sebagai file.

ProspectAlternateKey	Expression
34765109400	0.507257142857143
77789807300	0.507257142857143
22966404557	0.507257142857143
23045073700	0.507257142857143
23111850000	0.507257142857143
3958	0.507257142857143
3962	0.507257142857143
3964	0.507257142857143
81879768371	0.507257142857143
26233962774	0.507257142857143
81095297336	0.507257142857143
26223351955	0.507257142857143
81782693862	0.507257142857143
6980	0.507257142857143
6990	0.507257142857143
6992	0.507257142857143
6995	0.507257142857143
6973	0.507257142857143
8269	0.507257142857143
8273	0.507257142857143
8271	0.507257142857143
1978	0.507257142857143
1987	0.507257142857143



## ***REFERENSI***

Scott, R. S. (2007). Experiences of using Data Mining in a Banking Application. Department of Computation, UMIST, UK. ,ESPRIT HPCN project no. 22693.

Sharma, S. &. (2010). Toward an integrated knowledge discovery and data mining process model. Department of Information Systems, the Information Systems Research Institute, Virginia Commonwealth University, USA, The Knowledge Engineering Review, Vol. 25:1, 49–67.

Vinekar, V. T. (2009). The Interaction of Business Intelligence and Knowledge Management in Organizational Decision-Making. Journal of International Technology and Information Management, Volume 18, Number 2.