

PENERAPAN METODA RESIKO BAYESIAN DALAM PEMBUATAN APLIKASI ANALISIS PROYEK DATA MINING

Oleh : Ayuningtyas*

ABSTRACT

Although at present the data mining has been used widely, but there are not many researches done recently that give the significant contribution to the people that will use the algorithm. Most of the researchers merely yield the decision of whether or not it is the time of the data mining to use in the company, how the financial condition is, the effect that will be given to the firm if the decision is taken using data mining, and how the data quality condition to the owned account.

The Bayesian method is commonly used to analyze it because this method can do some improvements toward the old data owned by the firm and the produced new data will give the simple probability value that can be used to take the decision.

This thesis will try to test a mechanism to do the evaluation if a firm is indeed reasonable to do the data mining activity by making an application. The produced scores from Bayesian method will be used to measure the risk of whether or the firm should do the data mining activity.

Keywords : Data mining, Bayesian risk method.

LATAR BELAKANG

Data Mining telah banyak dipergunakan pada berbagai bidang, mulai melacak tindak kejahatan sampai dengan pembangunan komunitas pengetahuan sampai dengan transaksi bisnis. Banyak algoritma tentang penerapan *Data Mining* yang dibahas dalam berbagai penelitian, tapi tidak banyak yang menghitung apakah sudah layak suatu perusahaan untuk melakukan aktifitas *Data Mining*.

Manfaat dari nilai yang dihasilkan oleh *Data Mining* dapat terganggu oleh beberapa faktor yaitu (1) kualitas data yang rendah, termasuk kontaminasi data oleh error atau tidak lengkapnya data dikarenakan terbatasnya bandwidth untuk melakukan akuisisi data dan (2) tidak tersedianya model data untuk menangkap relasi probabilitas yang rumit pada data. (Nie, G., Zhang, L., Liu, Y., Zheng, X., Shi, Y, 2008). Kualitas data yang baik memiliki empat dimensi, yaitu (1) Keakuratan data (*accuracy*) , (2) Ketepatan waktu (*Timeliness*) , (3) Kelengkapan data (*Completeness*) dan (4) Konsistensi data (*Consistency*).

Meskipun saat ini *Data Mining* telah digunakan secara luas, tapi tidak banyak penelitian yang dikerjakan saat ini memberikan masukan yang berarti pada orang-orang

* Dosen STIKOM

yang akan menggunakan algoritma tersebut. Kebanyakan para-peneliti hanya menghasilkan keputusan apakah sudah saatnya *Data Mining* tersebut digunakan atau tidak pada perusahaan tersebut. Bagaimana keadaan finansialnya, efek yang akan diberikan kepada perusahaan apabila diambil keputusan menggunakan *Data Mining*, bagaimana keadaan kualitas data terhadap *account* yang dimiliki tidak banyak dijelaskan. (Nie, G., Zhang, L., Liu, Y., Zheng, X., Shi, Y, 2008).

RUMUSAN MASALAH

Pengambilan keputusan yang baik tentu harus didasarkan atas data yang valid. Untuk dapat menghasilkan data yang valid tentu harus menggunakan metode yang tepat untuk mengukurnya. Permasalahan umum yang muncul dan akan diselesaikan adalah :

1. Pemilihan metode yang tepat yang akan digunakan untuk mengukur parameter yang telah ditentukan, karena banyak metode yang dapat digunakan untuk melakukan pengukuran, salah satunya adalah metode resiko Bayesian
2. Setelah pengukuran selesai dilakukan, tentu harus dilakukan implementasi dan pengujian terhadap hasil pengukuran.

Berdasarkan permasalahan umum yang telah dijelaskan pada bagian sebelumnya, maka penulis dapat menetapkan suatu rumusan masalah, yaitu :

1. Mengapa model analisis Bayesian yang digunakan untuk menentukan layak atau tidak suatu perusahaan melakukan aktifitas data mining
2. Bagaimana implementasi aplikasi analisis proyek data mining berbasis metoda resiko bayesian
3. Bagaimana menguji dan mengevaluasi implementasi aplikasi analisis proyek data mining berbasis metoda resiko Bayesian.

TUJUAN

1. Mengetahui model analisis Bayesian yang digunakan untuk menentukan layak atau tidak suatu perusahaan melakukan aktifitas data mining
2. Mengetahui implementasi aplikasi analisis proyek data mining berbasis metoda resiko bayesian
3. Untuk menguji dan mengevaluasi implementasi aplikasi analisis proyek data mining berbasis metoda resiko Bayesian.

MANAJEMEN PROYEK PENGEMBANGAN DATA MINING

Untuk melakukan *Data Mining*, terdapat beberapa hal yang harus menjadi perhatian sebelum benar-benar menerapkan *Data Mining* pada suatu perusahaan. Hal-hal yang perlu diperhatikan tersebut adalah kualitas dari suatu data yang akan dilihat, dimensi yang berhubungan dengan sumber daya manusia dan financial, dan yang paling penting adalah dukungan dari para executive perusahaan untuk melakukan *Data Mining* tersebut.

Berikut adalah penjelasan mengenai hal-hal yang perlu diperhatikan sebelum memutuskan untuk melakukan *Data Mining*.

1. Faktor-faktor yang Mempengaruhi Data Mining

Manfaat dari nilai yang dihasilkan oleh *Data Mining* dapat terganggu oleh beberapa faktor yaitu (1) kualitas data yang rendah, termasuk kontaminasi data oleh error atau tidak lengkapnya data dikarenakan terbatasnya bandwidth untuk melakukan akuisisi data dan (2) tidak tersedianya model data untuk menangkap relasi probabilitas yang rumit pada data.

Biaya yang dikeluarkan untuk *Data Mining* termasuk untuk pembayaran proyek dan biaya yang harus dibayarkan karena kesalahan pandangan dari model *Data Mining* yang digunakan.

Jadi terdapat beberapa hal dasar yang harus dilakukan pada *Data Mining* yaitu yang berhubungan trade-off antara efisiensi dan efisiensi biaya dari *distributed data mining application* serta akurasi dan *reliability* dari hasil yang diprediksikan.

2. Data Quality

Data yang mempunyai kualitas yang baik tentu akan memberikan hasil analisis yang baik pula. Ada banyak faktor yang mempengaruhi kualitas dari suatu data. Menurut Wang, R., Strong, D., & Guarascio, L. pada papernya yang berjudul *Beyond accuracy: What data quality means to data consumers. Total Data Quality Management Research Program*, menyebutkan bahwa terdapat beberapa dimensi yang dapat digunakan untuk mengukur kualitas suatu data yang dikelompokkan menjadi 20 dimensi. Dari kedua puluh dimensi tersebut, hanya ada empat kategori yang disimpulkan dari penelitian yang dikerjakan oleh Wang, R., Strong, D., & Guarascio, L. yaitu : (1) *Intrinsic Data Quality* yang terdiri dari *accuracy, objectivity, believability, dan reputation*; (2) *Contextual Data Quality* yang terdiri dari *value-added, relevancy, timeliness, completeness, dan appropriate amount of data*; (3) *Representational Data Quality* yang terdiri dari *interpretability, ease of understanding, representational consistency, and concise representation*; serta (4) *Accessibility Data Quality* yang terdiri dari *accessibility and access security*, seperti yang terlihat pada gambar

1. Meskipun dimensi-dimensi tersebut ditujukan untuk Information System, tapi dimensi-dimensi tersebut dapat digunakan untuk melakukan evaluasi terhadap kualitas suatu data pada *Data Mining*. Pada analisis pengambilan keputusan suatu *Data Mining*, ditambahkan atribut *data size*. Dimensi-dimensi tersebut adalah sebagai berikut :

a) Keakuratan data (*Accuracy*)

Data atau nilai yang tersimpan berisi dengan nilai yang sesungguhnya. Dimensi akurasi adalah dimensi yang paling jelas terlihat kebenaran suatu data yang digunakan.

b) Ketepatan waktu (*Timeliness*)

Data atau nilai yang tersimpan tidak ketinggalan jaman atau kadaluwarsa.

Setiap data yang tersimpan tentu akan menjadi tidak berguna apabila sudah terlalu lama tersimpan.

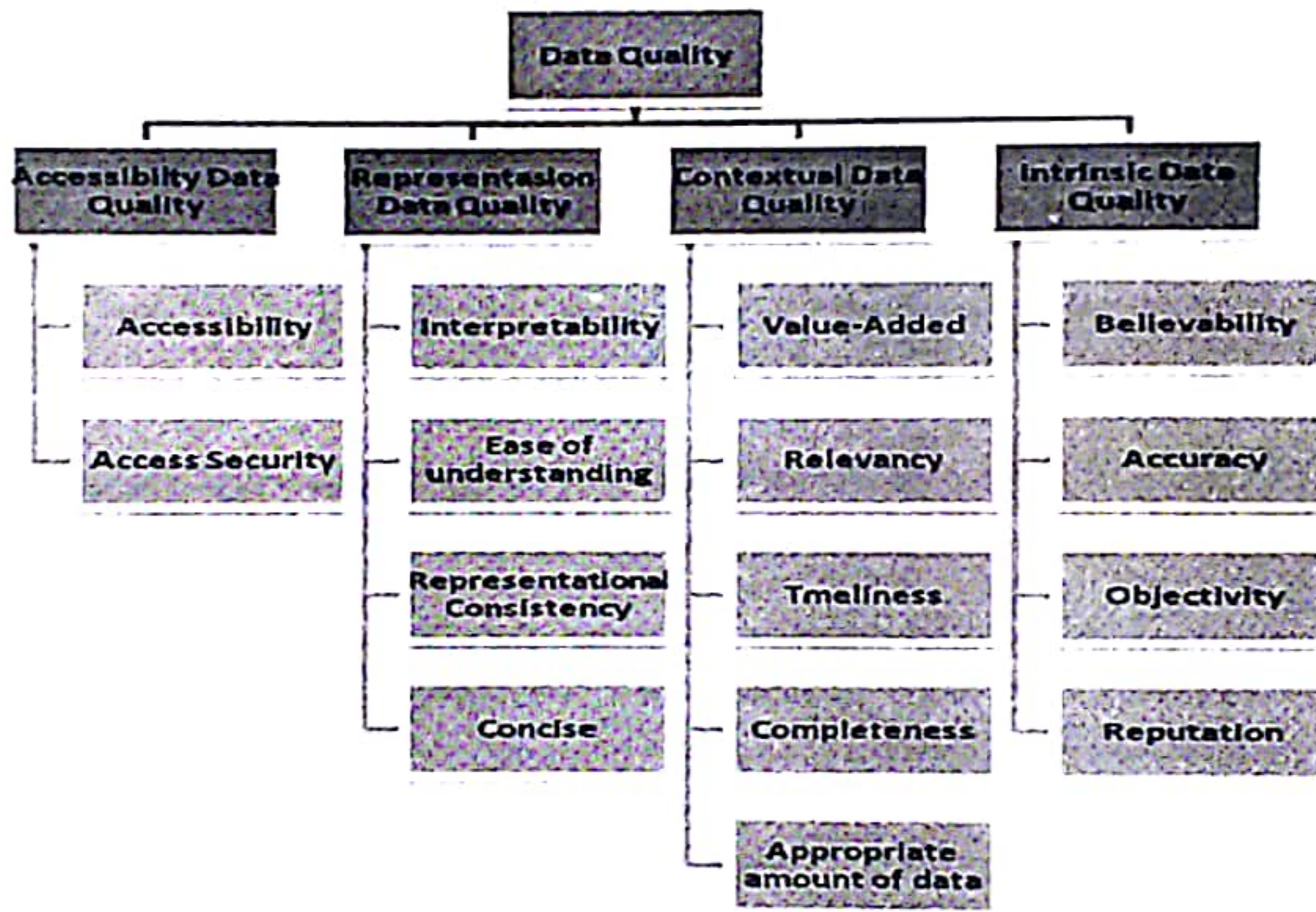
c) Kelengkapan data (*Completeness*)

Semua data atau nilai dari suatu variable khusus telah tersimpan. Kelengkapan data dapat diberlakukan sebagai tingkat kepuasan. Sebagai contoh, nilai default dari suatu data atau nilai estimasi dapat ditetapkan sebagai pengisi nilai yang hilang (kosong). Bagaimanapun juga pengisian data seperti itu pasti akan berdampak pada keakuratan data yang dihasilkan.

d) Konsistensi data (*Consistency*)

Nilai suatu data yang tersimpan harus tetap sama pada semua kasus.

Gambar Framework Konseptual Data Quality



Lima dimensi yang dijelaskan diatas seperti terlihat pada tabel 1 berikut:

Tabel Score Data Quality

Dimensi evaluasi	Score				
	1	2	3	4	5
Keakuratan data (<i>Accuracy</i>)	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100
Ketepatan waktu (<i>Timeliness</i>)	<i>Worst</i>	<i>Bad</i>	<i>Somewhat</i>	<i>Good</i>	<i>Best</i>
Kelengkapan data (<i>Completeness – number of missing values</i>)	<i>Lots of missing values</i>	<i>Many missing values</i>	<i>Several missing values</i>	<i>Few missing values</i>	<i>No missing values</i>
Konsistensi data (<i>Consistency</i>)	<i>Very Low</i>	<i>Somewhat Low</i>	<i>Somewhat</i>	<i>Somewhat High</i>	<i>Very High</i>
Ukuran Data (<i>Data Size</i>)	<i>Not at All</i>	<i>Somewhat not suitable</i>	<i>Somewhat</i>	<i>Suitable</i>	<i>Very Suitable</i>

Score Total didapat dari persamaan berikut :

$$Squality = (Saccuracy + Stimeliness + Scompleteness + Sconsistency + Syears)/5.....P-1$$

3. Human and Finance factor

Data Mining project tidak bisa secara independent terlepas dari strategi dan kondisi perusahaan. Faktor Sumber Daya Manusia (SDM) memegang peranan penting dalam hal ini. *Data Mining* membutuhkan seorang ahli untuk memilih atribut yang sesuai untuk model dan menilai tingkat ketertarikan (*interestingness*) dan tingkat kegunaan (*usefulness*) dari rule yang dikemukakan. Selain hal tersebut, dukungan financial juga sangat penting untuk berhasilnya suatu kegiatan. Evaluasi dari seorang ahli tentang factor SDM dan financial tergantung dari permintaan perusahaan.

Faktor-faktor SDM yang harus diperhatikan pada bagian ini adalah seperti yang dijelaskan berikut :

- Sponsor*: berpengaruh, fokus pada nilai bisnis, antusias
- User group*: pemilik usaha dan evaluator dari tingkat kesuksesan project
- Business analyst*: merupakan orang yang berpengalaman pada bidang dan aplikasi tertentu
- Data analyst*: merupakan orang yang berpengalaman pada *exploratory data analysis* (EDA) dan *Data Mining*.
- Data management specialist*: merupakan orang yang berpengalaman pada database administration, mempunyai hak akses ke physical data.
- The project manager*: merupakan orang yang berpengalaman pada bidang manajemen proyek

Selain keenam faktor tersebut, peran dari *knowledge manager* juga diperlukan. *Knowledge manager* harus berpengalaman pada *knowledge interpreting* dan *knowledge implementation*. Daftar orang yang telah disebutkan sebelumnya masih harus melalui proses training tentang proses dari *Data Mining*. Penilaian untuk tujuh factor SDM terlihat pada Table 2.

Tabel 1 Score Human Factor

SDM yang dibutuhkan	Score				
	1	2	3	4	5
<i>Sponsor</i>	Lake	Weak	Somewhat	Good	Influential/ enthusiastic
<i>User Group</i>	Lake	Weak	Regular	Good	Experienced
<i>Business analyst</i>	Lake	Weak	Somewhat	Familiar	Experienced
<i>Data analyst</i>	Lake	Weak	Somewhat	Familiar	Experienced
<i>Data management specialist</i>	Lake	Weak	Somewhat	Familiar	Experienced
<i>The project manager</i>	Lake	Weak	Somewhat	Familiar	Experienced

Score Total didapat dari persamaan berikut :

$$Shuman = (Ssponsor + Suser\ group + Sbusiness\ analyst + Sdata\ analyst + Sdata\ management + Sproject\ manager)/6 \dots\dots\dots P-2$$

Penilaian finansial diberikan oleh seorang ahli finance evaluation. Semua nilai yang diberikan adalah bilangan bulat (integer). Nilai maximum finance score adalah lima (5) dan nilai minimum adalah satu (1). Nilai financial dinotasikan sebagai *Sfinance*.

4. The support of the top executives

Tidak mungkin suatu kegiatan yang berhubungan dengan kepentingan perusahaan tidak memperoleh dukungan dari orang-orang yang bekerja diperusahaan tersebut, terutama para pimpinan perusahaan (*top executives*), begitu juga untuk melakukan *Data Mining*. Hal ini akan menentukan apakah project yang akan dikerjakan mendapatkan sumber daya dan dukungan yang layak. Pertanyaan-pertanyaan yang berhubungan dengan dukungan *top executives* terhadap pelaksanaan *Data Mining* terlihat pada tabel 2-4 berikut :

Tabel 1 Score Support Top Executives

No.	Pertanyaan	Score		
		1	2	3
1.	<i>How long have you known data mining (years)</i>	<0.5	0.5–2	>2
2.	<i>Do you know the aim of DM</i>	<i>Do not know</i>	<i>Somewhat</i>	<i>Quite clearly</i>
3.	<i>Do you think the analysis of the past data will improve your decision making</i>	<i>Do not think so</i>	<i>Somewhat</i>	<i>Yes</i>
4.	<i>Do you know the process of DM</i>	<i>Do not know</i>	<i>Somewhat</i>	<i>Quite clearly</i>
5.	<i>Could you accept if the preprocess of the data cost more than 60% of the planed time</i>	<i>Not at all</i>	<i>Somewhat</i>	<i>Yes</i>
6.	<i>Do you think human is more important than algorithm and machine in DM</i>	<i>Don't think so</i>	<i>Somewhat</i>	<i>Yes</i>
7.	<i>If DM project needs the confidential data of your company, would you provide</i>	<i>Reject</i>	<i>Decide based on the condition</i>	<i>Yes</i>
8.	<i>If DM meet a huge setback, accuracy is low for example, would you stop the project soon</i>	<i>Yes</i>	<i>Decide based on the condition</i>	<i>Going on support the project</i>
9.	<i>The management level of the project manager you intend to choose</i>	<i>Department level</i>	<i>Medium level</i>	<i>Top level</i>
10.	<i>Would you mind paying a lot for the DM software</i>	<i>Yes</i>	<i>Neutral</i>	<i>Not at all</i>
11.	<i>Would you support the payment for the DM after the project</i>	<i>No</i>	<i>Neutral</i>	<i>Yes</i>

* Seseorang yang tidak mempunyai keahlian atau pengetahuan tentang data mining

Nilai (score) yang didapat dari pertanyaan-pertanyaan yang diajukan ke *Top Executives (original score)* kemudian dinormalkan agar nilai yang didapat benar-benar setara antara *Executive* yang satu dengan yang lainnya. Persamaan yang digunakan untuk melakukan normalisasi *original score* adalah sebagai berikut :

$$v' = \frac{v-1}{3-1} (5-1) + 1 = 2(v-1) + 1 \dots\dots\dots P-3$$

$$S_{support} = \sum_{i=1}^{11} V_i' / 11 \dots\dots\dots P-4$$

Sehingga jika didapat *original score* seperti terlihat pada tabel 4 maka nilai hasil normalisasi adalah sebagai berikut:

Tabel 1 Score of the support from the executives

Attribute	1	2	3	4	5	6	7	8	9	10	11
Original	2	3	3	1	2	1	1	3	2	3	3
Normalized	3	5	5	1	3	1	1	5	3	5	5

Jika seluruh score yang diperoleh (*Squality, Shuman, Sfinance, Ssupport*) digabungkan maka akan diperoleh persamaan :

$$Stotal = (Squality + Shuman + Sfinance + Ssupport) / 4 \dots\dots\dots P-5$$

Karena kualitas data merupakan dasar dari pelaksanaan *Data Mining*, maka diasumsikan bahwa jika nilai *Squality* yang diperoleh lebih kecil dari 3 maka perusahaan tersebut tidak layak untuk melaksanakan *Data Mining*, dan demikian sebaliknya jika nilai *Squality* yang diperoleh lebih besar atau sama dengan 3 maka perusahaan tersebut layak untuk melaksanakan *Data Mining*. Sebagai alternatif dipergunakan nilai *Stotal*, jika nilai *Stotal* lebih besar atau sama dengan 3 maka perusahaan tersebut layak untuk melaksanakan *Data Mining*.

5. Model Pengambilan Keputusan

Model yang digunakan dalam pengambilan keputusan adalah dengan menggunakan metode **Resiko Bayes**ian.

Action set dinotasikan dengan $A = \{a_1, a_2\}$, dimana a_1 adalah melakukan data mining dan a_2 adalah tidak melakukan data mining. *Natural condition* dinotasikan dengan $= \{, \}$, dimana $,$ adalah perusahaan yang secara nyata layak untuk melakukan data mining, dan $,$ adalah yang sebaliknya.

Diasumsikan bahwa pengambil keputusan adalah orang rasional, yang menghindari resiko (*risk averse person*). Asumsi yang dibangun tentang persamaan matematikanya adalah dengan menggunakan persamaan *exponensial*, seperti yang terlihat pada **P-8** berikut :

$$u(y) = -c.b^\alpha + c(y+b)^\alpha \quad 0 < \alpha < 1 \dots\dots\dots P-6$$

Fungsi persamaan untuk *the loss function* adalah seperti terlihat pada P-9 berikut :

$$l(\theta, a) = \sup_{\theta \in \Theta} \sup_{a \in N} u(\theta, a) - u(\theta, a) \dots\dots\dots P-7$$

Walau dengan evaluasi yang cukup sederhana dari kualitas data dan dukungan dari para executive dapat ditetapkan bahwa propabilitas awal adalah $\pi(\theta_1)$ dan $\pi(\theta_2)$.

Pengamatan yang dilakukan terhadap pengambilan keputusan adalah apakah suatu perusahaan layak atau tidak untuk melakukan data mining. Pengamatan dari permasalahan analisis keputusan adalah $X = \{x_1, x_2\}$. Penjelasan mengenai metode komputasi dari X telah dijelaskan pada bagian 2.2.

Nilai x_1 dinotasikan untuk hasil evaluasi yang layak untuk melakukan data mining, dan x_2 dinotasikan untuk hasil evaluasi yang tidak layak. Persamaan P-10 adalah model komputasi dari Bayesian risk (Ting, 1987)

$$r(\pi, \delta^\pi) = \inf_{\delta \in \Delta} \sum_{\theta_i \in \Theta} \sum_{x_j \in X} l(\theta_i, \delta(x_j)) f(x_j | \theta_i) \pi(\theta_i) \dots\dots\dots P-8$$

Jika $p_{ij} = f(x_j | \theta_i)$, maka diambil contoh untuk p_{11} dimana diasumsikan sebagai kemungkinan dari perusahaan yang benar-benar layak untuk melakukan data mining.

Kemungkinan sebelumnya dinotasikan dengan persamaan berikut:

$$\pi(\theta_i | x_j) = \frac{f(x_j | \theta_i) \pi(\theta_i)}{\sum f(x_j | \theta_i) \pi(\theta_i)} \dots\dots\dots P-9$$

Pada proyek ini menggunakan persamaan:

$$\begin{aligned} m(x_1) &= p_{11}\pi(\theta_1) + p_{21}\pi(\theta_2), \\ m(x_2) &= p_{12}\pi(\theta_1) + p_{22}\pi(\theta_2) \dots\dots\dots P-10 \end{aligned}$$

$$\pi(\theta_1 | x_1) = \frac{p_{11}\pi(\theta_1)}{m(x_1)},$$

$$\pi(\theta_1 | x_2) = \frac{p_{12}\pi(\theta_1)}{m(x_2)} \dots\dots\dots P-11$$

$$\pi(\theta_2 | x_1) = \frac{p_{21}\pi(\theta_2)}{m(x_1)},$$

$$\pi(\theta_2 | x_2) = \frac{p_{22}\pi(\theta_2)}{m(x_2)} \dots\dots\dots P-12$$

Notasi untuk pengambilan keputusan menggunakan $\delta(x)$. Jika hasil evaluasi yang diperoleh adalah layak (x_1) dan perusahaan memutuskan untuk melakukan data mining (a_1). Jika hasil evaluasi yang diperoleh adalah tidak layak (x_2) maka perusahaan tersebut memutuskan tidak melakukan data mining (a_2). Persamaan untuk kriteria pengambilan keputusan adalah sebagai berikut:

$$a_1 = \delta_1(x_1), a_2 = \delta_1(x_2) \dots\dots\dots P-13$$

Persamaan untuk *expectation lost* adalah sebagai berikut :

$$E^0 l(\Theta, \delta_1(x_1)) = l(\theta_1, \delta_1(x_1)) * \pi(\theta_1 | x_1) + l(\theta_2, \delta_1(x_1)) * \pi(\theta_2 | x_1) \dots\dots\dots P-14$$

$$E^0 l(\Theta, \delta_1(x_2)) = l(\theta_1, \delta_1(x_2)) * \pi(\theta_1 | x_2) + l(\theta_2, \delta_1(x_2)) * \pi(\theta_2 | x_2) \dots\dots\dots P-15$$

Persamaan untuk metode *Bayesian Risk* adalah sebagai berikut :

$$r(\pi, \delta_1) = E^0 l(\Theta, \delta_1(x_1)) * m(x_1) + E^0 l(\Theta, \delta_1(x_2)) * m(x_2) \dots\dots\dots P-16$$

Untuk cara yang sama, komputasi dilakukan untuk kriteria pengambilan keputusan dengan menggunakan metode *Bayesian Risk*. $\delta_1(x)$.

Dengan membandingkan nilai dari kedua hasil persamaan tersebut dapat dipilih kriteria dengan hasil yang paling rendah. Para executive dapat menggunakan model ini untuk membantu mereka mengambil keputusan. Jika kriteria pengambilan keputusan tersebut memberikan hasil $\delta_1(x)$ maka perusahaan tersebut dapat melakukan data mining hanya jika nilai yang diperoleh memenuhi syarat. Perusahaan tersebut tidak dapat melakukan data mining jika mendapatkan nilai yang tidak sesuai syarat. Jika kriteria pengambilan keputusan adalah $\delta_2(x)$, maka pilihan perusahaan adalah $\delta_1(x)$, yaitu tetap melakukan data mining meskipun tidak layak untuk melakukannya.

HASIL DAN PEMBAHASAN

Hasil dari penerapan metode adalah sebuah aplikasi yang dapat menerima inputan parameter dan menampilkan hasil keputusan yang didapat dari inputan parameter tersebut.

Semua nilai parameter yang digunakan pada aplikasi ini diperoleh dari *case study* paper *Decision analysis of data mining project based on Bayesian risk* (Nie, G., Zhang, L., Liu, Y., Zheng, X., Shi, Y, 2008). Inputan data parameter pertama adalah parameter Data Kualitas. Hasil dari inputan parameter ini adalah Keputusan Awal dengan memperhatikan SQuality. Tampilan dari inputan seperti terlihat pada gambar 2.

NILAI PARAMETER DATA KUALITAS

Nilai Data Accuracy: 4
 Nilai Data Timeliness: 4
 Nilai Data Completeness: 3
 Nilai Data Consistency: 5
 Nilai Data Data Size: 4

Hasil perhitungan parameter
 Hasil perhitungan (Squality) adalah 4
 Perusahaan bisa melakukan Data Mining

PERINGATAN
 Hasil perhitungan (Squality) adalah 4
 Perusahaan bisa melakukan Data Mining

SIMPAN DATA TUTUP

IDQp	ScAcc	ScTm	ScCmp	ScCns	ScSz	Sdr
1	4	3	3	3	4	340
2	3	2	3	3	4	200

Gambar Tampilan Form Nilai Parameter Data Kualitas

Inputan data parameter yang kedua adalah parameter Human Factor. Hasil dari inputan ini adalah Shuman yang menjadi inputan untuk nilai total. Tampilan dari inputan seperti terlihat pada gambar 3.

NILAI PARAMETER HUMAN FACTOR

Nilai Sponsor: 4
 Nilai User Group: 2
 Nilai Business analyst: 4
 Nilai Data analyst: 3
 Nilai Data management specialist: 5
 Nilai The project manager: 5

Keterangan
 Hasil perhitungan nilai parameter Human Factor (Shuman) adalah 3.8333333333333333

SIMPAN TUTUP

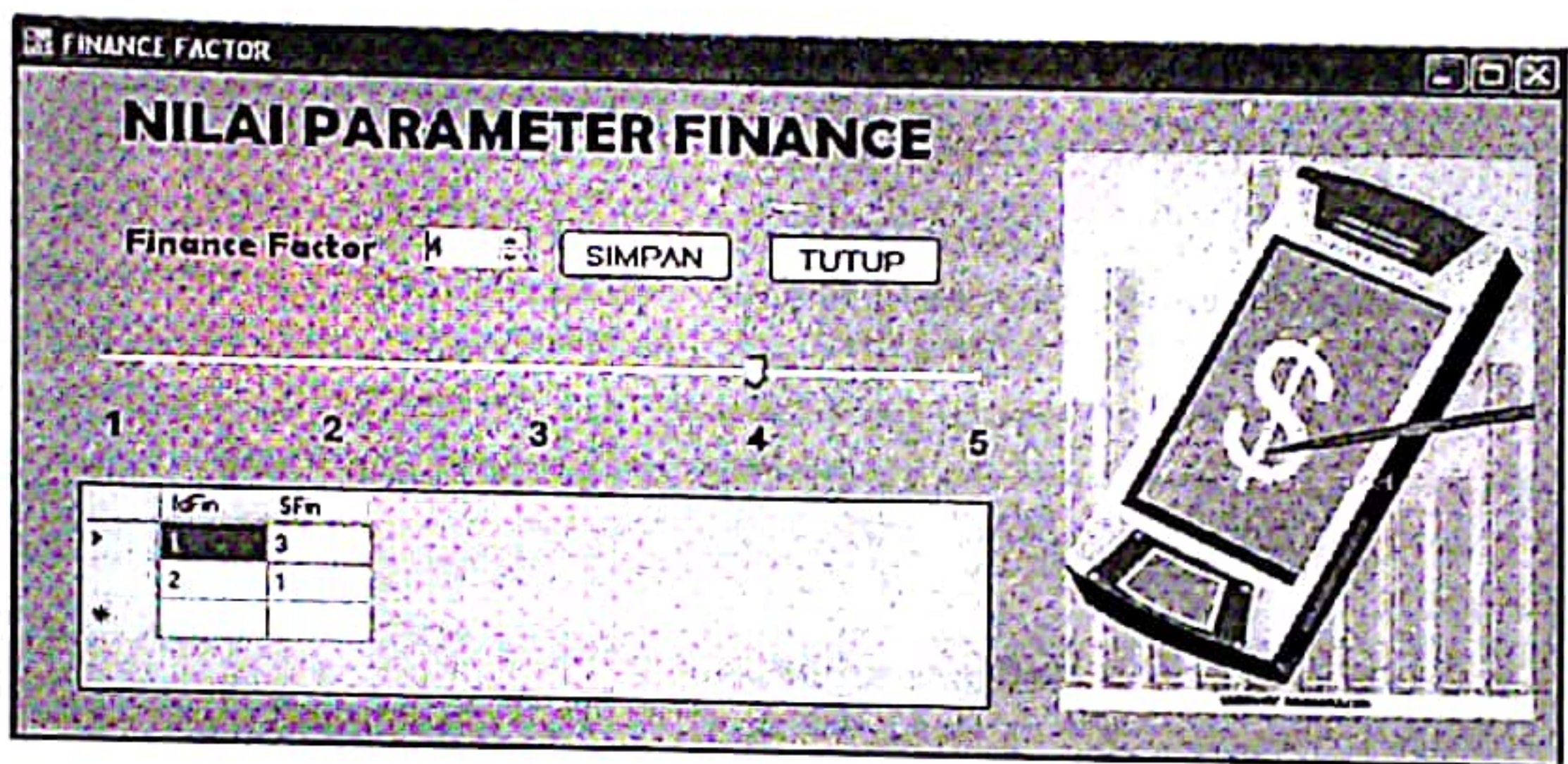
IDHm	ScSps	ScUGr	ScBAn	ScDAn
1	2	2	3	3
2	2	3	4	3

Gambar Tampilan Form Nilai Parameter Human Factor

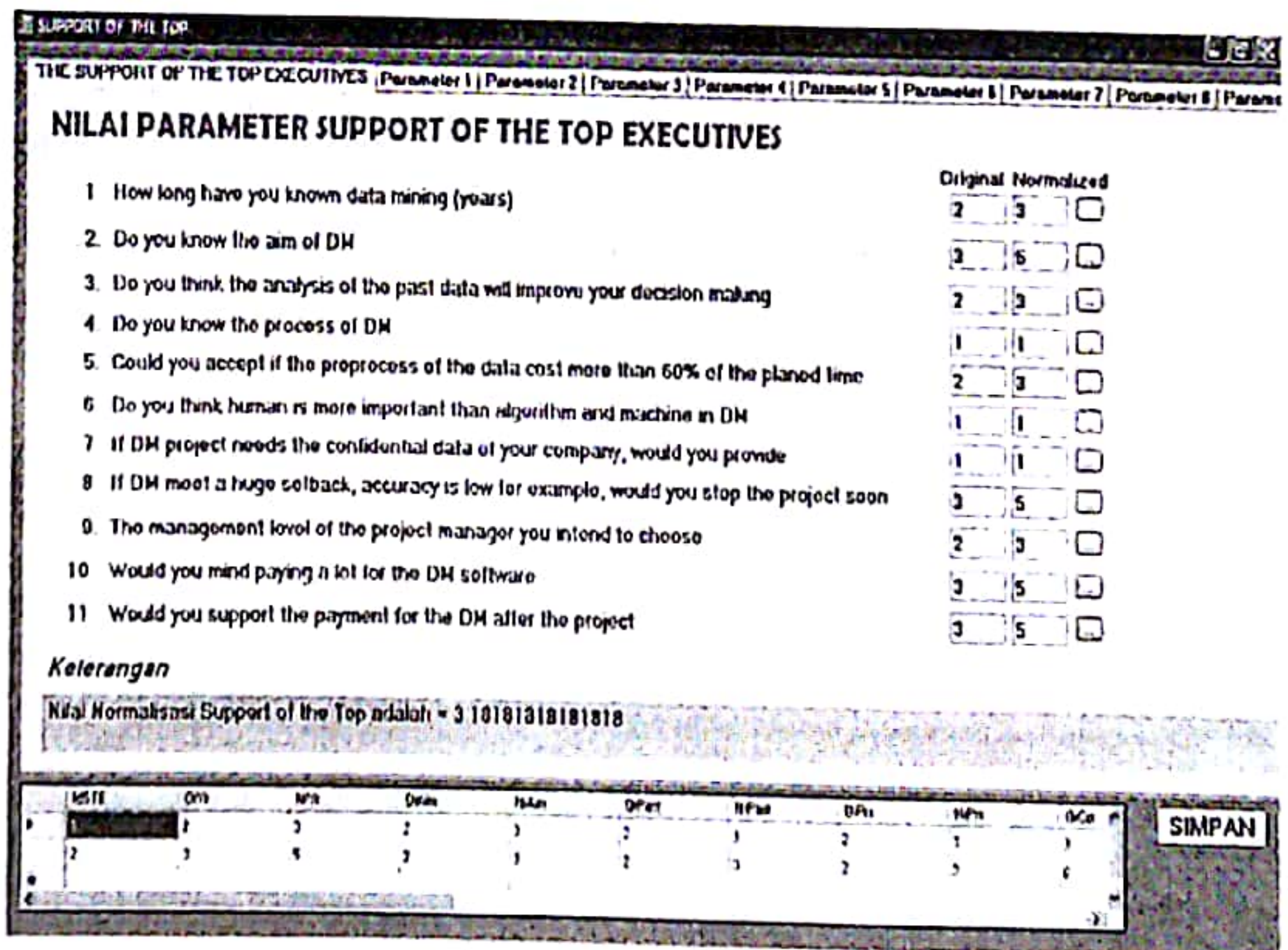
Inputan data parameter yang ketiga adalah parameter Finance. Nilai parameter langsung diisi oleh seorang ahli finance evaluation. Tampilan dari inputan seperti terlihat pada gambar 4.

Inputan data parameter yang terakhir adalah parameter *Support Of The Top Executives*. Hasil dari inputan ini adalah Ssupport yang menjadi inputan untuk nilai total. Tampilan dari inputan seperti terlihat pada gambar 5.

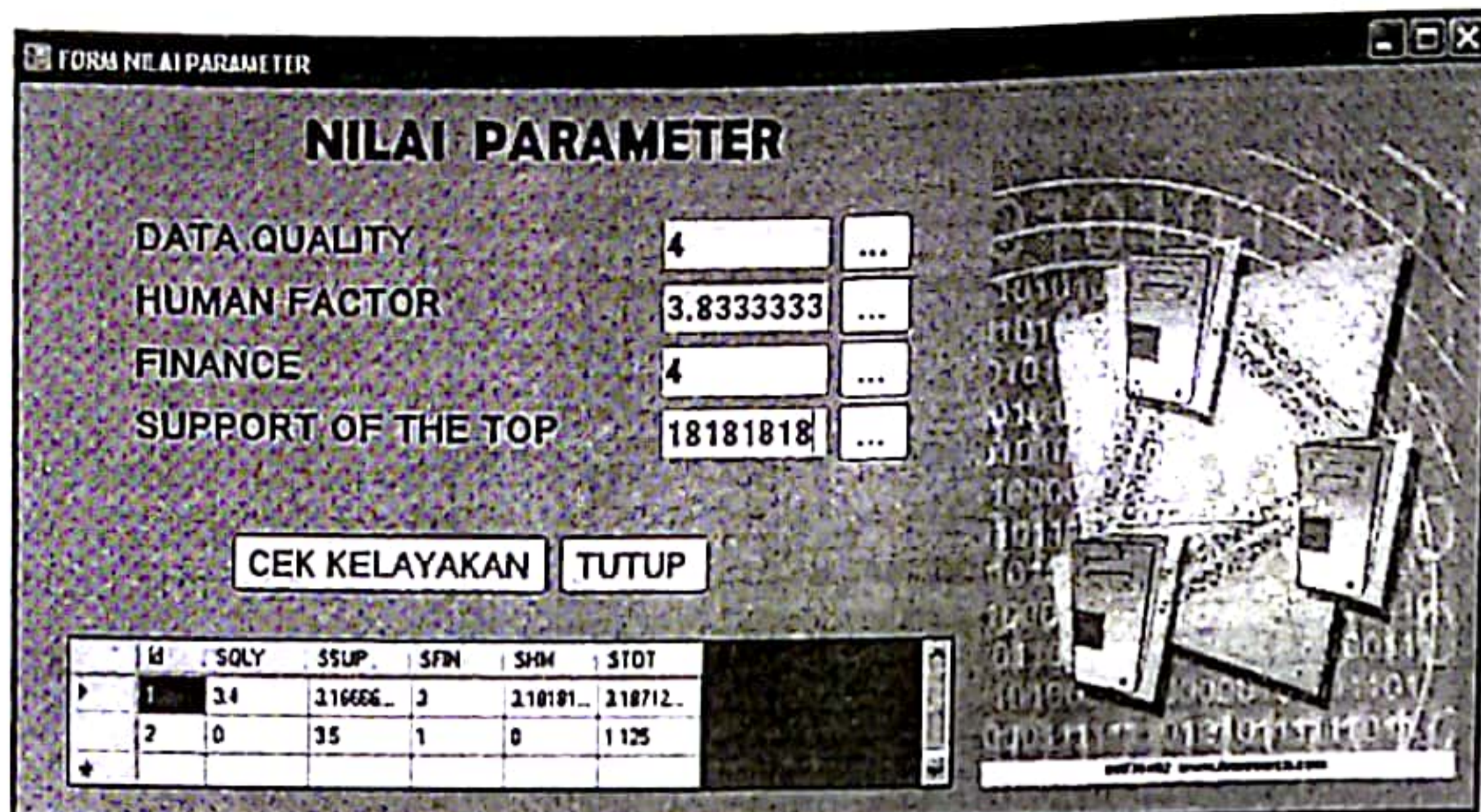
Setelah semua inputan parameter lengkap, maka dapat dihitung nilai Stotal yang merupakan nilai penguat untuk menghasilkan keputusan apakah dapat dilakukan *Data Mining* atau tidak sesuai dengan syarat yang berlaku. Tampilan dari inputan seperti terlihat pada gambar 6.



Gambar Tampilan Form Nilai Parameter Human Factor



Gambar Tampilan Form Nilai Parameter Human Factor



Gambar Tampilan Form Nilai Parameter Human Factor

Setelah diketahui bahwa dengan melihat kesiapan sumber daya yang dimiliki selanjutnya untuk proses pengambilan keputusan apakah pimpinan menyetujui pelaksanaan data mining ini perlu dilakukan perhitungan tersendiri.

Notasi untuk pengambilan keputusan menggunakan $\delta(x)$. Jika hasil evaluasi yang diperoleh adalah layak (x_1) dan perusahaan memutuskan untuk melakukan data mining (a_1). Jika hasil evaluasi yang diperoleh adalah tidak layak (x_2) maka perusahaan tersebut memutuskan tidak melakukan data mining (a_2).

Jika setelah perusahaan tersebut melaksanakan suatu penelitian tentang keberhasilan pelaksanaan data mining sehingga diperoleh suatu *prior probability* dan jika diasumsikan bahwa N adalah perusahaan tersebut layak untuk melakukan data mining dengan *prior probability* sebesar 0.6, maka

- $\pi(\theta_1)$ adalah kemungkinan perusahaan memutuskan melaksanakan data mining = 0.6
- $\pi(\theta_2)$ adalah kemungkinan perusahaan memutuskan tidak melaksanakan data mining = 0.4

Jika diperkirakan tingkat akurasi dari evaluasi adalah sebesar 0,8 maka nilai yang diperoleh adalah sebagai berikut:

- $f(x_1|\theta_1)=p_{11}=0.15$
- $f(x_1|\theta_2)=p_{12}=0.85$
- $f(x_2|\theta_1)=p_{21}=0.85$
- $f(x_2|\theta_2)=p_{22}=0.15$

Menurut persamaan P-12 sampai dengan P-14 maka diperoleh nilai untuk

- $m(x_1) = p_{11} \cdot \pi(\theta_1) + p_{21} \cdot \pi(\theta_2) = (0.15 \cdot 0.6) + (0.85 \cdot 0.4) = 0.43$
 - $m(x_2) = p_{12} \cdot \pi(\theta_1) + p_{22} \cdot \pi(\theta_2) = (0.85 \cdot 0.6) + (0.15 \cdot 0.4) = 0.57$
- dan nilai

- $\pi(\theta_1|x_1) = (0.85 \cdot 0.4) / 0.43 = 0.79$
- $\pi(\theta_1|x_2) = (0.15 \cdot 0.4) / 0.57 = 0.11$
- $\pi(\theta_2|x_1) = (0.15 \cdot 0.6) / 0.43 = 0.21$
- $\pi(\theta_2|x_2) = (0.85 \cdot 0.6) / 0.57 = 0.89$

Untuk nilai finansial yang diperkirakan akan diperoleh, maka diperkirakan perolehan keuntungan yang akan didapat seperti terlihat pada tabel 5 berikut. Dari tabel tersebut dapat dijelaskan sebagai berikut:

- a) jika perusahaan tersebut memang layak dan perusahaan tersebut melaksanakan project data mining maka perkiraan keuntungan yang diperoleh adalah sebesar \$ 559,000
- b) jika perusahaan tersebut memang layak dan perusahaan tersebut tidak melaksanakan project data mining maka perkiraan keuntungan yang diperoleh adalah sebesar \$ 464,000
- c) jika perusahaan tersebut tidak layak dan perusahaan tersebut melaksanakan project data mining maka perkiraan keuntungan yang diperoleh adalah sebesar \$ 321,000
- d) jika perusahaan tersebut tidak layak dan perusahaan tersebut tidak melaksanakan project data mining maka perkiraan keuntungan yang diperoleh adalah sebesar \$ 400,100

Tabel 1 Perkiraan keuntungan yang diperoleh

Action	Status	
	Actually Suitable	Actually unsuitable
Do data mining project	\$ 559,000	\$ 321,000
Not Do data mining project	\$ 464,000	\$ 400,100

Dengan Asumsi bahwa utility function dari para executive dari perusahaan tersebut adalah $U(y) = \frac{1}{0,936} (1 - e^{-0,0000025y})$ maka diperoleh nilai *lose function* adalah

$$l(y) = 1 - u(y) = \frac{1}{0,936} (e^{-0,0000025y} - 0,064)$$

Apabila menerapkan persamaan P-14 dan P-15 maka akan diperoleh nilai untuk :

- $E^0 l(\Theta, \delta_1(x_1)) = (((\exp(-0.0000025 \cdot 559000) - 0.064) \cdot 0.79) / 0.936) + (((\exp(-0.0000025 \cdot 321000) - 0.064) \cdot 0.21) / 0.936) = 0.2402$
- $E^0 l(\Theta, \delta_1(x_2)) = (((\exp(-0.0000025 \cdot 464000) - 0.064) \cdot 0.11) / 0.936) + (((\exp(-0.0000025 \cdot 400100) - 0.064) \cdot 0.89) / 0.936) = 0.3181$

sehingga akan diperoleh nilai untuk Bayesian Risk adalah

$$\delta_1(x) = r(\pi, \delta_1) = (0,3181 \cdot 0,57) + (0,2404 \cdot 0,43) = 0,289$$

$$\delta_2(x) = r(\pi, \delta_2) = (0,3181 \cdot 0,43) + (0,57 \cdot 0,2404) = 0,2737$$

Karena $r(\pi, \delta_1) > r(\pi, \delta_2)$ maka kriteria untuk δ_2 harus diterima. Artinya para executive harus membuat keputusan untuk mulai mengerjakan project data mining meskipun evaluasi yang dihasilkan kurang layak.

Setelah dimasukkan kedalam proses program maka dihasilkanlah nilai seperti yang ditunjukkan pada frame bagian kanan form, seperti terlihat pada gambar 7.

Gambar 7 Tampilan Hasil Perhitungan Resiko

INPUTKAN DATA-DATA BERIKUT		HASIL PERHITUNGAN RESIKO	
phiT1	0.8	m(x1)	0.43
phiT2	0.4	m(x2)	0.57
p11	0.15	phiT1x1	0.7906976
p12	0.85	phiT1x2	0.1052631
p21	0.85	phiT2x1	0.2093023
p22	0.15	phiT2x2	0.8947368
y1	559000	Ed1x1	0.2408362
y2	321000	Ed1x2	0.3181777
y3	464000	rd1	0.2849208
y4	400100	rd2	0.2740930

KESIMPULAN

Untuk penerapan data mining, setiap perusahaan mempunyai analisa sendiri yang didukung oleh beberapa parameter. Hasil dari perhitungan parameter tersebut akan menunjukkan apakah perusahaan tersebut layak atau tidak untuk melakukan data mining. Parameter utama yang digunakan pada penelitian ini adalah parameter Data Quality dan data yang lain adalah data pendukung yang memperkuat hasil keputusan.

DAFTAR PUSTAKA

- Nie, G., Zhang, L., Liu, Y., Zheng, X., Shi, Y. *Decision analysis of data mining project based on Bayesian risk*. Elsevier Ltd. (2008).
- Pollard, W. E. *Bayesian Statistics for Evaluation Research, An Introduction*. Contemporary Evaluation Research Series. SAGE Publications Ltd. California. (1986).
- Wang, R., Strong, D., & Guarascio, L. (1994). Beyond accuracy: What data quality means to data consumers. *Total Data Quality Management Research Program, Journal of Management Information Systems*; Spring 1996; 12, 4; ABI/INFORM Global pg. 5
- Rabin, Matthew. *Risk Aversion and Expected-Utility Theory*. *Econometrica* 68(5) (2000)