

DATA MINING UNTUK MENGANALISIS GAGAL SERAH DANA PADA TRANSAKSI JUAL BELI SAHAM

**Eka Miranda; Hans Yonathan; Ivander G. Lokanatha;
Josep Benedictus Suryawan**

Information Systems Department, School of Information Systems, Binus University
Jl. K.H. Syahdan No. 9, Palmerah, Jakarta Barat 11480
ekamiranda@yahoo.com

ABSTRACT

Competitive environment among investment firms encourage them to be able to conduct analysis related to investments activities with the potential failure of transfers' funds. The purpose of this writing is to analyze the needs of the information required in the process of analysis the potential failure of funds transferring in a process of buying and selling stocks and using data mining techniques (Clustering) to support analysis process of the failure of fund transferring in a process of buying and selling stocks . Data collection methods used in this study consists of: Literature study, interviews and observations. While the implementation phase of data mining techniques consist of: Data Cleaning, Data Integration, Data Selection, Data Transformation, Data Mining, Pattern Evaluation and Knowledge Presentation. Data mining technique is used to analyze the data in order to know which transactions can lead to failed transfers funds and group them based on Receiving Failure. The clustering result will produce knowledge that is useful in the decision making especially to the top management level. The conclusions of this writing is using data mining techniques that have been designed in this writing, help the management take decisions with more precise so as to avoid losses for company.

Keywords: data mining, clustering, failed transfers of funds

ABSTRAK

Persaingan yang kompetitif di antara perusahaan investasi mendorong mereka untuk dapat melakukan analisis terkait dengan investasi yang akan dilakukan oleh nasabah berikut dengan potensi kegagalan serah dana nasabah. Tujuan penulisan ini adalah menganalisis kebutuhan informasi yang diperlukan dalam proses analisis potensi akan kegagalan serah dana pada suatu transaksi jual beli saham dan menggunakan teknik data mining (Clustering) untuk mendukung proses analisis akan kegagalan serah dana pada suatu transaksi jual beli saham. Metode pengumpulan data yang digunakan pada penulisan ini terdiri atas: Studi kepustakaan, wawancara dan observasi. Sedangkan tahapan implementasi teknik data mining terdiri atas: Data Cleaning, Data Integration, Data Selection, Data Transformation, Data Mining, Pattern Evaluation dan Knowledge Presentation. Teknik data mining digunakan untuk memecahkan permasalahan yaitu bagaimana menganalisis data agar dapat mengetahui transaksi yang dapat menyebabkan gagal serah dana lalu mengelompokkannya berdasarkan Receiving Failure. Hasil pengelompokan menghasilkan pengetahuan yang berguna dalam pengambilan keputusan untuk jajaran top level management perusahaan. Kesimpulan dari penulisan ini adalah dengan menggunakan teknik data mining yang telah dirancang pihak management dapat mengambil keputusan dengan lebih tepat sehingga menghindarkan kerugian untuk perusahaan.

Kata kunci: data mining, clustering, gagal serah dana

PENDAHULUAN

Investasi merupakan penanaman sejumlah dana dalam bentuk uang ataupun barang yang diharapkan akan memberikan hasil yang lebih dikemudian hari. Investasi dalam bentuk surat berharga (sekuritas) biasanya dapat dilakukan melalui pasar uang atau pasar modal (Resty Aditia, 2009). Pasar Modal bertindak sebagai penghubung antara para investor dengan perusahaan ataupun institusi pemerintah melalui perdagangan instrumen melalui jangka panjang seperti obligasi, saham, dan lainnya. Fungsi pasar modal adalah meningkatkan dan menghubungkan aliran dana jangka panjang dengan "kriteria pasarnya" secara efisien yang akan menunjang pertumbuhan riil ekonomi secara keseluruhan (Pandji Anoraga, 2008). Dalam melakukan suatu transaksi jual beli saham di pasar modal, transaksi yang dilakukan tidak bisa langsung antara pembeli dengan penjual, diperlukan adanya suatu lembaga khusus yaitu lembaga kliring dan penjaminan sebagai mediator penengah untuk memastikan transaksi yang diproses aman.

Gagal serah dana merupakan suatu keadaan di mana tidak terpenuhi kewajiban serah dana termasuk kewajiban serah efek yang diganti dengan serah uang pada suatu transaksi jual beli saham. Sedangkan permasalahan lain adalah kesulitan menganalisis data dalam jumlah yang sangat besar dan dengan data yang besar tersebut maka akan muncul masalah baru yaitu kesulitan menampilkan hasil analisis secara visual yang dialami di departemen pengendalian resiko. Prediksi potensi akan terjadinya gagal serah dana pada suatu transaksi jual beli saham dilakukan dengan memanfaatkan teknologi *data mining*. Teknik *data mining neural network* untuk prediksi dalam pasar saham merupakan contoh aplikasi *next generation data mining* karena data yang ditangani dalam *data mining* untuk prediksi pasar saham bersifat *time series* (Siahaan Ernestasia, 2010).

Data mining merupakan sebuah metode untuk pemanfaatan kembali data-data lama yang berlimpah. Data-data tersebut bisa sebagai sumber informasi untuk menemukan informasi-informasi yang tersembunyi yang berguna untuk pengambilan keputusan serta perencanaan strategi. Teknik *data mining* yang digunakan pada penulisan kali ini adalah *Clustering*, di mana histori transaksi dikumpulkan dan dianalisa, sehingga hasil dari analisa itu membagi data itu menjadi tiga *cluster*, yaitu *cluster* yang mempunyai potensi yang tinggi terhadap gagal serah dana, *cluster* yang mempunyai potensi terhadap gagal serah dana dan *cluster* yang mempunyai potensi yang rendah terhadap gagal serah dana pada suatu transaksi jual beli saham.

Masalah yang diformulasikan pada penulisan ini adalah: Kenyataan bahwa perusahaan dalam hal ini Departemen Pengendalian Resiko memiliki banyak data namun kurang akan informasi yang berkualitas. Selain itu perusahaan memiliki kebutuhan untuk dapat menganalisis potensi suatu transaksi akan kegagalan serah dana pada suatu transaksi jual beli saham dan menampilkan hasil analisis secara visual. Diperlukan adanya metode analisis resiko lain sebagai referensi dalam penentuan analisis pengendalian resiko yang telah ada. Tujuan dari penulisan ini adalah untuk menganalisis kebutuhan informasi yang diperlukan dalam proses analisis potensi akan kegagalan serah dana pada suatu transaksi jual beli saham dan Menggunakan teknik *data mining Clustering* untuk keperluan analisis.

Studi Pustaka

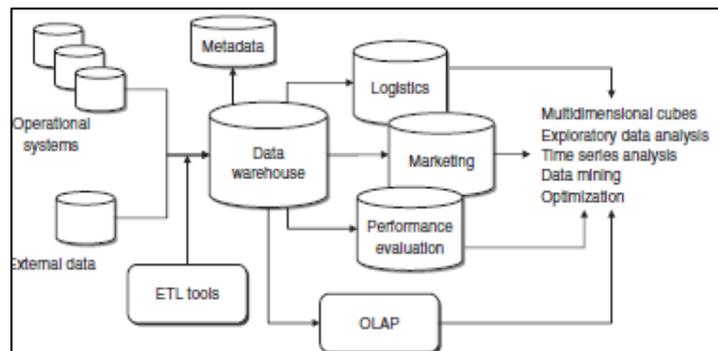
Data warehouse

Data warehouse merupakan tempat penyimpanan untuk ringkasan dari data historis yang seringkali diambil dari *database* departemen atau perusahaan terpisah. *Data warehouse* mengumpulkan semua data perusahaan dalam satu tempat untuk memberikan pandangan yang lebih baik dari proses bisnis dan dapat meningkatkan kinerja organisasi. *Data warehouse* mendukung proses pembuatan keputusan manajemen, proses ini *subject-oriented*, terintegrasi, waktu yang bervariasi dan

permanen (karakteristik data warehouse). *Data warehouse* berfokus pada konsep bisnis daripada proses bisnis dan menampung semua informasi yang relevan yang dikumpulkan dari beberapa sistem pemroses. Pendekatan *data warehouse* dan BI telah disetujui sebagai *middleware layer* untuk sistem pendukung pengambilan keputusan (*Decision Support Systems*) (Seufert dan Schiefer, 2005)

Menurut Inmon (2005), *Data warehouse* adalah kumpulan dari *database* yang terintegrasi dan berorientasi subjek serta dirancang untuk mendukung fungsi pengambilan keputusan, dimana setiap unit data relevan terhadap satu kejadian pada waktu tertentu. *Data warehouse* bertanggung jawab untuk menyimpan, mengintegrasikan, dan mendistribusikan kebutuhan akses (Steve dan Nancy Williams, 2007,). Hal ini dipertegas oleh Hammergren dan Simon (2009) menjelaskan kemampuan *data warehouse* dapat mendukung hal berikut ini, pertama, menjalankan data bisnis, data yang diproduksi oleh aplikasi perusahaan, kedua, mengintegrasikan data bisnis, untuk meningkatkan kualitas dan sinkronisasi dua atau lebih aplikasi perusahaan, bahkan yang tidak dirancang untuk bekerja bersama satu sama lain, ketiga, memonitor data bisnis, mengetahui hubungan antar data, disampaikan kepada pengguna akhir dalam pelaporan dan dukungan data dalam menunjang pengambilan keputusan

Berikut adalah gambar arsitektur *data warehouse* (Vercellis, 2009):



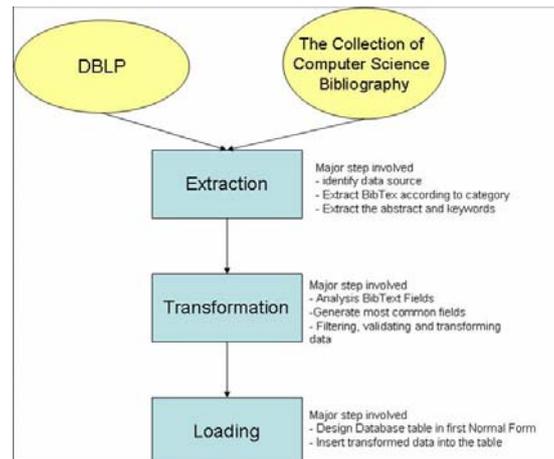
Gambar 1 *Architecture and functions of a data warehouse*

Tahapan untuk membangun data warehouse terdiri atas, pertama, Identifikasi sumber data; Langkah pertama sebelum mulai mengembangkan *data warehouse* yaitu identifikasi sumber data. Perlu mencari tahu apa data yang dibutuhkan untuk diletakkan ke dalam *data warehouse*. Kedua, membangun ETL (*Extraction, Transformation, Loading tool*) Setiap *data warehouse* memiliki data yang berbeda persyaratan, oleh karena itu, *ETL tool* yang sesuai adalah solusi yang lebih baik untuk memenuhi persyaratan.

Ketiga, ekstraksi; Perlu menentukan sistem *database* yang akan digunakan dan juga mencari tahu data apa yang diperlukan sebelum mendapatkannya. Penurunan biaya perangkat keras dan penyimpanan telah mengatasi masalah duplikasi data dan juga kekhawatiran pada kurangnya ruang penyimpanan sebagai tempat penyimpanan data yang berlebihan atau yang tidak diperlukan. Namun, mungkin tidak ada alasan untuk menyimpan data yang tidak diperlukan dan telah diidentifikasi tidak berguna dalam proses pengambilan keputusan. Oleh karena itu, diperlukan untuk mendapatkan hanya ekstrak data yang relevan sebelum membawa ke *data warehouse*.

Keempat, transformasi; Setelah penggalian data dari berbagai sumber, transformasi dibutuhkan untuk menjamin konsistensi data. Agar mengubah data ke dalam *data warehouse* dengan benar, perlu mengetahui cara pemetaan bidang sumber data eksternal ke *data warehouse*. Transformasi dapat dilakukan selama ekstraksi data atau saat memuat data ke dalam *data warehouse*. Integrasi ini bisa menjadi masalah yang kompleks ketika jumlah sumber data menjadi lebih besar.

Kelima, loading; Setelah proses penggalian, mengubah dan pembersihan telah dilakukan, data diambil dan dimasukkan ke dalam *data warehouse*. Pemuatan data dapat dikategorikan ke dalam dua jenis; pemuatan data yang saat ini ada dalam *database* operasional dan pemuatan pembaruan *data warehouse* dari perubahan yang telah terjadi dalam *database* operasional. Untuk menjamin kesegaran data, *data warehouse* perlu untuk terus memperbaharui datanya. Banyak persoalan yang perlu untuk dipertimbangkan terutama saat memuat data yang sudah diperbaharui ke *data warehouse*. Sementara melakukan pemutakhiran *data warehouse*, perlu memastikan bahwa tidak ada data yang dilepaskan dan juga untuk memastikan *overhead* minimum selama proses *scanning file* yang ada.



Gambar 2 Tahap-tahap membangun data warehouse

Sumber : Wah, Teh Ying, Peng, Ng Hooi, Hok, Ching Sue, Building Data Warehouse, *Proceedings of the 24th South East Asia Regional Computer Conference, 2007*, Bangkok, Thailand

Data Mining

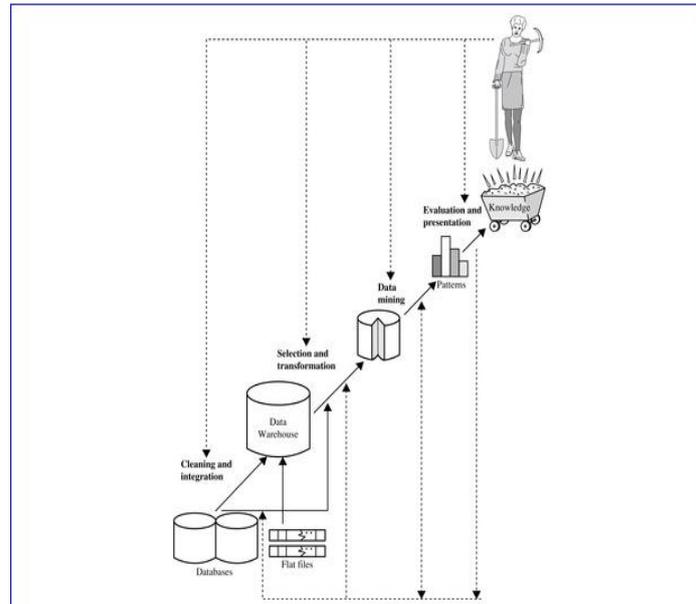
Data Mining adalah proses menganalisa data yang berjumlah besar untuk menemukan suatu pola yang akan digunakan untuk pengambilan keputusan. Hal ini diperkuat oleh teori dari Han Jiawei (2011:p36) yang mengatakan *data mining* adalah proses menemukan pola yang menarik dan pengetahuan dari data yang berjumlah besar.

Kelebihan *Data Mining* sebagai alat analisis adalah (a) *Data mining* mampu menangani *data* dalam jumlah besar dan kompleks. (b) *Data mining* dapat menangani *data* dengan berbagai macam tipe atribut. (c) *Data mining* mampu mencari dan mengolah data secara semi-otomatis. Disebut semi-otomatis karena dalam beberapa teknik data mining, diperlukan parameter yang harus di-*input* oleh user secara manual. (d) *Data mining* dapat menggunakan pengalaman ataupun kesalahan terdahulu untuk meningkatkan kualitas dan hasil analisa sehingga mendapat hasil yang terbaik

Kekurangan *data mining* sebagai alat analisis adalah (a) Terkadang ada *data* yang lebih mudah diselesaikan dengan *statistic manual* dibandingkan dengan menggunakan teknik *data mining*. (b) *Data mining* tidak dapat menemukan *knowledge* secara *instant*. (c) Algoritma *data mining* cukup kompleks. (d) Hasil dari *data mining* tidak dapat langsung digunakan, harus dianalisa dan diinterpretasi terlebih dahulu.

Langkah menemukan pengetahuan pada *data mining* terdiri dari: (a) *Data Cleaning*; memperbaiki data yang salah, menghapus data yang rusak dan tidak konsisten. (b) *Data Integration*;

mengintegrasikan data dari berbagai macam sumber dan menyatukan agar mudah dipilih dan diproses nantinya. (c) *Data Selection*; memilih data yang dibutuhkan pada *database* dan digunakan untuk proses analisis. (d) *Data Transformation*; mengubah dan menggabungkan data dari berbagai macam bentuk menjadi satu bentuk yang sama agar mudah diproses. (e) *Data Mining*; tahap untuk menerapkan metode dalam proses *modeling* data yang akan digunakan pada proses *data mining*. (f) *Pattern Evaluation*; melakukan evaluasi akan *patern* yang telah diproses, aspek-aspek yang dievaluasi adalah hasil *output* yang didapat setelah proses *data mining* dilakukan. (g) *Knowledge Presentation*; melakukan penyajian hasil dari proses *data mining* yang sudah diproses.



Gambar 3. Proses *Knowledge Discovery*

Sumber: Jiawei, H. (2011). *Data Mining : Concepts and Techniques*.

Konsep dasar *Clustering*

Clustering adalah proses pengelompokan kumpulan data menjadi beberapa kelompok sehingga objek di dalam satu kelompok memiliki banyak kesamaan dan memiliki banyak perbedaan dengan objek di kelompok lain. Perbedaan dan persamaannya biasanya berdasarkan nilai atribut dari objek tersebut dan dapat juga berupa perhitungan jarak. Proses *clustering* akan menghasilkan *cluster* yang baik apabila pertama, tingkat kesamaan yang tinggi dalam satu *class* dan kedua, tingkat kesamaan yang rendah antar *class*. Kesamaan yang dimaksud merupakan pengukuran secara *numeric* terhadap dua buah objek. Nilai kesamaan ini akan semakin tinggi bila dua objek yang dibandingkan memiliki kemiripan yang tinggi. Perbedaan kualitas hasil *clustering* bergantung pada metode yang dipakai.

Metode *Clustering* juga harus dapat mengukur kemampuannya dalam usaha untuk menemukan suatu pola tersembunyi pada data yang tersedia. Dalam mengukur nilai kesamaan ini ada beberapa metode yang dapat dipakai. Salah satu metodenya adalah *weighted Euclidean Distance*. Dalam metode ini dua buah point dapat dihitung jaraknya bila diketahui nilai dari masing-masing atribut pada kedua point tersebut, berikut rumusnya:

$$Distance(p, q) = \left(\sum_{k=1}^n \mu_k |P_k - q_k|^r \right)^{1/r}$$

Keterangan:

N = Jumlah *record* data

K= Urutan *field* data

r= 2

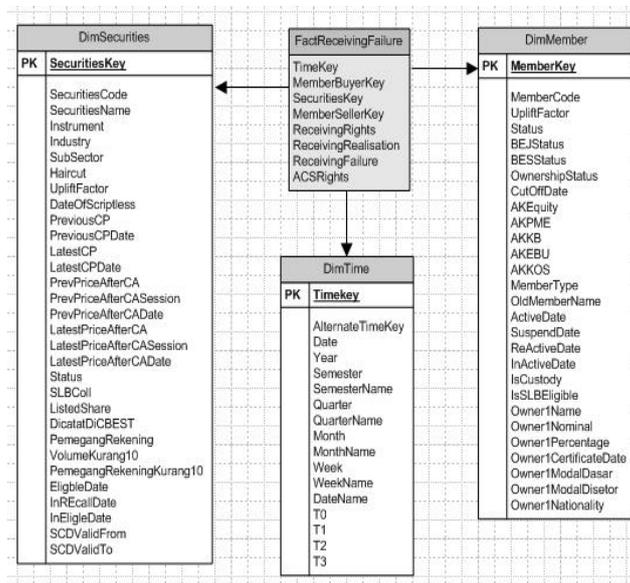
μ_k = Bobot *field* yang diberikan *user*

METODE

Metode pengumpulan data yang digunakan pada penulisan ini terdiri atas: Studi kepustakaan, Wawancara dan Observasi. Sedangkan tahapan implementasi teknik *data mining* terdiri atas: *Data Cleaning*, *Data Integration*, *Data Selection*, *Data Transformation*, *Data Mining*, *Pattern Evaluation* dan *Knowledge Presentation*

HASIL DAN PEMBAHASAN

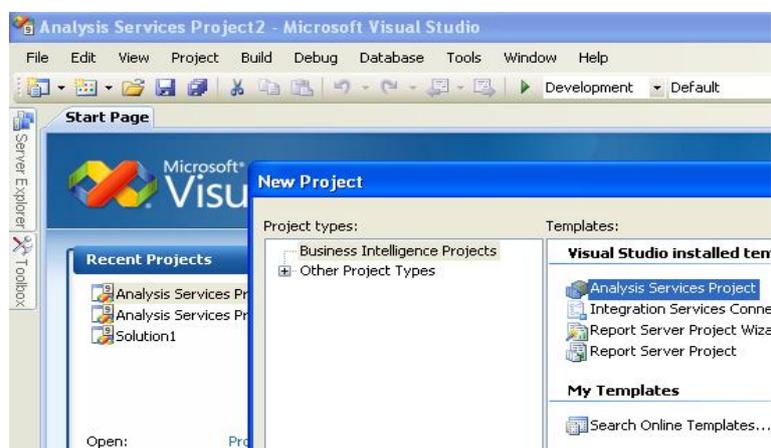
Proses utama yang akan dianalisa dalam tulisan ini adalah proses pengendalian resiko. Pada proses ini secara lebih spesifik akan dianalisa tiga hal yaitu: Gagal serah dana, Tipe sekuritas yang beresiko dan Member yang beresiko. Data pada *table Fact Receiving Failure* digunakan untuk menganalisa akan suatu potensi kegagalan serah dana pada suatu transaksi dari data yang sudah dikumpulkan dalam tabel fakta tersebut. Gambar *table Fact Receiving Failure* adalah sebagai berikut:



Gambar 4 Star Schema Fact Receiving Failure

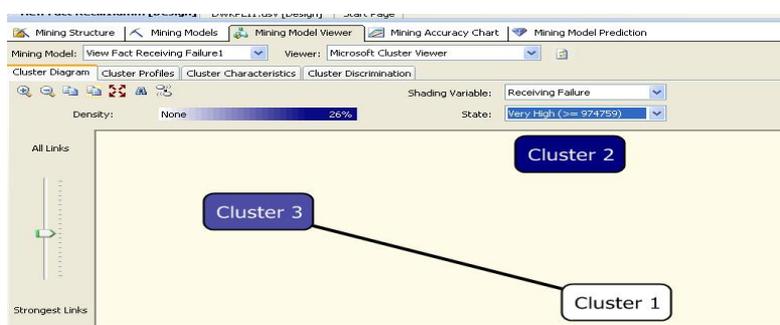
Tahapan dalam melakukan data mining adalah pertama, *Data Cleaning*; data dari *data warehouse* lebih berkualitas daripada data yang ada pada *database* biasa karena data dari *data warehouse* telah melewati proses *ETL*. Sehingga data sudah bersih dan berkualitas untuk dilakukan proses *mining*. Kedua, *Data Integration*; Karena data yang dipakai berada pada *data warehouse*, maka

tidak perlu melakukan proses ini lagi karena data yang berada pada *data warehouse* disimpan secara terpusat. Ketiga, *Data Selection*; tahap *data selection* merupakan tahap pemilihan data yang *relevant* yang berada di dalam *data warehouse*, seperti yang terlihat hasilnya pada Gambar 4. *Star scheme Fact Receiving Failure*. Keempat, *Data Transformation*; pada tahap ini, data yang sudah diseleksi pada tahap *data selection* akan digunakan untuk pembuatan model *data mining*. Model *data mining* ini berguna untuk memproses data tersebut agar dapat digunakan untuk proses penganalisaan yang menghasilkan informasi tersembunyi di dalamnya dan sangat bermanfaat dalam pengambilan keputusan. Berikut merupakan *view* yang dibentuk dari *fact receiving failure* yang sudah melewati tahap *data selection*: *View Fact Receiving Failure*. Kelima, *Data Mining*; tahap ini merupakan proses untuk menganalisa data yang sudah melewati tahap-tahap sebelumnya sehingga dari data-data tersebut dapat ditemukan informasi tersembunyi di dalamnya yang sangat berguna bagi perusahaan untuk membuat suatu keputusan. Pada tahap ini akan dilakukan proses perancangan model *data mining*, yang terdiri dari teknik-teknik yang biasanya dipakai dalam proses *data mining*, dalam penulisan ini teknik yang dipakai adalah *Clustering* pada *SQL Server Business Intelligence Development Studio 2008*.



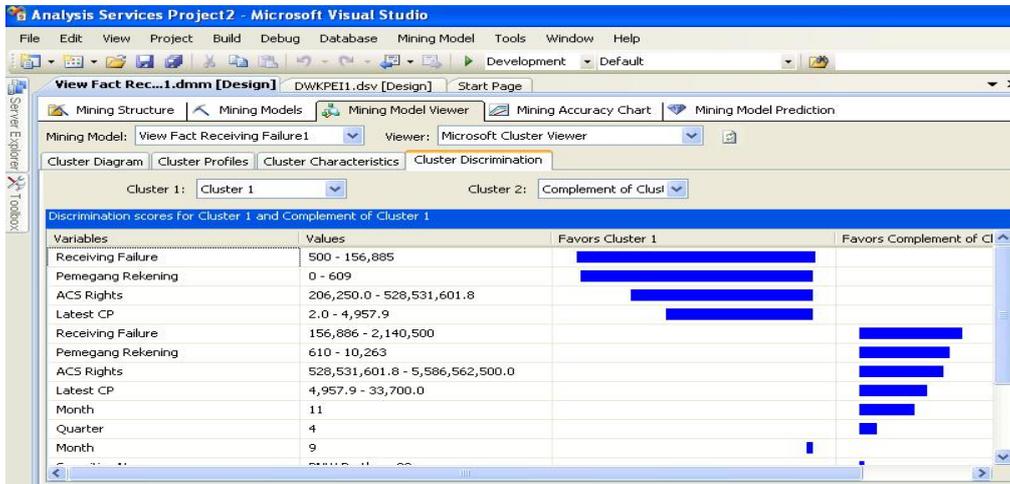
Gambar 5 New Project pada *SQL Server Business Intelligence Development Studio 2008*

Gambar 6. menggambarkan tiga pembentukan *cluster* dari data yang dimiliki. *Cluster* tersebut terdiri dari *cluster* yang mempunyai potensi yang tinggi terhadap gagal serah dana, *cluster* yang mempunyai potensi terhadap gagal serah dana dan *cluster* yang mempunyai potensi yang rendah terhadap gagal serah dana pada suatu transaksi jual beli saham.



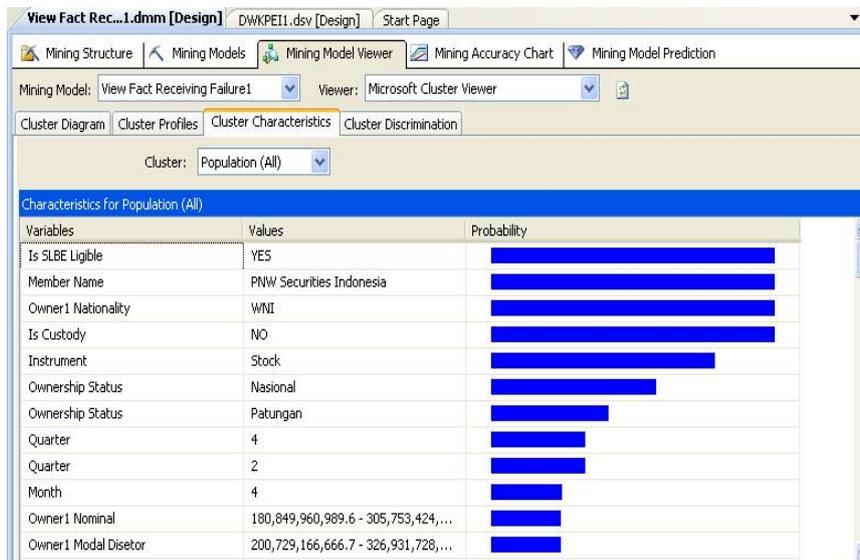
Gambar 6 Clustering the data

Gambar 7. menggambarkan hasil *Cluster Discrimination* terhadap tiga *cluster* yang terbentuk.



Gambar 7 *Cluster discrimination*

Gambar 8. menggambarkan karakteristik dari tiap *cluster* yang terbentuk.



Gambar 8 *Cluster characteristic*

Keenam, *Pattern Evaluation*; setelah melakukan proses data mining, tahapan selanjutnya adalah melakukan pattern evaluation, pattern evaluation dilakukan untuk mengevaluasi model data mining yang telah dibuat beserta hasil output yang dihasilkan dari proses data mining yang telah dilakukan, ketujuh, *Knowledge Presentation*; setelah semua tahap proses *data mining* sebelumnya dilakukan semuanya, tahap terakhir dalam perancangan sistem *data mining* merupakan tahap *knowledge presentation*. Tahap ini bertujuan untuk visualisasi dan merepresentasikan pengetahuan yang sudah dihasilkan dari proses *data mining* sebelumnya.



Gambar 9 Data Mining Model Dashboard

SIMPULAN

Informasi yang diperlukan dalam proses analisis potensi akan kegagalan serah dana pada suatu transaksi jual beli saham terdiri atas: informasi akan fungsi perusahaan, tugas dan wewenang tiap divisi, sistem kliring yang sedang dijalankan, dan masalah yang sedang dihadapi. Adapun tabel yang digunakan adalah tabel fakta dan dimensi sebagai berikut: tabel fakta *receiving failure*, tabel dimensi *time*, tabel dimensi *member*, dan tabel dimensi *securities*. Serta atribut yang digunakan untuk melakukan analisis gagal serah dana pada transaksi jual beli saham adalah: *memberName*, *OwnershipStatus*, *IsCustody*, *IsSLBEligible*, *OwnerINominal*, *OwnerIPercentage*, *OwnerIModalDasar*, *OwnerIModalDisetor*, *OwnerINationality*, *SecuritiesName*, *Instrument*, *Industry*, *SubSector*, *PemegangRekening*, *PemegangRekeningKurang10*, *UpliftFactor*, *LatestCP*, *quarter*, *month*, *ReceivingRights*, *ReceivingRealtisation*, *ReceivingFailure*, dan *ACSRight*. Teknik *data mining* yang digunakan adalah teknik *clustering*. Adapun hasil *cluster* yang dihasilkan adalah membentuk tiga *cluster*, yaitu: *cluster* yang sangat berpotensi terjadinya gagal serah dana (*high risk*), *cluster* yang mempunyai potensi terjadinya gagal serah dana (*normal risk*), *cluster* yang mempunyai sedikit potensi terjadinya gagal serah dana (*low risk*).

DAFTAR PUSTAKA

- Hammergren, T.C., Simon A.R. (2009). *Data Warehousing For Dummies*. 2nd Edition. Indianapolis, Indiana: Wiley Publishing, Inc.
- Inmon, W.H. (2005). *Building the Data Warehouse*. Fourth Edition. Indianapolis: Wiley Publishing Inc.
- Jiawei, H. (2011). *Data Mining : Concepts and Techniques*. Waltham: Morgan Kaufmann.
- Pandji Anoraga. (2009). *Pengantar Pasar Modal*. Jakarta: Rineka Cipta.

- Resty Aditia Safitri. (2009). *Analysis Of The Formation Of Portofolio Securities In Three Mining Company Of Record As Index Lq 45 In Indonesia Stock Exchange*, 1-3. Jakarta: Universitas Gunadarma.
- Seufert, A., Schiefer, J. (2005), Enhanced business intelligence- supporting business processes with real-time business analytics. *Proceedings of the 16th International Workshop on Database and Expert System Applications*. DEXA'05
- Siahaan, Ernestasia. (2010). *Penggunaan Teknik Data Mining Neural Network Untuk Prediksi Dalam Pasar Saham*, 1-2. Bandung: Institut Teknologi Bandung.
- Vercellis, Carlo, (2009), *Business Intelligence: Data Mining and Optimization for Decision Making*. United Kingdom: John Wiley and Sons.
- Wah, Teh Ying, Peng, Ng Hooi, Hok, Ching Sue. (2007). Building Data Warehouse. *Proceedings of the 24th South East Asia Regional Computer Conference*, Bangkok, Thailand
- Williams, Steve., Williams, Nancy. (2007). *The Profit Impact of Business Intelligence*. Morgan Kaufmann.