

**PENERAPAN DATA MINING DENGAN ASSOCIATION RULES UNTUK
MELIHAT HUBUNGAN TERTANGGUNG, PEMILIHAN PRODUK
DAN PERILAKU NASABAH
(Studi Kasus : PT. Prudential Life Assurance)**

Awan Setiawan

Universitas Langlangbuana
awans2425@gmail.com

Erwin Yulianto

Universitas Langlangbuana
rwinyulianto@yahoo.com

ABSTRACT

The world of insurance business that is full of competition makes the perpetrators must always think about breakthrough strategies that can guarantee the continuity of their insurance business. One of the main assets owned by insurance companies is business data in an extraordinary amount.

Data mining is a new technology that is very useful to help insurance companies find very important information from business data as the main asset they have. Data mining can predict trends and traits of business behavior that are very useful to support important decision making. Automated analysis carried out by data mining exceeds that carried out by traditional support systems. Apriori and FP-Growth are the most famous algorithms for finding high frequency patterns, these algorithms are part of the Rule Association used in this study.

Keywords: *Data Mining, Association Rules, FP Growth, Apriori Algorithm, Customer Behavior.*

ABSTRAK

Dunia bisnis asuransi yang penuh persaingan membuat para pelakunya harus selalu memikirkan strategi-strategi terobosan yang dapat menjamin kelangsungan bisnis asuransi mereka. Salah satu aset utama yang dimiliki oleh perusahaan asuransi adalah data bisnis dalam jumlah yang luar biasa banyak.

Data mining merupakan teknologi baru yang sangat berguna untuk membantu perusahaan asuransi menemukan informasi yang sangat penting dari data bisnis sebagai aset utama yang mereka miliki. *Data mining* bisa meramalkan tren dan sifat-sifat perilaku bisnis yang sangat berguna untuk mendukung pengambilan keputusan penting. Analisis yang diotomatisasi yang dilakukan oleh *data mining* melebihi yang dilakukan oleh sistem pendukung tradisional. Apriori dan *FP-Growth* adalah algoritma paling terkenal untuk menemukan pola frekuensi tinggi, algoritma tersebut merupakan bagian dari *Association Rule* yang digunakan dalam penelitian ini.

Kata Kunci: *Data Mining, Association Rules, FP Growth, Algoritma Apriori, Perilaku Nasabah.*

1. PENDAHULUAN

1.1 Latar Belakang Masalah

Dunia bisnis asuransi yang penuh persaingan membuat para pelakunya harus selalu memikirkan strategi-strategi terobosan yang dapat menjamin kelangsungan bisnis asuransi mereka. Salah satu aset utama yang dimiliki oleh perusahaan asuransi adalah data bisnis dalam jumlah yang luar biasa banyak. Ini melahirkan kebutuhan akan adanya teknologi yang dapat memanfaatkannya untuk membangkitkan pengetahuan-pengetahuan baru, yang dapat membantu dalam pengaturan strategi bisnis asuransi.

Minat nasabah nasabah terekam dalam data *questioner* yang biasanya diisi nasabah saat mengajukan permohonan pembuatan asuransi, data ini tersimpan sejak 5 tahun yang lalu di dalam database *Core System* yang digunakan di Asuransi Prudential. Jumlah data juga cukup besar yaitu sebanyak 3 juta *record* sehingga bisa mewakili jika dijadikan sebagai sumber data mining. Data *questioner* selama ini hanya digunakan untuk dasar dalam pembuatan polis saja dan sesudah itu nyaris tidak digunakan, padahal bisa dimanfaatkan untuk mendapatkan informasi yang mungkin dibutuhkan perusahaan.

Pengolahan data minat nasabah akan memberikan gambaran pada perusahaan mengenai keinginan pasar secara global tentunya hal ini sangat berguna dalam menentukan produk yang dapat diterima dan dibutuhkan oleh nasabah. Produk memiliki pengertian yang luas yaitu segala sesuatu yang ditawarkan, dimiliki, digunakan, atau dikonsumsi sehingga dapat memuaskan keinginan dan kebutuhan. Karena dengan adanya produk yang ditawarkan oleh perusahaan dengan kualitas yang baik tentunya akan menarik minat nasabah maupun calon nasabah untuk membeli produk tersebut maupun untuk menjadi nasabah tetap dalam produk tersebut. Karena, semakin baik kualitas produk yang ditawarkan suatu perusahaan, tentu akan mempengaruhi minat nasabah dalam berasuransi sehingga akan laku dipasar dan akan berdampak pada profit perusahaan.

Data mining merupakan teknologi baru yang sangat berguna untuk membantu perusahaan asuransi menemukan informasi yang sangat penting dari data bisnis sebagai aset utama yang mereka miliki. Data mining bisa meramalkan tren dan sifat-sifat perilaku bisnis yang sangat berguna untuk mendukung pengambilan keputusan penting dalam menentukan strategi-strategi yang dapat menjamin kelangsungan bisnis asuransi mereka. Analisis yang diotomatisasi yang dilakukan oleh data mining melebihi yang dilakukan oleh sistem pendukung keputusan tradisional yang sudah banyak digunakan pada umumnya, sehingga mendorong penulis untuk mengambil penelitian tentang penerapan *Data Mining* dengan *Association Rules* Untuk Melihat Hubungan Tertanggung, Pemilihan *Product* dan Perilaku Nasabah di PT Prudential Life Assurance.

1.2 Tinjauan Pustaka

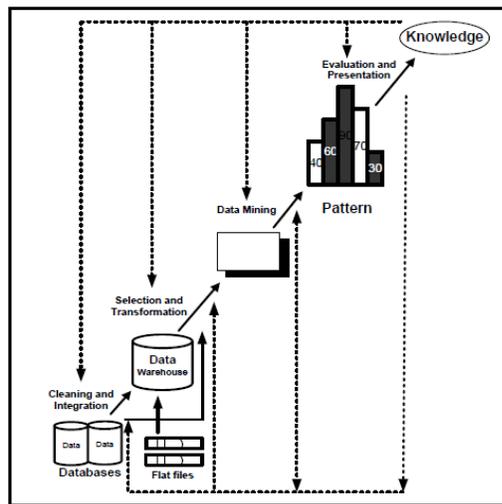
1.2.1 Data Mining

Data mining adalah kombinasi secara logis antara pengetahuan data, dan analisa statistik yang dikembangkan dalam pengetahuan bisnis atau suatu proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, tiruan dan *machine-learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat bagi pengetahuan yang terkait dari berbagai database besar (Therling K, 2006). *Data mining* juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari

suatu kumpulan data (Pramudiono, 2007). *Data mining*, sering juga disebut sebagai *Knowledge Discovery in Database (KDD)*. KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santosa, 2007).

Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. *Data mining* berkaitan dengan bidang ilmu-ilmu lain, seperti *database system*, *data warehousing*, statistik, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu, *data mining* didukung oleh ilmu lain seperti *neural network*, pengenalan pola, *spatial data analysis*, *image database*, *signal processing* (Han, 2006).

Sebagai suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap yang diilustrasikan di Gambar 1. Tahapan-tahapan tersebut antara lain Pembersihan Data (*Data Cleaning*), Integrasi Data (*Data Integration*), Seleksi Data (*Data Selection*), Transformasi Data (*Data Transformation*), Proses Mining, Evaluasi Pola (*Pattern Evaluation*), dan Presentasi Pengetahuan (*Knowledge Presentation*) yang bersifat interaktif, pemakai terlibat langsung atau dengan perantara *knowledge base*.



Gambar 1 Tahapan Data Mining
Sumber: Han, 2006

1.2.2 Perilaku Konsumen

Perilaku konsumen merupakan studi tentang bagaimana individu, kelompok dan organisasi dan proses yang dilakukan untuk memilih, mengamankan, menggunakan dan menghentikan produk, jasa, pengalaman atau ide untuk memuaskan kebutuhan dan dampaknya terhadap konsumen dan masyarakat (Kotler & Lan Keller, 2008).

Dalam penelitian ini perilaku nasabah yang akan diteliti antara lain pembelian polis baru, *Top Up*, *Freelook*, *Partial Withdraw* dan *Surrender*. Beberapa istilah asuransi yang digunakan, yaitu (Sanabila, 2016) :

1. Tertanggung adalah seseorang yang atas jiwanya diadakan perjanjian Asuransi Jiwa.
2. *Top Up* adalah penambahan dana investasi di luar pembayaran premi dasar asuransi

3. *Freelook* adalah masa selama 14 hari saat pemegang polis dapat membatalkan polis apabila tidak menyetujui syarat dan ketentuan dalam ketentuan polis karena alasan apapun. Perusahaan akan mengembalikan premi yang telah dibayarkan dikurangi biaya pembatalan polis.
4. *Partial Withdraw* adalah penarikan dana investasi sebagian.
5. *Surrender* adalah permintaan tertulis dari pemegang polis untuk mengakhiri polis sebelum akhir masa asuransi.

1.2.3 Association Rules

Aturan asosiasi (*Association rules*) atau analisis afinitas (*affinity analysis*) berkenaan dengan studi tentang “apa bersama apa”. Sebagai contoh dapat berupa transaksi di supermarket, misalnya seseorang yang membeli susu bayi juga membeli sabun mandi. Pada kasus ini berarti susu bayi bersama dengan sabun mandi. Karena awalnya berasal dari studi tentang *database* transaksi pelanggan untuk menentukan kebiasaan suatu produk dibeli bersama produk apa, maka aturan asosiasi juga sering dinamakan *market basket analysis*. Aturan asosiasi ingin memberikan informasi dalam bentuk hubungan “if-then” atau “jika-maka”. Aturan ini dihitung dari data yang sifatnya probabilistik (Santosa, 2007). Analisis asosiasi dikenal juga sebagai salah satu metode data mining yang menjadi dasar dari berbagai metode data mining lainnya. Salah satu tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (*frequent pattern mining*) banyak menarik perhatian para peneliti untuk menghasilkan algoritma yang efisien. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, *support* (nilai penunjang) yaitu prosentase kombinasi item tersebut dalam *database* dan *confidence* (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan asosiatif. Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan asosiatif yang memenuhi syarat minimum untuk *support* (*minimum support*) dan syarat minimum untuk *confidence* (*minimum confidence*) (Pramudiono, 2007).

Terdapat beberapa algoritma yang sudah dikembangkan mengenai aturan asosiasi, akan tetapi ada satu algoritma klasik yang sering dipakai yaitu algoritma apriori. Ide dasar dari algoritma ini adalah dengan mengembangkan *frequent* itemset. Dengan menggunakan satu item dan secara rekursif mengembangkan *frequent* itemset dengan dua item, tiga item dan seterusnya, hingga *frequent* itemset dengan semua ukuran. Untuk mengembangkan *frequent* set dengan dua item, dapat menggunakan *frequent* set item. Alasannya adalah apabila set satu item tidak melebihi *support* minimum, maka sembarang ukuran itemset yang lebih besar tidak akan melebihi *support* minimum tersebut.

Secara umum, mengembangkan set dengan *fc*-item menggunakan *frequent* set dengan $k-1$ item yang dikembangkan dalam langkah sebelumnya. Setiap langkah memerlukan sekali pemeriksaan ke seluruh isi database. Dalam asosiasi terdapat istilah *antecedent* dan *consequent*, *antecedent* untuk mewakili bagian “jika” dan *consequent* untuk mewakili bagian “maka”. Dalam analisis ini, *antecedent* dan *consequent* adalah sekelompok item yang tidak punya hubungan secara bersama (Santosa, 2007). Dari jumlah besar aturan yang mungkin dikembangkan, perlu memiliki aturan-aturan yang cukup kuat tingkat ketergantungan antar item dalam *antecedent* dan *consequent*. Untuk mengukur kekuatan aturan asosiasi ini, digunakan ukuran *support* dan *confidence*. *Support* adalah rasio antara jumlah transaksi yang memuat *antecedent* dan *consequent* dengan jumlah transaksi. *Confidence* adalah rasio antara jumlah transaksi yang

meliputi semua item dalam *antecedent* dan *consequent* dengan jumlah transaksi yang meliputi semua item dalam *antecedent* yang dijelaskan pada 2 persamaan berikut.

$$S = \frac{\Sigma(Ta+Tc)}{\Sigma(T)}$$

Keterangan:

S = Support

$\Sigma(Ta + Tc)$ = Jumlah transaksi yang mengandung *antecedent* dan *consequent*

$\Sigma(T)$ = Jumlah transaksi

$$C = \frac{\Sigma(Ta+Tc)}{\Sigma(Ta)}$$

Keterangan:

C = Confidence

$\Sigma(Ta + Tc)$ = Jumlah transaksi yang mengandung *antecedent* dan *consequent*

$\Sigma(Ta)$ = Jumlah transaksi yang mengandung *antecedent*

Langkah pertama algoritma apriori adalah, *support* dari setiap item dihitung dengan men-scan database. Setelah *support* dari setiap item didapat, item yang memiliki *support* lebih besar dari minimum *support* dipilih sebagai pola frekuensi tinggi dengan panjang 1 atau sering disingkat 1-*itemset*. Singkatan k-*itemset* berarti satu set yang terdiri dari k item. Iterasi kedua menghasilkan 2-*itemset* yang tiap set-nya memiliki dua item. Pertama dibuat kandidat 2-*itemset* dari kombinasi semua 1-*itemset*. Lalu untuk tiap kandidat 2-*itemset* ini dihitung support-nya dengan men-scan database. *Support* artinya jumlah transaksi dalam database yang mengandung kedua item dalam kandidat 2-*itemset*. Setelah *support* dari semua kandidat 2-*itemset* didapatkan, kandidat 2-*itemset* yang memenuhi syarat minimum *support* dapat ditetapkan sebagai 2-*itemset* yang juga merupakan pola frekuensi tinggi dengan panjang 2 (Pramudiono, 2007).

Untuk selanjutnya iterasi ke-k dapat dibagi lagi menjadi beberapa bagian sebagai berikut :

1. Pembentukan kandidat k-*itemset* dibentuk dari kombinasi (k-1)-*itemset* yang didapat dari iterasi sebelumnya. Satu ciri dari algoritma apriori adalah adanya pemangkasan kandidat k-*itemset* yang subset-nya berisi k-1 item dan tidak termasuk dalam pola frekuensi tinggi dengan panjang k-1.
2. Penghitungan *support* dari tiap kandidat k-*itemset* Support dari tiap kandidat k-*itemset* didapat dengan men-scan database untuk menghitung jumlah transaksi yang memuat semua item di dalam kandidat k-*itemset* tersebut. Ini adalah juga ciri dari algoritma apriori yaitu diperlukan penghitungan dengan scan seluruh database sebanyak k-*itemset* terpanjang.
3. Tetapkan pola frekuensi tinggi yang memuat k item atau k-*itemset* ditetapkan dari kandidat k-*itemset* yang *support*-nya lebih besar dari minimum *support*. Kemudian dihitung *confidence* masing-masing kombinasi item. Iterasi akan berhenti ketika semua item telah dihitung sampai tidak ada kombinasi item lagi (Pramudiono, 2007)

Secara ringkas algoritma apriori sebagai berikut:

```
Create L1 = set of supported itemsets of cardinality one
Set k to 2
while ((k-1) != 0)
{
    Create Ck from Lk-1
    Prune all the itemsets in Ck that are not
    supported, to create Lk
    Increase k by 1
}
The set of all supported itemsets is L1 U L2 U . . . U Lk
```

Selain algoritma apriori, terdapat juga algoritma lain seperti *FP-Grwoth*. Perbedaan algoritma *apriori* dengan *FP-Growth* pada banyaknya scan *database*. Algoritma *apriori* melakukan scan *database* setiap kali iterasi sedangkan algoritma *FP-Growth* hanya melakukan sekali di awal (Bramer, 2007).

Berdasarkan beberapa pengertian di atas studi kasus dari *Assosiation Rule* adalah sebuah studi pada suatu tempat yang ditentukan untuk melihat hubungan antara *antecedent* dan *consequent* yang ingin diketahui dengan menggunakan algoritma yang ditentukan.

1.2.4 Rapid Miner

RapidMiner merupakan perangkat lunak yang bersifat terbuka (*open source*). *RapidMiner* adalah sebuah solusi untuk melakukan analisis terhadap data mining, *text mining* dan analisis prediksi. *RapidMiner* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *RapidMiner* memiliki kurang lebih 500 operator data *mining*, termasuk operator untuk input, output, data *preprocessing* dan visualisasi. *RapidMiner* merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin data *mining* yang dapat diintegrasikan pada produknya sendiri. *RapidMiner* ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi.

RapidMiner sebelumnya bernama YALE (*Yet Another Learning Environment*), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh RalfKlinkenberg, Ingo Mierswa, dan Simon Fischer di *Artificial Intelligence Unit* dari *University of Dortmund*. *RapidMiner* didistribusikan di bawah lisensi AGPL (*GNU Affero General Public License*) versi 3. Hingga saat ini telah ribuan aplikasi yang dikembangkan menggunakan *RapidMiner* di lebih dari 40 negara. *RapidMiner* sebagai *software open source* untuk data *mining* tidak perlu diragukan lagi karena *software* ini sudah terkemuka di dunia. *RapidMiner* menempati peringkat pertama sebagai *Software data mining* pada polling oleh KDnuggets, sebuah portal data-mining pada 2010-2011.

2. METODOLOGI

Metode penelitian yang diterapkan pada penelitian ini adalah Deskriptif Kuantitatif. Metode deskriptif studi kasus ini mengemukakan masalah dengan mengumpulkan data dan menyajikan data terhadap suatu objek penelitian. Dalam penelitian ini penulis melakukan analisa terhadap data *questioner* nasabah yang tersimpan dalam *database core system* di PT Prudential Life Assurance. Selain data itu juga dikombinasikan dengan master data polis dan *log* aktifitas polis. Semua data tersebut diproses dengan bantuan *RamitMiner*. Tujuan dari penelitian dengan metode deskriptif kuantitatif ini adalah untuk membuat deskripsi atau gambaran secara sistematis, faktual, dan akurat mengenai fakta-fakta, sifat-sifat serta hubungan antara fenomena yang diselidiki untuk mengambil kesimpulan dari pembahasan yang dilakukan.

Dalam penelitian ini penulis ingin mengetahui hubungan antara tertanggung, produk asuransi yang dipilih dan perilaku nasabah setelah polis dibeli. Adapun tahapannya adalah:

1. Data yang diperoleh dari dataabse perusahaan pertama kali di hilangkan dari *noise*, data yang tidak konsisten, atau data yang tidak relefan dengan hipotesa data mining yang akan dilakukan.
2. Karena data yang diperoleh tidak hanya dari satu file maka tahap selanjutnya adalah, menggabungkan data database questioner, master polis, dan log aktifitas polis kedalam satu databse baru.
3. Memilih atribut apa yang akan diproses pada data mining.
4. Menggubah data atau digabung ke dalam format yang sesuai untuk proses data mining.
5. Melakukan proses terhadap data yang melewati 4 tahap sebelumnya dengan metode/rule yang sudah ditentukan.
6. Melakukan evaluasi apakah hasil yang didapat telah sesuai dengan yang kaidah-kaidah yang diinginkan atau belum.

3. ANALISA DAN PERANCANGAN SISTEM

3.1 Analisis Data Mining

Dalam penelitian ini akan dicari nilai *support* dan *confidence* dari hubungan latar belakang pembelian polis dengan melihat tertanggung, pembelian produk dan perilaku nasabah. Sedangkan perilaku dapat dikategorikan sebagai tindakan nasabah seperti pembatakan polis sebelum 30 hari, penarikan dana sebagian dan penutupan polis.

Data yang terlalu acak akan membuat proses mining memakan waktu lama dan tingkat hubungannya rendah, maka yang akan diproses mining meliputi:

1. Hubungan tertanggung dengan pembelian produk polis
2. Hubungan tertanggung, produk yang dipilih dan pemeblian polis baru.
3. Hubungan tertanggung, produk yang dipilih dengan penutupan polis sebelum 30 hari
4. Hubungan tertanggung, produk yang dipilih dengan penarikan penarikan dana sebagian
5. Hubungan tertanggung, produk yang dipilih dengan penutupan polis

3.2 Analisis Sumber Data

Data yang digunakan dalam penelitian ini terdiri dari tiga sumber data, yaitu data *questioner* nasabah, master polis dan data *log* aktifitas polis. Data *Questioner* Nasabah adalah data isian nasabah yang harus diisi oleh nasabah saat mengajukan permohonan pembuatan polis asuransi. Atribut yang ada dapat dilihat pada tabel 1 di bawah ini.

Tabel 1 Data *Questioner* Nasabah

Atribut	Keterangan
<i>CHDRNUM</i>	No Polis
<i>LIFCNUM</i>	No Nasabah
<i>LIFE</i>	Kode Life
<i>UWQSDATA</i>	Data Jawaban U/W QUESTIONNAIRE
<i>UWQSM01A</i>	Tinggi Badan
<i>UWQSM02A</i>	Berat Badan

Atribut	Keterangan
<i>UWQSM03A</i>	Flag ID
<i>UWQSM04A</i>	Kode Ketertarikan
<i>UWQSM05A</i>	Type Perlindungan
<i>UWQSM06A</i>	Pekerjaan Pasangan
<i>UWQSM07A</i>	Pekerjaan Pasangan
<i>UWQSM08A</i>	Pasangan diatas 4m
<i>UWQSM09A</i>	Type Kunjungan Sosial
<i>UWQSM10A</i>	Pass Kunjungan Sosial

Data Master Polis adalah data Polis yang dimiliki nasabah setelah pengajuan nasabah disetujui oleh perusahaan. Atribut yang ada dapat dilihat pada tabel 2 berikut.

Tabel 2 Data Master Polis

Atribut	Keterangan
<i>CHDRNUM</i>	NO POLIS
<i>CLNTNUM</i>	NO NASABAH PEMEGANG POLIS
<i>CHDRTYPE</i>	KODE PRODUK
<i>HISSDTE</i>	TANGGAL MULAI BERLAKU POLIS
<i>STATCODE</i>	STATUS POLIS
<i>BILLODATE</i>	TANGGAL PENAGIHAN
<i>PAIDTODATE</i>	TANGGAL JATUH TEMPO

Data *log* aktifitas Polis adalah data semua aktifitas polis dari sejak diterbitkan sampai polis tersebut habis masa pertanggungannya. Atribut yang ada dapat dilihat pada Tabel 3 berikut ini.

Tabel 3 Data Log Aktivitas Polis

Atribut	Keterangan
<i>CHDRPFX</i>	<i>CONT HEADER PREFIX</i>
<i>CHDRCOY</i>	<i>CONT HEADER COMPANY</i>
<i>CHDRNUM</i>	<i>CONTRACT NUMBER</i>
<i>RECODE</i>	<i>RECORD CODE</i>
<i>TRANNO</i>	<i>TRANSACTION NUMBER</i>
<i>PTRNEFF</i>	<i>PTRN EFFECTIVE DATE</i>
<i>TRDT</i>	<i>TRANSACTION DATE</i>
<i>TRTM</i>	<i>TRANSACTION TIME</i>
<i>TERMID</i>	<i>TERMINAL ID</i>
<i>USER</i>	<i>USER NUMBER</i>
<i>BATCPFX</i>	<i>BATCH</i>
<i>BATCCOY</i>	<i>BATCH</i>
<i>BATCBRN</i>	<i>BATCH</i>

Atribut	Keterangan
<i>BATCTYR</i>	<i>BATCH</i>
<i>BATCTMN</i>	<i>BATCH</i>
<i>BATCTRCDE</i>	<i>BATCH TRANSACTION</i>
<i>BATCBATCH</i>	<i>BATCH BATCH</i>
<i>PRTFLG</i>	<i>PRINT FLAG</i>
<i>VALIDFLAG</i>	<i>VALID FLAG</i>
<i>USRPRF</i>	<i>USER PROFILE</i>
<i>JOBNM</i>	<i>JOB NAME</i>
<i>DATIME</i>	<i>TIMESTAMP</i>
<i>DATESUB</i>	<i>DATE SUBMITTED</i>

Dalam penelitian ini diasumsikan bahwa data yang diambil sudah berupa tabel-tabel dalam satu server. Untuk proses mining, data questioner, master polis dan log aktifitas polis digabungkan dengan primary key CHDRNUM. Setelah itu baru dilakukan proses mining. Proses integrasi data dilakukan ketika proses ETL (*ekstract, transform, and Load*) ketika membangun data warehouse, dalam proses ETL data dalam data source digabungkan menjadi satu dalam data warehouse dengan key CHDRNUM atau CLNTNUM.

3.3 Penggunaan Algoritma

Algoritma Apriori dan FP-Growth adalah algoritma yang paling terkenal untuk menemukan pola frekuensi tinggi. Pola frekuensi tinggi adalah pola-pola item di dalam suatu database yang memiliki frekuensi atau support di atas ambang batas tertentu yang disebut dengan istilah minimum *support* atau *threshold*. *Threshold* adalah batas minimum transaksi. Jika jumlah transaksi kurang dari *threshold* maka item atau kombinasi item tidak akan diikuti dalam perhitungan selanjutnya.

Penggunaan *threshold* dapat mempercepat proses perhitungan. Algoritma apriori dibagi menjadi beberapa tahap yang disebut iterasi. Tiap iterasi menghasilkan pola frekuensi tinggi dengan panjang yang sama, dimulai dari pass pertama yang menghasilkan pola frekuensi tinggi dengan panjang satu. Pada iterasi pertama ini, support dari setiap item dihitung dengan men-scan *database*. Setelah *support* dari setiap item didapat, item yang memiliki *support* lebih besar dari minimum *support* dipilih sebagai pola frekuensi tinggi dengan panjang 1 atau sering disingkat 1-itemset. Singkatan k-itemset berarti satu set yang terdiri dari k item. Tahapan-tahapan pengolahan *data mining*, contoh data masuk dapat dilihat pada Tabel 4 di bawah ini.

Tabel 4 Data Awal

No Polis	Tertanggung	Kode Produk	Prilaku Nasabah
A0000001	Diri Sendiri	U1B	Beli polis baru
A0000002	Diri Sendiri	U1B	Topup
A0000003	Ayah	U1H	Beli polis baru
A0000004	Ibu	U1B	Freelook

A0000005	Istri	U1H	Freelook
A0000006	Diri Sendiri	U1H	Beli polis baru
A0000007	Diri Sendiri	U1B	Beli polis baru
A0000008	Ibu	U1B	Surrender
A0000009	Ibu	U1B	Surrender
A0000010	Diri Sendiri	U1H	Freelook

Dari data awal tersebut didapat kandidat pertama (C1) seperti pada Tabel 5.

Tabel 5 Kandidat Pertama (C1)

<i>ItemSet</i>	<i>Count</i>
Diri Sendiri	5
Ayah	1
Ibu	3
Istri	1
U1B	6
U1H	4
<i>ItemSet</i>	<i>Count</i>
Beli polis baru	4
<i>Top up</i>	1
<i>Freelook</i>	3
<i>Surrender</i>	2

Ditetapkan *threshold* = 3, maka kandidat yang nilainya kurang dari 3 akan dihapus. Sehingga, didapat hasil seperti pada Tabel 6 (L1).

Tabel 6 Hasil Setelah *Threshold* Ditetapkan (L1)

<i>ItemSet</i>	<i>Count</i>
Diri Sendiri	5
Ibu	3
U1B	6
U1H	4
Beli polis baru	4
<i>Freelook</i>	3

Setelah iterasi ke 2 didapat kandidat kedua (C2) seperti pada tabel 7.

Tabel 7 Kandidat Kedua (C2)

<i>ItemSet</i>	<i>Count</i>
Diri Sendiri,U1B	3

Diri Sendiri, U1H	2
Ibu, U1B	3
Ibu, U1H	0

Setelah iterasi ke 3 didapat hasil seperti pada Tabel 8.

Tabel 8 Hasil Iterasi Ke-3

<i>ItemSet</i>	<i>Count</i>
Diri Sendiri, U1B, Beli Polis baru	2
Diri Sendiri, U1H, Beli Polis baru	1
Diri Sendiri, U1B, <i>Freelook</i>	0
Diri Sendiri, U1H, <i>Freelook</i>	1
Ibu, U1B, Beli Polis Baru	0
Ibu, U1B, <i>Feelook</i>	0
Ibu, U1H, Beli Polis Baru	0
Ibu, U1H, <i>Feelook</i>	0
Diri Sendiri, U1B, Beli Polis baru	2
Diri Sendiri, U1H, Beli Polis baru	1

Berdasarkan Tabel 7 dan 8 di atas, jika dimasukkan ke dalam persamaan *Convidence* dan *Support*, dapat diambil hasil sebagai berikut:

- $$S = \frac{\sum(Ta+Tc)}{\sum(T)}$$

$$\text{Support Diri Sendiri, U1B, Beli Polis baru} = \frac{\text{count(DiriSendiri,U1B,BeliPolis baru)}}{\text{jumlah transaksi}} = \frac{2}{10}$$
- $$C = \frac{\sum(Ta+Tc)}{\sum(Ta)}$$

$$\text{Confidence Diri Sendiri, U1B, Beli Polis} = \frac{\text{count(Diri Sendiri,U1B,Beli Polis baru)}}{\text{count (Diri Sendiri, U1B)}} = \frac{2}{3}$$
- $$S = \frac{\sum(Ta+Tc)}{\sum(T)}$$

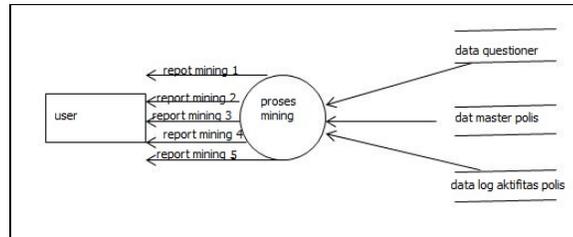
$$\text{Support Diri Sendiri, U1H, Beli Polis} = \frac{\text{count(DiriSendiri,U1H,BeliPolis baru)}}{\text{jumlah transaksi}} = \frac{1}{10}$$
- $$C = \frac{\sum(Ta+Tc)}{\sum(Ta)}$$

$$\text{Confidence Diri Sendiri, U1H, Beli Polis} = \frac{\text{count(Diri Sendiri,U1H,Beli Polis baru)}}{\text{count (Diri Sendiri, U1H)}} = \frac{1}{2}$$

3.4 Pemodelan Fungsi

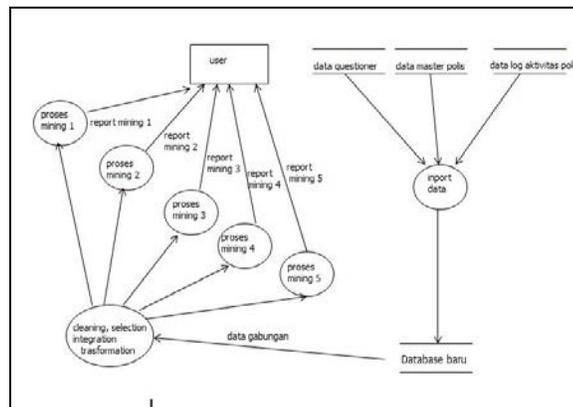
Pemodelan fungsi digambarkan dengan *Data Context Diagram (DCD)*, *Data Flow Diagram (DFD)*. Gambar 2 menunjukkan DCD/DFD level-0 pada Aplikasi Data Mining yang

terdiri dari 3 input dan 4 output. *External entity* berupa pengguna atau *user* dan tiga database yaitu database *questioner*, master polis dan logaktivitas polis.



Gambar 2 Data Flow Diagram Level 0

Gambar 3 berikut merupakan DFD Level-1 dari Aplikasi Data Mining yang dipecah menjadi beberapa proses kecil guna menjelaskan fungsi-fungsi dan arus data yang mengalir pada Aplikasi Data Mining.



Gambar 3 Data Flow Diagram Level 1

Berikut proses-proses yang terdapat pada Aplikasi *Data Mining*:

1. *Import Data*, yaitu proses *load* data dari *database* ke dan *database questioner*, master polis dan data *log* aktifitas polis ke database baru. Semua data akan dimasukkan tanpa adanya proses filtering.
2. *Cleaning, Integrasi, Selection, dan Transformasi*
 - a. *Cleaning* data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data yang tidak relevan.
 - b. Integrasi data (*data integration*) merupakan penggabungan data dari berbagai database ke dalam satu *database* baru. *questioner*, data master polis maupun data log aktivitas polis disimpan dalam satu *database*.
 - c. *Selection* data adalah proses menyeleksi atribut apa yang akan diproses pada mining selanjutnya.
 - d. Transformasi data merupakan proses mengubah data atau digabungkan ke dalam format yang sesuai untuk diproses dalam *data mining*.
3. Proses *mining* produk merupakan proses *mining* untuk mengetahui hubungan pemilihan produk dengan tertanggung

4. Proses mining T679, merupakan proses *mining* untuk mengetahui hubungan tertanggung, pemilihan produk dengan penambahan *Top Up*.
5. Proses *mining* T512 (*surrender*), merupakan proses *mining* untuk mengetahui hubungan tertanggung, pemilihan produk dengan penutupan polis.
6. Proses *mining* T510 (*partial withdraw*) merupakan proses *mining* untuk mengetahui hubungan tertanggung, pemilihan produk dengan pengambilan dana sebagian.
7. Proses *mining* TI46 (*freelook*) merupakan proses *mining* untuk mengetahui hubungan tertanggung, pemilihan produk dengan penutupan polis setelah review 1 bulan.

3.6 Implementasi

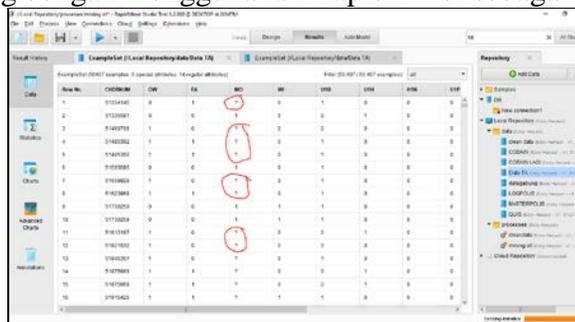
3.6.1 Data Input Tools

Data hasil ETL dalam *database* didapat 3 *file* yaitu *questioner*, Master polis, dan *log* aktifitas polis. Ketiga *file* itu kemudian diseleksi item yang akan dipakai proses *data mining*, dilakukan *cleansing*, integrasi serta transformasi data sehingga siap untuk diproses oleh RapidMiner. Berikut Gambar 4 yang berisi sumber data yang akan dioleh:

Row No.	CHDRNUM	OW	FA	MO	WI	UTB	UTH	HEN	UTP
1	51334145	0	1	?	?	1	0	0	0
2	51335581	0	0	1	0	0	1	0	0
3	51489706	1	0	?	0	0	0	0	0
4	51485392	1	1	?	0	1	0	0	0
5	51485392	1	1	?	0	1	0	0	0
6	51593085	0	0	1	0	1	0	0	0
7	51608828	1	0	?	0	1	0	0	0
8	51623898	1	1	?	0	1	0	0	0
9	51730259	0	0	1	1	1	0	0	0
10	51730259	0	0	1	1	1	0	0	0
11	51813187	1	0	?	0	0	1	0	0
12	51821632	1	0	?	0	0	0	0	0
13	51845357	1	0	?	0	1	0	0	0
14	51875699	1	1	?	0	0	1	0	0
15	51875699	1	1	?	0	0	1	0	0
16	51915425	1	1	?	1	1	0	0	0

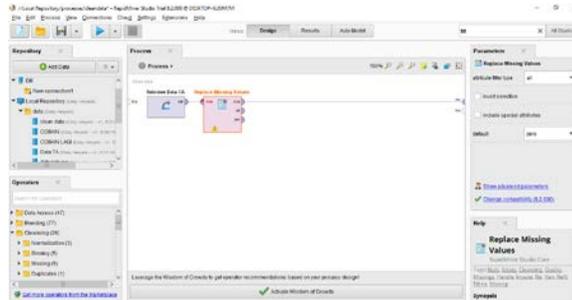
Gambar 4 Input Data RapidMiner

Setelah diperiksa ternyata masih ada *field* yang memuat data yang kosong, sehingga akan dilakukan proses *cleansing* dengan menggunakan RapidMiner sebagaimana Gambar 5 berikut.



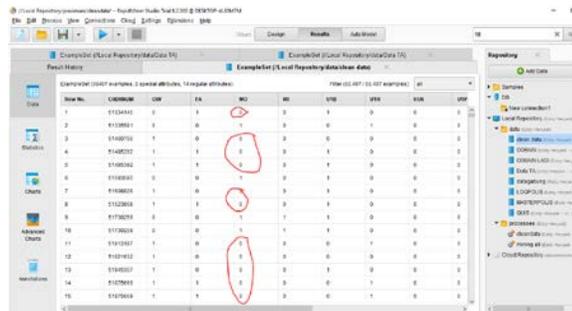
Gambar 5 Data Missing

Proses *cleansing* di RapidMiner dapat dilihat pada Gambar 6 berikut.



Gambar 6 Proses Pembetulan Data Missing

Setelah di-run didapat hasil sudah tidak ada *missing* data lagi. Hasil *cleansing* dapat dilihat pada Ggaambar 7 di bawah ini.

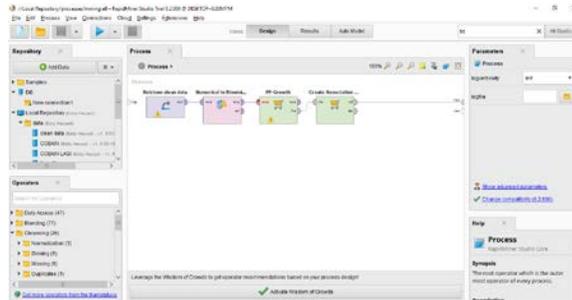


Gambar 7 Data Input Yang Sudah Clean

3.6.2 Proses Data Mining

Urutan proses untuk data mining ini bisa dilihat sebagai berikut:

1. Seperti pada Gambar 8 Proses *Data Mining* di RapidMiner pertama kali yaitu melakukan *retrive* file yang berisi data ta yang sudah bersih, kemudian *convert* numerik ke binominal, kemudia algoritma apriori dan terakhir *create assosiation rule*



Gambar 8 Proses Data Mining

Adapun hasil yang diperoleh dapat dilihat pada Gambar 9 sebagai berikut:

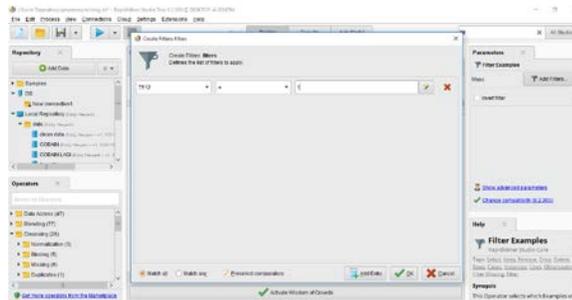
No	Premise	Consequence	Support	Confidence	LiftRatio	Gain	p-Value	LR
3	FA_T912	OP	0.003	0.999	0.999	-0.014	0.999	3.9
4	FA_T91	OP	0.003	0.997	0.999	-0.011	0.999	3.9
5	U98_51	OP	0.007	0.997	0.999	-0.012	0.999	1.9
6	U98	OP	0.447	0.997	0.992	-0.073	0.919	1.9
7	FA	OP	0.020	0.998	0.975	-0.028	0.996	1.9
8	T91	OP	0.004	0.994	1.009	-0.008	0.999	1.9
9	U98_T91	OP	0.003	0.993	1.009	-0.013	0.999	1.9
10	U98	OP	0.108	0.997	0.993	-0.027	0.919	1.9
11	FA_U98	OP	0.006	0.970	0.997	-0.009	0.999	1.9
12	T912	OP	0.006	0.999	1.009	-0.008	0.999	1.9
13	U98_T912	OP	0.003	0.924	1.009	-0.011	0.999	1.9
14	U98_51_51	OP	0.004	0.933	1.009	-0.011	0.999	1.9
15	FA_T912	OP	0.001	0.948	1.009	-0.011	0.999	1.9
16	U98_T912	OP	0.004	0.996	1.009	-0.011	0.999	1.9
17	U98_T912	OP	0.002	0.995	1.009	-0.012	0.999	1.9

Gambar 9 Hasil Proses Data Mining

Agar menghasilkan laporan-laporan *data mining* yang diinginkan, dipecah menurut perilaku nasabah dan harus ditambahkan filter data sebelum proses data mining. Penambahan filter tersebut bisa dilihat pada Gambar 10 dan 11 di bawah ini.



Gambar 10 Penambahan Filter



Gambar 11 Kondisi Filter Proses Data Mining

Contoh hasil proses *data mining* setelah penambahan filter seperti di Gambar 12.

No	Premise	Consequence	Support	Confidence	LiftRatio	Gain	p-Value	LR
2	FA	OP	0.075	0.999	0.992	-0.193	0.999	1.9
3	FA_T912	OP	0.079	0.949	0.992	-0.193	0.999	1.9
4	T912_51	OP	0.076	0.949	0.992	-0.193	0.999	1.9
5	U98	OP	0.277	0.998	0.992	-0.226	0.999	1.9
6	U98_T912	OP	0.077	0.998	0.992	-0.226	0.999	1.9
7	T912_51	OP	0.077	0.998	0.992	-0.226	0.999	1.9
8	U98_51	OP	0.173	0.949	0.949	-0.193	0.999	1.9
9	U98	OP	0.174	0.949	0.949	-0.193	0.999	1.9
10	U98_T912	OP	0.174	0.949	0.949	-0.193	0.999	1.9
11	T912_U98	OP	0.174	0.949	0.949	-0.193	0.999	1.9
12	U98_51	OP	0.146	0.971	0.949	-0.179	0.999	1.9
13	OP	OP	0.020	1	1	-0.020	0	1
14	U98	T912	0.020	1	1	-0.020	0	1
15	U98	T912	0.020	1	1	-0.020	0	1
16	U98	T912	0.020	1	1	-0.020	0	1
17	U98	T912	0.020	1	1	-0.020	0	1

Gambar 12 Contoh Report Data Mining

4. PENUTUP

4.1 Kesimpulan

Kesimpulan yang dapat diambil dalam penelitian ini adalah proses *Data Mining* dapat digunakan untuk menampilkan informasi hubungan antara produk yang dipilih, tertanggung, dan perilaku nasabah PT. Prudential Life Assurance sebagai berikut:

1. Keterkaitan hubungan antara tertanggung dan pemilihan produk. Untuk nilai tertanggung diri sendiri dan produk UIB memiliki tingkat *Confidence* dan *Support* tertinggi.
2. Hubungan Tertanggung, pemilihan produk dan perilaku positif nasabah Asuransi Prudential dengan polis yang dibelinya. Hasil yang didapatkan kecenderungan polis yang dibeli untuk diri sendiri mempunyai perilaku positif dibandingkan dengan untuk orang tua.
3. Hubungan Tertanggung, pemilihan produk, dan perilaku negatif nasabah Asuransi Prudential dengan polis yang dibelinya. Hasil yang didapatkan kecenderungan polis yang dibeli untuk orang tua mempunyai perilaku negatif dibandingkan dengan untuk diri sendiri.

4.2 Saran

Untuk pengembangan Aplikasi *Data Mining* yang lebih baik, maka perlu pengembangan lebih lanjut dengan menggunakan jumlah data nasabah yang lebih banyak dan periode waktu yang lebih lama. Hubungan yang diteliti dapat bervariasi dengan item lain, misalnya perilaku nasabah dalam pengajuan claim polis, baik kesehatan, kecelakaan, maupun meninggal dunia dihubungkan dengan berat badan/tinggi badan, kebiasaan merokok, dan sebagainya.

DAFTAR PUSTAKA

- Al Fata, Hanif Universitas Amikom, (2007),” Analisis dan Perancangan Sistem Informasi untuk Keunggulan Bersaing Perusahaan dan Organisasi Modern”, Penerbit Andi, Jogjakarta
- Bramer, Max, (2007), “Principles of Data Mining”, Springer, London
- Elmasri, Ramez & Shamkant B. Navathe, (2000), “*Fundamentals of Database Systems. Third Edition*”, Addison Wesley Publishing Company, New York.
- Han, J. & Kamber, M, (2006), “*Data Mining Concepts and Techniques Second Edition*”. Morgan Kauffman, San Francisco.
- Kadir, Abdul, (1999), “Konsep dan Tuntunan Praktis Basis Data”, Penerbit Andi, Yogyakarta.
- Kotler, Philip & Lan Keller, Kevin, (2008). *Manajemen Pemasaran*, Indonesia, PT Macanan Jaya Cemerlang,
- Rainardi, Vincent, (2008), “*Building a Data Warehouse with Examples in SQL Server*”, Springer, New York

- Santosa, Budi, (2007), “Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis”, Graha Ilmu, Yogyakarta
- Setiawan, Awan & Margono, Kuntjahjo SL, (2015), “*Predicting Success Study Using Students GPA Category*”, The Asian Journal of Technology Management, Vol. 8, No. 1, Institut Teknologi Bandung, School of Business and Management
- Setiawan, Awan & Rusmawan, Dadan, (2017), “*Graduate Rate Analysis of Student Using Data Mining and Algorithm Apriori*”, International Journal of Soft Computing, Vol. 12, No. 5, p287-293
- Suryana, Ase., Yulianto, Erwin., Pratama, & Khrisna Dea, (2017), “Perancangan Sistem Pendukung Keputusan Penilaian Prestasi Pegawai Terbaik Menggunakan Metode SAW, AHP & TOPSIS”, Jurnal Ilmiah Teknologi Informasi Terapan (Jitter), Vol. 3, No. 2, p130-139, Universitas Widyatama, Bandung
- Therling K. (2006). “*An Introduction to DataMining: Discovering hidden value in your data warehouse*”, www.thearling.com, diakses tanggal 15 April 2018