

**ANALISIS DATA MINING UNTUK DENDA KETERLAMBATAN PENGIRIMAN
BATUBARA DENGAN PENDEKATAN METODE *DECISION TREE C5***
(Studi kasus: Aplikasi Batubara Online (BBO) – PT. Indonesia Comnets Plus)

Oleh :
ACEP SOPIAN FIRMANSYAH
NPM : 2015210082

Bandung, 12 Agustus 2017
Menyetujui,

Dr. Djajasukma Tjahjadi, S.E., MT.
Pembimbing

**PROGRAM STUDI PASCASARJANA
MAGISTER SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER LIKMI BANDUNG
2017**

ABSTRAK

ANALISIS DATA MINING UNTUK DENDA KETERLAMBATAN PENGIRIMAN BATUBARA DENGAN PENDEKATAN METODE *DECISION TREE C5*

(Studi kasus: Aplikasi Batubara Online (BBO) – PT. Indonesia Comnets Plus)

Oleh :

ACEP SOPIAN FIRMANSYAH

NPM : 2015210082

Perkembangan teknologi telah memberikan pengaruh yang sangat besar di dalam kehidupan masyarakat, dan memiliki peran penting dalam pendorong lahirnya inovasi-inovasi baru. Terdapat banyak perusahaan yang menjadi kompetitor pemasok batubara ke PT PLN. Hal tersebut menimbulkan persaingan bisnis mitra yang ingin memasok batubara ke PT PLN. Untuk menghadapi persaingan bisnis tersebut PT Indonesia Comnets Plus (ICON+) sebagai anak perusahaan, dituntut agar memberikan pelayanan dan informasi akurat, yang kedepannya data tersebut menjadi bahan pertimbangan dan referensi PT PLN untuk menentukan kebijakan, yang dapat memaksimalkan segala aktivitas yang berhubungan pengiriman batubara. Penelitian ini melakukan analisa *data mining*, dengan membangun model pohon keputusan atau sering di sebut *decision tree c5.0* yang dapat menghasilkan aturan-aturan untuk mengidentifikasi keterlambatan pengiriman batubara.

Sumber data yang digunakan dalam penelitian ini berasal dari aplikasi Batubara Online (BBO) di PT ICON+, data tersebut berupa data transaksi mulai dari penjadwalan sampai pembongkaran. Metode *decision tree c5.0* memilih beberapa tahapan integrasi data, menghilangkan reduksi data, membersihkan data yang kosong, melakukan tranformasi data, kemudian dibuat model klasifikasi algoritma C.5.0 sampai evaluasi dan validasi hasil.

Atribut-atribut yang sangat berpengaruh dalam keterlambatan batubara adalah tanggal tiba batubara, spek yang digunakan, wilayah pengiriman batubara, volume batubara yang dikirim, sampai jumlah konfirmasi yang harus dipenuhi mitra setiap bulannya. Atribut yang berpengaruh cukup besar adalah tanggal tiba batubara ke PLTU.

Kata Kunci : *Data Mining*, model klasifikasi, pengiriman batubara, pohon keputusan, *decision tree c5.0*, keterlambatan batubara, denda keterlambatan pengiriman batubara.

ABSTRACT

DATA MINING ANALYSIS FOR THE DELIVERY OF COAL BY APPROACH TO DECISION TREE C5 METHOD

(Case Study: Aplikasi Batubara Online (BBO) – PT. Indonesia Comnets Plus)

By :

ACEP SOPIAN FIRMANSYAH

NPM : 2015210082

Technological developments have had a huge impact on people's lives, and have an important role in driving new innovations. There are many companies that compete coal suppliers to PT PLN. This creates competition for business partners who want to supply coal to PT PLN. To face the business competition PT Indonesia Comnets Plus (ICON +) as a subsidiary, is required to provide accurate service and information, which in the future the data becomes the consideration and reference of PT PLN to determine the policy, which can maximize all activities related to coal delivery. This research performs data mining analysis, by constructing decision tree model or often called decision tree c5.0 which can generate rules to identify the delay of delivery of coal.

Source of data used in this research comes from BBO application in PT ICON +, the data in the form of transaction data from scheduling until demolition. The decision tree c5.0 method selects several stages of data integration, eliminates data reduction, clears empty data, performs data transformation, and then creates a model classification algorithm C.5.0 until evaluation and validation of results.

The most influential attributes in coal delay are the arrival date of the coal, the spec used, the coal delivery area, the volume of coal delivered, to the amount of confirmations that the partner must meet each month. The attributes that are of considerable influence are the date of arrival of coal to the steam power plant

Keywords : Data Mining, classification model, coal delivery, decision tree, decision tree c5.0, coal delay, penalty delay of coal delivery.

KATA PENGANTAR

Puji syukur ke hadirat Tuhan YME, atas segala karuniannya sehingga tesis ini dapat terselesaikan dengan baik. Penulis juga mengucapkan terima kasih kepada Bapak Dr.Djajasukma Tjahjadi, S.E., M.T atas segala bimbingannya.

Pada kesempatan ini penulis ingin mengucapkan terima kasih banyak kepada :

1. Ibu Hera Episcia Amban selalu SPV SBS Ketenagalistrikan ICON+ yang telah memberikan kesempatan dan mendukung penelitian.
2. Bapak Wachyu Setiyono selaku PLH SPV SBS Ketenagalistrikan ICON+ yang telah memberikan dukunganya dan ijinnya jika ada keperluan kampus.
3. Para Staf dan rekan kerja di PT.ICON+
4. Orang tua yang sudah mendukung dan memberikan semangat.
5. Rekap-rekan Program Magister Sistem Informasi atas doa dan dukungannya.
6. Teman-teman dan sahabat yang sudah memberikan dukungan dan semangatnya.
7. Dan semua pihak yang tidak bisa disebutkan satu persatu.

Kritik dan saran sangat diharapkan demi perbaikan-perbaikan selanjutnya. Akhir kata, semoga Tesis ini dapat memberikan manfaat dan menambahkan pengetahuan bagi kita semua.

Daftar Isi

ABSTRAK.....	ii
ABSTRACT	iii
KATA PENGANTAR.....	iv
Daftar Isi	v
Daftar Tabel.....	viii
Daftar Gambar.....	ix
Daftar Lampiran.....	x
BAB I PENDAHULUAN	11
1.1 Latar Belakang	11
1.2 Rumusan Masalah	15
1.3 Tujuan Penelitian	15
1.4 Ruang Lingkup Penelitian	15
BAB II TINJAUAN PUSTAKA.....	16
2.1 Data Mining	16
2.2 Klasifikasi	16
2.3 Tahap-Tahap <i>Data Mining</i>	17
2.4 <i>Decision Tree</i> (Pohon Keputusan).....	18
2.5 Algoritma C5.0	19
2.6 Evaluasi Model Klasifikasi.....	22
2.7 Analisis Data	23
2.7.1 Macam – Macam Analisis Data	23
2.7.2 Tujuan Analisis Data	24
2.7.3 Langkah dan Prosedur <i>Data Mining</i>	24
2.8 Penelitian Sebelumnya	25

2.8.1	Analisis Data Keterlambatan bahan baku berdasarkan pendekatan data <i>wirehouse</i> dan <i>decision tree</i> C5.0	25
2.8.2	Data Mining Jasa Pengiriman Titipan Kilat di PT Citra Van Titipan Kilat (TIKI) dengan metode Decision tree	25
2.8.3	<i>Flight Delay Prediction</i>	26
2.8.4	Penggunaan <i>Educational Data Mining</i> (EDM) Untuk Mendeteksi Keterlambatan Masa Studi Mahasiswa (Studi Kasus Udinus).....	27
BAB III OBJEK DAN METODE PENELITIAN		28
3.1	Latar Belakang Perusahaan.....	28
3.2	Visi dan Misi PT INDONESIA COMNETS PLUS (ICON+).....	29
3.3	Nilai-nilai Perusahaan	29
3.3.1	<i>INTEGRITY</i>	29
3.3.2	<i>CARE</i>	29
3.3.3	<i>OPEN MIND</i>	29
3.3.4	<i>INNOVATION</i>	30
3.3.5	<i>TEAM WORK</i>	30
3.3.6	<i>EXCELLENCE</i>	30
3.4	Objek Penelitian	30
3.4.1	Proses Bisnis Pegiriman	31
3.5	Metode Penelitian	36
3.5.1	Identifikasi Masalah	36
3.5.2	Praproses Data	36
3.5.3	Pembuatan Model Klasifikasi C5.0	37
3.5.4	Pengujian Model Klasifikasi C5.0	37
3.5.5	Evaluasi Dan Validasi Hasil	37
3.6	Lingkungan Penelitian.....	37
BAB IV IMPLEMENTASI PENELITIAN		38
4.1	Rencana Penelitian	38
4.1.1	Pemahaman dari sisi bisnis	38

4.1.2	Pemahaman dari sisi data	38
4.2	Data Collection.....	38
BAB V KESIMPULAN DAN SARAN.....		48
5.1	Kesimpulan	48
5.2	Saran.....	48
DAFTAR PUSTAKA		50
LAMPIRAN		52

Daftar Tabel

Tabel 1.1 Evaluasi dua kelas prediksi (Fawcett 2006)	23
Table 4.1 Data Master Mitra Pemasok Batubara	40
Table 4.2 Data Master Pembangkit	40
Table 4.3 Data Jadwal Konfirmasi Rakor	40
Table 4.4 Data Jadwal Pengiriman	40
Table 4.5 Data Pembongkaran	40
Table 4.6 Data Master Spek	41
Table 4.7 Data Master Moda	41
Table Lampiran 1.9 Data Pasokan Batubara	41

Daftar Gambar

Gambar 2.1 Langkah-langkah dalam Proses Data Mining (Maimon & Rokach, 2010)	16
Gambar 3.1 Proses Bisnis Batubara (Schaffer, Agusti and Earle.2009)	31
Gambar 3.2 Flow CIF dari Rakor sampai approval dokument NOR (DITDAN. 2014)	33
Gambar 3.3 Flow CIF dari permohonan ijin sandar upload coa dan cow (DITDAN. 2014)	34
Gambar 3.4 Flow CIF dari Hitung tagihan pemasok sampai penandatanganan berita acara serah terima batubara, perhitungan dan denda keterlambatan (jika ada) (DITDAN. 2014)	34
Gambar 3.5 Flow CIF dari input tagihan sampai pembayaran (DITDAN. 2014)	35
Gambar 3.6 Tahapan Penelitian	36
Gambar 4.1 Diagram Database Setelah Transformasi	41
Gambar 4.2 Hasil <i>Decision Tree C5.0</i> Prediksi Faktor-faktor Keterlambatan	43
Gambar 4.3 Hasil <i>Decision Tree C5.0</i> Prediksi Faktor-faktor Keterlambatan Lanjutan	44
Gambar 4.5 Jumlah pengiriman batubara oleh mitra	46
Gambar 4.6 Presentase Jumlah pengiriman batubara oleh mitra	46
Gambar 4.7 Presentase Pasokan batubara Mitra	47

Daftar Lampiran

LAMPIRAN A : Pengambilan data dan Pemrosesan data awal	52
2. Table Data Master Pembangkit.....	53
3. Table Data Jadwal Konfirmasi Rakor	54
4. Table Data Jadwal Pengiriman.....	55
5. Table Data Kedatangan.....	56
6. Table Data Pembongkaran.....	57
7. Table Data Master Spek.....	58
8. Table Data Master Moda	58
9. Table Data Saldo Pasokan Batubara	59
LAMPIRAN B : Tahapan-tahapan pembuatan model prediksi	61
1. Data yang sudah di export ke microsoft excel.....	61
2. Pembuatan deklarasi <i>variable</i> di IBM SPSS Statistics	62
3. Masukan Data ke dari excel ke SPSS.....	62
4. Pemilihan <i>source</i>	63
5. Pemilihan <i>source</i> untuk di <i>import</i>	63
6. Pemilihan penentuan <i>Type Direction</i>	63
7. Hasil Penentuan <i>Type Direction</i>	64
8. Penentuan <i>field ops</i> , untuk penentuan variable yang akan di proses	64
9. Tampilan sesudah melakukan pemilihan model prediksi C5.....	64
10. Exsekusi dan pemilihan model.....	65
11. Hasil	65

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era globalisasi, perkembangan teknologi telah memberikan pengaruh yang sangat besar di dalam kehidupan masyarakat, dan memiliki peran penting dalam pendorong lahirnya inovasi-inovasi baru. Salah satu pengaruh tersebut di bidang teknologi informasi yaitu pengetahuan yang bisa di ambil dari *database*, dengan begitu sebuah organisasi atau perusahaan dapat mengidentifikasi pengetahuan yang tersimpan di *database*, sehingga membantu dalam menentukan arah perkembangan bisnis di masa yang mendatang.

Sebagai contoh dalam dunia bisnis pada aplikasi Batubara Online (BBO), setiap hari terjadinya proses transaksi batubara. Hal ini memungkinkan data yang diperoleh akan menjadi banyak dan menumpuk, jika dibiarkan maka data batubara tersebut hanya akan menjadi sebuah data biasa, diharapkan dengan adanya analisa *data mining*, perusahaan dapat memaksimalkan layanan di segala aktivitas yang berhubungan dengan proses batubara. Selain satu upaya dalam memenuhi kebutuhan tersebut adalah dengan melakukan analisa terhadap kegiatan pengiriman batubara yang di butuhkan oleh PT PLN terkait penerimaan batubara dari mitra, dengan mengambil data pengiriman batubara dari jadwal pengiriman sampai pembongkaran yang mengalami keterlambatan ataupun tidak mengalami keterlambatan pengiriman batubara. Semakin cepat pemenuhan kebutuhan tersebut, maka PT PLN dapat memaksimalkan segala aktivitas yang berhubungan dengan penerimaan batubara di masing-masing PLTU (Pembangkit Listrik Tenaga Uap) yang di bawah naungan organisasi PT PLN.

PT Indonesia Comnets Plus (ICON+) sendiri adalah anak perusahaan PLN yang berfokus pada penyediaan jaringan, jasa, dan Konten telekomunikasi, khusus untuk mendukung teknologi dan sistem informasi PT PLN (Persero) dan publik. Untuk Bussiness Process Owner (BPO) dari aplikasi batubara online (BBO), ini sendiri adalah satuan batubara PT. PLN sedangkan untuk *implementor* dan pemeliharaan aplikasi di tugaskan kepada PT.ICON+ selaku anak perusahaan.

Terdapat banyak perusahaan yang menjadi kompetitor pemasok batubara ke PT PLN. Hal tersebut menimbulkan persaingan bisnis yang ketat antara pemasok batubara yang ingin memasok batubara ke PT PLN. Untuk menghadapi persaingan bisnis tersebut PT ICON+ dituntut agar memberikan pelayanan dan informasi yang akurat tentang kinerja pemasok dalam pengiriman batubara, yang kedepannya menjadi bahan pertimbangan dan referensi PT PLN untuk menentukan kebijakan, dalam pengadaan batubara di *eprocurement* PLN, sehingga kedepannya stok cadangan batubara PLTU tidak dalam status kekurangan batubara, sehingga tidak ada pengalihan pasokan batubara. Pengalihan pasokan terjadi dikarenakan cadangan pasokan batubara di PLTU mulai menipis, jika stok cadangan batubara di PLTU habis mengakibatkan akan terjadi pemadaman lampu, baik itu pemadaman lampu bergilir atau keseluruhan. Ketersediaan bahan baku batubara ke PLTU, merupakan hal penting bagi produktifitas kerja PLTU. Maka semakin optimal penggunaan batubara yang di dimanfaatkan oleh PLTU, target dan kinerja PLTU dapat tercapai tanpa adanya hambatan oleh permasalahan yang disebabkan oleh ketersediaan batubara tersebut.

Oleh karena itu, analisis data telah menjadi kebutuhan penting untuk meningkatkan daya saing dalam organisasi bisnis atau perusahaan, pengambilan keputusan yang cenderung mengandalkan intuisi menjadi kurang relevan di tengah persaingan yang semakin kompleks, sehingga dengan demikian manajemen satuan batubara PLN harus mengambil keputusan berdasarkan fakta aktual dan tidak hanya mengandalkan intuisi dan pengalaman kuantitatif saja, kualitas perusahaan dalam pemanfaatan teknologi dapat mempengaruhi daya saing dalam berbagai aspek. Pada level manajerial maupun pihak penentu kebijakan membutuhkan informasi yang berfungsi untuk mendukung proses analisis, tidak hanya dalam kebutuhan transaksional, untuk melakukan analisis maka perusahaan perlu ngumpulkan atau harus memiliki data yang banyak secara kuantias dan baik secara kualitas, sehingga analisa yang di dihasilkan akan semakin baik, namun untuk memenuhi hasil analisis yang baik akan tergantung pada jumlah dan kualitas yang digunakan, sehingga dibutuhkan seuatu teknologi yang

dapat membantu pihak manajemen satuan batubara PLN dalam melakukan analisis terhadap kumpulan data tersebut dalam berbagai faktor.

Maka diperlukan sebuah aplikasi yang mampu mengelola data, yang tadinya hanya tumpukan data biasa, menjadi data yang bisa bermanfaat dalam proses batubara, sehingga bisa diperoleh informasi yang bermanfaat bagi penggunanya dan dimanfaatkan secara optimal. Pemanfaatan informasi dan pengetahuan yang terkandung di dalam banyaknya data tersebut, disebut dengan *data mining*. *Data mining* dimaksudkan untuk memberikan solusi nyata bagi para pengambil keputusan di dunia bisnis, untuk mengembangkan bisnis. *Data mining* adalah sebuah istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database* atau sering disebut *Knowledge Discovery in Database (KDD)* (Kusrini, 2009).

Perkembangan penggunaan data mining pada perusahaan juga saat ini dapat meningkatkan kualitas pelayanan informasi dan keakuratan data, yang diambil dari hasil pengelolaan *data mining*. Dengan meningkatnya jumlah transaksi batubara di aplikasi BBO, semakin banyak data yang terkumpul dan tersimpan, maka banyak data yang dapat di gali dan semakin banyak pengetahuan yang dapat diambil.

Penelitian-penelitian terkait yang telah dilakukan antaranya adalah Rendy (2016) yang menggunakan pendekatan data *warehouse* dan *decision tree c5.0*, untuk mengelola pembelian barang dan bahan baku. Model yang terbentuk menghasilkan aturan-aturan yang dapat membantu mengklasifikasikan atribut yang berpengaruh terhadap keterlambatan penerimaan barang, sehingga dapat penunjang dalam pengambilan keputusan yang strategis. Pada penelitian yang dilakukan oleh Ibnu Graha (2014) tujuan dari penelitian ini Klasifikasi data pengiriman titipan dengan data mining metode *decision tree* digunakan dengan tujuan memberikan rencana strategis bagi perusahaan untuk mengetahui karakteristik pasar. Sehingga dengan demikian dapat dianalisis pasar yang sudah ada ataupun menemukan peluang-peluang yang baru serta menemukan rencana strategis untuk meningkatkan keuntungan. Penelitian yang dilakukan Martinez, Vincent (2012) mencoba menjajaki berbagai model prediksi dan berbagai metode evaluasi Dengan mengukur kinerja model dengan menggunakan data nyata, dan melihat hasil

yang menarik mengenai prediktabilitas penundaan. Metode prediksi penundaan terbaik tampaknya paling spesifik yang memperhitungkan semua kombinasi parameter kategoris dan kondisi pada jam kedatangan. Kinerja model sangat menarik untuk dievaluasi, karena Berbagai ukuran yang digunakan, dan parameter yang berbeda disesuaikan. Penelitian yang dilakukan oleh Defri Kurniawan, Wibowo Wicaksono dan Yani Parti Astuti (2016) menggunakan *educational data mining* untuk mendeteksi keterlambatan masa studi mahasiswa, untuk memprediksi mahasiswa yang tepat waktu dan tidak tepat waktu agar pengelolaan akademik di dalam memetakan mahasiswa yang berpotensi mengalami keterlambatan masa studi bisa di tanggulangi sejak awal. Untuk mendapatkan solusi dari permasalahan yang di dapat ICON+ sebagai memberi sarana informasi kepada PT PLN, maka penelitian ini melakukan analisa *data mining*, dengan membangun model pohon keputusan atau sering disebut *decision tree c5* yang dapat menghasilkan aturan-aturan untuk mengidentifikasi keterlambatan pengiriman batubara, sehingga dapat dimanfaatkan dalam kegiatan analisa kinerja mitra, dalam memasok batubara ke PLTU.

Perkembangan penggunaan data mining pada perusahaan juga saat ini dapat meningkatkan kualitas pelayan informasi dan keakuratan data, yang diambil dari hasil pengelolaan *data mining*. Dalam studi penelitian ini akan diteliti tentang proses keterlambatan batubara dengan metode *decision tree c5*.

Dengan meningkatnya jumlah transaksi batubara di aplikasi BBO semakin banyak data yang terkumpul dan tersimpan, maka banyak data yang dapat di gali dan semakin banyak pengetahuan yang dapat diambil.

Sesuai uraian yang telah dipaparkan diatas, maka judul tesis yang dapat diambil adalah **“ANALISIS DATA MINING UNTUK DENDA KETERLAMBATAN PENGIRIMAN BATUBARA DENGAN PENDEKATAN METODE *DECISION TREE C5* (STUDI KASUS : APLIKASI BATUBARA ONLINE (BBO) – PT. INDONESIA COMNET PLUS)”**

1.2 Rumusan Masalah

Dari latar belakang dipaparkan diatas, maka dirumuskan beberapa permasalahan sebagai berikut:

1. Bagaimana cara memanfaatkan *data Mining* untuk menghasilkan informasi dalam mengola informasi data denda keterlambatan batubara di aplikasi BBO?
2. Bagaimana membangun model *data mining* pada aplikasi BBO?
3. Bagaimana melakukan analisa denda keterlambatan batubara yang bersumber dari data pengiriman dan pembongkaran batubara menggunakan model *decision tree* c5?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

1. Menerapkan teknis *Data Mining* dalam menampilkan informasi mengenai data denda keterlambatan berdasarkan database yang ada di PT ICON+.
2. Menghasilkan infomasi mengenai berapa persen pemasok yang mengalami denda dan tidak mengalami denda.
3. Melakukan analisa keterlambatan batubara, dengan menggunakan data pengiriman dan pembongkaran batubara dengan menggunakan decision tree c5.
4. Menghasilkan faktor atau *variabel* yang mempengaruhi denda keterlambatan batubara.

1.4 Ruang Lingkup Penelitian

Penulis membatasi ruang lingkup penelitian dengan tujuan penelitian tepat sasaran :

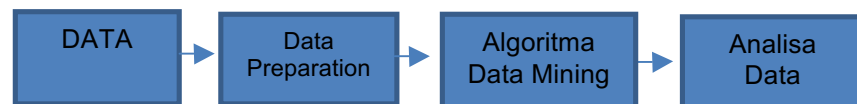
1. Penelitian ini menggunakan metode *Decision tree* c5.0.
2. Hanya mencakup denda keterlambatan batubara,
3. Penelitian dilakukan hanya pada apikasi BBO PT.ICON +.
4. Data yang digunakan pada penelitian ini yaitu data transaksi di januari 2016 sampai juni 2017 pada aplikasi BBO.
5. Analisis denda keterlambatan batubara diambil dari proses bisnis CIF (Cost, Insurance and Freight)

BAB II TINJAUAN PUSTAKA

2.1 Data Mining

Data Mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam basis data. *Data Mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data besar (Turban, 2009). Proses dalam tahap data mining (Gambar 2.1) terdiri dari tiga langkah Utama (Sumathi, 2006), yaitu:

1. *Data Preparation* Pada langkah ini, data dipilih, dibersihkan, dan dilakukan pemrosesan mengikuti pedoman dan *knowledge* dari ahli domain yang menangkap dan mengintegrasikan data internal dan eksternal kedalam tinjauan organisasi secara menyeluruh.
2. *Algoritma data mining* Penggunaan algoritma data mining dilakukan pada langkah ini untuk menggali data yang terintegrasi untuk memudahkan identifikasi informasi bernilai
3. *Fase analisa data* Keluaran dari data mining dievaluasi untuk melihat apakah *knowledge* domain ditemukan dalam bentuk rule yang telah diekstrak dari jaringan.



Gambar 2.1
Langkah-langkah dalam Proses Data Mining (Maimon & Rokach, 2010)

2.2 Klasifikasi

Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Han, 2006). Algoritma klasifikasi yang banyak digunakan secara luas, yaitu *Decision/classification trees*, *Bayesian classifiers/ Naïve Bayes classifiers*, *Neural networks*, *Analisa Statistik*,

Algoritma Genetika, *Rough sets*, *k-nearest neighbor*, Metode Rule Based, *Memory based reasoning*, dan *Support vector machines (SVM)*. Klasifikasi data terdiri dari 2 langkah proses. Pertama adalah learning (*fase training*), dimana algoritma klasifikasi dibuat untuk menganalisa data training lalu direpresentasikan dalam bentuk rule klasifikasi. Proses kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari rule klasifikasi (Han, 2006). Proses klasifikasi didasarkan pada empat komponen (Gorunescu, 2011): a. Kelas Variabel dependen yang berupa kategorikal yang merepresentasikan label yang terdapat pada objek. Contohnya: resiko penyakit jantung, resiko kredit, *customer loyalty*, jenis gempa. b. *Predictor* Variabel independen yang direpresentasikan oleh karakteristik (atribut) data. Contohnya: merokok, minum alkohol, tekanan darah, tabungan, aset, gaji. c. Training dataset Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*. d. Testing dataset Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

2.3 Tahap-Tahap Data Mining

Istilah *data mining* dan *knowledge discovery in databases (KDD)*, sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining. Proses KDD secara garis besar dapat dijelaskan sebagai berikut (Fayyad, 1996) :

1. Data Selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan suatu berkas, terpisah dari basis data operasional.

2. Pre-processing/Cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain

membuang duplikasi data, memeriksa data yang inkonsisten, memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses *enrichment*, yaitu proses memperkaya data yang sudah ada dengan data atau informasi yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

3. *Transformation*

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Interpretation/ Evaluation*

Pola informasi yang dihasilkan dari proses data mining, perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.4 **Decision Tree (Pohon Keputusan)**

Induksi pohon keputusan secara sederhana adalah struktur pohon yang merepresentasikan sebuah pembelajaran yang dimulai dari kelas (*level*) yang disusun dari urutan elemen-elemen yang akan dievaluasi. Sebuah pohon keputusan adalah sebuah *flowchart* yang sama dengan struktur pohon, dimana *internal node* (bagian batang pohon / *non leaf node*) yang merepresentasikan dari atribut-atribut yang diuji coba. Setiap cabang mempresentasikan nilai atau hasil dari atribut yang diuji coba.

Bagian daun (*terminal node*) mempresentasikan kelas dari sebuah label. Bagian terakhir dari sebuah pohon adalah akar (*root node*). Metodologi ini menggunakan pemisahan (*split*) *univariate*, sehingga mudah dipahami oleh pemakai dikarenakan bentuknya yang sederhana. Akan tetapi batasan-batasan yang diterapkan pada representasi aturan dan pohon tertentu dapat secara signifikan membatasi bentuk fungsional dari model (Han *et al*, 2012).

2.5 Algoritma C5.0

Algoritma C5.0 adalah perluasan dari algoritma C4.5 dan juga ID3 (Patil *etal.* 2012). C5.0 adalah algoritma klasifikasi yang dapat menangani kumpulan data besar. C5.0 lebih baik daripada C4.5 dalam hal kecepatan, memori dan efisiensi. Model C5.0 bekerja dengan memisahkan sampel berdasarkan bidang yang menyediakan keuntungan informasi dengan maksimum. Model C5.0 dapat membagi sampel berdasarkan nilai *information gain* terbesar. Atribut yang memiliki *information gain* terbesar akan dipilih sebagai *parent* atau untuk *node* selanjutnya.

Sebuah pohon keputusan adalah *classifier* yang direpresentasikan sebagai struktur pohon, di mana masing-masing simpul adalah simpul daun. Klasifikasi yang berlaku untuk semua kasus yang mencapai daun atau *node non-leaf*, beberapa tes dilakukan pada nilai atribut tunggal, dengan satu cabang dan subpohon untuk setiap kemungkinan hasil tes. *Node* dalam pohon keputusan melibatkan pengujian atribut tertentu. Biasanya, tes pada *node* membandingkan nilai atribut dengan konstan. Namun, beberapa pohon membandingkan dua atribut satu sama lain, atau menggunakan beberapa fungsi dari satu atau lebih atribut. *Node* daun memberikan klasifikasi yang berlaku untuk semua kasus yang mencapai daun, atau satu set klasifikasi, atau distribusi probabilitas atas semua klasifikasi yang mungkin (Witten *et al.* 2011). *Algoritma generate decision tree* adalah sebagai berikut (Han *et al.* 2012):

1. Partisi data, *D*, data latih yang telah ditentukan label kelasnya
2. *Attribute_list*, himpunan yang terdiri dari kandidat atribut
3. *Attribute_selection_method*, prosedur untuk menentukan kriteria pemotongan yang partisi *tuple* data terbaik ke kelas masing-masing.

Algoritma klasifikasi pohon keputusan adalah sebagai berikut (Han *et al.* 2012):

1. Membuat simpul N;
2. Jika semua tupel di D memiliki kelas yang sama yaitu C, Maka simpul N sebagai simpul daun dan diberi label dengan kelas C.
3. Jika *attribute list* kosong, Maka Jadikan simpul N sebagai simpul daun dan diberi label = nilai kelas terbanyak pada sampel.
4. Menerapkan *attribute selection method* (D, *attribute list*) untuk memperoleh atribut uji terbaik.
5. Beri label simpul N dengan atribut data uji.
6. Jika atribut bernilai diskret dan diperbolehkan untuk dipisah, maka
7. *Attribute list* <- atribut uji
8. Untuk setiap nilai j dari atribut uji yang diketahui
 - a. Buat D_j menjadi kumpulan data *tuple* D untuk memenuhi hasil j
 - b. Jika D_j kosong maka Tambahkan simpul daun yang diberi label = nilai kelas yang terbanyak pada D ke simpul N.
 - c. Selainnya, tambah cabang baru di bawah dengan memanggil fungsi *generate decision tree* (D_j *attribute list*) ke simpul N;
 - d. Kembalikan N;

Algoritma C5.0 memiliki tiga parameter masukan: D, *attribute_list*, dan *attribute_selection_method*. D merupakan satu set *tuple* pelatihan dan label kelas yang terkait. *Attribute list* menggambarkan suatu *tuple*. *Attribute selection method* menentukan prosedur untuk memilih atribut yang mengolah *tuple* menurut kelasnya. Prosedur ini membutuhkan ukuran atribut dengan *information gain* dan *gain indeks*. Algoritma pohon keputusan *generate decision tree* dimulai dari simpul tunggal, N, yang merepresentasikan data latih di D.

Model klasifikasi yang digunakan yaitu *tree* dan *rule-based*. *Tree* dimulai sebagai *node* tunggal, N mewakili *tuple* pelatihan D. *Tree* memiliki struktur pohon seperti *flowchart* yang masing-masing simpul internal *non leaf node* menunjukkan pengujian pada atribut, masing-masing cabang mewakili hasil dari pengujian dan masing-masing

simpul daun merupakan label kelas. *Root* merupakan simpul paling atas pada struktur *tree*. Sedangkan *rule-based* cara yang baik untuk mewakili informasi atau pengetahuan. Aturan *classifier* menggunakan aturan *IF* (kondisi) –*THEN* (kesimpulan) untuk klasifikasi. *IF* merupakan bagian (*or left side*) dari aturan ini dikenal sebagai aturan prasyarat. Sedangkan *THEN* (*or right side*) merupakan bagian konsekuen (Han et al. 2012). *Information gain* digunakan sebagai ukuran pemilihan atribut. Atribut dengan *information gain* terbesar ditentukan sebagai atribut pemisah untuk simpul *N*. Ukuran pemilihan atribut didefinisikan sebagai berikut (Han et al. 2012).

$$\text{Info}(D) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (1)$$

Dengan *info (D)* merupakan informasi yang dibutuhkan untuk mengklasifikasi label kelas sebuah *tuple* di *D*. p_i adalah peluang bukan nol dengan sebuah *tuple* acak di *D* memiliki kelas C_i dan ditentukan dengan $|C_i D|/|D|$. Fungsi log menggunakan basis 2, karena informasi yang dikodekan dalam bit. *Info (D)* juga dikenal sebagai *entropy*. Partisi *tuple* di *D* pada beberapa atribut *A* memiliki nilai *v* yang berbeda $\{a_1, a_2, \dots, a_v\}$ dari data latih. Atribut *A* digunakan untuk memisahkan *D* ke dalam *v* partisi atau sub himpunan $\{D_1, D_2, \dots, D_v\}$. $|D_j|/|D|$ merupakan bobot partisi ke-*j*. Nilai *entropy* yang dihasilkan untuk mengklasifikasi *tuple* dari *D* berdasarkan partisi oleh *A* adalah (Han et al. 2012):

$$\text{Info}_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times \text{Info}(D_j)$$

Information gain yang diperoleh pada atribut *A*

adalah: $\text{Gain}(A) = \text{Info}(D) - \text{Info}_A(D)$

Gain (A) menyatakan berapa banyak cabang yang akan diperoleh pada *A*. Atribut *A* dengan *information gain* tertinggi informasi, *Gain (A)*, dipilih sebagai atribut pada node *N* (Han et al. 2012). Algoritma C4.5 dan C5.0 dapat menghasilkan pengklasifikasi yang dinyatakan sebagai pohon keputusan atau set aturan (aturan-aturan). Metode set aturan C4.5 ini terlalu lambat dan memakan banyak memori (RAM). C5.0 merupakan algoritma

baru untuk memberikan aturan-aturan (*ruleset*), dan memperbaiki secara substansial. C5.0 memiliki beberapa jenis data baru selain yang tersedia di C4.5, termasuk tanggal, waktu, *timestamp* (tanggal dan waktu), atribut diskrit, dan label kasus. Beberapa perbedaan algoritma C4.5 dan C5.0 (Rulequest 2012) :

1. Akurasi: C5.0 memiliki tingkat kesalahan yang lebih rendah. C4.5 dan C5.0 memiliki akurasi prediksi yang sama, tetapi set aturan C5.0 lebih kecil.
2. Kecepatan: C5.0 jauh lebih cepat, menggunakan algoritma yang berbeda dan sangat dioptimalkan.
3. *Memory*: C5.0 umumnya lebih ringan dari C4.5.

Perbedaan utama antara C4.5 dan C5.0 adalah *boosting*, *winnowing*. *Boosting* adalah teknik untuk menghasilkan dan menggabungkan beberapa pengklasifikasi untuk meningkatkan akurasi prediksi. *Winnowing* adalah langkah seleksi fitur yang dilakukan sebelum pemodelan (Rulequest, 2012).

2.6 Evaluasi Model Klasifikasi

Evaluasi model merupakan tahapan yang juga dilakukan dalam penelitian dengan tujuan untuk memperoleh informasi yang terdapat pada hasil klasifikasi. Dalam penelitian ini dipilih evaluasi model berupa *confusion matrix* dengan tujuan untuk mempermudah dalam menganalisis performa algoritma karena *confusion matrix* memberikan informasi dalam bentuk angka sehingga dapat dihitung rasio keberhasilan klasifikasi. *Confusion matrix* adalah salah satu model evaluasi berbentuk matrik 2×2 yang digunakan untuk mendapatkan jumlah ketepatan klasifikasi *dataset* terhadap kelas data pada kedua algoritma yang dipakai. Dalam kasus dengan dua klasifikasi data keluaran dengan contoh: ya dan tidak, pinjam atau tidak pinjam, atau contoh lainnya, tiap kelas yang diprediksi memiliki empat kemungkinan keluaran yang berbeda, yaitu *true positives* (TP) dan *true negatives* (TN) menunjukkan ketepatan klasifikasi seperti pada Tabel 1.1 Jika prediksi keluaran bernilai positif sedangkan nilai aslinya adalah negatif maka disebut dengan *false positive* (FP) dan jika prediksi keluaran bernilai negatif sedangkan nilai aslinya adalah positif maka disebut dengan *false negative* (FN).

Tabel 1.1
Evaluasi dua kelas prediksi (Fawcett 2006)

		<i>Predicted Class</i>	
		<i>Yes</i>	<i>No</i>
<i>Actual Class</i>	<i>Yes</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
	<i>No</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Keakuratan model klasifikasi akan diuji dengan menghitung akurasi. Akurasi merupakan proporsi jumlah prediksi yang tepat dengan menghitung jumlah prediksi yang tepat dibagi total prediksi (Fawcett 2006).

2.7 Analisis Data

Analisis data diartikan sebagai upaya mengolah data menjadi informasi, sehingga karakteristik atau sifat-sifat data tersebut dapat dengan mudah dipahami dan bermanfaat untuk menjawab masalah-masalah yang berkaitan dengan kegiatan penelitian.

Dengan demikian, teknik analisis data dapat diartikan sebagai cara melaksanakan analisis terhadap data, dengan tujuan mengolah data tersebut menjadi informasi, sehingga karakteristik atau sifat-sifat datanya dapat dengan mudah dipahami dan bermanfaat untuk menjawab masalah-masalah yang berkaitan dengan kegiatan penelitian, baik berkaitan dengan deskripsi data maupun untuk membuat induksi, atau menarik kesimpulan tentang karakteristik populasi (parameter) berdasarkan data yang diperoleh dari sampel (statistik).

2.7.1 Macam – Macam Analisis Data

Teknik analisis data dalam penelitian, dibagi menjadi dua, yaitu teknik analisis data *deskriptif* dan teknik analisis data *inferensial*. Teknik analisis data penelitian secara *deskriptif* dilakukan melalui statistika *deskriptif*, yaitu statistik yang digunakan untuk menganalisis data dengan cara mendeskripsikan atau menggambarkan data yang telah terkumpul sebagaimana adanya tanpa bermaksud membuat generalisasi hasil penelitian.

Temasuk dalam teknik analisis data statistik deskriptif antara lain penyajian data melalui tabel, grafik, diagram, persentase, frekuensi, perhitungan mean, median atau modus.

Sementara itu teknik analisis data inferensial dilakukan dengan statistik *inferensial*, yaitu statistik yang digunakan untuk menganalisis data dengan membuat kesimpulan yang berlaku umum. Ciri analisis data *inferensial* adalah digunakannya rumus statistik tertentu (misalnya uji t, uji F, dan lain sebagainya). Hasil dari perhitungan rumus statistik inilah yang menjadi dasar pembuatan generalisasi dari sampel bagi populasi.

2.7.2 Tujuan Analisis Data

Adapun tujuan dari analisis data ialah untuk mendeskripsikan data, biasanya dalam bentuk frekuensi, ukuran tendensi sentral maupun ukuran dispersi, sehingga dapat dipahami karakteristik datanya. Dalam statistik kegiatan mendeskripsikan data ini dibahas pada statistika deskriptif. Tujuan lainnya adalah untuk membuat induksi atau menarik kesimpulan tentang karakteristik populasi, atau karakteristik populasi berdasarkan data yang diperoleh dari sampel (statistik). Kesimpulan yang diambil ini biasanya dibuat berdasarkan pendugaan (estimasi) dan pengujian hipotesis. Dalam statistika, kegiatan membuat induksi atau menarik kesimpulan tentang karakteristik populasi atau sampel ini dibahas pada statistika inferensial.

2.7.3 Langkah dan Prosedur Data Mining

Adapun langkah-langkah dalam *data mining*, yang diantaranya sebagai berikut ini:

1. Tahap mengumpulkan data, dilakukan melalui instrumen pengumpulan data.
2. Tahap editing, yaitu memeriksa kejelasan dan kelengkapan pengisian instrumen pengumpulan data.
3. Tahap koding, yaitu proses identifikasi dan klasifikasi dari setiap pertanyaan yang terdapat dalam instrumen pengumpulan data menurut variabel-variabel yang diteliti.
4. Tahap tabulasi data, yaitu mencatat atau entri data ke dalam tabel induk penelitian.
5. Tahap pengujian kualitas data, yaitu menguji validitas dan reliabilitas instrumen pengumpulan data.

6. Tahap mendeskripsikan data, yaitu tabel frekuensi dan/atau diagram, serta berbagai ukuran tendensi sentral, maupun ukuran *dispersi*. tujuannya memahami karakteristik data sampel penelitian.
7. Tahap pengujian hipotesis, yaitu tahap pengujian terhadap proposisi-proposisi yang dibuat apakah proposisi tersebut ditolak atau diterima, serta bermakna atau tidak. Atas dasar pengujian hipotesis inilah selanjutnya keputusan dibuat.

2.8 Penelitian Sebelumnya

2.8.1 Analisis Data Keterlambatan bahan baku berdasarkan pendekatan data *wirehouse* dan *decision tree* C5.0

Pada penelitian yang dilakukan oleh Rendy Septiandi (2016) tujuannya aplikasi OLAP dapat menyajikan informasi pembelian barang dan bahan baku secara multi dimensi, sehingga mempermudah pengguna dalam menganalisis data dan membuat laporan, dan menampilkan atribut yang berpengaruh terhadap keterlambatan penerimaan barang. Aturan yang dihasilkan yaitu 45 aturan, yang terdiri dari 19 aturan kelas ontime dan 26 aturan kelas delay. Pada aturan kelas delay maka didapatkan atribut-atribut penjelas pada model klasifikasi keterlambatan penerimaan barang yaitu atribut departemen yang meminta barang (ccid), periode quartal (quartalid), vendor penyedia barang (vendorname), petugas purchasing (buyername), kategori kuantitas (qtystatus). Dengan memanfaatkan sistem OLAP yang telah dibangun menggunakan aplikasi SpagoBI, maka atribut-atribut penjelas model klasifikasi keterlambatan penerimaan barang akan ditampilkan sehingga membentuk representasi informasi transaksi pembelian barang berdasarkan atribut yang berpengaruh terhadap keterlambatan penerimaan barang.

2.8.2 Data Mining Jasa Pengiriman Titipan Kilat di PT Citra Van Titipan Kilat (TIKI) dengan metode *Decision tree*

Pada penelitian yang dilakukan oleh Ibnu Graha(2014) tujuan dari penelitian ini Klasifikasi data pengiriman titipan dengan *data mining* metode *decision tree* digunakan dengan tujuan memberikan rencana strategis bagi perusahaan untuk mengetahui

karakteristik pasar. Sehingga dengan demikian dapat dianalisis pasar yang sudah ada ataupun menemukan peluang-peluang yang baru serta menemukan rencana strategis untuk meningkatkan keuntungan. Teknik *data mining* yang digunakan dalam klasifikasi data pengiriman titipan menggunakan metode *decision tree* dengan algoritma C5.0. Atribut yang digunakan untuk klasifikasi terdiri dari Service, Wilayah, Tonase, Harga, dan Waktu. Dengan pengimplemetasikan data mining menggunakan *decision tree* dapat diketahui bahwa pada WIB customer memiliki kecenderungan atau karakteristik lebih memilih layanan One Night Service jika tonase ≤ 10 dan harga ≤ 50000 . Kemudian pada wilayah WITA customer memiliki kecenderungan atau karakteristik lebih memilih layanan Reguler jika tonase > 10 .

2.8.3 Flight Delay Prediction

Penelitian yang dilakukan Martinez, Vincent tahun 2012 mencoba menjajaki berbagai model prediksi dan berbagai metode evaluasi Dengan mengukur kinerja model dengan menggunakan data nyata, dan melihat hasil yang menarik mengenai prediktabilitas penundaan. Metode prediksi penundaan terbaik tampaknya paling spesifik yang memperhitungkan semua kombinasi parameter kategoris dan kondisi pada jam kedatangan. Kinerja model sangat menarik untuk dievaluasi, karena Berbagai ukuran yang digunakan, dan parameter yang berbeda disesuaikan. Namun prediksi yang didapat tampaknya lebih baik dari yang ada terlihat dalam literatur yang digunakan oleh *Flight Caster* (Prediksi jangka Panjang), Hanya didasarkan pada distribusi kumulatif empiris. Metode estimasi adalah metode yang sangat menarik, untuk belajar dan memanipulasi Dengan metode data-centered, bisa digunakan masing-masing waktu, merekonstruksi model probabilistik dari beberapa pengamatan.

2.8.4 Penggunaan *Educational Data Mining* (EDM) Untuk Mendeteksi Keterlambatan Masa Studi Mahasiswa (Studi Kasus Udinus)

Pada penelitian yang dilakukan oleh Defri Kurniawan, Wibowo Wicaksono dan Yani Parti Astuti tujuan tujuannya dari data mining ini, untuk memprediksi mahasiswa yang tepat waktu dan tidak tepat waktu agar pengelolaan akademik di dalam memetakan mahasiswa yang berpotensi mengalami keterlambatan masa studi bisa di tanggulasi sejak awal dengan metode *decision tree* c4.5.

BAB III OBJEK DAN METODE PENELITIAN

Dalam bab ini akan dipaparkan metode atau langkah-langkah yang akan digunakan untuk membahas permasalahan yang diambil dalam penelitian. Pada bagian ini juga akan menjelaskan mengenai algoritma, kemudian alat yang akan dipakai dalam pengerjaan penelitian.

3.1 Latar Belakang Perusahaan

Studi kasus yang akan dilakukan PT.ICON+ , didirikan pada tanggal 3 Oktober 2000, PT Indonesia Comnets Plus (ICON+) berfokus pada penyediaan jaringan, jasa, dan content telekomunikasi, khusus untuk mendukung teknologi dan system informasi PT PLN (Persero) dan publik. Untuk itu perseroan mengadakan berbagai layanan unggulan seperti *clear channel*, *multi protocol label switching* (MPLS), akses internet broadband, *Voice over Internet Protocol* (VoIP), dan aplikasi perbankan.

Sebagai anak perusahaan yang dimiliki sepenuhnya oleh PLN, pada awalnya ICON+ berfokus untuk melayani kebutuhan PLN akan jaringan telekomunikasi. Seiring dengan kebutuhan industri akan jaringan telekomunikasi dengan tingkat *availability* dan *reliability* yang konsisten, ICON+ melihat peluang baru untuk mengembangkan usahanya yaitu dengan mengkomersialkan kelebihan kapasitas jaringan telekomunikasi ketenagalistrikan serat optik milik PLN di Jawa dan Bali.

Berdasarkan pemikiran tersebut, ICON+ mulai menjalin kerjasama dengan berbagai perusahaan, terutama yang kegiatan operasionalnya membutuhkan jaringan telekomunikasi yang ekstensif dan handal. Hingga saat ini Perseroan melayani lebih dari 1127 perusahaan di Indonesia, di industri-industri utama yaitu telekomunikasi, perbankan, keuangan, pemerintahan dan manufaktur.

3.2 Visi dan Misi PT INDONESIA COMNETS PLUS (ICON+)

Visi

Menjadi penyedia solusi teknologi informasi dan komunikasi terkemuka di Indonesia berbasis jaringan melalui pemanfaatan aset strategis.

Misi

1. Memberikan layanan Teknologi Informasi dan Komunikasi terbaik di kelasnya kepada pelanggan guna meningkatkan nilai perusahaan.
2. Memenuhi kebutuhan dan harapan PLN secara proaktif dengan menyediakan solusi-solusi TIK yang inovatif dan memberikan nilai tambah.
3. Membangun organisasi pembelajar yang berkinerja tinggi untuk mendorong perusahaan mencapai bisnis yang unggul dan menjadi pilihan bagi talenta-talenta terbaik.
4. Memberikan kontribusi terhadap perkembangan telekomunikasi nasional.

3.3 Nilai-nilai Perusahaan

3.3.1 INTEGRITY

Perilaku yang harus dimiliki: berketuhanan Yang Maha Esa, berdedikasi tinggi, Menjalankan apa yang saya katakan, Setia terhadap perusahaan, Bertanggung jawab, Dapat Dipercaya, Menolak Suap, dan Menggunakan fasilitas perusahaan hanya untuk kepentingan perusahaan.

3.3.2 CARE

Perilaku yang harus dimiliki: melayani dengan sepenuh hati, bertindak cepat dan tanggap, terlibat aktif dalam usaha mencapai tujuan stakeholders sesuai dengan kompetensinya (keterlibatan dan kontribusi), Berpikir dan bertindak untuk melayani.

3.3.3 OPEN MIND

Perilaku yang harus dimiliki : berpikir positif, mau menerima saran, percaya Diri, pembelajar, profesional, saling menghargai dan menghormati, serta membuka diri.

3.3.4 INNOVATION

Perilaku yang harus dimiliki : bersikap kreatif, analisa sebelum bertindak, selalu ingin tahu, selalu aktif meningkatkan pengetahuan, berpikir ke depan, dan lebih baik dari kompetitor.

3.3.5 TEAM WORK

Perilaku yang harus dimiliki : fokus, toleransi, proaktif, saling mengisi dan menyemangati, Kemampuan melakukan sebagai pelaku tim maupun pemain tim, kompak (kebersamaan), semangat gotong royong, mau berbagi, berpikir menang-menang.

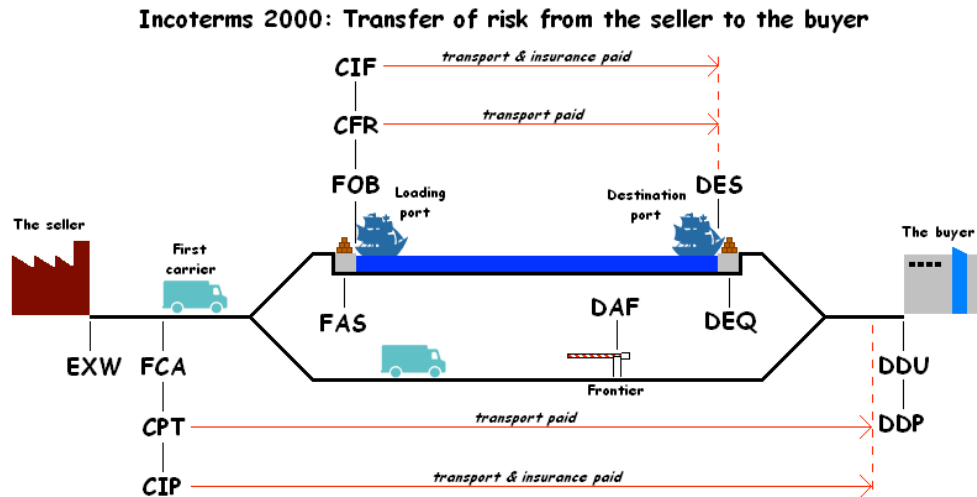
3.3.6 EXCELLENCE

Perilaku yang harus dimiliki : meningkatkan kompetensi, optimis dan pantang menyerah, selalu mau lebih maju, bekerja dengan cerdas, antusias, ingin menjadi yang terbaik, berani mencoba, menghargai waktu dan momentum, fokus kepada tujuan akhir.

3.4 Objek Penelitian

Sumber data yang digunakan dalam penelitian ini berasal dari aplikasi BBO di PT ICON+, data tersebut berupa data transaksi proses transaksi batubara mulai dari penjadwalan sampai data pembongkaran. Operasional aplikasi batubara di mulai dari 2011 sampai sekarang. Tapi efektifnya denda keterlambatan di 2015 pertengahan, karena sebelum tahun 2015 aplikasi batubara di peruntukan untuk monitoring kapal saja tidak sampai perhitungan denda. Di tahun 2017 ada beberapa aturan dan kontrak amandemen baru antara pemasok batubara dengan PLN, maka data yang akan diolah yaitu data dari Januari 2017 sampai Juli 2017 jumlah datanya ada 884 row.

3.4.1 Proses Bisnis Pegiriman



Gambar 3.1
Proses Bisnis Batubara (Schaffer, Agusti and Earle.2009)

Incoterms atau *International Commercial Terms* adalah kumpulan istilah yang dibuat untuk menyamakan pengertian antara penjual dan pembeli dalam perdagangan internasional. Dan juga menjelaskan hak dan kewajiban pembeli dan penjual yang berhubungan dengan pengiriman barang.

Hal-hal yang dijelaskan meliputi proses pengiriman barang, penanggung jawab proses ekspor-impor, penanggung biaya yang timbul dan penanggung risiko bila terjadi perubahan kondisi barang yang terjadi akibat proses pengiriman.

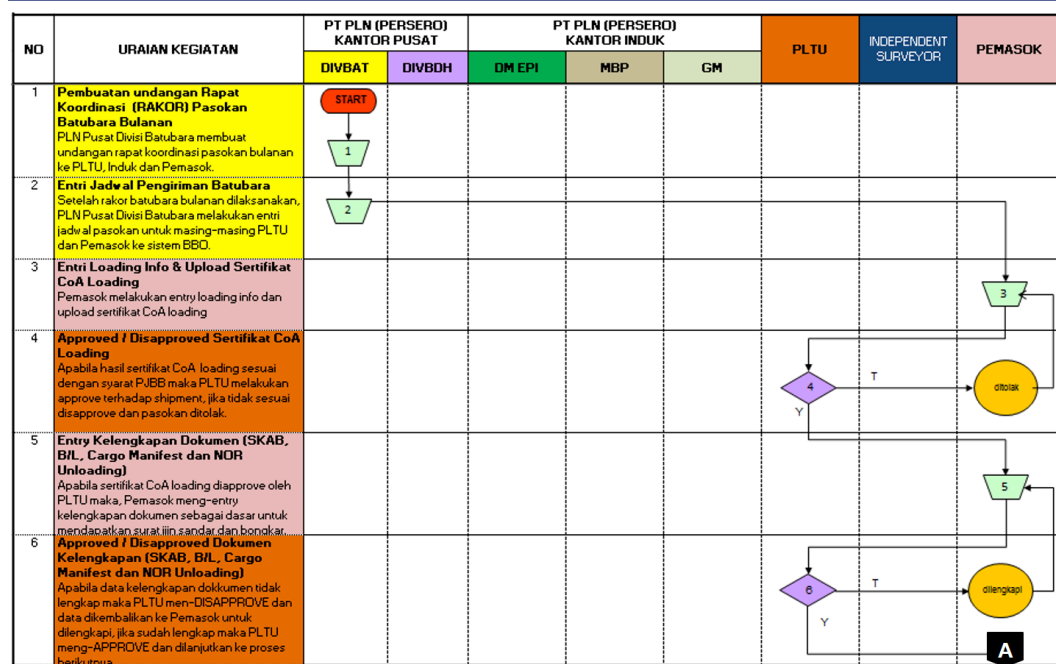
Tiga belas istilah dalam Incoterms 2000:

1. EXW (nama tempat): *Ex Works*, pihak penjual menentukan tempat pengambilan barang.
2. FCA (nama tempat): *Free Carrier*, pihak penjual hanya bertanggung jawab untuk mengurus izin ekspor dan menyerahkan barang ke pihak pengangkut di tempat yang telah ditentukan.

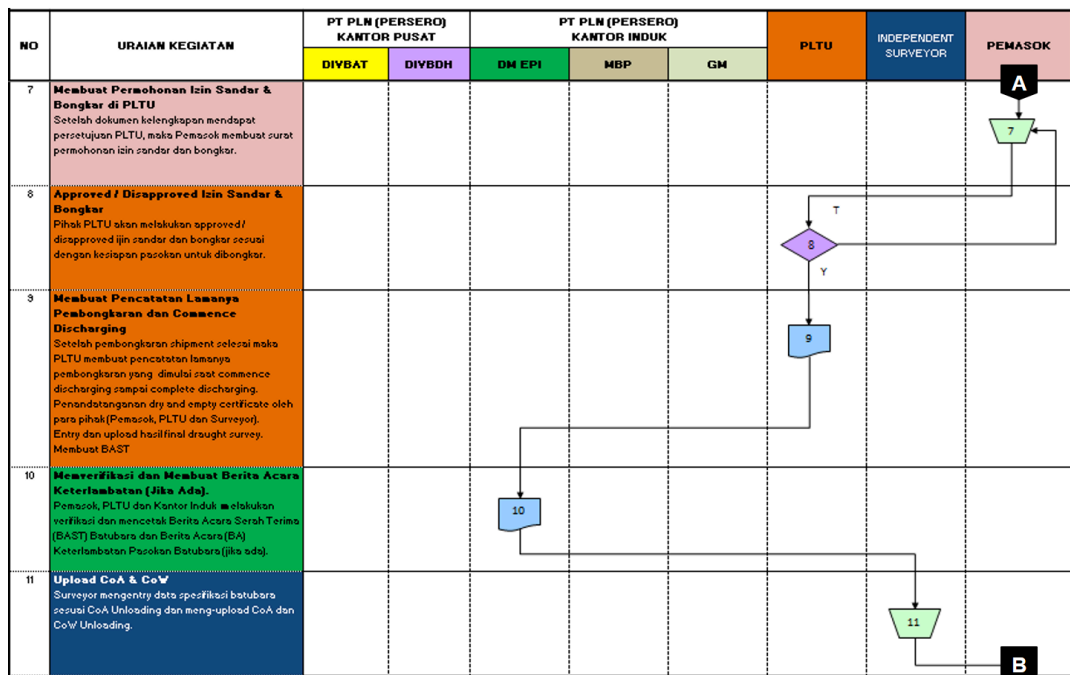
3. FAS (nama pelabuhan keberangkatan): *Free Alongside Ship*, pihak penjual bertanggung jawab sampai barang berada di pelabuhan keberangkatan dan siap disamping kapal untuk dimuat. Hanya berlaku untuk transportasi air.
4. FOB (nama pelabuhan keberangkatan): *Free On Board*, pihak penjual bertanggung jawab dari mengurus izin ekspor sampai memuat barang di kapal yang siap berangkat. Hanya berlaku untuk transportasi air.
5. CFR (nama pelabuhan tujuan): *Cost and Freight*, pihak penjual menanggung biaya sampai kapal yang memuat barang merapat di pelabuhan tujuan, namun tanggung jawab hanya sampai saat kapal berangkat dari pelabuhan keberangkatan. Hanya berlaku untuk transportasi air.
6. CIF (nama pelabuhan tujuan): *Cost, Insurance and Freight*, sama seperti CFR ditambah pihak penjual wajib membayar asuransi untuk barang yang dikirim. Hanya berlaku untuk transportasi air.
7. CPT (nama tempat tujuan): *Carriage Paid To*, pihak penjual menanggung biaya sampai barang tiba di tempat tujuan, namun tanggung jawab hanya sampai saat barang diserahkan ke pihak pengangkut.
8. CIP (nama tempat tujuan): *Carriage and Insurance Paid to*, sama seperti CPT ditambah pihak penjual wajib membayar asuransi untuk barang yang dikirim.
9. DAF (nama tempat): *Delivered At Frontier*, pihak penjual mengurus izin ekspor dan bertanggung jawab sampai barang tiba di perbatasan negara tujuan. Bea cukai dan izin impor menjadi tanggung jawab pihak pembeli.
10. DES (nama pelabuhan tujuan): *Delivered Ex Ship*, pihak penjual bertanggung jawab sampai kapal yang membawa barang merapat di pelabuhan tujuan dan siap untuk dibongkar. izin impor menjadi tanggung jawab pihak pembeli. Hanya berlaku untuk transportasi air.
11. DEQ (nama pelabuhan tujuan): *Delivered Ex Quay*, pihak penjual bertanggung jawab sampai kapal yang membawa barang merapat di pelabuhan tujuan dan barang telah dibongkar dan disimpan di dermaga. Izin impor menjadi tanggung jawab pihak pembeli. Hanya berlaku untuk transportasi air.

12. DDU (nama tempat tujuan): *Delivered Duty Unpaid*, pihak penjual bertanggung jawab mengantar barang sampai di tempat tujuan, namun tidak termasuk biaya asuransi dan biaya lain yang mungkin muncul sebagai biaya impor, cukai dan pajak dari negara pihak pembeli. Izin impor menjadi tanggung jawab pihak pembeli.
13. DDP (nama tempat tujuan): *Delivered Duty Paid*, pihak penjual bertanggung jawab mengantar barang sampai di tempat tujuan, termasuk biaya asuransi dan semua biaya lain yang mungkin muncul sebagai biaya impor, cukai dan pajak dari negara pihak pembeli. Izin impor juga menjadi tanggung jawab pihak penjual.

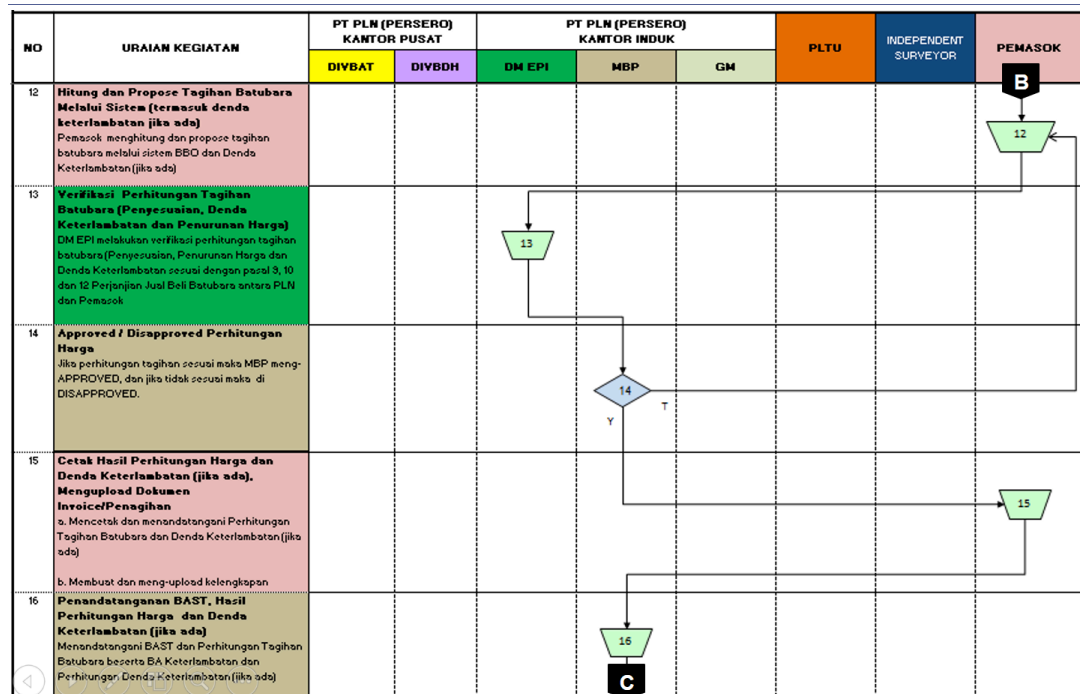
berikut ini pada gambar 3.2 sampai dengan gambar 3.5 merupakan proses bisnis CIF yang di pakai PLN



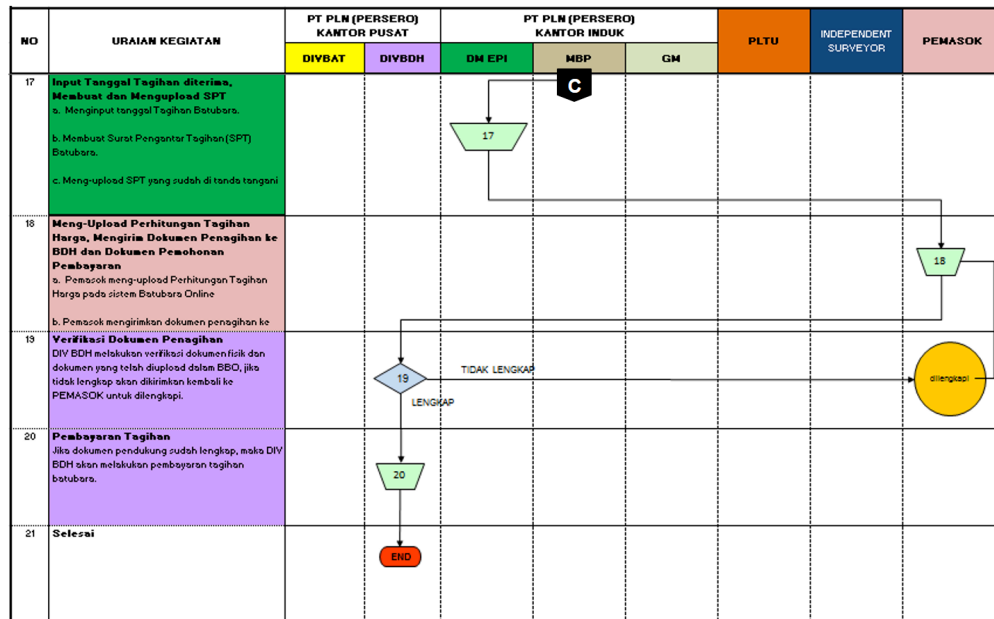
Gambar 3.2
Flow CIF dari Rakor sampai approval dokument NOR (DITDAN. 2014)



Gambar 3.3
Flow CIF dari permohonan izin sandar upload coa dan cow (DITDAN. 2014)



Gambar 3.4
Flow CIF dari Hitung tagihan pemasok sampai penandatanganan berita acara serah terima batubara, perhitungan dan denda keterlambatan (jika ada) (DITDAN. 2014)



Gambar 3.5
Flow CIF dari input tagihan sampai pembayaran (DITDAN. 2014)

Untuk jenis Kalori batubara terdiri dari 3 yaitu :

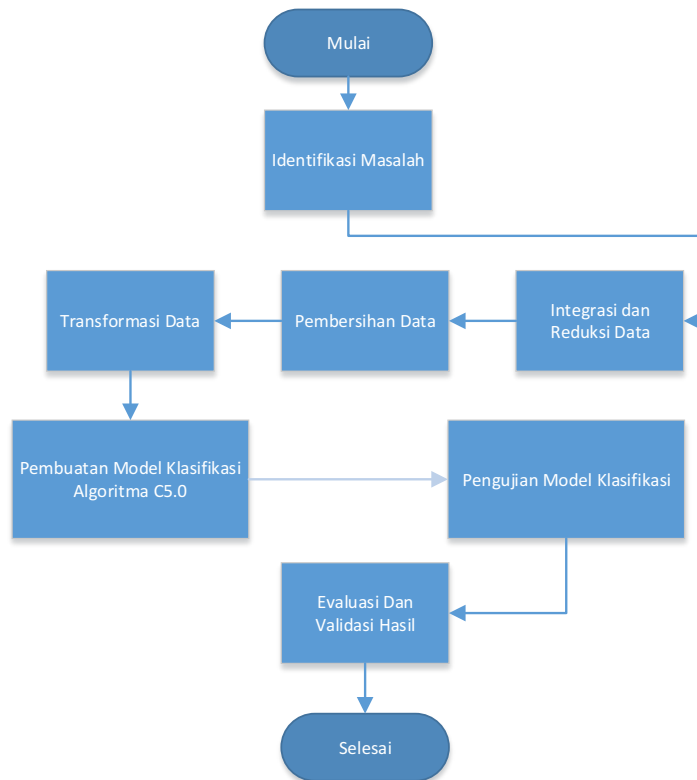
1. *High calories coal*
2. *Medium calories coal*
3. *Low calories coal.*

Sedangkan untuk nilai kalorinya batubara : 3900, 7000, 4000, 4200, 4300, 4400, 4600, 4800, 4900, 5000, 5100, 5400, 5700, 6100, 6300, 4250.

Di Aplikasi BBO yang digunakan sebagai speknya adalah MRC (*Medium calories coal*), LRC (*Low calories coal*), 3900, 7000, 4000, 4200, 4300, 4400, 4600, 4800, 4900, 5000, 5100, 5400, 5700, 6100, 6300, 4250. Jadi gabungan antara nilai kalori dan jenis kalori batubara, di karena belum semua pemasok batubara ke PLN menggunakan parameter nilai kalori dari batubara, karena ada beberapa pemasok kontraknya masih menggunakan jenis kalori batubara.

3.5 Metode Penelitian

Pada penelitian ini, tahapan penelitian yang akan dilakukan indentifikasi masalah, kemudian melakukan model klasifikasi menggunakan decision tree c5. Tahap penelitian ditunjukan pada gambar 3.6 :



Gambar 3.6
Tahapan Penelitian

3.5.1 Identifikasi Masalah

Tahap indentifikasi masalah dilakukan untuk mengetahui kebutuhan informasi yang pengguna butuhkan, sehingga hasil dari penelitian sesuai dengan yang di harapkan pengguna.

3.5.2 Praproses Data

Praproses data adalah kegiatan mempersiapkan data yang akan digunakan, agar informasi atau pengetahuan yang ingin diambil dari data tersebut dapat diambil dengan lebih mudah. Tahapan ini dilakukan setelah data yang akan diproses didapatkan dari basis data, beserta informasi mengenai atribut penting dari data. Ada beberapa proses

yang dilakukan pada tahapan ini, yaitu penanganan missing value, integrasi data, reduksi dimensi data, konsistensi data, pembersihan data, dan transformasi data. Data batubara di ambil dari aplikasi BBO yang tersimpan di database, karena data yang di peroleh dari tahun 2012 sampai dengan 2017. Kemudian data di filter berdasarkan efektifnya penggunaan denda keterlambatan di aplikasi, maka di pilihlah tahun 2017 januari-juli 2017 pengambilan datanya.

3.5.3 Pembuatan Model Klasifikasi C5.0

Pada tahap ini dibangun model klasifikasi untuk membentuk pohon keputusan. Pembentukan pohon keputusan ini menggunakan algoritma C5.0 yang menerapkan ukuran informasi dalam membuat pohon keputusan. Model klasifikasi dibuat dengan menggunakan aplikasi clementine 10.1.

3.5.4 Pengujian Model Klasifikasi C5.0

Pengujian model klasifikasi merupakan tahap akhir pada pembangunan model klasifikasi, dimana pada tahap ini semua pola yang telah dilakukan akan diuji.

3.5.5 Evaluasi Dan Validasi Hasil

Evaluasi dilakukan dengan mengamati hasil prediksi menggunakan Algoritma C5.0 dan validasi dilakukan dengan mengukur hasil prediksi dibandingkan dengan data asal.

3.6 Lingkungan Penelitian

Alat-alat bantu yang digunakan dalam penelitian ini adalah sebagai berikut :

1. Komputer personal dengan spesifikasi prosesor Intel® Core i5-5200U CPU @ 2.20 GHz, Memory (RAM) 8 GB.
2. Sistem Operasi Windows 10.
3. Oracle 11g.
4. Toad For Oracle 11g.
5. IMB SPSS Statistics 23.
6. Clemententine.
7. Microsoft Excel untuk membantu melakukan praproses terhadap data.

BAB IV IMPLEMENTASI PENELITIAN

4.1 Rencana Penelitian

4.1.1 Pemahaman dari sisi bisnis

Salah satu tujuan *Data Mining* sendiri adalah untuk mencari pola tertentu yang dipergunakan untuk memenuhi tujuan bisnis itu sendiri. Pada umumnya tujuan dari bisnis batubara online, adalah meningkatkan kinerja supaya suplay pasokan batubara ke PLTU, agar tidak mengalami keterlambatan pasokan, dan mengurangi harga transportasi, sehingga PT.PLN bisa mengefesiensikan pengeluaran pembelian batubara, dan menjadi referensi dalam penentuan kebijakan.

4.1.2 Pemahaman dari sisi data

Pada proses ini dilakukan pengumpulan data yang akan digunakan untuk memproses pengelolaan *data mining*. Data-datanya dari table digunakan untuk pengolahan data :

1. TB_JADWAL
2. TB_REAL_PENGIRIMAN
3. TB_BONGKAR
4. TB_PASOKAN_DT
5. TB_MASTER_SPEK
6. TB_MASTER_MODA
7. TB_MASTER_PEMASOK

Isi field pada table bisa di lihat di lampiran A

4.2 Data Collection

Data yang akan digunakan harus dikumpulkan terlebih dahulu sesuai dengan kriteria proses pengelolaan data. Berikut kriteria data yang akan digunakan untuk proses *data mining* :

- a. Data jadwal konfirmasi pasokan mitra(Pemasok) ke PLTU, sesuai rapat kordinasi yang telah di sepakati di setiap bulan. Yang di ambil dari table TB_JADWAL dengan melihat nilai volume confirm > 0 (lebih besar dari nol) atau volume confrim tidak sama dengan null atau tgl_ta_confirm tidak sama dengan null.
- b. Data mitra pemasok batubara, berisikan data-data mitra mulai dari nama mitra, alamat, no telepon dan lain-lain yang berhubungan dengan data master, akan tetapi data mintra yang di tampilkan, adalah data mitra yang pemasok batubara kontrak CIF.
- c. Data pembangkit batubara, berisikan data-data nama pembangkit, alamat pembangkit dan pembangkit tersebut pelimpahan wewenang unit mana. Pembangkit yang di tampilkan yang mempunyai kontrak CIF dengan pemasok batubara.
- d. Data jadwal pengiriman batubara berisikan data-data batubara yang di mulai dari tanggal dia muat, data transportasi yang di gunakan untuk mengirim batubara baik menggunakan kapal/tongkang/vessel, kemudian tanggal saat kapal sandar,selesai muat dan jumlah batubara yang di maut dengan satuan Metrik Ton (MTON).
- e. Data kedatangan berikan data-data saat proses kedatangan kapal mulai dari datang kapal semua di catat, jika ada mengalami gangguan kapal atau gangguan akibat bencana alam, semua di catat di table kedatangan.
- f. Data Pembongkaran berisikan data-data pembongkaran mulai dari tanggal bongkar, selesai bongkar,volume yang di bongkar.

4.3 Reduksi Data dan Pembersihan Data

Pada tahap ini data yang diperoleh kemudian disederhakan atau lebih ringkas lagi agar mempermudah proses *data mining* dan membersihkan data-data yang di anggap error.

- a. Proses yang pertama adalah memilih atribut-atribut yang akan dipakai pada proses *data mining*.
- b. Proses kedua menghilangkan *field* atau kolom yang sama, akan tidak terjadi duplikasi.

dikarenakan atribut yang berasal dari table sangatlah banyak dan tidak semua dipakai untuk proses *data mining*.

Table 4.1

Data Master Mitra Pemasok Batubara

No	Field	Tipe Data
1	ID_PEMASOK	INTEGER
2	NAMA_PEMASOK	VARCHAR2(200 BYTE)

Table 4.2

Data Master Pembangkit

No	Field	Tipe Data
1	ID_PLTU	INTEGER
2	NAMA_PEMBANGKIT	VARCHAR2(200 BYTE)

Table 4.3

Data Jadwal Konfirmasi Rakor

No	Field	Tipe Data
1	ID_PEMASOK	INTEGER
2	ID_PLTU	INTEGER
3	NO_JADWAL	VARCHAR2(20 BYTE)
4	TGL_TA_CONFIRM	DATE
5	VOLUME_CONFIRM	NUMBER

Table 4.4

Data Jadwal Pengiriman

No	Field	Tipe Data
1	ID_PENGIRIMAN	INTEGER
2	NO_JADWAL	VARCHAR2(200 BYTE)
3	TA	DATE
4	ID_SPEK	VARCHAR2(20 BYTE)
5	ID_MODA	INTEGER

Table 4.5

Data Pembongkaran

No	Field	Tipe Data
1	ID_PENGIRIMAN	INTEGER
2	VOLUME_DRAUGHT_SURVEY	NUMBER

Table 4.6

Data Master Spek

No	Field	Tipe Data
1	ID_SPEK	INTEGER
2	NAMA_SPEK	VARCHAR2(200 BYTE)

Table 4.7

Data Master Moda

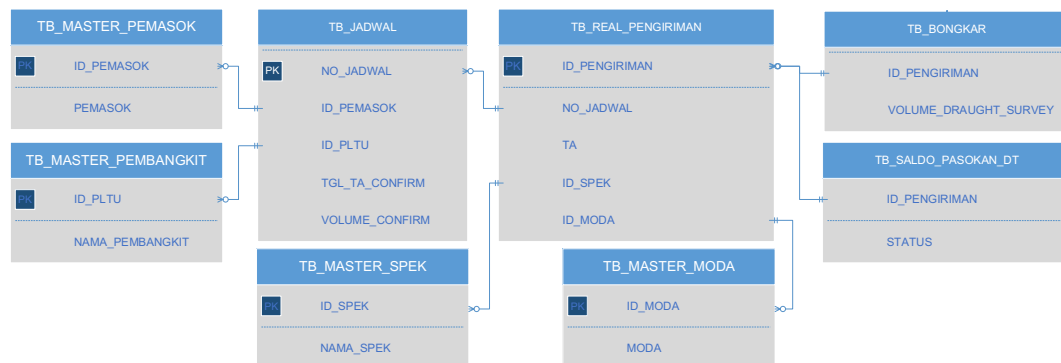
No	Field	Tipe Data
1	ID_MODA	INTEGER
2	MODA	VARCHAR2(200 BYTE)

Table Lampiran 1.9

Data Pasokan Batubara

No	Field	Tipe Data
1	ID_PENGIRIMAN	INTEGER
2	STATUS	CHAR(1 BYTE)

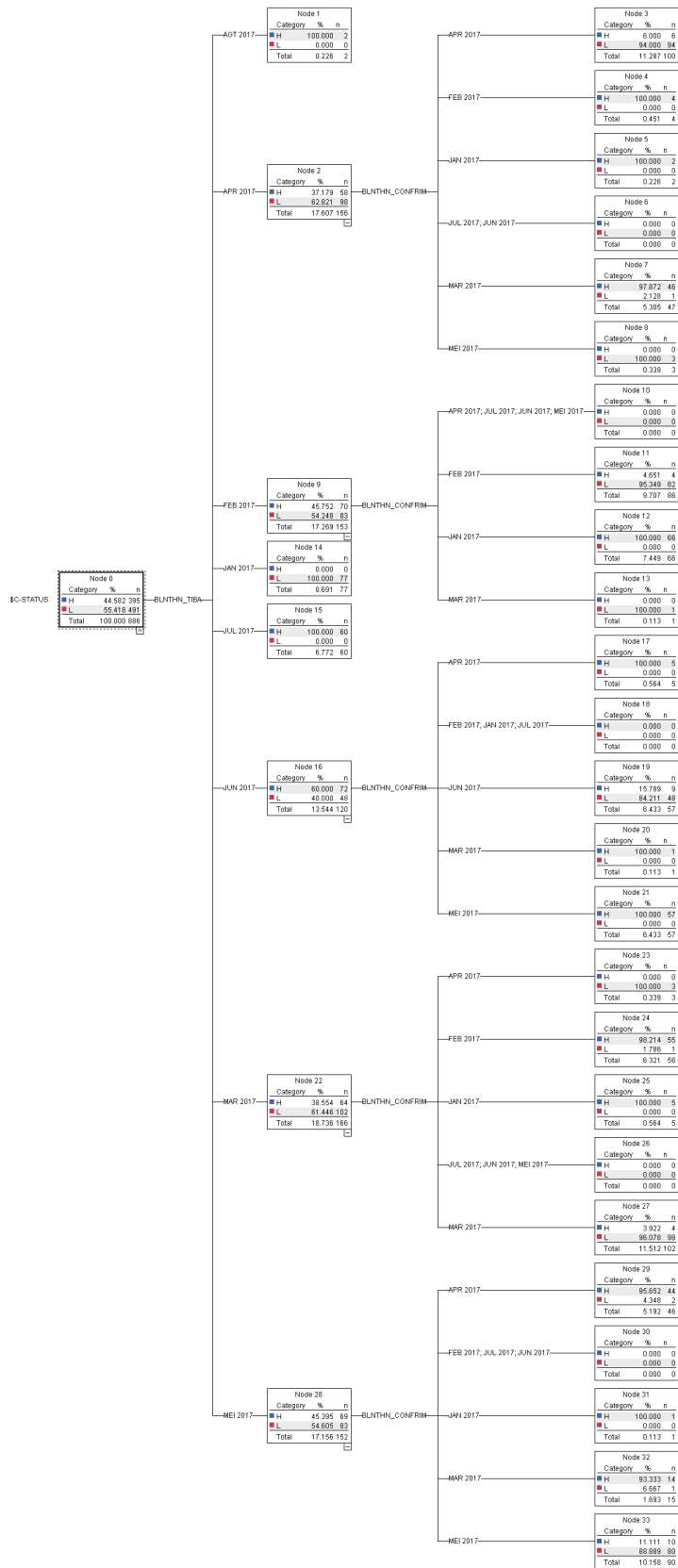
4.4 Transformasi Data dan Clustering



Gambar 4.1 Diagram Database Setelah Transformasi

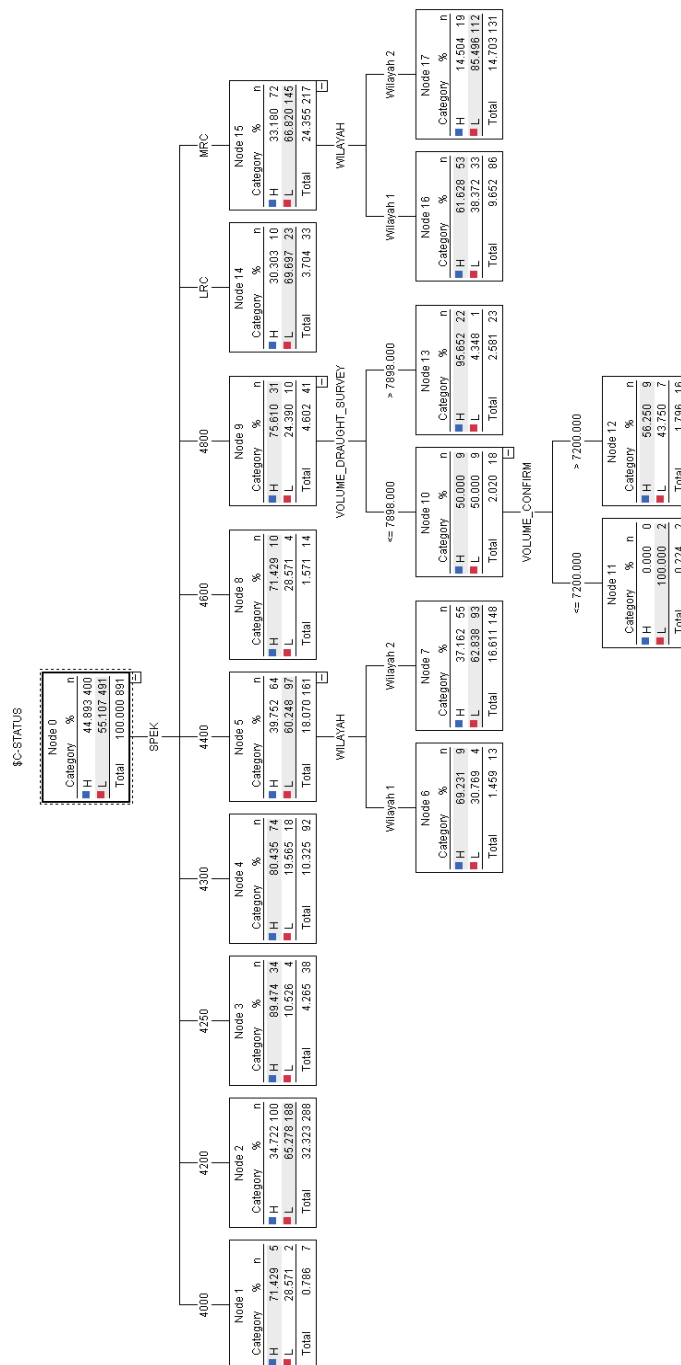
Pada Fase ini mempersiapkan data menjadi dataset. Dataset adalah suatu data digunakan sebagai dasar untuk membentuk model yang akan digunakan dan dapat disesuaikan selama proses *data mining* karena dimungkinkan kembali ke tahap yang sebelumnya jika ada informasi yang kurang lengkap. Model yang di pakai menggunakan

model *decision tree c5*, untuk software yang digunakan dalam proses pembentukan model ini menggunakan TOAD 11.5.0.56 dan export datanya menggunakan Microsoft Excel, kemudian variable diformat sesuai aturan spss, kemudian data yang sudah diformat tersebut di import di clementine. Total record yang siap untuk diprediksi dan yang sudah di bersihkan jika ada redundansi data atau data kosong, hasilnya menjadi 884 record.



Gambar 4.2 Hasil Decision Tree C5.0 Prediksi Faktor-faktor Keterlambatan

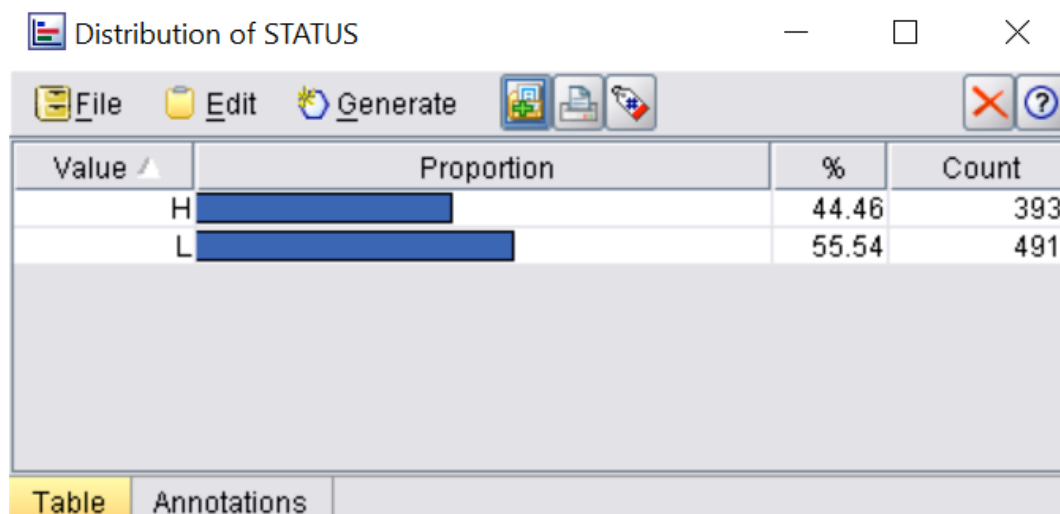
Pada gambar 4.2 faktor keterlambatan dipengaruhi oleh bulan tiba kapal, jika bulan tiba sama dengan (=) bulan konfirmasi saat rapat kordinasi maka pasokan dianggap lancar, akan tetapi untuk mencari faktor-faktor lain bisa mengidentifikasi faktor keterlambatan, maka dibuatlah *data mining* kembali agar menjelaskan lebih detail selain bulan tiba kapal, faktor lain yang dapat prediksi keterlambatan batubara.



Gambar 4.3 Hasil *Decision Tree C5.0* Prediksi Faktor-faktor Keterlambatan Lanjutan

Berdasarkan hasil decision tree c5.0 pada gambar 4.3 bahwa analisis keterlambatan batubara berdasarkan aturan-aturan yang dihasilkan oleh model klasifikasi C5.0. Model tersebut menghasilkan 11 aturan. aturan batubara tersebut tersebut status lancar terdiri 4 aturan, sedangkan untuk status hutang aturan batubara tersebut ada 7 aturan. Maka didapat 5 atribut yang digunakan dalam pembentukan aturan pada klasifikasi keterlambatan batubara. Atribut spek adalah nilai kalori batubara, wilayah yaitu wilayah batubara itu di pasok, atribut volume draught survey adalah pasokan batubara yang di terima oleh PLTU, atribut volume confirm adalah pasokan yang harus dipenuhi mitra sesuai rapat kordinasi setiap bulan.

Pada gambar 4.2 terlihat bahwa atribut Spek berpengaruh paling tinggi untuk menentukan klasifikasi data pengiriman batubara, hal ini juga ditunjukan dengan atribut spek menempati sebagai *root node*. Kemudian atribut wilayah dan volume draught survey menempati internal node yang pertama.



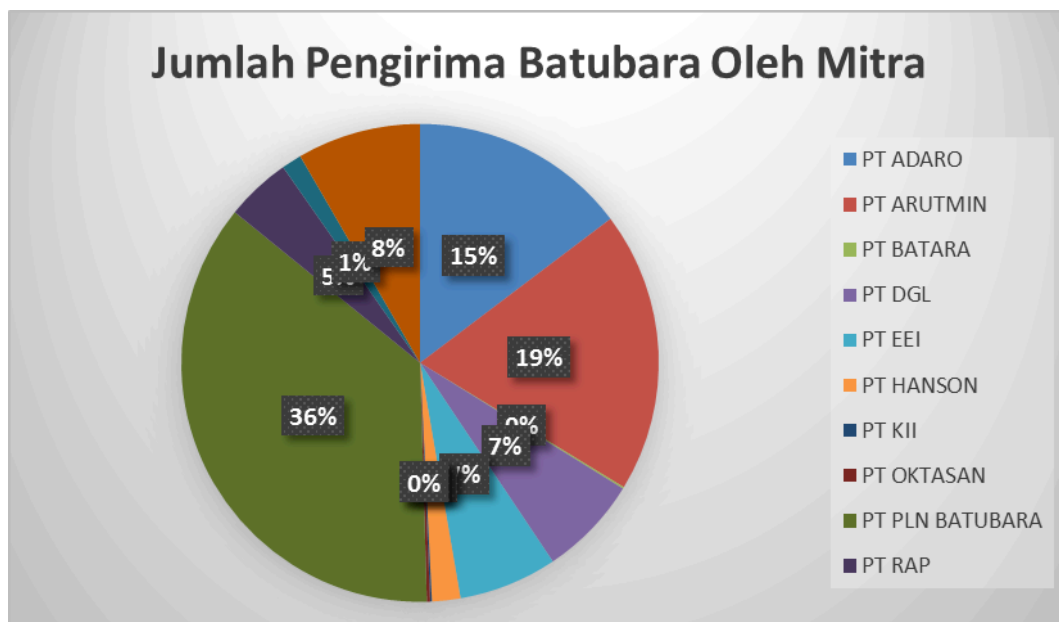
Gambar 4.4 Jumlah transaksi Hutang dan lancar

Pada gambar 4.4 jumlah data keseluruhan adalah 884 data pada januari 2017 sampai dengan juli 2017, dengan presentase 44.46 yang pasokan yang mengalami keterlambatan, sedangkan presentase 55.54 data pasokan yang lancar.

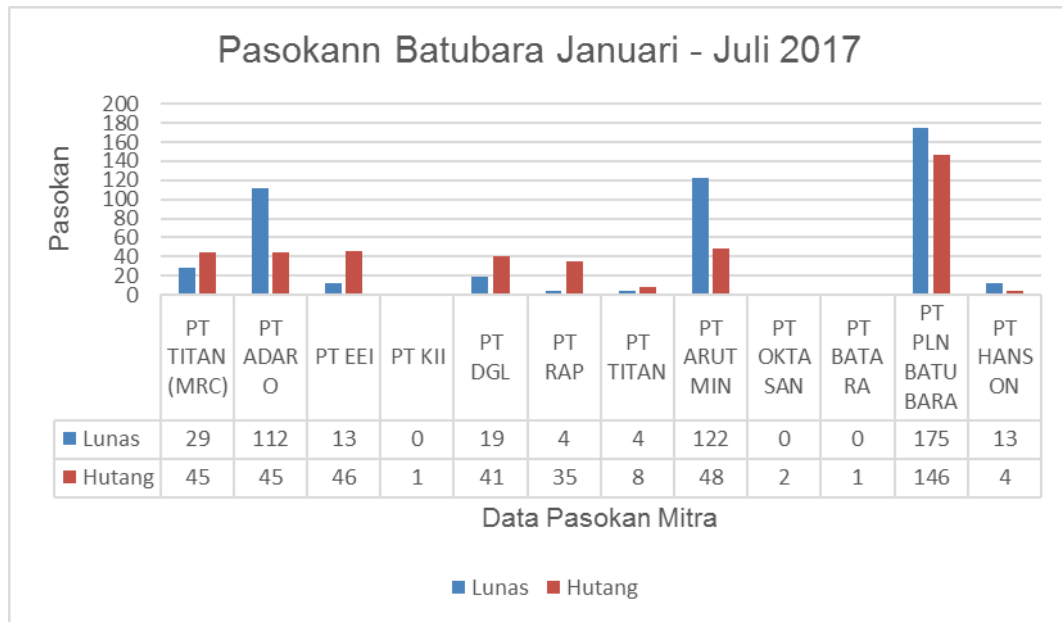
Distribution of PEMASOK #4			
Value	Proportion	%	Count
PT ADARO		14.71	130
PT ARUTMIN		19.0	168
PT BATARA		0.11	1
PT DGL		6.79	60
PT EEI		6.67	59
PT HANSON		1.92	17
PT KII		0.11	1
PT OKTASAN		0.23	2
PT PLN BATUBARA		36.31	321
PT RAP		4.41	39
PT TITAN		1.36	12
PT TITAN (MRC)		8.37	74

Gambar 4.5 Jumlah pengiriman batubara oleh mitra

Pada gambar 4.5 menunjukan jumlah pasokan yang dikirim oleh mitra ke PLN.



Gambar 4.6 Presentase Jumlah pengiriman batubara oleh mitra



Gambar 4.7 Presentase Pasokan batubara Mitra

Pada gambar 4.7 terlihat bahwa pasokan mitra keseluruhan yang mengalami keterlambatan dan pasokan yang lancar.

BAB V

KESIMPULAN DAN SARAN

Setelah melakukan proses pembahasan hasil, maka pada bab ini akan dikemukakan oleh penulis beberapa kesimpulan dan saran yang berkaitan dengan proses *data mining*. Kesimpulan dan saran tersebut semoga dapat dijadikan masukan bagi PLN, dalam mempertimbangkan calon pemasok yang akan memasok batubara.

5.1 Kesimpulan

Kesimpulan yang didapat dari pengetahuan model prediksi C5.0, bahwa bulan tahun tiba pasokan batubara jika tidak sama dengan bulan tahun konfirmasi akan mengakibatkan keterlambatan pasokan batubara. Sedangkan atribut pendukung lainnya dalam mengidentifikasi keterlambatan batubara yaitu spek batubara, volume batubara, wilayah batubara yang mengirim, karena beberapa variable tersebut berpengaruh terhadap keterlambatan yang mengakibatkan hutang atau denda yang harus di bayar pemasok batubara.

5.2 Saran

Terbukanya pengetahuan-pengetahuan dari penggunaan *data mining* ini bahwa masih banyak informasi yang dapat digali dari data yang ada di aplikasi BBO. Supaya lebih bermanfaat untuk proses data mining berikutnya perlu di perhatikan hal-hal berikut:

1. Ketersediaan data lebih variatif, untuk kedepannya perlu dipikirkan tidak hanya bersifat dari data transaksi saja. Tapi dari kebiasaan user pemasok dalam memproses atau menjalankan aplikasi sesuai SLA yang sudah ditentukan saat kesepatan PLN dan pemasok batubara, jadi bisa terlihat kinerja pemasok dalam menjalankan aplikasi sesuai SLA, agar mempercepat pengiriman sampai dengan pembayaran.

2. Membandingkan data dari sisi konfirmasi jadwal sampai dengan batubara samapai di PLTU, rata-rata pengiriman pemasok melakukan pengiriman batubara berapa hari.

DAFTAR PUSTAKA

- Bramer, Max. 2007. *Principles of Data Mining*. London : Springer.
- DIT(DAN). 2014. mengenai Perubahan Proses Bisnis Aplikasi Batubara Online (BBO)
Surat : No. 1810/121/DIT(DAN)/2014. PT PLN(Persero).
- Fawcett T. 2006. *An introduction to ROC analysis*. Pattern Recognit. Lett.27:861–874.
- Fayyad, Usama. 1996. *Advances in Knowledge Discovery and Data Mining*. MIT Press.
- Graha, Ibnu.2016. Data Mining Jasa Pengiriman Titipan Kilat di PT Citra Van Titipan Kilat (TIKI) dengan metode Decision tree.Surakarta. Universitas Muhammadiyah Surakarta.
- Han, J.,&Kamber, M. 2006. *Data Mining Concept and Tehniques*.San Fransisco : Morgan Kauffman.
- Han J, Kamber M, Pei J. 2012. *Data Mining: Concepts and Techniques*. 3rd ed. Massachusetts (US): Morgan Kaufmann.
- Kurniawan, Defri & Wiraksono, Wibowo & Astuti, Yani Parti.2016.Penggunaan Educational Data Mining(EDM) Untuk Mendeteksi Keterlambatan Masa Studi Mahasiswa.Semarang: Universitas Dian Nuswantoro.
- Kusrini dan E. T. Luthfi. 2009. *Algoritma Data Mining*” Yogyakarta : Andi Offset.
- Larose, D. T. 2005. *Discovering Knowledge in Data*. New Jersey: John Willey & Sons, Inc.
- Maimon, Oded & Rokach, Lior. 2005. *Data Mining and Knowledge Discovey Handbook*. New York: Springer.
- Martinez, Vincent.2012.Flight Delay Prediction., Department of Computer Science, ETH Zurich.German
- Rulequest. 2012. “Data Mining Tools See5 and C5.0”. [Internet]. [Diunduh tanggal 13/05/2017]. Dapat diunduh dari: <http://rulequest.com/see5-comparison.html>.
- Schaffer, Agusti and Earle.2009. *International Business Law and its Environment, 7e*. Mason, OH: Cengage Learning.
- Septiandi, Rendi.2016. Analisis Data Keterlambatan Bahan Baku Berdasarkan Pendekatan Data Warehouse dan Pohon Keputusan C5.0.Bogor : Institut Pertanian Bogor.

- Sumathi,& S. Sivanandam, S.N. 2006. Introduction to Data Mining and its Applications.
Berlin Heidelberg New York: Springer.
- Sunarto, 2006, Pengantar Manajemen,Bandung : CV Alfabeta.
- Witten, I. H., Frank, E., & Hall, M. A. 2011. Data Mining: Practical Machine Learning and
Tools. Burlington : Morgan Kaufmann Publisher.

LAMPIRAN

LAMPIRAN A : Pengambilan data dan Pemrosesan data awal

Beberapa table yang digunakan dalam pemrosesan data :

1. Table Data Master Mitra Pemasok Batubara

No	Field	Tipe Data
1	ID_PEMASOK	INTEGER
2	NAMA_PEMASOK	VARCHAR2(200 BYTE)
3	ALAMAT	VARCHAR2(200 BYTE)
4	PIC_PRES DIR	VARCHAR2(200 BYTE)
5	PIC_PRES DIR_HP	VARCHAR2(200 BYTE)
6	PIC_DIR MARKETING	VARCHAR2(200 BYTE)
7	PIC_DIR MARKETING_HP	VARCHAR2(200 BYTE)
8	PIC_MAN_OPR	VARCHAR2(200 BYTE)
9	PIC_MAN_OPR_HP	VARCHAR2(200 BYTE)
10	STATUS_PEMASOK	VARCHAR2(200 BYTE)
11	NAMA_TAMBANG	VARCHAR2(200 BYTE)
12	PRODUKSI	VARCHAR2(200 BYTE)
13	LOADING_PORT	VARCHAR2(200 BYTE)
14	FASILITAS_LOADING	VARCHAR2(200 BYTE)
15	JARAK_HAULING	VARCHAR2(200 BYTE)
16	FLAG_BLACKLIST	VARCHAR2(200 BYTE)
17	STATUS_AKTIF	VARCHAR2(200 BYTE)
18	KETERANGAN	VARCHAR2(200 BYTE)
19	CREATEBY	VARCHAR2(200 BYTE)
20	CREATEDATE	DATE
21	UPDATEBY	VARCHAR2(200 BYTE)
22	UPDATEDATE	DATE
23	APP_ACCESS	VARCHAR2(200 BYTE)
24	ID_PUSAT	VARCHAR2(22 BYTE)
25	STATUS	VARCHAR2(200 BYTE)
26	TLP	VARCHAR2(50 BYTE)
27	FAX	VARCHAR2(50 BYTE)
28	EMAIL	VARCHAR2(50 BYTE)
29	JBT_MAN_OPR	VARCHAR2(22 BYTE)
30	NAMA_SINGKAT	VARCHAR2(20 BYTE)
31	ID_WILAYAH	VARCHAR2(50 BYTE)
32	NO_NPWP	VARCHAR2(200 BYTE)
33	NPWP	BLOB
34	NO_FAKTUR_PAJAK	VARCHAR2(200 BYTE)
35	FAKTUR_PAJAK	BLOB
36	NO_PKP	VARCHAR2(200 BYTE)
37	PKP	BLOB

38	NO_PEMBAYARAN_ROYALTI	VARCHAR2(200 BYTE)
39	PEMBAYARAN_ROYALTI	BLOB
40	NO_KWITANSI	VARCHAR2(200 BYTE)
41	KWITANSI	BLOB
42	KODE_BANK	VARCHAR2(3 BYTE)
43	NO_REKENING	VARCHAR2(100 BYTE)
44	TIPIKAL_RUMUS	INTEGER
45	NAMA_ALIAS	VARCHAR2(225 BYTE)
46	NAMA_PANGGILAN	VARCHAR2(200 BYTE)

Berisikan data-data mitra yang memasok batubara ke PLN, menggunakan Aplikasi BBO.

2. Table Data Master Pembangkit

No	Field	Tipe Data
1	ID_PLTU	INTEGER
2	NAMA_PEMBANGKIT	VARCHAR2(200 BYTE)
3	NAMA_PIC	VARCHAR2(200 BYTE)
4	ALAMAT	VARCHAR2(200 BYTE)
5	KEPEMILIKAN	VARCHAR2(200 BYTE)
6	LATITUDE	VARCHAR2(20 BYTE)
7	LONGITUDE	VARCHAR2(20 BYTE)
8	WILAYAH	VARCHAR2(200 BYTE)
9	JENIS_PEMBANGKIT	VARCHAR2(200 BYTE)
10	TELP	VARCHAR2(200 BYTE)
11	DAYA_TERPASANG	VARCHAR2(200 BYTE)
12	SFC	VARCHAR2(200 BYTE)
13	CAPACITY_FACTOR	VARCHAR2(200 BYTE)
14	HEAT_RATE	VARCHAR2(200 BYTE)
15	CREATEBY	VARCHAR2(200 BYTE)
16	CREATEDATE	DATE
17	UPDATEBY	VARCHAR2(200 BYTE)
18	UPDATEDATE	DATE
19	APP_ACCESS	VARCHAR2(200 BYTE)
20	GM_NAMA	VARCHAR2(200 BYTE)
21	GM_JABATAN	VARCHAR2(200 BYTE)
22	GM_ALAMAT	VARCHAR2(200 BYTE)
23	SHIPPER_NAME	VARCHAR2(200 BYTE)
24	SHIPPER_JABATAN	VARCHAR2(200 BYTE)
25	SHIPPER_ALAMAT	VARCHAR2(200 BYTE)
26	NAMA_SINGKAT	VARCHAR2(20 BYTE)
27	ID_UNITPENGELOLA	VARCHAR2(50 BYTE)
28	KAPASITAS_STOCKPILE	NUMBER
29	MINIMUM_STOK	NUMBER
30	HARI_OPERASI	NUMBER
31	NAMA_PENDEK	VARCHAR2(200 BYTE)
32	ID_INSTANSI	VARCHAR2(50 BYTE)

33	APV_BY	VARCHAR2(2 BYTE)
34	KOTA	VARCHAR2(200 BYTE)
35	UNIT_WILAYAH	VARCHAR2(20 BYTE)
36	ID_SPP	NUMBER
37	TOTAL_DAYA	NUMBER
38	KD_PEMBANGKIT	VARCHAR2(5 BYTE)
39	NM_PEMBANGKIT	VARCHAR2(200 BYTE)
40	KD_ENTITAS	VARCHAR2(4 BYTE)
41	TIPIKAL	NUMBER
42	ID_SITE	NUMBER

Berisikan data-data PLTU yang sudah menggunakan aplikasi BBO.

3. Table Data Jadwal Konfirmasi Rakor

No	Field	Tipe Data
1	ID_JADWAL	INTEGER
2	ID_PASOKAN	INTEGER
3	ID_PEMASOK	INTEGER
4	ID_TRANSPORTIR	INTEGER
5	ID_BAHAN_BAKAR	INTEGER
6	ID_PLTU	INTEGER
7	ID_RAKOR	INTEGER
8	NO_RAKOR	VARCHAR2(20 BYTE)
9	NO_JADWAL	VARCHAR2(20 BYTE)
10	BULAN_TAHUN	VARCHAR2(10 BYTE)
11	TANGGAL	DATE
12	TGL_TA	DATE
13	TGL_TD	DATE
14	VOLUME	INTEGER
15	CREATEBY	VARCHAR2(200 BYTE)
16	CREATEDATE	DATE
17	UPDATEBY	VARCHAR2(200 BYTE)
18	UPDATEDATE	DATE
19	APP_ACCESS	VARCHAR2(200 BYTE)
20	NO_PJBB	VARCHAR2(50 BYTE)
21	TGL_TA_CONFIRM	DATE
22	VOLUME_CONFIRM	NUMBER
23	TGL_ETD	DATE
24	ID_PJBB	NUMBER
25	TA	DATE
26	FLAG_PENGALIHAN	NUMBER
27	ID_VESSEL	NUMBER
28	ID_JADWAL_VESSEL	NUMBER

Data-data rapat kordinasi setiap bulan, yang berisikan data jadwal pengiriman batubara dan jumlah batubara yang harus dipasok oleh mitra, sesuai kesepakatan yang disepakati

antara PLN, anak perusahaan, PLTU dan mitra yang memasok ke PLTU, saat rapat kordinasi.

4. Table Data Jadwal Pengiriman

No	Field	Tipe Data
1	ID_PENGIRIMAN	INTEGER
2	ID_PEMASOK	INTEGER
3	ID_PLTU	INTEGER
4	ID_TRANSPORTIR	INTEGER
5	ID_BAHAN_BAKAR	INTEGER
6	ID_PASOKAN	INTEGER
7	ID_JADWAL	INTEGER
8	ID_KONTRAK	INTEGER
9	ID_PJBB	INTEGER
10	ID_RAKOR	INTEGER
11	NO_RAKOR	VARCHAR2(20 BYTE)
12	VOLUME	NUMBER(12,3)
13	TGL_TA	VARCHAR2(20 BYTE)
14	TGL_TD	VARCHAR2(20 BYTE)
15	NO_MANIFEST	VARCHAR2(200 BYTE)
16	NO_BL	VARCHAR2(200 BYTE)
17	TANGGAL	DATE
18	PELABUHAN	VARCHAR2(200 BYTE)
19	NO_SHIPMENT	VARCHAR2(200 BYTE)
20	NO_PENGAJUAN	VARCHAR2(200 BYTE)
21	NO_PENGIRIMAN	VARCHAR2(200 BYTE)
22	NO_JADWAL	VARCHAR2(200 BYTE)
23	PIC	VARCHAR2(200 BYTE)
24	TELP	VARCHAR2(200 BYTE)
25	KETERANGAN	VARCHAR2(200 BYTE)
26	TAHAP	VARCHAR2(200 BYTE)
27	STATUS_KETERLAMBATAN	VARCHAR2(200 BYTE)
28	BULAN_TAHUN	VARCHAR2(20 BYTE)
29	CREATEBY	VARCHAR2(200 BYTE)
30	CREATEDATE	DATE
31	UPDATEBY	VARCHAR2(200 BYTE)
32	UPDATEDATE	DATE
33	APP_ACCESS	VARCHAR2(200 BYTE)
34	POSTING	NUMBER
35	ID_KAPAL	NUMBER
36	ID_TONGKANG	NUMBER
37	ID_QUICKSAMPLING	VARCHAR2(20 BYTE)
38	ID_BA_PENERIMAAN	VARCHAR2(20 BYTE)
39	ETA_RAKOR	DATE
40	ETD_SHIPMENT	DATE

41	TD	DATE
42	ETA_REVISI	DATE
43	TA	DATE
44	PEL_MUAT	VARCHAR2(200 BYTE)
45	NO_SAMPLING	VARCHAR2(20 BYTE)
46	NO_BONGKAR	VARCHAR2(20 BYTE)
47	NO_SANDAR	VARCHAR2(20 BYTE)
48	NO_INVOICE	VARCHAR2(20 BYTE)
49	NO_SIL	VARCHAR2(200 BYTE)
50	STATUS_SIL	VARCHAR2(300 BYTE)
51	KETERANGAN_SIL	VARCHAR2(300 BYTE)
52	TGL_SIL	DATE
53	FILE_SIL	BLOB
54	NO_MANUVER	VARCHAR2(200 BYTE)
55	TGL_SIL_RESPONSE	DATE
56	NO_SAMPLING_TERIMA	VARCHAR2(20 BYTE)
57	NO_PSA_LOADING	VARCHAR2(300 BYTE)
58	TGL_PSA_LOADING	DATE
59	KETERANGAN_PSA_LOADING	VARCHAR2(300 BYTE)
60	NO_PSA_UNLOADING	VARCHAR2(300 BYTE)
61	TGL_PSA_UNLOADING	DATE
62	KETERANGAN_PSA_UNLOADING	VARCHAR2(300 BYTE)
63	NO_PIL	VARCHAR2(200 BYTE)
64	FLAG_PSA_UNLOADING	VARCHAR2(100 BYTE)
65	ID_PJAB	NUMBER
66	NO_BA_LOADING	VARCHAR2(200 BYTE)
67	TGL_BA_LOADING	DATE
68	VOLUME_BL	NUMBER
69	TGL_BL	DATE
70	TGL_MULAI_MUAT	DATE
71	TGL_SELESAI_MUAT	DATE
72	FLAG_BLN_KONFIRMASI	DATE
73	FLAG_PENGALIHAN	NUMBER
74	ID_PEL_MUAT	NUMBER
75	ID_SPEK	VARCHAR2(20 BYTE)
76	ID_MODA	INTEGER

Berisikan jadwal pengiriman yang harus dikirim pemasok ke pltu sesuai jadwal yang sudah di tentukan saat rapat kordinasi.

5. Table Data Kedatangan

No	Field	Tipe Data
1	ID_KEDATANGAN	INTEGER
2	ID_PENGIRIMAN	VARCHAR2(20 BYTE)
3	ID_UNIT_JETTY	INTEGER
4	ID_PLTU	VARCHAR2(200 BYTE)

5	NO_KEDATANGAN	VARCHAR2(80 BYTE)
6	NO_PENGIRIMAN	VARCHAR2(80 BYTE)
7	NO_SHIPMENT	VARCHAR2(80 BYTE)
8	NO_RAKOR	VARCHAR2(80 BYTE)
9	TIME	VARCHAR2(80 BYTE)
10	TGL_TD	VARCHAR2(80 BYTE)
11	PELAPOR	VARCHAR2(200 BYTE)
12	CREATEBY	VARCHAR2(200 BYTE)
13	CREATEDATE	DATE
14	UPDATEBY	VARCHAR2(200 BYTE)
15	UPDATEDATE	DATE
16	APP_ACCESS	VARCHAR2(200 BYTE)
17	TA	DATE
18	ETA	DATE
19	TD	DATE
20	FLAG_PSA	VARCHAR2(1 BYTE)
21	FLAG_PIS	VARCHAR2(1 BYTE)
22	NO_SIS	VARCHAR2(300 BYTE)
23	HR_GANGGUAN	NUMBER
24	KET_GANGGUAN	VARCHAR2(2000 BYTE)
25	HR_FORCEMAJURE	NUMBER

Data kedatangan kapal berisikan data-data kapal mulai dari kedatangan kapal, gangguan saat kedatangan atau gangguan karena akibat bencana alam, semua data di catat dan masuk ke table kedatangan.

6. Table Data Pembongkaran

No	Field	Tipe Data
1	ID_BONGKAR	INTEGER
2	ID_PENGIRIMAN	INTEGER
3	ID_KEDATANGAN	INTEGER
4	ID LENGKAP_DOK	VARCHAR2(30 BYTE)
5	ID_TRANSPORTIR	INTEGER
6	ID_UNIT_JETTY	INTEGER
7	ID_JENIS_ALAT	INTEGER
8	ID_PLTU	VARCHAR2(20 BYTE)
9	NO_PENGIRIMAN	VARCHAR2(40 BYTE)
10	NO_RAKOR	VARCHAR2(20 BYTE)
11	NO_SHIPMENT	VARCHAR2(40 BYTE)
12	NO_BONGKAR	VARCHAR2(40 BYTE)
13	TIME_MERAPAT	DATE
14	TANGGAL	VARCHAR2(20 BYTE)
15	TIME_BONGKAR	DATE
16	ALAT_BONGKAR	VARCHAR2(200 BYTE)
17	CREATEBY	VARCHAR2(200 BYTE)
18	CREATEDATE	DATE

19	UPDATEBY	VARCHAR2(200 BYTE)
20	UPDATEDATE	DATE
21	APP_ACCESS	VARCHAR2(200 BYTE)
22	TIME_SELESAI	DATE
23	VOLUME	NUMBER
24	DRAUGHT_SURVEY_FILE	BLOB
25	BOL	VARCHAR2(10 BYTE)
26	SKAB	VARCHAR2(10 BYTE)
27	MANIFEST	VARCHAR2(10 BYTE)
28	KET_SANDAR	VARCHAR2(255 BYTE)
29	NO_SIB	VARCHAR2(100 BYTE)
30	TGL_SIB	DATE
31	VOLUME_DRAUGHT_SURVEY	NUMBER
32	NO_BA_BONGKAR	VARCHAR2(100 BYTE)
33	TGL_BA_BONGKAR	DATE
34	TENDERED	DATE
35	ACCEPTED	DATE
36	COMMENDED	DATE
37	VOLUME_DRY_EMPTY	NUMBER
38	PEJABAT_BONGKAR	VARCHAR2(100 BYTE)
39	RPSURVEYOR	NUMBER
40	HR_GANGGUAN	VARCHAR2(20 BYTE)
41	CASE_OFF	DATE

Berisikan data pembongkaran mulai dari kapal mulai bongkar sampai selesai dan jumlah batubara yang dibongkar.

7. Table Data Master Spek

No	Field	Tipe Data
1	ID_SPEK	INTEGER
2	NAMA_SPEK	INTEGER
3	CREATEBY	VARCHAR2(20 BYTE)
4	CREATEDATE	DATE

Berisikan data master Nilai kalori batubara yang mulai dari kalori 3900, 7000, 4000, 4200, 4300, 4400, 4600, 4800, 4900, 5000, 5100, 5400, 5700, 6100, 6300, 4250.

8. Table Data Master Moda

No	Field	Tipe Data
1	ID_MODA	NUMBER
2	MODA	VARCHAR2(20 BYTE)
3	KETERANGAN	VARCHAR2(100 BYTE)
4	CREATEBY	VARCHAR2(20 BYTE)
5	CREATEDATE	DATE
6	VOLUME_MIN	NUMBER

7	VOLUME_MAX	NUMBER
---	------------	--------

Berisikan data master ukuran tongkang yang masing-masing memiliki kapasitas maksimal daya angkut. Mulai dari ukuran tongkang 230 ft, 270 ft, 300 ft, 330 ft.

9. Table Data Saldo Pasokan Batubara

No	Field	Tipe Data
1	ID_PENGIRIMAN	INTEGER
2	ID_PEMASOK	INTEGER
3	ID_PLTU	INTEGER
4	STATUS	CHAR(1 BYTE)
5	NO_URUT	NUMBER
6	TAHAP_KE	NUMBER
7	VOL_DS	NUMBER
8	VOL_CONFIRM	NUMBER
9	VOL_SALDO_RAKOR	NUMBER
10	VOL_SISA	NUMBER
11	VOL_REALISASI	NUMBER
12	TA	DATE
13	BLNTHN_CONFRIM	VARCHAR2(8 BYTE)
14	BLNTHN_TIBA	VARCHAR2(8 BYTE)
15	BLNTHN_HUT	VARCHAR2(8 BYTE)
16	THBL_CONFIRM	NUMBER
17	THBL_TIBA	NUMBER
18	THBL_HUT	NUMBER
19	NO_RAKOR	VARCHAR2(20 BYTE)
20	IS_APPROVAL	NUMBER
21	CREATEDATE	DATE
22	IS_CSR	VARCHAR2(20 CHAR)
23	VOL_ALIH	NUMBER
24	IS_SPLIT	VARCHAR2(1 BYTE)
25	IS_LAST_SHIPMENT	VARCHAR2(1 BYTE)
	IS_TOLERANSI	VARCHAR2(1 BYTE)

Data saldo pasokan batubara berisikan log data pasokan batubara dari mitra ke PLN mulai dari pasokan yang lancar dan pasokan yang mengalami keterlambatan, mulai dari keterlambatan pengiriman sampai volume batubara yang dikirim kurang.

Kemudian table-tabel tersebut digabungkan, dan diambil beberapa variable yang akan dijadikan parameter untuk pengolahan data menggunakan *data mining*.

The screenshot shows the Toad for Oracle interface. The SQL editor contains a query that selects data from various tables in the BBOPROD28 schema. The Data Grid displays the results of this query, showing columns for ID_JADWAL, ID_PASOKAN, ID_PEMASOK, ID_TRANSPORTIR, ID_BAHAN_BAKAR, ID_PLTU, ID_RAKOR, NO_RAKOR, NO_JADWAL, BULAN_TAHUN, TANGGAL, TGL_TA, TGL_TD, VOLUME, CREATEBY, CREATEDATE, and UP.

ID_JADWAL	ID_PASOKAN	ID_PEMASOK	ID_TRANSPORTIR	ID_BAHAN_BAKAR	ID_PLTU	ID_RAKOR	NO_RAKOR	NO_JADWAL	BULAN_TAHUN	TANGGAL	TGL_TA	TGL_TD	VOLUME	CREATEBY	CREATEDATE	UP
14292	5361				3	1105	RRK02161105	SCH03164292	16/03/2016	06:43:35	20/03/2016		MAARIF PKB	16/03/2016	06:43:35	
14707	244				5	1106	RRK03161106	SCH04164707	12/04/2016	08:45:10	05/04/2016		MAARIF PKB	12/04/2016	08:45:10	
14578	244				1	1106	RRK03161106	SCH04164578	11/04/2016	15:05:42	12/04/2016		MAARIF PKB	11/04/2016	15:05:42	
15155	2923				24	1107	RRK04161107	SCH05165155	18/05/2016	00:39:06	07/05/2016		MAARIF PKB	18/05/2016	00:39:06	
15511	159				8	1109	RRK06161109	SCH06165511	07/06/2016	10:12:43	02/07/2016	23/06/2016	7200	MAARIF PKB	07/06/2016	10:12:43
15764	2923				24	1109	RRK06161109	SCH06165764	07/06/2016	17:13:36	28/07/2016	13/07/2016	7500	MAARIF PKB	07/06/2016	17:13:36
15965	159				3	1110	RRK07161110	SCH07165965	29/07/2016	14:14:04	03/08/2016	26/07/2016	7200	MAARIF PKB	29/07/2016	14:14:04
15992	2923				24	1109	RRK06161109	SCH07165992	21/07/2016	11:23:09	13/07/2016		MAARIF PKB	21/07/2016	11:23:09	
15950	2923				5	1110	RRK07161110	SCH07165950	29/07/2016	09:46:57	04/08/2016	24/07/2016	7500	MAARIF PKB	29/07/2016	09:46:57

Contoh bentuk data yang sudah di proses cleansing dan di transformasi sesuai aturan *data mining*. Query ke database :

```

SELECT I.NAMA_PALINGILAN AS PEMASOK,
TO_CHAR(A.TGL_TA_CONFIRM,'MON YYYY') AS BLTH_CONFIRM, TO_CHAR
(D.TA,'MON YYYY') AS BLTH_TIBA, C.WILAYAH, G.NAMA_SPEK AS SPEK, N.MODA,
A.VOLUME_CONFIRM, M.VOLUME_DRAUGHT_SURVEY, K.STATUS FROM
BBOPROD28.TB_JADWAL A, BBOPROD28.TB_PJBB B,
BBOPROD28.TB_MASTER_WILAYAH C, BBOPROD28.TB_REAL_PENGIRIMAN D,
BBOPROD28.TB_MASTER_SPEK G, BBOPROD28.TB_MASTER_PEMASOK I,
BBOPROD28.TB_SALDO_PASOKAN_DT K, BBOPROD28.TB_BONGKAR M,
BBOPROD28.TB_MASTER_MODA N WHERE A.NO_JADWAL = D.NO_JADWAL
AND D.ID_PJBB = B.ID_PJBB(+) AND D.ID_PEMASOK = B.ID_PEMASOK AND
D.ID_PLTU = B.ID_PLTU(+) AND B.ID_WILAYAH = C.ID AND D.SPEK = G.ID_SPEK
AND I.ID_PEMASOK(+) = D.ID_PEMASOK AND A.TGL_TA_CONFIRM BETWEEN
TO_DATE ('012017', 'MMYYYY') AND TO_DATE ('072017', 'MMYYYY') AND D.TGL_BL
>= (TO_DATE ('2016', 'YYYY')) AND D.ID_PEMASOK IN (SELECT ID_PEMASOK
FROM BBOPROD28.TB_MASTER_PEMASOK WHERE KETERANGAN NOT LIKE
'%FOB%') AND D.ID_PENGIRIMAN = K.ID_PENGIRIMAN AND D.ID_PENGIRIMAN =
M.ID_PENGIRIMAN AND NVL(D.ID_MODA,12) = N.ID_MODA AND
K.BLNTHN_CONFRIM IS NOT NULL AND K.BLNTHN_CONFRIM NOT IN('NOV

```

2016','DEC 2016') AND K.BLNTHN_TIBA IS NOT NULL ORDER BY D.ID_PENGIRIMAN ASC.

The screenshot shows a database query tool interface. The top panel displays a SQL query:


```
SELECT I.NAMA_PALINGILAN AS PEMASOK,
       TO_CHAR(A.TGL_TA_CONFIRM, 'MON YYYY') AS BLTH_CONFIRM,
       TO_CHAR(D.TA, 'MON YYYY') AS BLTH_TIBA,
       C.WILAYAH,
       G.NAMA_SPEK AS SPEK,
       N.MODA,
       A.VOLUME_CONFIRM,
       M.VOLUME_DRAUGHT_SURVEY,
       K.STATUS
FROM BBOPROD28.TB_JADWAL A,
     BBOPROD28.TB_PJBB B,
     BBOPROD28.TB_MASTER_WILAYAH C,
     BBOPROD28.TB_REAL_PENGIRIMAN D,
```

 The bottom panel shows the results in a Data Grid with the following columns: PEMASOK, BLTH_CONFIRM, BLTH_TIBA, WILAYAH, SPEK, MODA, VOLUME_CONFIRM, VOLUME_DRAUGHT_SURVEY, STATUS. The data rows are as follows:

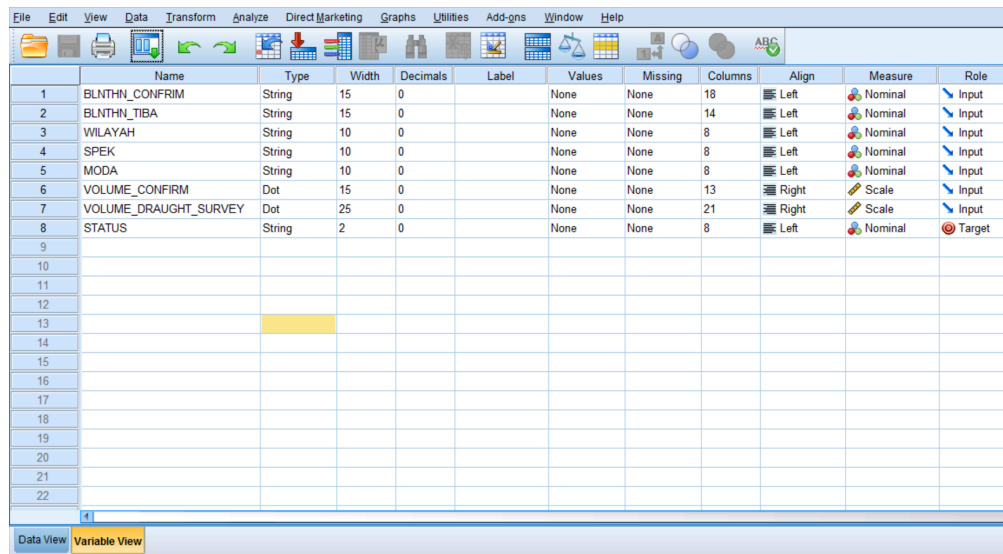
PEMASOK	BLTH_CONFIRM	BLTH_TIBA	WILAYAH	SPEK	MODA	VOLUME_CONFIRM	VOLUME_DRAUGHT_SURVEY	STATUS
PT ARUTMIN	JAN 2017	JAN 2017	Wilayah 2	4200	330 ft	9500	15241,201 L	
PT ADARO	JAN 2017	JAN 2017	Wilayah 2	MRC	-	10000	10349,9 L	
PT ADARO	JAN 2017	JAN 2017	Wilayah 2	MRC	-	10000	10314,922 L	
PT ADARO	JAN 2017	JAN 2017	Wilayah 2	MRC	-	7500	9084,583 L	
PT DGL	JAN 2017	JAN 2017	Wilayah 2	4200	-	7500	7528,845 L	
PT DGL	JAN 2017	JAN 2017	Wilayah 2	LRC	-	7500	7529,744 L	
PT EEI	JAN 2017	JAN 2017	Wilayah 2	4300	300 ft	7500	7355,677 L	

LAMPIRAN B : Tahapan-tahapan pembuatan model prediksi

1. Data yang sudah di export ke microsoft excel

1	PEMASOK	BLTH_CONFIRM	BLTH_TIBA	WILAYAH	SPEK	MODA	VOLUME_CONFIRM	VOLUME_DRAUGHT_SURVEY	STATUS
2	PT ARUTMIN	JAN 2017	JAN 2017	Wilayah 2	4200	330 ft	9.500	15.241 L	
3	PT ADARO	JAN 2017	JAN 2017	Wilayah 2	MRC	-	10.000	10.350 L	
4	PT ADARO	JAN 2017	JAN 2017	Wilayah 2	MRC	-	10.000	10.315 L	
5	PT ADARO	JAN 2017	JAN 2017	Wilayah 2	MRC	-	7.500	9.085 L	
6	PT DGL	JAN 2017	JAN 2017	Wilayah 2	4200	-	7.500	7.529 L	
7	PT DGL	JAN 2017	JAN 2017	Wilayah 2	LRC	-	7.500	7.530 L	
8	PT EEI	JAN 2017	JAN 2017	Wilayah 2	4300	300 ft	7.500	7.356 L	
9	PT EEI	JAN 2017	FEB 2017	Wilayah 2	4300	300 ft	7.500	7.498 H	
10	PT PLN BATUBARA	JAN 2017	JAN 2017	Wilayah 2	LRC	-	28.000	28.001 L	
11	PT ARUTMIN	JAN 2017	JAN 2017	Wilayah 2	4200	330 ft	9.500	12.513 L	
12	PT ARUTMIN	JAN 2017	JAN 2017	Wilayah 2	4200	330 ft	9.500	9.521 L	
13	PT ADARO	JAN 2017	JAN 2017	Wilayah 2	MRC	-	7.500	7.608 L	
14	PT ADARO	JAN 2017	JAN 2017	Wilayah 2	MRC	-	7.500	7.527 L	
15	PT DGL	JAN 2017	JAN 2017	Wilayah 2	LRC	-	7.500	7.518 L	
16	PT ADARO	JAN 2017	JAN 2017	Wilayah 2	MRC	-	10.000	10.000 L	
17	PT ADARO	JAN 2017	JAN 2017	Wilayah 2	MRC	-	7.500	9.036 L	
18	PT PLN BATUBARA	JAN 2017	JAN 2017	Wilayah 2	4400	300 ft	7.500	7.563 L	
19	PT ADARO	JAN 2017	JAN 2017	Wilayah 2	MRC	-	7.500	7.533 L	
20	PT PLN BATUBARA	JAN 2017	JAN 2017	Wilayah 2	LRC	-	5.000	4.836 L	
21	PT ADARO	JAN 2017	JAN 2017	Wilayah 2	MRC	-	10.000	11.557 L	
22	PT ADARO	JAN 2017	JAN 2017	Wilayah 2	MRC	-	10.000	10.330 L	
23	PT ADARO	JAN 2017	JAN 2017	Wilayah 2	MRC	-	10.000	10.072 L	
24	PT ADARO	JAN 2017	JAN 2017	Wilayah 2	4200	330 ft	9.500	12.001 L	

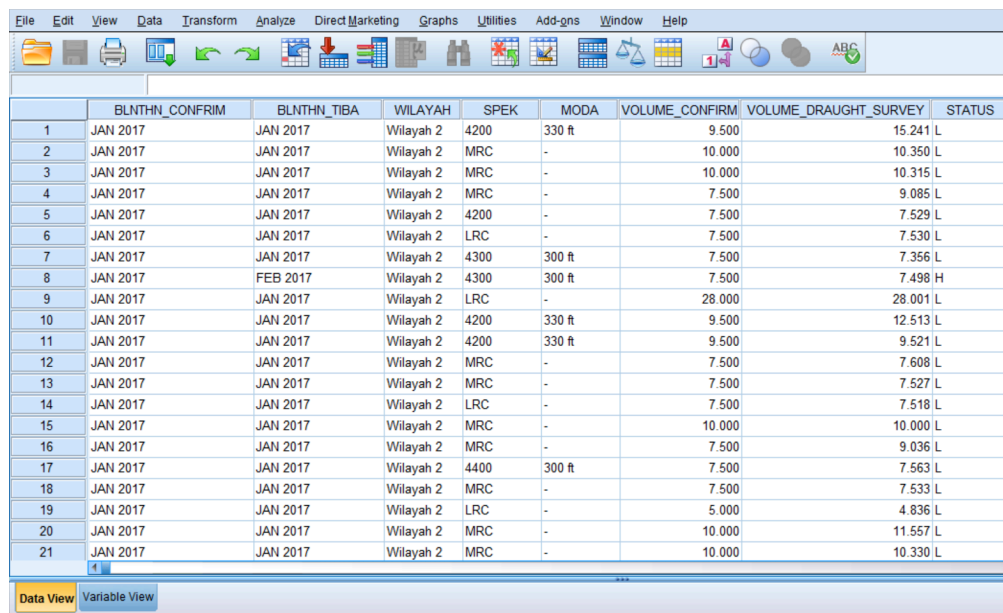
2. Pembuatan deklarasi *variable* di IBM SPSS Statistics



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	BLNTHN_CONFIRM	String	15	0		None	None	18	Left	Nominal	Input
2	BLNTHN_TIBA	String	15	0		None	None	14	Left	Nominal	Input
3	WILAYAH	String	10	0		None	None	8	Left	Nominal	Input
4	SPEK	String	10	0		None	None	8	Left	Nominal	Input
5	MODA	String	10	0		None	None	8	Left	Nominal	Input
6	VOLUME_CONFIRM	Dot	15	0		None	None	13	Right	Scale	Input
7	VOLUME_DRAUGHT_SURVEY	Dot	25	0		None	None	21	Right	Scale	Input
8	STATUS	String	2	0		None	None	8	Left	Nominal	Target
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											

Data View Variable View

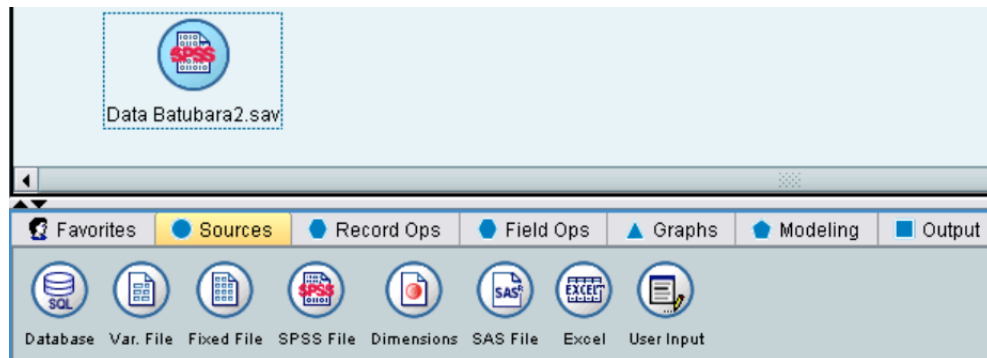
3. Masukkan Data ke dari excel ke SPSS



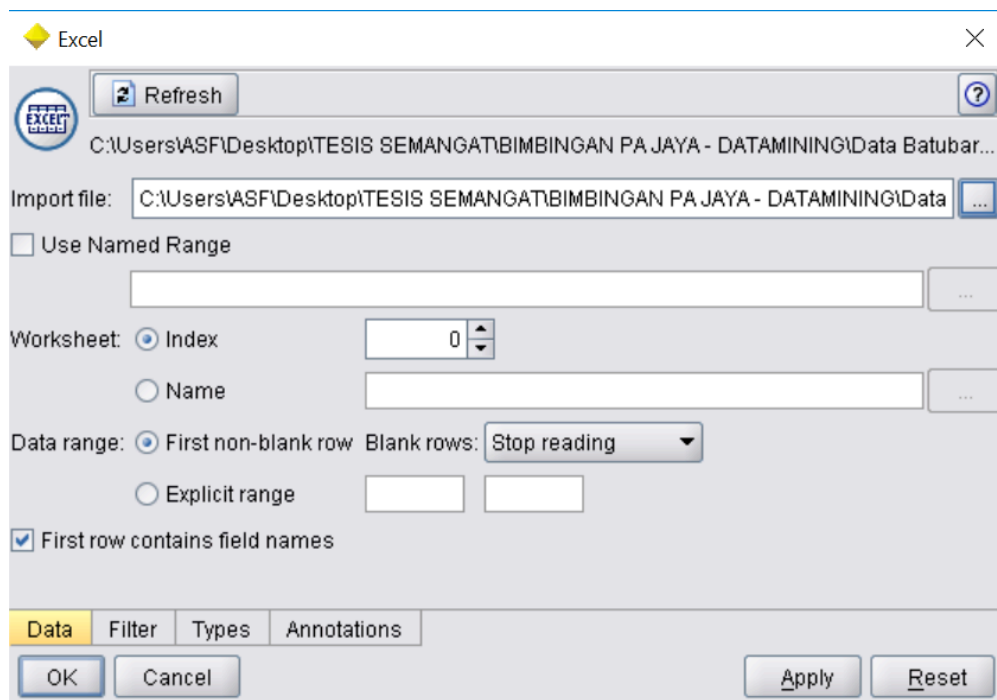
	BLNTHN_CONFIRM	BLNTHN_TIBA	WILAYAH	SPEK	MODA	VOLUME_CONFIRM	VOLUME_DRAUGHT_SURVEY	STATUS
1	JAN 2017	JAN 2017	Wilayah 2	4200	330 ft	9.500	15.241 L	
2	JAN 2017	JAN 2017	Wilayah 2	MRC	-	10.000	10.350 L	
3	JAN 2017	JAN 2017	Wilayah 2	MRC	-	10.000	10.315 L	
4	JAN 2017	JAN 2017	Wilayah 2	MRC	-	7.500	9.085 L	
5	JAN 2017	JAN 2017	Wilayah 2	4200	-	7.500	7.529 L	
6	JAN 2017	JAN 2017	Wilayah 2	LRC	-	7.500	7.530 L	
7	JAN 2017	JAN 2017	Wilayah 2	4300	300 ft	7.500	7.356 L	
8	JAN 2017	FEB 2017	Wilayah 2	4300	300 ft	7.500	7.498 H	
9	JAN 2017	JAN 2017	Wilayah 2	LRC	-	28.000	28.001 L	
10	JAN 2017	JAN 2017	Wilayah 2	4200	330 ft	9.500	12.513 L	
11	JAN 2017	JAN 2017	Wilayah 2	4200	330 ft	9.500	9.521 L	
12	JAN 2017	JAN 2017	Wilayah 2	MRC	-	7.500	7.608 L	
13	JAN 2017	JAN 2017	Wilayah 2	MRC	-	7.500	7.527 L	
14	JAN 2017	JAN 2017	Wilayah 2	LRC	-	7.500	7.518 L	
15	JAN 2017	JAN 2017	Wilayah 2	MRC	-	10.000	10.000 L	
16	JAN 2017	JAN 2017	Wilayah 2	MRC	-	7.500	9.036 L	
17	JAN 2017	JAN 2017	Wilayah 2	4400	300 ft	7.500	7.563 L	
18	JAN 2017	JAN 2017	Wilayah 2	MRC	-	7.500	7.533 L	
19	JAN 2017	JAN 2017	Wilayah 2	LRC	-	5.000	4.836 L	
20	JAN 2017	JAN 2017	Wilayah 2	MRC	-	10.000	11.557 L	
21	JAN 2017	JAN 2017	Wilayah 2	MRC	-	10.000	10.330 L	

Data View Variable View

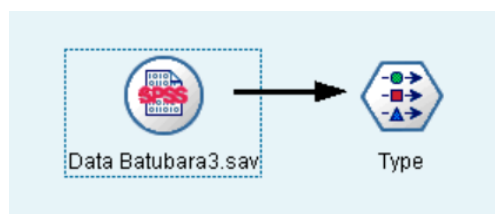
4. Pemilihan *source*



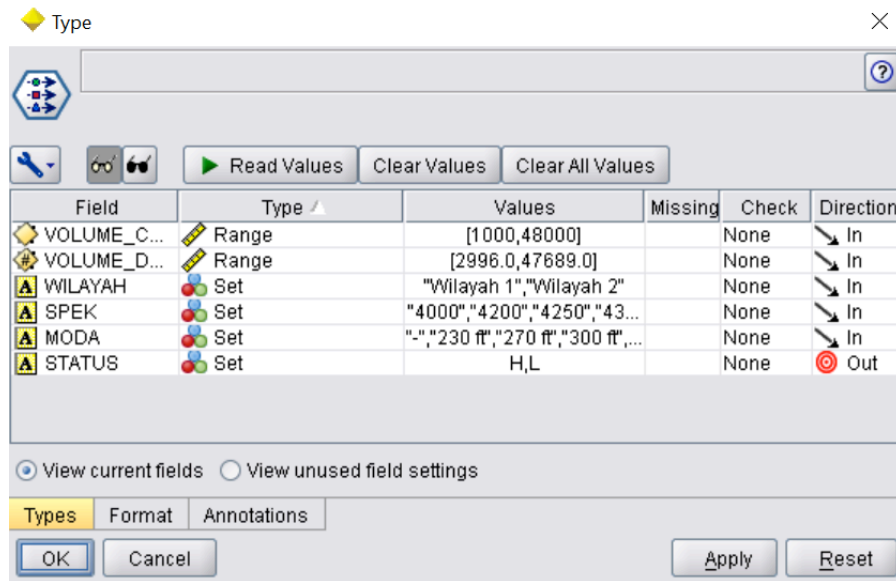
5. Pemilihan *source* untuk di *import*



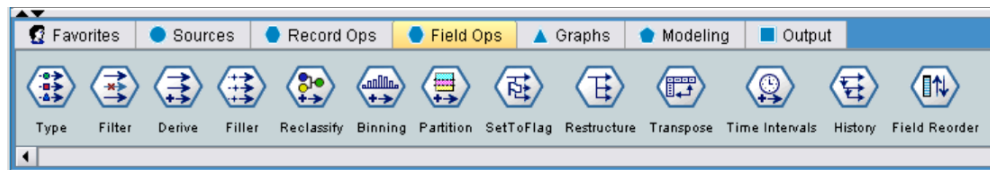
6. Pemilihan penentuan *Type Direction*



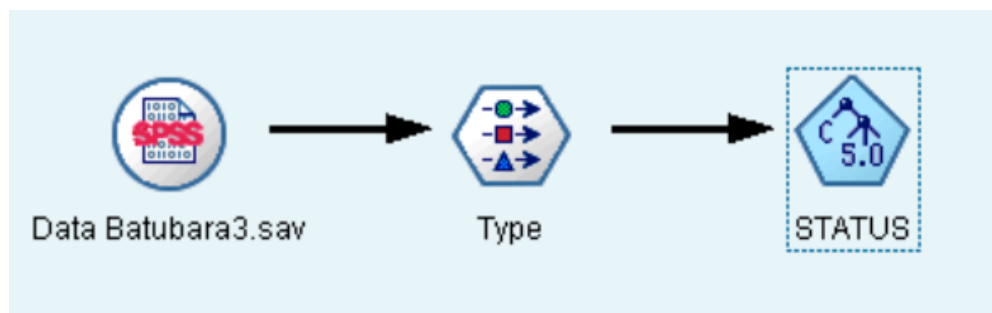
7. Hasil Penentuan *Type Direction*



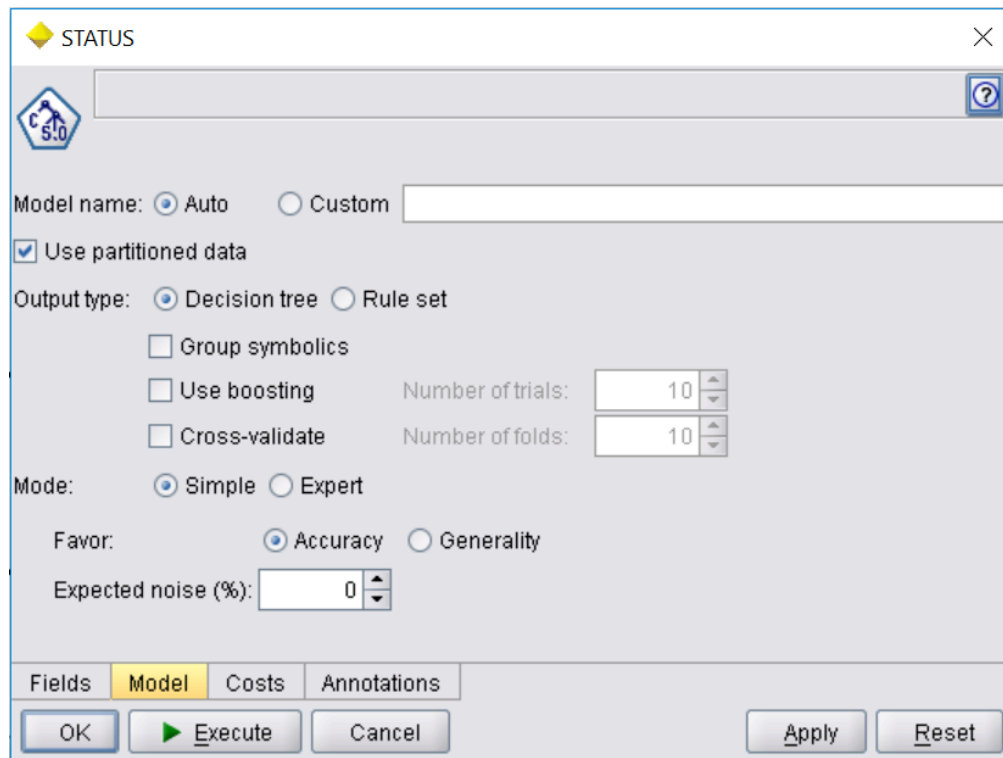
8. Penentuan *field ops*, untuk penentuan variable yang akan di proses



9. Tampilan sesudah melakukan pemilihan model prediksi C5



10. Exsekusi dan pemilihan model



The STATUS dialog box is shown with the following settings:

- Model name:** ☒ Auto ☐ Custom
- ☒ Use partitioned data
- Output type:** ☒ Decision tree ☐ Rule set
 - ☐ Group symbolics
 - ☐ Use boosting
 - ☐ Cross-validate
- Mode:** ☒ Simple ☐ Expert
- Favor:** ☒ Accuracy ☐ Generality
- Expected noise (%):** 0
- Number of trials:** 10
- Number of folds:** 10

At the bottom, there are tabs for Fields, Model (selected), Costs, and Annotations. Below the tabs are buttons for OK, Execute, Cancel, Apply, and Reset.

11. Hasil

