

Realtime Business Intelligence Menggunakan Algoritma Apriori dengan Data Stream Mining **(Studi Kasus: Penjadwalan PT. Citra Tiara Global)**

Fakhrian Fadlia Adiwijaya¹, Ana Hadiana²

¹Magister Sistem Informasi, Fakultas Pasca Sarjana
Universitas Komputer Indonesia

Jalan Dipati Ukur no. 112 - 116, Bandung, Jawa Barat, Indonesia

✉ : tkj.fakhrian@gmail.com

²Pusat Penelitian Informatika Lembaga Ilmu Penelitian (LIPI)

Jl. Cisitu, Sangkuriang, Bandung Jawa Barat, Indonesia

✉ : anahadiana68@gmail.com

Abstrak — Proses penjadwalan yang dilakukan oleh PT. Citra Tiara Global saat ini masih mengandalkan kedatangan kendaraan untuk setiap keberangkatannya, hal ini dikarenakan keterbatasan lahan parkir yang mengakibatkan jumlah kendaraan yang dapat ditampung di setiap cabang terbatas. Hal tersebut memaksa manajemen untuk memaksimalkan jadwal keberangkatan, agar tidak terjadi penumpukan kendaraan di setiap cabangnya. Data keberangkatan yang ada di PT. Citra Tiara Global saat ini hanya digunakan untuk melakukan rekapitulasi dan evaluasi terhadap keberangkatan. Dengan menggunakan algoritma apriori, data keberangkatan yang ada dapat digunakan untuk menggali informasi prediksi keterlambatan dan prediksi jumlah penumpang berdasarkan kriteria tertentu. Informasi prediksi yang diberikan akan diberikan secara realtime, dengan proses update data menggunakan metode CDC Push dan proses data mining menggunakan data stream mining. Pengimplementasian realtime business intelligence menggunakan algoritma apriori dengan data stream mining dapat membantu proses penjadwalan di PT. Citra Tiara Global dengan memberikan prediksi keterlambatan kendaraan dan prediksi jumlah penumpang. Berdasarkan hasil pengujian yang telah dilakukan, keakurasian prediksi berada antara 44% hingga 79% dengan minimum support yang digunakan bernilai 5%.

Kata Kunci — Association Rule, Apriori, Business Intelligence, Change Data Capture, Penjadwalan

I. PENDAHULUAN

PT. Citra Tiara Global atau lebih dikenal dengan Cititrans Travel merupakan perusahaan yang bergerak di bidang travel dengan tujuan Bandung-Jakarta yang berfokus kepada penjualan tiket keberangkatan. Cititrans berkantor pusat di Jl. Dipatiukur No.53 memiliki lahan parkir yang terbatas sehingga jumlah kendaraan yang dapat ditampung hanya sedikit dan sebagian kendaraan yang tidak digunakan di parkir di lokasi yang jauh, jumlah kendaraan yang sedikit juga berpengaruh terhadap keberangkatan untuk setiap tujuannya. Dengan kondisi tersebut membuat manajemen perlu untuk memaksimalkan penjadwalan agar tidak ada penumpukan kendaraan dan jumlah kendaraan yang ada dapat memenuhi kebutuhan untuk setiap tujuan keberangkatan. Jumlah penumpang yang menggunakan travel pada setiap keberangkatannya sangat beragam yang dipengaruhi oleh hari libur kantor dan hari-hari besar, akan lebih optimal apabila pada setiap keberangkatan dapat menutup pengeluaran untuk setiap keberangkatannya.

Data keberangkatan yang ada di PT. Citra Tiara Global saat ini hanya digunakan untuk melakukan rekapitulasi dan evaluasi terhadap keberangkatan untuk setiap tujuannya. Agar data tersebut dapat membantu dalam proses penjadwalan diperlukan *business intelligence* yang digunakan untuk mendapatkan pengetahuan dari operasi bisnis yang dilakukan,

pada kasus ini menggunakan *data mining* untuk menggali informasi prediksi keterlambatan berdasarkan kriteria waktu keberangkatan, keterlambatan keberangkatan, keterlambatan tiba, asal keberangkatan, tujuan keberangkatan dan jumlah penumpang untuk setiap keberangkatan. metode data mining yang tepat adalah *association rule* menggunakan algoritma *apriori* untuk mencari keterkaitan antar atribut tersebut.

Algoritma *apriori* membutuhkan waktu proses yang lama serta membutuhkan kapasitas *memory* yang besar, hal ini disebabkan proses mengeksplorasi data yang dilakukan dengan cara mencari hubungan antar atribut (G. Piatetsky, 1992). Algoritma *apriori* dapat menemukan kombinasi antar atribut dengan cepat apabila nilai *minimum support* yang ditentukan cenderung besar (R. Agrawal, 1994), namun dengan *minimum support* yang besar akan menghasilkan *rulebase* yang sedikit. Penentuan nilai *minimum support* yang kecil dapat menyebabkan waktu proses yang lama. *Data stream mining* memiliki tujuan untuk mengatasi masalah dari pemrosesan jumlah data yang banyak dan diproduksi secara terus menerus, sehingga membutuhkan waktu lama dan dapat menghabiskan *memory* komputer (Albert Bifet, 2009). Dengan metode *data stream mining*, diharapkan sistem *business intelligence* dapat menghasilkan analisis BI dengan tingkat keterlambatan yang rendah dengan membantu analisis

II. TINJAUAN PUSTAKA

Bagian ini menjelaskan dasar-dasar landasan teori yang digunakan dalam penelitian ini

A. Business Intelligence

Business intelligence (BI) adalah sebuah proses analisis yang mengubah data internal dan data eksternal menjadi informasi tentang kapabilitas, posisi pasar, aktifitas dan tujuan yang harus dicapai oleh perusahaan untuk tetap kompetitif. BI berdiri berdasarkan konsep sistem informasi seperti *Online Analytical Processing* (OLAP), pembuatan *query*, pelaporan, dan *data mining* yang menyediakan metode-metode berbeda untuk sebuah analisis data bisnis (J. Schiefer, 2005).

BI bertujuan untuk menyajikan berbagai informasi yang disesuaikan dengan kebutuhan setiap penggunanya. Informasi tersebut bisa berasal dari mana saja, kemudian data tersebut diolah dan disajikan dalam bentuk informasi yang dapat dimengerti oleh penggunanya untuk memenuhi tujuan perusahaan.

B. Jenis-jenis Business Intelligence

Terdapat lima jenis atau kategori dari business intelligence, yaitu :

1. Enterprise Reporting, digunakan untuk menghasilkan laporan-laporan statis yang di distribusikan ke banyak orang. Jenis laporan ini sangat sesuai untuk laporan operasional dan dashboard.
2. Cube Analysis, digunakan untuk menyediakan analisis OLTP multidimensional yang ditujukan untuk manajer bisnis dalam lingkungan terbatas.
3. Ad-Hoc Query and Analysis, digunakan untuk memberikan akses kepada pengguna agar dapat melakukan query pada basis data, dan menggali informasi sampai pada tingkat paling dasar dari informasi transaksional. Query ini berfungsi untuk menjelajahi informasi yang dilakukan oleh pengguna.
4. Statistical Analysis and Data Mining, digunakan untuk melakukan analisis prediksi atau menentukan korelasi sebab akibat diantara dua matrik.
5. Delivery Report and Alert, digunakan secara proaktif untuk mengirimkan laporan secara lengkap atau memberikan peringatan kepada populasi pengguna yang besar atau banyak.

Pada penelitian ini digunakan jenis atau kategori Statistical Analysis and Data Mining. Proses yang digunakan untuk menganalisa dan menjelajahi informasi mengenai prediksi delay kendaraan berdasarkan korelasi antara jadwal keberangkatan, delay tiba, delay berangkat, dan tujuan.

C. Real-Time System

Real-time system (sistem waktu nyata) harus menghasilkan respon yang tepat dalam batas waktu yang telah ditentukan. Jika respon melewati batas waktu tersebut, maka terjadi degradasi performansi atau kegagalan sistem. Sebuah sistem real-time adalah sistem yang kebenarannya secara logis

didasarkan pada kebenaran hasil keluaran sistem dan ketepatan waktu hasil tersebut saat dikeluarkan (Willey, 2004). menurut fungsinya sistem ini terbagi menjadi hard real-time dan soft real-time.

Sistem hard real-time dibutuhkan untuk menyelesaikan critical task dengan jaminan waktu tertentu. Jika kebutuhan waktu tidak terpenuhi, maka aplikasi akan gagal. Secara umum, sebuah proses di kirim dengan sebuah pernyataan jumlah waktu yang dibutuhkan untuk menyelesaikan atau menjalankan proses input / output.

Sistem *soft real-time* memiliki sedikit kelonggaran waktu dibandingkan dengan sistem *hard real-time*, karena perangkat lunak hanya mendapatkan prioritas dalam menjalankan prosesnya, sehingga risiko kegagalan tidak berakibat fatal. Adanya keterlambatan waktu respon akan mengakibatkan penurunan kinerja dan mengakibatkan waktu tunda yang lebih lama.

D. Algoritma Apriori

Algoritma apriori merupakan salah satu algoritma yang digunakan untuk melakukan pencarian jumlah kemunculan atribut. Algoritma ini menggunakan pengetahuan kemunculan suatu atribut yang telah diketahui sebelumnya, untuk melakukan penggalan informasi pada proses selanjutnya. Pada algoritma apriori, menentukan kandidat yang mungkin muncul dengan cara memperhatikan minimum support dan minimum confidence. Nilai support adalah nilai pengujung atau persentase kombinasi atribut dalam basis data.

Untuk mendapatkan nilai support untuk sebuah atribut, digunakan rumus sebagai berikut :

$$\text{Support}(A) = \left(\frac{\text{Jumlah transaksi mengandung } A}{\text{Total Transaksi}} \right) \times 100\% \quad (1)$$

Sedangkan untuk mencari nilai support untuk dua item, digunakan rumus sebagai berikut :

$$\text{Support}(A, B) = \left(\frac{\text{Jumlah transaksi mengandung } A \text{ dan } B}{\text{Total Transaksi}} \right) \times 100\% \quad (2)$$

Sedangkan nilai confidence adalah nilai kepastian, yaitu kuatnya hubungan antar item. Nilai confidence dapat dicari setelah pola frekuensi munculnya sebuah item ditemukan. Rumus yang digunakan untuk mencari nilai confidence adalah sebagai berikut :

$$\text{Confidence} = \frac{\text{Jumlah transaksi yang mengandung } A \text{ dan } B}{\text{Jumlah transaksi yang mengandung } A} \quad (3)$$

Proses utama yang dilakukan dalam algoritma apriori untuk mendapatkan frequent itemset yaitu :

1. Join (Penggabungan)
Proses ini dilakukan dengan cara mengkombinasikan suatu item dengan item yang lainnya hingga tidak dapat terbentuk kombinasi lagi.
2. Prune (Pemangkasan)

Hasil dari item yang telah dikombinasikan, kemudian dipangkas dengan menggunakan minimum support yang telah ditentukan.

item A dan item B benar saling terkait dan muncul atau dibeli secara bersamaan.

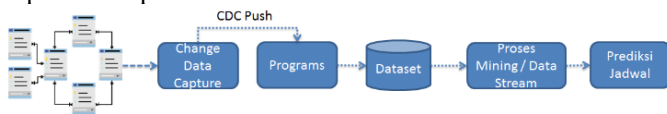
E. Change Data Capture (CDC)

CDC adalah pendekatan inovasi integrasi data, berdasarkan identifikasi, menangkap, dan mengirimkan perubahan yang dibuat oleh data sumber. Dengan memproses hanya perubahannya, CDC membuat proses integrasi data lebih efisien dan mengurangi biaya dengan mengurangi latensi (Attunity, 2006). Terdapat dua model skenario CDC (Corporation, 2005) yaitu (1) *CDC Pull* : secara periodik meminta perubahan data, setiap waktu menerima sekumpulan *record* yang merepresentasikan semua perubahan yang ditangkap sejak siklus permintaan terakhir. Skenario ini hanya menangkap dan memindahkan hanya data yang berubah saja. (2) *CDC Push* : mekanisme pengiriman perubahan data segera setelah perubahannya terjadi. Metode ini menggunakan *listeners* yang menunggu *event* perubahan dan *publisher* yang digunakan untuk mengirim dan memberikan notifikasi perubahan secara *real-time*.

Pada penelitian ini menggunakan metode *CDC push*, setiap kegiatan *insert*, *update*, atau *delete* yang terjadi pada tabel sumber akan dipindahkan ke tabel CDC yang tepat.

III. METODELOGI PENELITIAN

Metodologi yang digunakan dalam perancangannya sistem *real-time* BI untuk membantu proses penjadwalan kendaraan di PT. Citra Tiara Global terdapat beberapa tahapan yaitu (1) Analisis Sumber Data. Pada tahap ini dilakukan proses analisis data dari tabel sumber yang digunakan, dengan memilih tabel-tabel yang terlibat langsung dalam proses penjadwalan. (2) Preprocessing data. Pada tahap ini dilakukan proses pemindahan data menggunakan metode *CDC push* dari tabel sumber ke dalam dataset yang nantinya akan digunakan pada proses analisis *data mining*. (3) Proses *data mining* menggunakan algoritma *apriori* dengan *data stream mining* untuk mendapatkan pengetahuan pada penjadwalan yang telah terjadi sehingga menghasilkan prediksi keterlambatan kendaraan untuk setiap tujuan. Metodologi yang diusulkan dapat dilihat pada **Error! Reference source not found.**



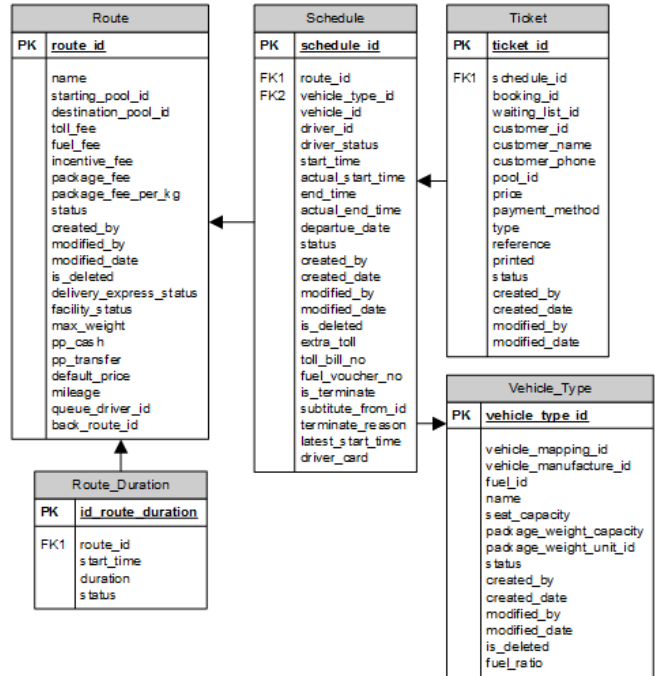
Gambar 1. Penelitian Usulan

IV. PEMBAHASAN DAN PENGUJIAN

Pada bagian ini akan membahas mengenai tahapan yang dilakukan pada metodologi penelitian dan pengujian yang dilakukan.

A. Analisis Sumber Data

Sumber data yang digunakan pada penelitian ini merupakan data penjadwalan yang telah dilakukan di PT. Citra Tiara Global sejak bulan Januari 2016 hingga Desember 2016. Terdapat beberapa tabel dari basis data yang akan digunakan dalam penelitian ini, skema relasi tabel yang terkait dapat dilihat pada **Error! Reference source not found.**



Gambar 2. Skema Relasi Tabel Sumber

B. Preprocessing Data

Tahapan *preprocessing data* diperlukan pada proses *data mining*, hal ini dikarenakan tidak semua atribut yang ada dapat digunakan pada proses *data mining*. Pada tahapan *preprocessing data* terdapat proses ekstraksi data yaitu proses pengambilan data dari sumber data, pemilihan atribut dengan memilih atribut yang akan digunakan dan proses pembersihan data untuk menghilangkan nilai *null* dan mengurangi keberagaman nilai pada data. Pada penelitian ini atribut yang akan digunakan dapat dilihat pada I.

TABEL I
PEMILIHAN ATRIBUT

Nama Atribut	Asal Tabel
seat_capacity	vehicle_type
starting_pool_id	route
destination_pool_id	route
duration	route_duration
start_time	schedule
actual_start_time	schedule
actual_end_time	schedule
type	ticket
starting_pool_id	route

Contoh data yang diambil dari sumber data dapat dilihat pada **Error! Reference source not found.**

TABEL II
CONTOH DATA SUMBER

Starting_pool_id	Destination_pool_id	Start_time	Duration (Menit)	Actual_start_time	Actual_end_time	Seat_capacity	Count (Type)
BN	DU	10:45:00	240	10:33:11	13:48:35	8	8
BN	DU	10:45:00	240	10:38:15	14:24:17	8	6
BN	DU	10:45:00	240	10:39:00	14:11:21	8	4
BN	DU	10:45:00	240	10:39:22	Null	8	2
BN	DU	10:45:00	240	10:39:26	Null	8	1
BN	DU	10:45:00	240	10:39:51	14:35:07	8	7
BN	DU	10:45:00	240	10:40:08	14:09:56	8	5
BN	DU	10:45:00	240	10:40:24	13:59:47	8	3
BN	DU	10:45:00	240	10:40:51	13:49:53	8	8
BN	DU	10:45:00	240	10:40:56	Null	8	8

Setelah proses ekstraksi dan pemilihan atribut telah dilakukan, proses selanjutnya adalah proses pembersihan data untuk menghilangkan nilai *null* dan mengurangi keberagaman data. Terdapat beberapa ketentuan yang digunakan untuk waktu *delay* dan penumpang.

Ketentuan *delay* perusahaan :

1. Kategori 0 : Keterlambatan kurang dari 16 menit
2. Kategori 1 : Keterlambatan 16 menit s/d 30 Menit
3. Kategori 2 : Keterlambatan 31 menit s/d 120 menit
4. Kategori 3 : Keterlambatan 121 menit s/d 240 menit
5. Kategori 4 : Keterlambatan lebih dari 241 menit.

Ketentuan penumpang :

1. Kategori 1 : Jumlah penumpang 30% dari jumlah kursi yang tersedia.
2. Kategori 2 : Jumlah penumpang 50% dari jumlah kursi yang tersedia.
3. Kategori 3 : Jumlah penumpang 80% atau lebih dari jumlah kursi yang tersedia

Contoh data yang telah melewati proses pembersihan data dapat dilihat pada **Error! Reference source not found.**

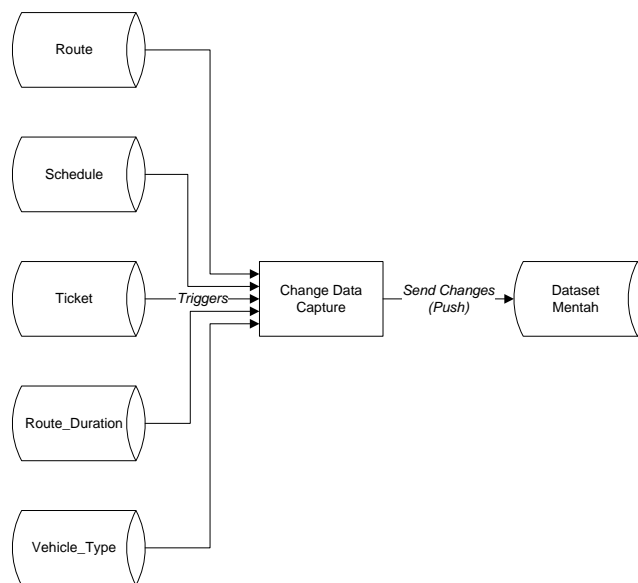
TABEL III
DATA CLEANSING

Starting_pool_id	Destination_pool_id	Penumpang	Jadwal	delay_berangkat	delay_tiba
DU	SM	3	6:00:00	0	4
DU	SM	2	8:00:00	1	2
DU	SM	3	9:00:00	1	4
DU	SM	3	10:00:00	1	4
DU	SM	1	11:00:00	1	2
DU	SM	2	12:00:00	1	2
DU	SM	3	13:00:00	1	1
DU	SM	2	14:00:00	1	0
DU	SM	2	15:00:00	2	0
DU	SM	1	16:00:00	2	0
DU	SM	2	5:00:00	0	4
DU	SM	3	6:00:00	1	0
DU	SM	2	11:00:00	1	2
DU	SM	1	12:00:00	1	0
DU	SM	3	13:00:00	1	2
DU	SM	2	16:00:00	1	0
DU	SM	2	17:00:00	2	0
DU	SM	3	18:00:00	1	0
DU	SM	3	20:00:00	1	0
DU	SM	1	21:00:00	0	0

Pada tahap *preprocessing data* diimplementasikan CDC *push* pada proses ekstraksi hingga pembersihan data sumber, proses CDC *push* yang digunakan dapat dilihat pada **Error! Reference source not found.**

Penggunaan teknik CDC pada penelitian ini adalah untuk mengurangi waktu proses query untuk pengambilan dataset pada basis data operasional, karena data yang diperlukan berada pada beberapa tabel. Langkah yang dilakukan pada proses CDC ini adalah sebagai berikut :

1. Melakukan proses query terhadap tabel route, schedule, ticket, route_duration, vehicle_type dengan pemicu petugas menekan tombol cetak manifest.
2. Daftar atribut yang akan terisi saat petugas menekan tombol cetak manifest adalah starting_pool_id, destination_pool_id, start_time, duration, actual_start_time, seat_capacity, dan jumlah penumpang, untuk atribut actual_end_time secara default akan diberikan nilai null.
3. Atribut actual_end_time akan terisi saat petugas menekan tombol sampai untuk jadwal keberangkatan yang kendaraannya telah sampai di tujuan.



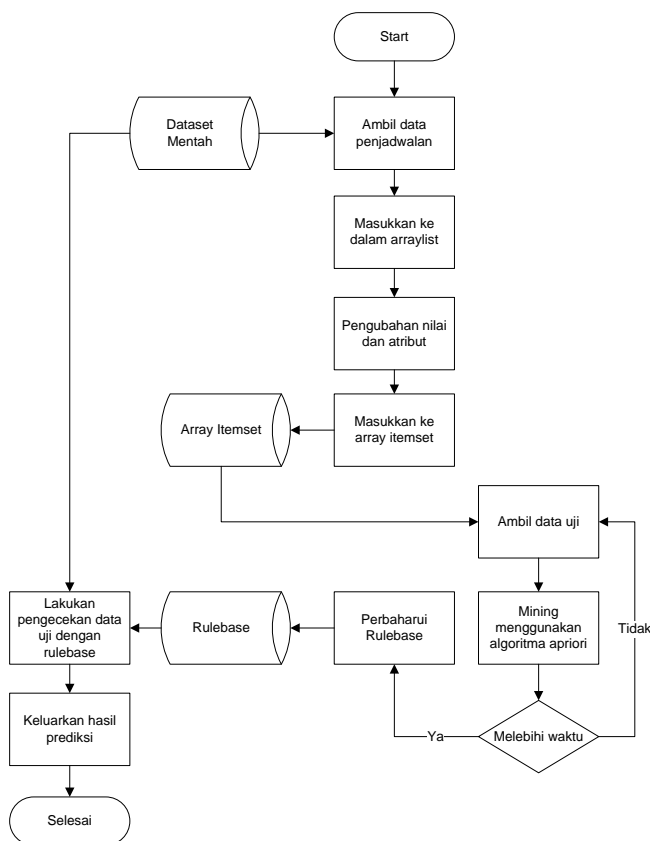
Gambar 3. Diagram Alir Change Data Capture

C. Proses Data Mining

Setelah dilakukan *preprocessing data*, tahap selanjutnya adalah *data mining* dengan *data stream mining*. Proses yang akan dilakukan sama dengan tahapan *preprocessing data* untuk mendapatkan *rulebasenya*, yang membedakan adalah adanya *threshold* waktu yang tidak boleh dilewati. *Data stream mining* memiliki karakteristik pemrosesan data yang mengalir dengan cepat dan apabila tidak segera di proses akan menghilang, karena proses yang dilakukan berada di dalam *memory* tanpa menggunakan basis data, hal ini bermanfaat untuk memproses data dengan cepat. Diagram alir proses data stream mining dapat dilihat pada **Error! Reference source not found.**

Pada diagram alir yang ada pada **Error! Reference source not found.** proses yang terjadi adalah sebagai berikut :

1. Proses pengambilan data dari basis data operasional yang telah melalui proses pemilihan atribut, dalam penelitian ini data yang diambil adalah data penjadwalan terbaru dengan rentang waktu empat bulan kebelakang.
2. Data yang telah diambil akan dimasukkan ke dalam array untuk dilakukan proses pengubahan sesuai dengan ketentuan yang telah ditetapkan oleh perusahaan perihal informasi yang dibutuhkan. Proses ini bertujuan untuk mengurangi keberagaman nilai yang ada pada basis data.
3. Data yang telah melalui proses perubahan terhadap nilai dan atribut akan dimasukkan kedalam basis data berbentuk array.
4. Proses selanjutnya adalah melakukan data mining menggunakan algoritma apriori, dalam proses ini diberikan batasan waktu proses dalam menemukan kombinasi item, hal ini bertujuan agar informasi yang dibutuhkan dapat diperoleh dengan cepat.
5. Apabila batas waktu yang telah ditentukan telah terlewati, proses selanjutnya adalah memasukkan hasil dari mining tersebut kedalam basis data rulebase.
6. Langkah selanjutnya adalah melakukan pengujian terhadap data penjadwalan terbaru, dan akan dilakukan pengecekan dengan data yang ada pada rulebase untuk mendapatkan informasi prediksi.



Gambar 4. Diagram Alir Data Stream Mining

D. Pengujian

Pada tahapan ini dilakukan serangkaian tes terhadap sistem yang telah dibuat, dengan tujuan untuk mengetahui apakah sistem tersebut telah mencapai tujuan yang diinginkan ataukah belum. Dalam penelitian ini, metode pengujian yang dilakukan adalah dengan menggunakan pengujian *eksperimental*, pengujian ini dilakukan dengan cara membandingkan aturan asosiasi yang dihasilkan oleh sistem dengan kondisi sebenarnya. Adapun data yang akan digunakan pada pengujian adalah sebagai berikut :

1. Data latih yang akan digunakan adalah data periode bulan September 2016 hingga Desember 2016 dengan total record penjadwalan sebanyak 6.505 record.
2. Data uji yang akan digunakan untuk membandingkan hasil dari sistem dengan keadaan sebenarnya, yaitu data periode bulan Januari 2017 hingga Februari 2017, dengan total record penjadwalan sebanyak 2.618 record.
3. Perhitungan keakurasian estimasi keterlambatan dengan menghitung total data prediksi yang dihasilkan data mining dibagi dengan data pada keadaan nyata yang memiliki kesamaan dengan prediksi.
4. Perhitungan keakurasian jumlah penumpang untuk setiap jadwal keberangkatan dengan menghitung total data prediksi yang dihasilkan data mining dibagi dengan data pada keadaan nyata yang memiliki kesamaan dengan prediksi.
5. Perhitungan tidak ditemukannya prediksi untuk data nyata atau data not found dengan cara membagi data yang tidak ditemukan dengan jumlah data penjadwalan nyata yang terjadi, hal ini dilakukan untuk mendapatkan minimum support dan minimum confidence yang sesuai.

Skenario pengujian yang digunakan pada penelitian ini adalah dengan mengubah nilai minimum support, minimum confidence, dan maksimal waktu pemrosesan (data stream mining). Ketentuan yang digunakan pada penelitian ini adalah sebagai berikut :

1. Minimum support yang digunakan adalah, rata-rata itemset, 1%, 3%, 5%, 10%, 15% dan 20% dari rata-rata itemset.
2. Minimum confidence yang digunakan adalah 0%, 10%, 20%, 30%, 40%, 50%, 60%, dan 70%.
3. Waktu pemrosesan yang digunakan dalam melakukan data mining adalah 1 detik kemudian diikuti dengan kelipatan 5 detik hingga 1 menit. Untuk melihat batas rata-rata waktu yang dibutuhkan untuk mendapatkan informasi.

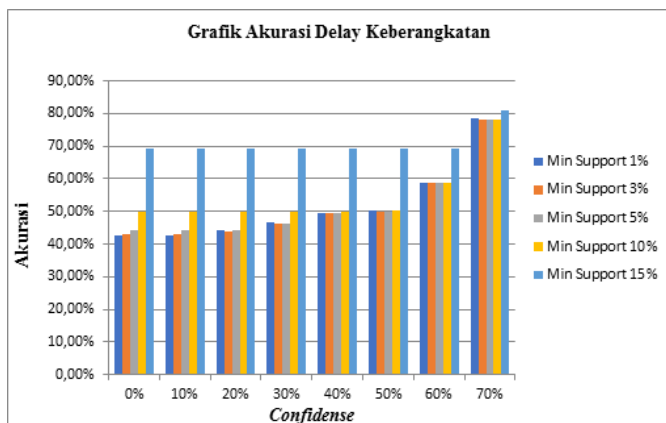
V. HASIL PENGUJIAN

Bagian ini berisi hasil pengujian yang dilakukan untuk mendapatkan prediksi keterlambatan kendaraan dan prediksi jumlah penumpang.

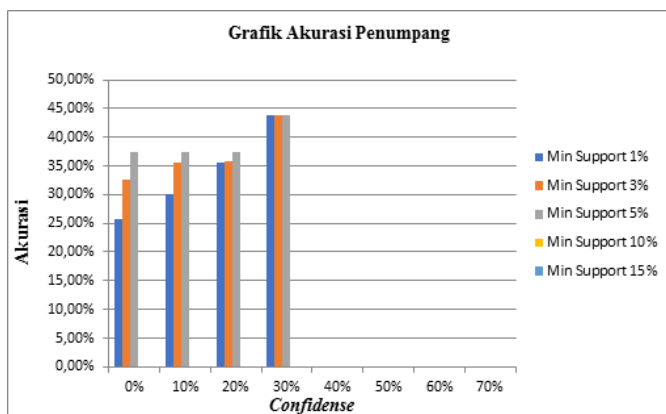
Pengujian dilakukan dengan membandingkan data prediksi dengan data telah terjadi, berdasarkan pengujian yang telah dilakukan *rulebase* yang akan terbentuk dapat ditemukan dengan *minimum confidence* yang ditentukan 30%, khususnya

untuk prediksi jumlah penumpang. Grafik hasil pengujian akurasi *delay keberangkatan* kendaraan dapat dilihat pada **Error! Reference source not found.**, untuk grafik akurasi prediksi penumpang dapat dilihat pada **Error! Reference source not found.**, dan grafik ketidakcocokan antara data prediksi dengan data nyata dapat dilihat pada **Error! Reference source not found.**

Berdasarkan hasil pengujian dapat dilihat bahwa semakin tinggi nilai *minimum support* dan *minimum confidence* yang ditentukan, maka tingkat *data not found* akan semakin besar, hal ini disebabkan berkurangnya ragam itemset dan juga aturan pada *rulebase* dan semakin tingginya *minimum confidence* juga mengakibatkan tingkat keakurasian untuk *delay keberangkatan* dan keakurasian prediksi penumpang semakin tinggi, hal ini disebabkan hanya terpilih kombinasi dengan persentase tinggi yang hampir mendekati data nyata yang akan terjadi.



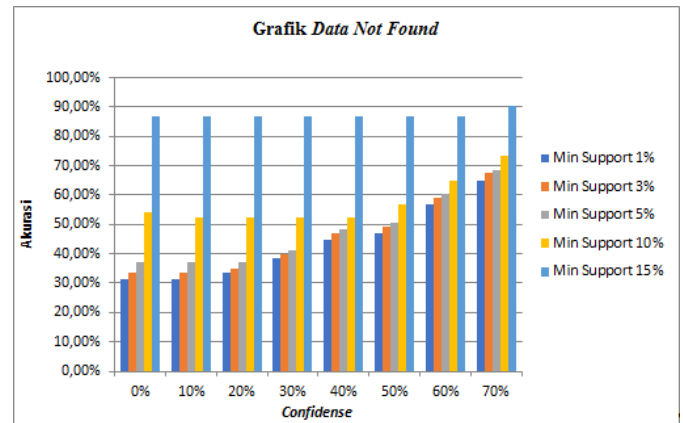
Gambar 5. Grafik Akurasi Delay Keberangkatan



Gambar 6. Grafik Akurasi Penumpang

VI. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan mengenai real-time business intelligence yang digunakan untuk membantu proses penjadwalan di PT. Citra Tiara Global ini, maka dapat diambil kesimpulan sebagai berikut :



Gambar 7. Grafik Data Not Found

1. Pengimplementasian real-time business intelligence menggunakan algoritma apriori untuk mendapatkan kombinasi antar item dapat diimplementasikan dengan metode data stream mining, dengan semakin tinggi minimum confidence dan minimum support dapat mengakibatkan :
 - a. Meningkatkan akurasi dari prediksi keterlambatan kendaraan dan akurasi penumpang.
 - b. Meningkatkan data yang tidak memiliki kecocokkan dengan rulebase yang ada.
 - c. Dengan pemberian batasan terhadap waktu, dapat memberikan informasi prediksi dengan cepat.
2. Informasi untuk membantu dalam proses penjadwalan dapat diberikan dengan beberapa pertimbangan diantaranya pertimbangan waktu tiba atau lama keberangkatan dan pertimbangan jumlah penumpang.
3. Algoritma apriori dapat diimplementasikan pada data stream mining dengan cara membatasi waktu pada proses mining yang dilakukan. Pada penelitian ini diperoleh hasil untuk mendapatkan prediksi keterlambatan kendaraan dan prediksi penumpang menggunakan data penjadwalan pada bulan september sampai desember 2016 diperoleh akurasi dengan range 44%-79% dengan batasan minimum support 5% dengan beragam nilai minimum support yang telah ditentukan dan waktu pemrosesan selama satu detik.

DAFTAR PUSTAKA

- [1] Shapiro G. Piatetsky, "Discovery, Analysis and Presentaton of Strong Rules," *AAA/MIT Press*, 1992.
- [2] and R. Srikant R. Agrawal., Chile, 1994, pp. 487-499
- [3] Richard Kirby Albert Bifet, *Data Stream Mining A Practical Approach.*: The University of Waikato, 2009.
- [4] A. Seufert J. Schiefer, "Enhanced Business Intelligence - Supporting Business Processes with Real-time Business Analytics," *The 16Th International Workshop on Dexa*, 2005.
- [5] Phillip A. Laplante & John Willey, *Real-Time System Design.*: IEEE Press, 2004.
- [6] Attunity. (2006, Agustus) Real Time Business Intelligence Enabling Effective Decision Making : Strategic, Real Time Data Integration Platform With Change Data Capture. Attunity White Paper.
- [7] Attachmate Corporation. (2005, Agustus) Capturing Changes to Host-Based Data Source. Attacmate Technical Paper.