
ANALISIS DATA MINING HOTEL BOOKING MENGGUNAKAN MODEL ID3

ANALYSIS OF DATA MINING HOTEL BOOKING USING ID3 MODEL

Agus Budiyantra^{1)*}, Andreanus Kevin Wijaya²⁾, Anthony Gunawan³⁾ dan Michael Rolland⁴⁾

¹⁾ Program Studi Teknik Informatika, STMIK Widuri, Jakarta
^{2,3,4)} Program Studi Sistem Informasi, Universitas Bunda Mulia Jakarta

Diterima 24 Februari 2021 / Disetujui 3 Maret 2021

ABSTRACT

The rapid development of information technology in this era makes it easier for someone to get information. Many business sectors are now promoting their products or services on the internet. An example is a hotel, in the technological era now we can easily find out about hotel information, ranging from location, price, and others. With the convenience to get information about this hotel, customers are indirectly increasing in a hotel. This of course causes the data contained in a hotel to increase as well. These data can be processed until we get an output and there is also data that is missing or cannot be processed. The data that can be processed can be analyzed until finally it becomes an information and prediction. In this journal, we will explain the Data Mining analysis in a hotel to analyze the success rate of a hotel. By doing this analysis, you will get insights about the level of success of the hotel and can also predict the future. Thus later the results of this analysis can be used by the hotel to assist in better decision making. Processing data in this study using the Rapid Miner application by entering data of customers who make hotel reservations.

Keywords: Data Mining, Data Mining Analytics, Hotel Analytics, Information System

ABSTRAK

Perkembangan teknologi informasi yang semakin pesat di era ini semakin memudahkan seseorang untuk mendapatkan informasi. Sektor usaha pun sekarang sudah banyak yang mempromosikan produk atau jasa mereka di internet. Contohnya adalah hotel, di era teknologi sekarang dapat dengan mudah mengetahui tentang informasi hotel, mulai dari lokasi, harga, dan lain-lain. Dengan kemudahan untuk mendapat informasi tentang hotel ini secara tidak langsung pelanggan pada sebuah hotel juga mengalami peningkatan. Hal ini tentunya menyebabkan data yang terdapat di sebuah hotel semakin meningkat juga. Data-data tersebut ada yang bisa penulis olah hingga mendapat sebuah output dan ada juga data yang *missing* atau tidak dapat diolah. Data yang dapat diolah tersebut dapat dianalisis hingga akhirnya menjadi sebuah informasi dan prediksi. Dalam jurnal ini, akan dijelaskan mengenai analisis Data Mining pada sebuah hotel untuk menganalisis tingkat keberhasilan suatu hotel. Dengan melakukan Analisis ini, maka akan diperoleh wawasan mengenai tingkat keberhasilan hotel tersebut dan juga dapat memprediksi kedepannya. Dengan demikian nantinya hasil dari Analisis ini bisa digunakan oleh pihak hotel untuk membantu dalam pengambilan keputusan yang lebih baik. Pengolahan data-data dalam penelitian ini menggunakan aplikasi Rapid Miner dengan memasukkan data pelanggan yang melakukan pemesanan hotel.

Kata Kunci: Data Mining, Analisis Data Mining, Analisis Hotel, Sistem Informasi

*Korespondensi Penulis:

E-mail: agusbudiyantra@yahoo.co.id

PENDAHULUAN

Perkembangan sistem informasi yang cepat telah membawa manusia ke era informasi (Deny & Andry, 2018). Di era ini kebutuhan akan sistem informasi (SI) semakin tinggi, karena SI menawarkan efisiensi dan efektivitas untuk mendukung organisasi dalam mencapai tujuan tersebut (Setiawan & Andry, 2019). Penggunaan teknologi informasi (TI) di dalam perusahaan mampu menciptakan keunggulan dalam persaingan yang kompetitif dan juga menjadi sasaran utama dari upaya penerapan SI/TI di perusahaan. Pembangunan perancangan yang akan dilaksanakan dengan menggunakan strategi SI dan TI dapat mencapai target yang memuaskan (Budiyantara & Leonardo, 2020). Penggunaan teknologi informasi akan dikatakan baik jika telah berjalan sesuai dengan standar teknologi informasi tersebut digunakan (Andry et al., 2019).

Dalam dunia ilmu komputer, dikenal luas Data Mining sebagai teknik untuk meringkas data dengan cara yang berbeda dengan yang biasa diterapkan, menemukan hubungan yang tidak diduga, menemukan pola yang dapat dipahami dan bermanfaat bagi pemilik data (Larose & Larose, 2014). Pengertian dari Data Mining adalah "Serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual" (Muzakir & Wulandari, 2016).

Data Mining sering kali digunakan oleh perusahaan besar untuk menggali data dengan tujuan untuk mendapatkan sebuah informasi yang akurat sehingga dapat menunjang dan meningkatkan proses bisnis perusahaan tersebut.

Penggunaan informasi teknologi dalam bisnis memudahkan setiap pelaku bisnis dari berbagai sector usaha dan menciptakan efisiensi serta efektivitas bagi perusahaan besar sehingga dengan penggunaan informasi teknologi perusahaan mampu menciptakan keunggulan kompetitif dan mampu bersaing dalam pasar global. Perkembangan organisasi

yang semakin kompleks mengakibatkan kebutuhan akan informasi dan data akan semakin meningkat.

Perkembangan pada dunia perhotelan dalam upaya penyediaan jasa akomodasi pariwisata khususnya di Indonesia juga semakin hari semakin menunjukkan perkembangan yang pesat (Sari & Hartama, 2018) dengan begitu data-data pada sebuah hotel terutama data pemesan tentunya akan menjadi sangat banyak, diantaranya ada data yang dapat digunakan untuk memprediksi, ada juga data-data yang hilang dan tidak berguna. Data-data yang tidak diperlukan disini tidak dapat digunakan dan dapat mengganggu ketika di proses. Maka dari itu penulis memilih terlebih dahulu data yang akan digunakan untuk Analisis nantinya.

Berdasarkan uraian di atas, pada penelitian ini akan dibahas mengenai bagaimana penulis akan melakukan sebuah analisis untuk melakukan data mining pada pemesanan hotel menggunakan aplikasi rapid miner dengan menggunakan data-data pemesan dari hotel yang kemudian diproses untuk mendapatkan output berupa tingkat keberhasilan suatu hotel dan juga prediksi dari data hotel booking?. Data yang digunakan dalam penelitian ini diperoleh dari situs Kaggle dengan nama hotel_bookings.

Manfaat penggunaan analisis Hotel booking yaitu dapat diperoleh pengetahuan dan wawasan mengenai tingkat keberhasilan suatu hotel sehingga hal ini bisa membantu pihak hotel dalam mengambil keputusan yang baik dan tepat dan dapat mengambil langkah atau tahapan selanjutnya.

TINJAUAN PUSTAKA

A. Data Mining

Data Mining atau Penambangan Data adalah sebuah proses yang dilakukan oleh perusahaan/instansi yang berguna untuk mengubah data mentah menjadi sebuah informasi yang bermanfaat dan dapat digunakan. Secara sederhana, data mining adalah istilah yang digunakan untuk

menggambarkan pengetahuan menemukan atau "menambang" dari sejumlah besar data (Loya & Carden, 2018). Data Mining dapat digunakan oleh perusahaan besar untuk menggali data dengan maksud mendapatkan informasi yang dapat menunjang dan meningkatkan proses bisnis perusahaan tersebut. Hasil dari data mining berguna untuk menganalisis masalah yang ada (Maulana & Fajrin, 2018). Proses Data Mining menggunakan software atau perangkat lunak untuk mencari pola didalam kumpulan data yang besar. Data Mining sangat berguna bagi pelaku bisnis karena dapat membuat si pengguna belajar lebih banyak tentang pelanggan mereka seperti mengembangkan strategi pemasaran yang lebih efektif, meningkatkan penjualan dan menekan biaya pengeluaran. Ada beberapa macam teknik data mining, namun ada 3 yang paling populer antara lain :

1. *Classification*

Klasifikasi merupakan teknik Data Mining klasik berdasarkan *machine learning* (Kaur & Singh, 2014). Klasifikasi merupakan teknik yang digunakan untuk mengambil informasi penting dan relevan tentang data dan juga metadata. Teknik ini digunakan untuk mengklasifikasikan data yang berbeda di kelas yang berbeda. Teknik *classification* sedikit mirip dengan *clustering* tetapi dalam klasifikasi, *data analyst* akan memiliki pengetahuan tentang kelas atau *cluster* yang berbeda. Secara sederhana dalam teknik klasifikasi anda akan menerapkan algoritma untuk memutuskan bagaimana data baru harus diklasifikasikan.

2. *Clustering*

Teknik *cluster* adalah sebuah kumpulan objek data, yaitu analoginya seperti benda-benda yang serupa dalam kelompok yang sama. Hal ini berarti benda-benda tersebut mirip satu sama lain dalam kelompok yang sama dan mereka berbeda atau tidak terkait dengan benda dalam kelompok lain. Teknik clustering ini adalah proses

untuk menemukan kelompok dan *cluster* didalam sebuah data.

3. *Association*

Asosiasi adalah metode yang digunakan untuk mengidentifikasi beberapa hubungan yang menarik antara variabel yang berbeda didalam database yang besar. Teknik ini dapat membantu membongkar beberapa pola tersembunyi di dalam sebuah data yang dapat digunakan untuk mengidentifikasi variabel dalam data. Asosiasi sangat berguna untuk memeriksa dan memprediksi perilaku pelanggan.

B. Pohon Keputusan (*Decision Tree*)

Pohon Keputusan adalah salah satu metode klasifikasi dan prediksi yang paling populer. Pohon Keputusan adalah sebuah flowchart yang berbentuk seperti struktur pohon yang digunakan untuk menentukan arah tindakan dan juga menunjukkan probabilitas statistik. Konsep dari pohon keputusan adalah mengubah data menjadi sebuah pohon keputusan dengan aturan-aturan keputusan. Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel yang berisi *atribut* dan *record*. Atribut disini menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon keputusan. Setiap cabang pada *Decision Tree* berisi keputusan, hasil, dan juga reaksi. Manfaat utama dengan menggunakan pohon keputusan adalah kemampuannya untuk *breakdown* proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambilan keputusan akan menjadi lebih menginterpretasikan solusi permasalahan (Haryati et al., 2015). *Decision Tree* dapat digunakan dalam berbagai situasi, mulai dari sesuatu hal yang sederhana dan bersifat pribadi, sampai ke hal yang lebih kompleks seperti pemilihan keputusan pada suatu perusahaan.

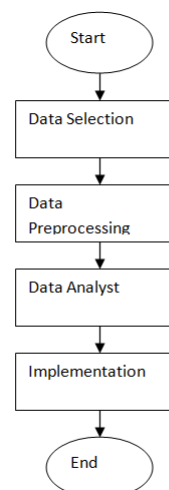
C. Rapid Miner

Rapid Miner adalah sebuah perangkat lunak berbasis Bahasa pemrograman Java yang digunakan untuk melakukan analisis data mining, text

mining, dan juga analisis prediksi. Rapid Miner dikembangkan oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer pada tahun 2001 dengan nama awal YALE (*Yet Another Learning Environment*). Rapid Miner merupakan salah satu perangkat lunak data mining yang mudah untuk digunakan, karena semua fasilitas untuk data *mining* sudah disediakan. Rapid Miner menyediakan GUI untuk merancang proses analitik (membaca data dari sumber, transformasi, menerapkan algoritma). Semua perubahan GUI disimpan dalam file XML (*eXtensible Markup Language*) dan kemudian file ini dibaca oleh Rapid Miner untuk menjalankan analisis. Rapid Miner memungkinkan koneksi ke Internet sumber data yang paling beragam seperti Oracle, Microsoft SQL Server, MySQL, dan akses ke Excel, Access serta berbagai format data lainnya (Krstevski et al., 2011). Metode yang disediakan oleh Rapid Miner antara lain *classification*, *clustering*, *association*, dan lain-lain. Software ini bersifat *open source*, Jadi orang lain dapat ikut untuk mengembangkan software ini.

METODE PENELITIAN

Metode penelitian adalah salah satu cara ilmiah yang digunakan untuk mendapatkan data yang jelas dan valid dengan tujuan untuk dapat ditemukan, dikembangkan, dan juga dibuktikan kejelasannya, suatu pengetahuan tertentu sehingga pada akhirnya dapat digunakan untuk memahami, memecahkan, dan mengantisipasi masalah yang ada dalam bidang tertentu. Tujuan dilakukan penelitian ini adalah untuk mempermudah pada saat transaksi hotel. Selain itu terdapat beberapa tahapan dalam penelitian ini dengan mengikuti pola umum penelitian ilmiah di perlihatkan pada gambar 1.



Gambar 1. Kerangka Kerja Penelitian
(Andry et al., 2020)

- Pemilihan Data:** mendapatkan sumber data dari berbagai sumber. Disini penulis mengumpulkan dan menggunakan data-data pelanggan hotel yang di dapatkan dari situs kaggle.
- Preprocessing Data :** adalah untuk membersihkan data yang tidak diinginkan dan mencari nilai yang hilang dan nilai data yang berisik dari dataset. Disini penulis melakukan penyortiran data-data untuk menghindari duplikasi data dan data yang missing.
- Analisis Data:** salah satu teknik penambangan data diterapkan untuk mendapatkan hasil. Penulis menganalisis data pelanggan dari sebuah hotel agar bisa menghasilkan informasi yang akurat dan berguna.
- Implementasi:** hasil dari teknik penambangan data yang diterapkan dalam aplikasi Rapid Miner.

Subjek Penelitian

Subjek dalam penelitian ini adalah para tamu yang booking hotel. Data pada penelitian ini diambil dari situs Kaggle dengan nama `hotel_bookings`.

Pengembangan Instrumen

Pengembangan instrumen merupakan hal yang sangat penting di dalam kegiatan penelitian. Instrumen penelitian mempunyai kontribusi yang amat sangat penting dalam menentukan kualitas pada sebuah penelitian, karena validity atau kejelasan data yang didapatkan akan sangat ditentukan oleh kualitas instrumen yang dipakai oleh si peneliti, di samping prosedur pengumpulan data yang digunakan. Pada pengembangan instrument ini, akan digunakan dalam penelitian adalah pedoman dokumentasi. Pedoman dokumentasi merupakan alat bantu yang digunakan untuk mengumpulkan data-data, dan arsip-arsip dokumentasi. Dokumentasi menggunakan data kaggle.

Teknik Analisis Data

Teknik analisis data yang digunakan dalam penelitian ini yaitu deskriptif. Ada dua jenis data dalam penelitian tindakan ini yaitu data kualitatif dan data kuantitatif. Namun yang digunakan adalah kuantitatif. kuantitatif adalah data yang berbentuk angka dan juga merupakan hasil dari pengukuran dan perhitungan. Perhitungan ini akan menggunakan dengan perangkat lunak Rapid Miner, dan juga dengan Decision Tree untuk klasifikasi dan prekdiksi yang paling populer. Untuk menghitung data klasifikasi dan prekdiksi membutuhkan data mining.

HASIL DAN PEMBAHASAN

Pada tahapan penelitian ini mengikuti tahapan Pohon Keputusan (*Decision Tree*). Dengan menggunakan seluruh perangkat lunak Rapid Miner dan MS Excel 2008. Untuk aplikasi RapiMiner merupakan salah satu aplikasi yang digunakan untuk menganalisis dataset untuk diubah sesuai kebutuhan pengguna. Tujuan penelitian ini untuk mengetahui *alternative* dan menunjukkan keputusan yang dipilih. Dan manfaat menggunakan *decision tree* untuk mencari berapa orang mencari type mereka.

A. Pengumpulan Data

Sumber untuk data utama yang digunakan dalam penelitian ini adalah dataset hotel *booking* pada sebuah format .xlsx. *Dataset* hotel terdiri dari atribut hotel, *is canceled*, *lead time*, *arrival date year*, *arrival date month*, *arrival date week number*, *arrival date day of month*, *stays in weekend night*, *stays in week nights*, *adults*, *children*, *Babies*, *Meal*, *country*, *market segment*, *distribution channel*, *is repeated guest*, *previous cancellations*, *previous bookings not canceled*, *reserved room type*, *assigned room type*, *booking changes*, *deposit type*, *agent*, *company*, *days in waiting list*, *customer type*, *adr*, *required car parking spaces*, *total of special requests*, *reservation status*, dan *reservation status date*. Lalu, jumlah data pada atribut tersebut berjumlah 10.000 dan terdapat 31 *attribute*. Di data ini akan menggunakan special attribute yaitu *customer type*.

Untuk mencari pengumpulan data menjadi data mining, dengan cara memasukan data ke Rapid Miner. Cara memasukanya dengan cara menggunakan *Retrieve*. Tujuan menggunakan *Retrieve* yaitu untuk *blok mode* yang digunakan mengakses data. Data yang sudah masuk akan muncul seperti pada table 1 dan 2.

Tabel 1. Hotel Bookings

Hotel	is_canceled	lead_time	arrival_date_year
Ressort Hotel	0	342	2015
Ressort Hotel	0	737	2015
Ressort Hotel	0	7	2015
Ressort Hotel	0	13	2015
Ressort Hotel	0	14	2015
Ressort Hotel	0	14	2015
Ressort	0	0	2015

Hotel	is_canceled	lead_time	arrival_date_year
Hotel			
Ressort Hotel	0	9	2015
Ressort Hotel	1	85	2015
Ressort Hotel	1	75	2015
Ressort Hotel	1	23	2015
Ressort Hotel	0	35	2015
Ressort Hotel	0	68	2015
Ressort Hotel	0	18	2015
Ressort Hotel	0	37	2015
Ressort Hotel	0	68	2015
Ressort Hotel	0	37	2015
Ressort Hotel	0	12	2015
Ressort Hotel	0	0	2015
Ressort Hotel	0	7	2015

Tabel 2. Hotel Booking

<i>Arrival_date_month</i>	<i>Arrival_date_week_Number</i>
July	27
July	27
July	27
July	27
July	27
July	27
July	27
July	27

<i>Arrival_date_month</i>	<i>Arrival_date_week_Number</i>
July	27
July	27
July	27
July	27
July	27
July	27
July	27
July	27
July	27
July	27
July	27
July	27
July	27

Setelah mengklik tombol tersebut, perlu mencari letak penyimpanan data mentah yang di miliki. Kemudian setelah data mentah tersebut sudah dipilih, maka akan muncul tampilan dari data mentah yang dipilih. Setelah itu, penulis ke Statistics dimana berapa statistic pada tiap artribut. Di Statistic terdapat berapa jenis dari name, type, missing, dan statistics. Name itu berada pada data setiap kategori sepeti meal, hotel, id, dan lain-lain. Lalu missing digunakan untuk mengecek apa ada data yang belum di isi. Lihat tabel 3.

Sebaiknya, jangan terdapat missing karena data missing dapat membuat analisis menjadi rumit. Lalu ada type dimana ada berapa type yaitu integer, polynominal, binominal, dan lain-lain. Yang terakhir statistics yang berguna untuk mengetahui jumlah terbesar, terkecil, dan rata-rata dari setiap kategori data bisa di lihat pada tabel 4.

Tabel 3. Statistics

Name	Type	Missing
reservation_status	Binominal	0
Hotel	Polynominal	0
is_canceled	Binominal	0
lead_time	Integer	0
arrival_date_year	Integer	0
arrival_date_month		0
arrival_date_week_number	Integer	0
arrival_date_day_of_month	Integer	0
Stays_in_weekend_night	Integer	0

Tabel 4. Statistics

Statistics		
Least	Most	Values
Checkout (3666)	Canceled (6174)	Canceled(6074), Checkout(3666)
city hotel(0)	resort hotel(9840)	Resort hotel(9840), City hotel(3666)
0(3666)	1(6174)	1(6174),0(3666)
Min	Max	Averages
0	737	115.833
Min	Max	Averages
2015	2017	2015.647

Least	Most	Value
Januari	October	October(1425)
Min	Max	Averages
1	53	31.244
Min	Max	Averages
1	31	15.855
Min	Max	Averages
0	16	1.284

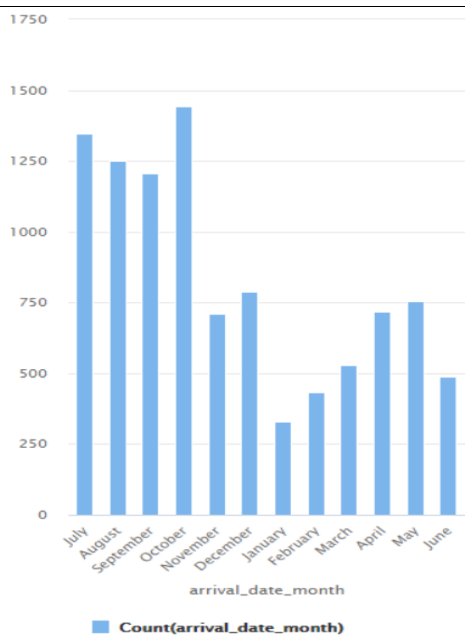
Pada setiap attribute terdapat values untuk menampilkan jumlah pada semua data. Berikut ini adalah data statistics pada setiap attribut tersebut:

Tabel 5. Meal

Index	Nominal value	Absolute count	Fraction
1	BB	7133	0.7133
2	HB	2098	0.2098
3	FB	421	0.0421
4	Undefined	326	0.0326
5	SC	22	0.0022

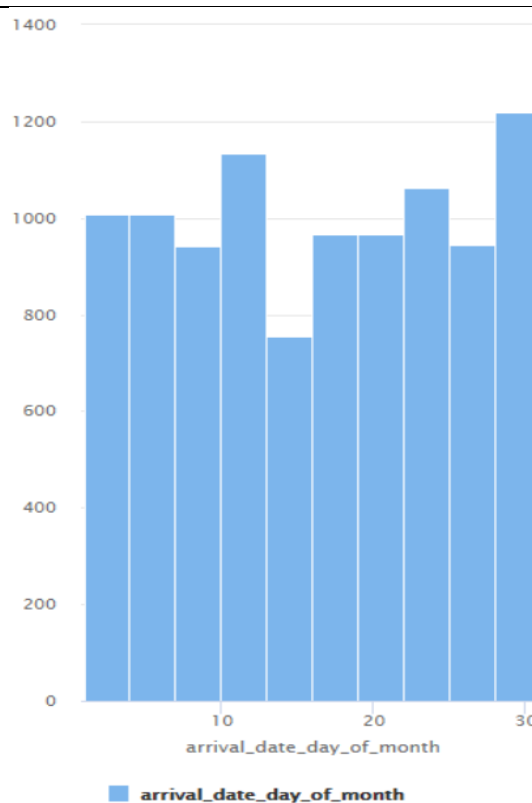
Pada tabel tersebut, penulis melihat yang paling banyak terdapat pada BB dengan sebanyak 7.133 orang. Lalu, dibawahnya yaitu HB dengan 2.098 orang, FB dengan sebanyak 421 orang, Undefined dengan sebanyak 326 orang, dan terakhir yaitu SC dengan sebanyak 22 orang. Lalu terdapat Fraction yang menghitung berapa pesenan orang yang memilih. Hasil Fraction harus 1 jika tika maka data tidak benar.

Lalu terdapat dengan namanya visualization. Tujuannya yaitu memunculkan grafik di setiap attribute. Untuk grafik akan di pilih sesuai dengan yang diinginkan. Berikut ini adalah grafik dari hasil data hotel tersebut :



Gambar 2. Grafik arrival date month

Pada Gambar 2 *Grafik arrival date month*, menjelaskan jumlah yang mengunjungi hotel tersebut berdasarkan bulan para tamu datang. Dari data tersebut, jumlah terbesar pada hotel tersebut ada pada bulan October dengan jumlah 1.443 orang. Dan berikutnya, terdapat pada bulan July yang jumlahnya 1.347 orang, lalu August yang jumlahnya 1.252 orang, bulan September dengan jumlah 1.205 orang, bulan December dengan jumlah 789 orang, bulan May dengan jumlah 754 orang, bulan April dengan jumlah 718 orang, bulan November dengan jumlah 712 orang, bulan March dengan jumlah 528 orang, bulan June dengan jumlah 490 orang, bulan February dengan jumlah 433 orang, dan terakhir bulan January dengan jumlah 329 orang. Dengan bergitu, Jumlah tamu yang menginap di hotel pada bulan October dengan jumlah 1.443 orang, dan yang paling terkecil tamu menginap di hotel yaitu pada bulan January dengan jumlah 329 orang.

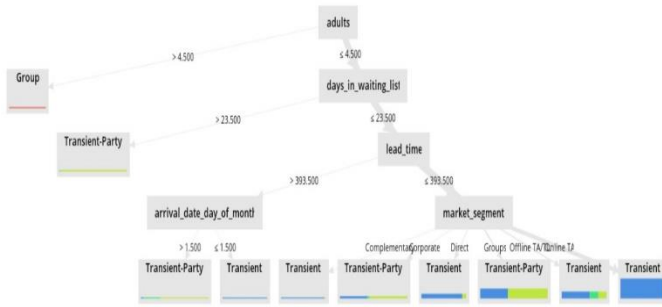


Gambar 3. Grafik arrival date day of month

Pada gambar 3 *Grafik arrival date day of month*, menjelaskan jumlah tamu yang mengunjungi hotel tersebut pada tanggal tertentu. Dari data tersebut, Jumlah para tamu datang pada tanggal 28-31 dengan jumlah 1.218 orang. Dan data terkecil pada tanggal 13-16 dengan jumlah 756 orang.

B. Pemodelan

Pemodelan merupakan tahap yang secara langsung melibatkan teknik data mining, yaitu dengan melakukan pemilihan teknik data mining dan menentukan algoritma yang akan digunakan. Algoritma yang digunakan dalam penelitian ini adalah algoritma. Dengan menggunakan pemodelan ID3, maka di dapatkan dengan pohon keputusan yang terbentuk adalah sebagai berikut :



Gambar 4. Hasil dari Decision Tree

Untuk melakukan aturan-aturan yang diperoleh dari pohon keputusan tersebut untuk menentukan jalur kelulusan berdasarkan atribut-atribut pada penelitian bisa dilihat pada Gambar 5.

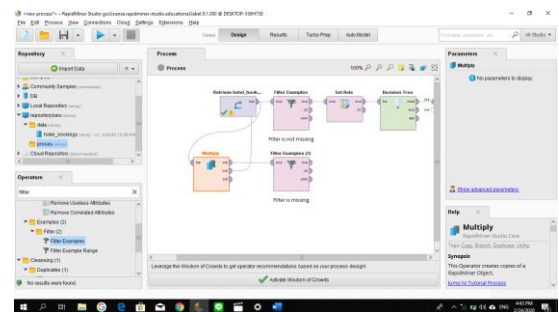
```

ts > 4.500: Group (Transient=0, Contract=0, Transient-Party=0, Group=16)
ts ≤ 4.500
  days_in_waiting_list > 23.500: Transient-Party (Transient=1, Contract=1, Transient-Party=147, Group=0)
  days_in_waiting_list ≤ 23.500
    lead_time > 393.500
      arrival_date_day_of_month > 1.500: Transient-Party (Transient=0, Contract=13, Transient-Party=37, Group=0)
      arrival_date_day_of_month ≤ 1.500: Transient (Transient=4, Contract=0, Transient-Party=0, Group=0)
    lead_time ≤ 393.500
      market_segment = Complementary: Transient (Transient=47, Contract=0, Transient-Party=1, Group=0)
      market_segment = Corporate
        lead_time > 14.500
          stays_in_week_nights > 0.500
            total_of_special_requests > 1.500: Contract (Transient=1, Contract=2, Transient-Party=0, Group=0)
            total_of_special_requests ≤ 1.500
              reservation_status_date > Aug 14, 2018: Transient-Party (Transient=17, Contract=4, Transient-Party=173, Group=0)
              reservation_status_date ≤ Aug 14, 2018: Transient (Transient=1, Contract=1, Transient-Party=0, Group=0)
            stays_in_week_nights ≤ 0.500: Transient (Transient=6, Contract=0, Transient-Party=0, Group=0)
          lead_time ≤ 14.500
            booking_changes > 1.500: Transient (Transient=1, Contract=1, Transient-Party=0, Group=0)
            booking_changes ≤ 1.500
              adr > 11.500
                arrival_date_week_number > 47.500
                  arrival_date_month = December: Transient (Transient=5, Contract=0, Transient-Party=0, Group=0)
                  arrival_date_month = November: Transient-Party (Transient=0, Contract=0, Transient-Party=3, Group=0)
                arrival_date_week_number ≤ 47.500: Transient (Transient=103, Contract=3, Transient-Party=10, Group=0)
              adr ≤ 11.500: Transient-Party (Transient=1, Contract=0, Transient-Party=0, Group=0)
            market_segment = Direct
              stays_in_weekend_nights > 6.500: Transient-Party (Transient=0, Contract=0, Transient-Party=2, Group=0)
              stays_in_weekend_nights ≤ 6.500
                adr > 295.665: Transient (Transient=1, Contract=0, Transient-Party=1, Group=0)
                adr ≤ 295.665
                  booking_changes > 3.500
                    stays_in_week_nights > 1.500: Transient (Transient=8, Contract=0, Transient-Party=1, Group=0)
                    stays_in_week_nights ≤ 1.500: Transient-Party (Transient=0, Contract=0, Transient-Party=2, Group=0)
  
```

Gambar 5. Rule atau aturan yang diperoleh

C. Validasi dan Evaluasi

Validasi Dan Evaluasi hasil merupakan tahanan menampilkan hasil dari data mining tersebut. Data hasil yang digunakan dengan menggunakan *Operator Performance*. Tujuan dari *Performance* yaitu menguji apa model *Decision Tree* bagus atau tidak. Untuk melakukannya, penulis menggunakan *Apply Model*. *Apply Model* bertujuan untuk mendapatkan prediksi pada data yang tidak terlihat atau mengubah data dengan menerapkan model *preprocessing*. Untuk melakukannya, penulis membutuhkan data, *filter examples*, *decision tree*, dan *apply model*. Untuk *filter examples* ada 2 buah untuk mengetahui yang mana *not missing*. Kemudian Tarik line dari retrieve ke filter example (2) kemudian Tarik kembali ke Filter Example (1) sampai ada logo sharing pada filter examples(1). Setelah mengklik logo sharing akan muncul tampilan blok multiply untuk menyalin data dan kemudian menyalurkannya ke filter (1) dan filter (2). Untuk bentuk desainya ada seperti pada Gambar 6.



Gambar 6. Set Role

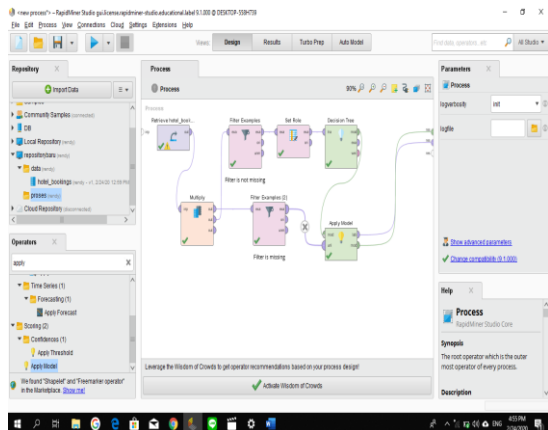
Kemudian klik filter example(2) dan add filters, kemudian pilih kategori (children) dan klik is missing. Setelah itu, hubungkan line filter examples(2) ke result dan kemudian di "run". Setelah di run disini terlihat bahwa di data children ada missing 4, data ini nanti bisa diprediksi menggunakan apply model. Dan contohnya pada Gambar 7.

The screenshot shows a data table with the following columns: arrival_date, stays_in_w, adults, children, hotel, and country. The table contains several rows of data, including some with missing values (NaN) in the 'children' column.

arrival_date	stays_in_w	adults	children	hotel	country
3	1	0	2	1	BB
5	0	2	2	1	BB
5	0	2	3	1	BB
13	2	0	2	1	BB

Gambar 7. Hasil Set Role

Setelah itu, penulis akan menggunakan apply model ke Decision Tree, dan kemudian hubungkan filter example(2) ke unlabel Apply Model. Untuk desain dari Apply Model ada pada Gambar 8 dan hasil dari Apply Model pada Gambar 9

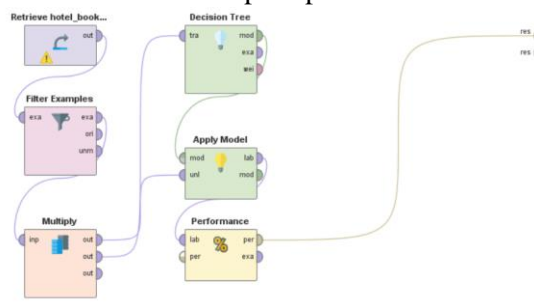


Gambar 8. Apply Model

Row No.	predicted(class)	label	is_cancelled	last_date	arrival_date	arrival_date_unix	arrival_date_iso_utc	start
1	0.172	Ch/Normal	1	2	2915	August	32	3
2	0.172	Ch/Normal	1	1	2915	August	32	5
3	0.172	Ch/Normal	1	1	2915	August	32	5
4	0.172	Ch/Normal	1	0	2915	August	33	13

Gambar 9. Hasil Apply Model

Lalu sekarang penulis akan menggunakan Operater Performance. Caranya adalah dengan drag Performace ke desain process dan dipasang dengan Apply Model. Untuk hasil desain seperti pada Gambar 10.



Gambar 10. Desain Performance

Setelah dijalankan, terdapat accuracy 100%. Apabila tidak 100% makan data kurang bagus. Accuracy memilik maksimal 100%. Hasil pada Performance ada pada Tabel 6.

Tabel 6. Hasil Performance

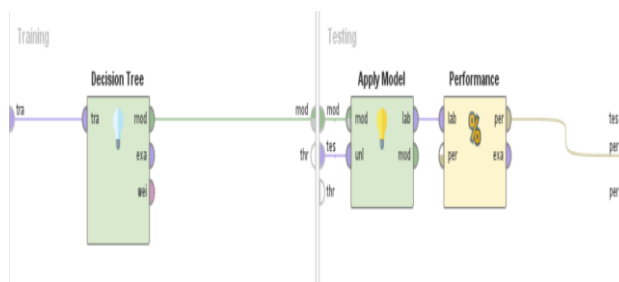
Accuracy : 100.00%

	<i>True Checkout</i>	<i>True Canceled</i>	<i>Class Precision</i>
<i>Pred. Checkout</i>	75166	0	100.00%
<i>Pred. Canceled</i>	0	43013	100.00%
<i>Class Recall</i>	100.00%	100.00%	

D. Mevalidasi Model

Pada proses ini penulis akan menggunakan Cross Validation. Cross Validation adalah proses untuk mevalidasi model pada sebuah data. Di dalam Cross Validation ada namanya number of folds yang digunakan untuk proses iterasi sesuai dengan jumlahnya dan akan diulang lalu akan diambil rata-rata yang terbaik. Setelah itu penulis akan klik 2 kali dan akan muncul sebuah lembar yang bernama Sub Process.

Sub Process terdapat 2 process yaitu Process Training dan Process Testing. Proses Training digunakan untuk proses pembelajaran sedangkan process testing digunakan untuk menguji modelnya pada prediksi laber. Pada Training, penulis menggunakan Decision Tree dan akan dihubungkan. Sedangkan pada Testing, akan digunakan Apply Model dan akan segera dihubungkan. Setelah Apply Model, akan di gunakan Performance dan akan juga dihubungkan. Untuk bentuk desain di Sub Process akan seperti pada Gambar 12.



Gambar 12.Desain Sub Process

Setelah itu, jalankan process dan akan muncul seperti pada Gambar 13. Pada Hasil tersebut, penulis mempunyai accuracy sebesar 100% dimana yang true check-out pada pred. Check-Out adalah 75166 sedangkan true canceled adalah 0. Lalu, true check-out pada pred. Canceled adalah 0 sedangkan true canceled adalah 43017. Sehingga class recallnya pada true check-Out dan true Canceled adalah 100%. Dan, sedangkan class precision pada pred. Check Out dan pred. Canceled adalah 100%.

Tabel 7. Hasil Cross Validation

Accuracy: 100.00% +/- 0.00% (micro average: 100.00%)

	<i>True Checkout</i>	<i>True Canceled</i>	<i>Class Precision</i>
<i>Pred. Checkout</i>	75166	0	100.00%
<i>Pred. Canceled</i>	0	43017	100.00%
<i>Class Recall</i>	100.00%	100.00%	

KESIMPULAN

Pada penelitian ini penulis melakukan analisis Data Mining pada sebuah hotel menggunakan data dari pelanggan hotel yang didapatkan dari situs kaggle dengan nama Hotel_Bookings. Penelitian ini menggunakan aplikasi Rapid Miner untuk membantu dalam melakukan pengolahan data pelanggan hotel sehingga pemrosesan data sampai menghasilkan output menjadi lebih mudah. Hasil yang didapatkan dari penelitian ini adalah mengetahui jumlah tamu hotel tiap bulannya dengan angka tertinggi berada di bulan oktober yaitu 1443 orang dan terkecil ada di bulan januari dengan jumlah 433 orang, Sedangkan untuk jumlah tamu berdasarkan tanggal check in didapatkan data tertinggi ada pada tanggal 28 sampai dengan 31 dengan jumlah 1.218 orang, sedangkan terendahnya ada di tanggal 13 sampai dengan 16 dengan jumlah 756 orang. Dengan demikian, penggunaan dari analisis Data Mining ini dapat memungkinkan pihak hotel untuk membuat decision making yang lebih baik lagi

berdasarkan informasi dan data-data yang telah dianalisis.

SARAN

Dalam melakukan penelitian ini masih terdapat beberapa kekurangan. Maka dari itu sekiranya kekurangan ini dapat diperbaiki pada penelitian berikutnya. Untuk itu, terdapat beberapa saran yang dapat dijadikan masukan untuk melakukan penelitian selanjutnya. Setelah melakukan analisis untuk mengukur tingkat keberhasilan suatu hotel, dan telah mendapatkan hasilnya, dengan demikian, diharapkan pada pengembangan selanjutnya dapat dilakukan analisis kembali mengenai faktor faktor apa saja yang dapat mempengaruhi tingkat keberhasilan suatu hotel tersebut. Agar nantinya pihak hotel bisa mengetahui kesalahan yang telah mereka lakukan, dan bisa meningkatkan kesalahan tersebut hingga pada akhirnya tingkat keberhasilan suatu hotel tersebut menjadi lebih meningkat.

DAFTAR PUSTAKA

- Andry, J. F., Geasela, Y. M., Wailan, A., Matjik, B. A., Kurniawan, A., & Junior, J. (2019). Penggunaan COBIT 4.1 Dengan Domain ME Pada Sistem Informasi Absensi (Studi Kasus: Universitas XYZ). *Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer*, 13(2), 97. <https://doi.org/10.30872/jim.v13i2.1152>
- Andry, J. F., Silaen, F. M., Tannady, H., & Saputra, K. H. (2020). *Electronic Health Record to Predict a Heart Attack Used Data Mining With Naive Bayes Method*. 9(2), 209–216. <https://doi.org/10.11591/ijict.v9i2.ppx-x-xx>
- Budiyantara, A., & Leonardo, J. (2020). Perancangan Enterprise Architecture Menggunakan Zachman Pada Pt . Sutura Indah Utama Design Enterprise Architecture Using Zachman In Pt . Sutura Indah Utama.

- Journal of Business and Audit Information Systems*, 3(1), 1–13.
- Deny., & Andry, J. F. (2018). Pengukuran Keberhasilan E-Learning dengan Mengadopsi Model Delone & Mclean. *Jurnal Sistem Informasi Bisnis*, 8(1), 68. <https://doi.org/10.21456/vol8iss1pp68-75>
- Haryati, S., Sudarsono, A., & Suryana, E. (2015). (2015). Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5. *Jurnal Media Infotama*, 11(2), 130–138.
- Kaur, B., & Singh, W. (2014). Review on Heart Disease Prediction System using Data Mining Techniques. *International Journal on Recent and Innovation Trends in Computing and Communication*, 2(10), 3003–3008.
- Krstevski, J., Mihajlov, D., & Chorbev, I. (2011). *Student Data Analysis with RapidMiner*. 19–28. <http://ictinnovations.org>
- Larose, D. T., & Larose, C. D. (2014). Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition. In *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition* (Vol. 9780470908747). <https://doi.org/10.1002/9781118874059>
- Loya, T., & Carden, G. (2018). Business intelligence and analytics. In *Higher Education Strategy and Planning*. <https://doi.org/10.4324/9781315206455-12>
- Maulana, A., & Fajrin, A. A. (2018). Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Transaksi Penjualan Spare Part Motor. *Klik - Kumpulan Jurnal Ilmu Komputer*, 5(1), 27. <https://doi.org/10.20527/klik.v5i1.100>
- Muzakir, A., & Wulandari, R. A. (2016). Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree. *Scientific Journal of Informatics*, 3(1), 19–26. <https://doi.org/10.15294/sji.v3i1.4610>
- Sari, R. W., & Hartama, D. (2018). Data Mining: Algoritma K-Means Pada Pengelompokan Wisata Asing ke Indonesia Menurut Provinsi. *Seminar Nasional Sains & Teknologi Informasi (SENSASI)*, 322–326.
- Setiawan, A. K., & Andry, J. F. (2019). IT Governance Evaluation using COBIT 5 Framework on the National Library. *Jurnal Sistem Informasi*, 15(1), 10–17. <https://doi.org/https://doi.org/10.21609/jsi.v15i1.790>