

## Prediksi Keterlambatan Penerbangan Menggunakan Metode Decision Tree Untuk Penentuan Premi Asuransi Perjalanan

Shella Lolitha<sup>1)</sup>, Samuel Lukas<sup>2)</sup>, & Frans Panduwinata<sup>3)</sup>

<sup>1,2,3)</sup> Program Studi Informatika, Fakultas Ilmu Komputer  
Universitas Pelita Harapan

MH Thamrin Boulevard 1100, Tangerang, Banten 15811

Telp (021) 546 0901

Mobile: <sup>1)</sup>0812-9697-9780, <sup>2)</sup>0817-6356-565, <sup>3)</sup>0813-9932-7582

Website <https://www.uph.edu/>, E-mail: SL2062@student.uph.edu<sup>1)</sup>, [samuel.lukas@uph.edu](mailto:samuel.lukas@uph.edu)<sup>2)</sup>, & [frans.panduwinata@uph.edu](mailto:frans.panduwinata@uph.edu)<sup>3)</sup>

**Abstrak** - Keterlambatan penerbangan menjadi suatu masalah yang tidak dapat dihindari pada penerbangan komersial. Kompensasi yang diatur dalam Peraturan Menteri dinilai tidak sebanding dengan keterlambatan yang terjadi. Sebaliknya, pihak perusahaan maskapai juga enggan untuk meningkatkan kualitas pelayanan, salah satu alasannya adalah membayar biaya kompensasi penumpang jauh lebih murah dibanding biaya yang diperlukan untuk meningkatkan pelayanan. Maka dari itu diperlukan sebuah sistem untuk penghitungan jumlah premi yang dapat dibayar oleh penumpang agar menguntungkan kedua belah pihak. Dengan menggunakan metode penghitungan statistik dan algoritma machine learning yaitu pohon keputusan, keterlambatan dapat diprediksi berdasarkan kategori keterlambatan yang tertulis dalam Peraturan Menteri dan premi asuransi dapat dihitung dengan sesuai dan saling menguntungkan kedua belah pihak. Tahap perancangan sistem adalah sebagai berikut: membaca data mentah penerbangan komersial di Indonesia dari tahun 2017 sampai 2019, melakukan preprocessing data, melakukan pembersihan data, melakukan pelatihan data, melakukan proses prediksi, melakukan kalkulasi premi dan membuat visualisasi untuk menampilkan hasil prediksi dan harga premi. Hasil pengujian berdasarkan matriks confusion menunjukkan bahwa model untuk memprediksi keterlambatan memiliki akurasi sebesar 72,76%. Kemudian dari proses validasi, didapatkan bahwa tingkat kemiripan hasil prediksi dengan hasil validasi tersebut adalah 96,14%. Hasil penghitungan premi memiliki nilai premi yang lebih masuk akal dan menguntungkan bagi penumpang penerbangan.

**Kata kunci:** Pohon Keputusan, Perhitungan Premi, Keterlambatan Penerbangan, Asuransi Perjalanan

**Abstract** - Flight delays become an inevitable issue on flight commercial. Compensation regulated in the Ministerial Regulation considered disproportionate with occurring delays. Otherwise, airline company party are also reluctant to improve the quality of service, one of the reason is pay compensation cost for passenger is much less expensive than the cost for improving services. Therefore, a system needed for calculating the amount of premium that can be paid by passengers to benefit both parties. By using statistical calculation method and machine learning algorithm, Decision Tree, delays can be predicted based on category of delays regulated in the Ministerial Regulation and insurance premium can be calculated accordingly and mutually beneficial to both parties. Phase of system design is as follows: read flight commercial in Indonesia from year 2017 to 2019 raw data, preprocess data, cleanse data, train data, process prediction, calculate premium and build visualization for presenting prediction result and premium price. Test result based on confusion matrix shows that model for predicting delays has an accuracy of 72.76%. Then from validation process, it obtained that similarity level of prediction result to validation result is 96.14%. The premium calculation result has premium value that is more reasonable and profitable for passenger flight.

**Keywords:** Decision Tree, Premium Calculation, Flight Information, Travel Insurance

### 1 PENDAHULUAN

Perkembangan transportasi penerbangan di Indonesia terus mengalami peningkatan sejak tahun 2003 [1]. Banyak masyarakat kini melakukan perjalanan

dengan menggunakan jasa penerbangan komersial di Indonesia. Terdapat banyak faktor mengapa transportasi penerbangan komersial se semakin populer dan diminati masyarakat. Salah satu faktor utama adalah waktu karena waktu sangat diperhitungkan dalam dunia

transportasi. Ketika menggunakan transportasi penerbangan menggunakan pesawat udara, waktu yang diperlukan untuk menempuh jarak tertentu menjadi lebih singkat. Faktor lainnya adalah keamanan yang erat kaitannya dengan kecelakaan. Tingkat kecelakaan fatal pesawat udara di seluruh dunia terus mengalami penurunan dari tahun ke tahun [2]. Hal ini menjadi salah satu bukti bahwa moda transportasi dengan menggunakan pesawat udara adalah transportasi yang paling aman saat ini. Namun di sisi lain, penerbangan komersial juga memiliki beberapa kekurangan yang membuat masyarakat memperhitungkan kembali dalam melakukan perjalanan menggunakan pesawat udara. Beberapa masyarakat berpendapat bahwa menggunakan transportasi pesawat udara berisiko lebih tinggi dibanding transportasi lainnya. Salah satu risiko dalam transportasi pesawat udara adalah cuaca, karena cuaca dapat menjadi ancaman keselamatan apabila terjadi perubahan cuaca secara signifikan dalam waktu yang singkat. Kemudian hal-hal yang sangat membuat masyarakat enggan menggunakan transportasi pesawat udara, yaitu keterlambatan dan pembatalan penerbangan. Terdapat banyak faktor yang dapat menyebabkan keterlambatan. Selain faktor cuaca, terdapat faktor lain seperti faktor manajemen penerbangan, keterlambatan pilot, co-pilot, atau awak kabin dan ketidaksiapan pesawat udara. Faktor lainnya adalah faktor teknis operasional, seperti bandar udara untuk keberangkatan dan tujuan tidak dapat digunakan untuk operasional pesawat udara, terjadinya antrian pesawat udara lepas landas atau mendarat, dan keterlambatan pengisian bahan bakar. Hal ini menyebabkan kerugian waktu yang cukup besar kepada penumpang pesawat udara. Peraturan Menteri Perhubungan Republik Indonesia Nomor 89 Tahun 2015 tentang Penanganan Keterlambatan (*Delay Management*) Pada Badan Usaha Angkutan Udara Niaga Berjadwal di Indonesia mengatur bahwa setiap badan usaha angkutan udara wajib memberikan kompensasi sesuai dengan kategori keterlambatan [3]. Namun kompensasi yang diatur dalam Peraturan Menteri tersebut dinilai tidak sebanding atau tidak memadai dengan keterlambatan yang terjadi. Sebaliknya, pihak badan usaha angkutan udara atau perusahaan maskapai juga enggan untuk meningkatkan kualitas pelayanan, salah satu alasannya adalah membayar biaya kompensasi penumpang jauh lebih ringan dibanding biaya yang diperlukan untuk meningkatkan pelayanan. Dilatarbelakangi oleh hal tersebut, diperlukan sebuah sistem untuk penghitungan jumlah premi yang dapat dibayar oleh penumpang agar menguntungkan kedua belah pihak. Ketika penumpang membeli tiket penerbangan disertai dengan pembelian premi asuransi perjalanan, pihak maskapai dapat

menerima dana tambahan untuk meningkatkan layanan maskapai. Dengan menggunakan metode penghitungan statistik dan algoritma *machine learning* yaitu *Decision Tree*, keterlambatan dapat diprediksi sesuai dengan kategori keterlambatan yang tertulis dalam Peraturan Menteri dan premi asuransi dapat dihitung dengan sesuai dan saling menguntungkan kedua belah pihak.

Dalam penelitian ini permasalahan mencakup bagaimana membuat model berbasis *web* dengan *Decision Tree* untuk memprediksi keterlambatan penerbangan pada penerbangan komersial di Indonesia dan bagaimana menghitung premi yang sesuai menggunakan metode statistik berdasarkan keterlambatan penerbangan pada penerbangan komersial di Indonesia. Adapun beberapa batasan dalam penelitian ini, yaitu metode yang digunakan untuk memprediksi keterlambatan penerbangan adalah *Decision Tree* dengan algoritma C5.0, metode yang digunakan untuk menghitung premi asuransi perjalanan adalah metode statistik, data yang diambil berasal dari [www.flightradar24.com](http://www.flightradar24.com), data yang diambil adalah data penerbangan (yang mencakup data maskapai, kota keberangkatan, kota tujuan, tanggal penerbangan, waktu keberangkatan dan tujuan (*standard dan actual*), dan juga status penerbangan) komersial di Indonesia pada tanggal 18 Juni 2017 sampai 4 Juli 2019 yang meliputi 8 maskapai penerbangan di Indonesia yaitu *Lion Air*, *Garuda Indonesia*, *Batik Air*, *Citilink*, *Sriwijaya*, *Wings Air*, *NAM Air*, dan *Express Air*, implementasi metode untuk memprediksi dan menghitung premi akan menggunakan bahasa pemrograman R, teknik evaluasi akan menggunakan *confusion matrix*, dan hasil yang ditampilkan berupa aplikasi berbasis *web* yang dibuat menggunakan *package R Shiny*. Tujuan utama dari penelitian ini adalah merancang sebuah model yang dapat memprediksi keterlambatan penerbangan pada penerbangan komersial di Indonesia. Hasil dari prediksi tersebut akan digunakan sebagai acuan dalam menghitung premi asuransi perjalanan untuk setiap penerbangan. Informasi mengenai premi asuransi perjalanan ini dapat digunakan untuk menentukan biaya yang harus dikeluarkan oleh penumpang penerbangan komersial di Indonesia saat melakukan pembelian tiket perjalanan.

## 2 LANDASAN TEORI

Pada bagian ini akan dibahas landasan teori yang digunakan dalam penelitian ini yang meliputi premi asuransi, penghitungan premi asuransi perjalanan, dan *decision tree*.

### 2.1 Premi Asuransi

Premi asuransi adalah sejumlah uang yang dibayarkan baik oleh individu maupun perusahaan yang dikeluarkan untuk membayar biaya polis asuransi selama beberapa periode tertentu [4]. Premi asuransi dibayarkan untuk polis yang mencakup perawatan kesehatan, mobil, rumah, maupun jiwa [5]. Setelah dibayarkan, premi merupakan pendapatan bagi perusahaan asuransi. Namun, selain menjadi pendapatan, premi juga merupakan kewajiban, karena perusahaan asuransi harus memberikan pertanggungan untuk klaim yang dibuat berdasarkan polis asuransi. Kegagalan membayar premi asuransi akan menyebabkan pembatalan terhadap polis asuransi.

### 2.2 Penghitungan Premi Asuransi Perjalanan

Terdapat beberapa langkah untuk menghitung premi asuransi menggunakan metode statistik khususnya standar deviasi. Pertama, tentukan *airline* yang preminya akan dihitung. Kedua, tentukan kota keberangkatan. Kemudian yang terakhir adalah menghitung premi. Terdapat dua variabel yang dibutuhkan untuk menghitung *delay*, yaitu rata-rata dari kerugian setiap penumpang yang menggunakan *airline* tersebut yang disimpan pada variabel *mean*, dan total kerugian dari setiap *airline* yang disimpan pada variabel *Tloss* [6].

$$Tloss = \sum_{i=1}^5 N(i) * loss(i) \tag{1}$$

$$Mean = Tloss/Passengers \tag{2}$$

Total kerugian (*Tloss*) dapat dihitung dengan persamaan (1) dan rata-rata (*Mean*) dapat dihitung dengan persamaan (2). *N* (1) menunjukkan jumlah penumpang yang memiliki *delay code* dan *loss* adalah kompensasi untuk penumpang yang mengalami *delay*. Huruf *i* mewakili tipe *delay* 0, 1, 2, 3, 4, 5 dan *Passengers* (2) adalah total jumlah penumpang.

Perusahaan asuransi bertanggung jawab untuk memberikan kompensasi dengan mencakup semua kerugian. Secara ideal, total premi bernilai sama dengan total kerugian, sehingga nilai premi yang didapatkan tidak terlalu merugikan bagi penumpang dan tidak terlalu menguntungkan bagi pihak maskapai. Namun, hal itu tidak berjalan demikian di dalam dunia bisnis.

$$Premium = \mu + \sigma \tag{3}$$

$$\mu = \sum_{i=0}^5 loss(i) \times p(i) \tag{4}$$

$$\sigma = \sqrt{\sum_{i=0}^5 (loss(i) - \mu)^2 \times p(i)} \tag{5}$$

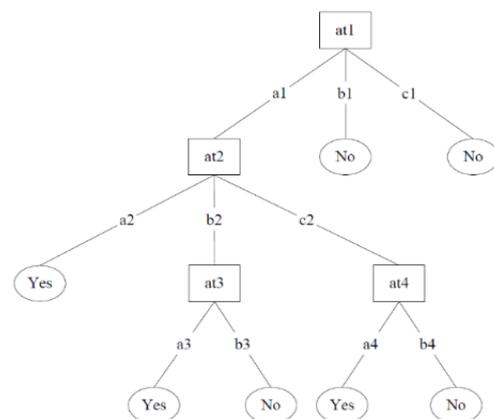
$$p(i) = \frac{f(i)}{\sum_{i=0}^5 f(i)} \tag{6}$$

Premi dihitung berdasarkan rata-rata (*mean*) dan standar deviasi kerugian, yang dapat dilihat pada persamaan (3). *Mean* dan standar deviasi dihitung dengan persamaan (4) dan (5). Kemungkinan dari setiap kode *delay* dihitung dengan persamaan (6) dan disimpan pada variabel *p(i)*(6), sedangkan *f(i)* adalah jumlah penumpang yang berangkat dari suatu kota dengan penerbangan yang memiliki tipe *delay* *i*. Berdasarkan hasil penghitungan premi asuransi yang dilakukan, terdapat kemungkinan untuk meningkatkan keuntungan penumpang ketika terjadi *delay*. Banyak kemungkinan yang dapat ditawarkan untuk keuntungan penumpang namun terdapat beberapa pertimbangan yang harus dipenuhi yaitu:

1. Total kerugian harus lebih rendah dari total premi. Idealnya total premi sama dengan total kerugian.
2. Nilai premi pada tipe *delay* 1, 2, 3, 4 harus ditingkatkan untuk menarik penumpang membeli asuransi.

### 2.3 Decision Tree

*Decision Tree* adalah salah satu metode dalam *machine learning* yang digunakan baik untuk klasifikasi maupun regresi (Rokach dan Maimon, 2008). Tujuan dari metode ini adalah untuk membuat model yang memprediksi nilai variabel target atau sering disebut sebagai *dependent variable* berdasarkan beberapa variabel *input* atau *independent variable*. *Decision Tree* dapat membantu untuk mengidentifikasi strategi mana yang paling cocok untuk mencapai tujuan. Metode *Decision Tree* yang digunakan untuk melakukan klasifikasi, disebut sebagai pohon klasifikasi. Sedangkan metode *Decision Tree* yang digunakan untuk melakukan regresi, disebut sebagai pohon regresi.



Gambar 1 Ilustrasi Decision Tree

Ilustrasi *Decision Tree* dapat dilihat pada Gambar 1. Ilustrasi tersebut menggunakan *data training* dengan menggunakan data yang terdapat pada Tabel 1. Pada Gambar 1, dapat dilihat data  $at1=a1$ ,  $at2=b2$ ,  $at3=a3$ ,  $at4=b4$  akan melakukan pemilihan terhadap node  $at1$ ,  $at2$ , dan kemudian  $at3$ , yang akan melakukan proses klasifikasi dengan hasil “yes” atau “no” [7].

**Tabel 1** Ilustrasi Data Training

at1	at2	at3	at4	Class
a1	a2	a3	a4	Yes
a1	a2	a3	b4	Yes
a1	b2	a3	a4	Yes
a1	b2	b3	b4	No
a1	c2	a3	a4	Yes
a1	c2	a3	b4	No
b1	b2	b3	b4	No
c1	b2	b3	b4	No

*Decision Tree* memiliki banyak algoritma yang dapat digunakan untuk membuat model *decision tree* tersebut. Beberapa algoritma yang populer digunakan adalah ID3, C4.5/C5.0, CART, CHAID, dan MARS. Setiap algoritma memiliki perbedaan dalam penggunaan metrik dalam mengukur pembagian yang terbaik [8].

### 3 METODOLOGI PENELITIAN

Tahap perancangan sistem adalah sebagai berikut: membaca *raw data* penerbangan komersial di Indonesia dari tahun 2017 sampai 2019, *data preprocessing*, *data cleansing*, *data training*, *prediction process*, dan *premium calculation*.

#### 3.1 Raw Data

*Raw data* atau data mentah yang diambil dari [www.flightradar24.com](http://www.flightradar24.com) merupakan data penerbangan komersial di Indonesia selama kurang lebih 2 tahun dari tanggal 18 Juni 2017 sampai dengan 4 Juli 2019. Data yang diambil merupakan data penerbangan dari 8 maskapai di Indonesia yaitu, *Lion Air*, *Garuda Indonesia*, *Batik Air*, *Citilink*, *Sriwijaya*, *Wings Air*, *NAM Air*, dan *Express Air*. Himpunan data penerbangan tersebut memiliki 1.019.557 *records* dengan *format file Excel Workbook (xlsx)*. Atribut-atribut yang ada di dalam himpunan data penerbangan tersebut dapat dilihat pada Tabel 2.

**Tabel 2** Atribut Raw Data

Nama	Deskripsi	Contoh
Date	Tanggal keberangkatan	5/22/2019 12/8/2018
From	Kota keberangkatan	Jakarta (CGK) Yogyakarta (JOG)
To	Kota tujuan	Samarinda (SRI) Jambi (DJB)
Flight	Nomor penerbangan	ID7281 GA266
Flight Time	Lama penerbangan	0:36

		0:49
STD	Scheduled Time of Departure (waktu keberangkatan yang dijadwalkan)	12:00:00 AM 6:25:00 AM
ATD	Actual Time of Departure (waktu keberangkatan yang sebenarnya)	11:59:00 PM 4:58:00 AM
STA	Standard Time of Arrival (waktu kedatangan)	12:40 AM 1:05 AM
Status	Keterangan status penerbangan	Landed 12:36 AM Landed 1:18 AM
Airlines	Nama maskapai	Citilink Sriwijaya

Pada setiap maskapai yang ada di dalam himpunan data penerbangan tersebut, terdapat jumlah *records* yang berbeda-beda. Jumlah *records* untuk setiap maskapai dapat dilihat pada Tabel 3. Himpunan data tersebut memiliki 1.798 unik penerbangan dengan kombinasi *airline-from-to*.

**Tabel 3** Jumlah Records Setiap Maskapai

Maskapai	Records
Lion Air	324701
Garuda Indonesia	271637
Batik Air	122006
Citilink	168929
Sriwijaya Air	65678
NAM Air	15001
XPRESS Air	839
WINGS Air	50266

Namun, himpunan data penerbangan tersebut tidak dapat langsung digunakan untuk melakukan proses prediksi dikarenakan kurangnya atribut yang dibutuhkan. Oleh karena itu, perlu dilakukan pengolahan data lebih lanjut atau *data preprocessing* dan juga pembersihan data atau *data cleansing*.

#### 3.2 Data Preprocessing

Atribut tambahan yang diperlukan dapat dilihat pada Tabel 4. Atribut-atribut tambahan tersebut dapat diperoleh dengan cara memanfaatkan atribut yang sudah ada. Oleh karena itu, perlu dilakukan pengolahan data atau *data preprocessing* untuk melengkapi himpunan data penerbangan tersebut.

**Tabel 4** Atribut Tambahan

Nama Atribut	Deskripsi	Format dan Contoh
TrueDelay	Total waktu keterlambatan dalam satuan menit	Didapatkan dengan cara mengurangi STD dengan ATD. Contoh: 173, 35
DelayType	Kategori keterlambatan berdasarkan total waktu (TrueDelay)	0 = tidak delay 1 = 30-60 menit delay 2 = 61-120 menit delay 3 = 121-180 menit delay 4 = 181-240 menit delay 5 = > 240 menit delay
Bdelay	Binary delay	0 = DelayType bernilai 0 1 = DelayType bernilai 1-5
Day	Nama hari penerbangan	1 = Minggu 2 = Senin 3 = Selasa 4 = Rabu 5 = Kamis 6 = Jumat 7 = Sabtu

<i>DeparType</i>	Kategori waktu penerbangan terjadwal (STD) berdasarkan pembagian waktu dalam satu hari	<i>Early Morning</i> = 00:00 – 5:59 <i>Morning</i> = 6:00 – 11:59 <i>Afternoon</i> = 12:00 – 17:59 <i>Night</i> = 18:00 – 23:59
<i>NumPass</i>	Jumlah penumpang di dalam satu Penerbangan	Didapatkan dengan cara mengeluarkan angka acak dengan <i>range</i> 80%-100% dari kapasitas maksimum (Data kapasitas maksimum dapat dilihat pada Tabel 5) Contoh: 138, 166, 116

Tabel 5 Jumlah Kapasitas Maksimum Setiap Maskapai

Maskapai	Kapasitas Maksimum
Lion Air	189
Garuda Indonesia	140
Batik Air	166
Citilink	166
Sriwijaya Air	153
NAM Air	120
XPRESS Air	128
WINGS Air	72

### 3.3 Data Cleansing

*Data cleansing* dilakukan untuk menghilangkan data-data yang tidak *valid* atau tidak dapat digunakan untuk melakukan *data training*. *Raw data* penerbangan yang tersedia memiliki 1.019.557 *records* yang terdiri dari 8 maskapai, 115 kota keberangkatan, dan 119 kota tujuan. Kriteria dari pembersihan data ini adalah setiap penerbangan untuk satu maskapai, satu kota keberangkatan, dan satu kota tujuan harus memiliki sekurang-kurangnya 104 penerbangan dengan asumsi memiliki paling sedikit satu kali penerbangan dalam satu minggu. Oleh karena itu, perlu dilakukan pengelompokan data berdasarkan maskapai, kota keberangkatan, dan kota tujuan. Proses *data cleansing* yang dilakukan akan sekaligus digunakan untuk mendapatkan satu atribut yang menyimpan nilai *delay* atau tidak *delay* berdasarkan perbandingan banyaknya jumlah penerbangan yang *delay* dan tidak *delay*. Atribut tersebut akan digunakan pada tahap berikutnya yaitu proses validasi. Atribut-atribut yang diperlukan untuk melakukan proses *data cleansing* terdapat pada Tabel 6.

Tabel 6 Atribut Pada Data Cleansing

Nama Atribut	Deskripsi	Format dan Contoh
<i>Airlines</i>	Nama maskapai	Citilink Garuda Indonesia
<i>From</i>	Kota keberangkatan	Jakarta (HLP) Semarang (SRG)
<i>To</i>	Kota tujuan	Samarinda (SRI) Denpasar (DPS)
<i>sum0</i>	Total penerbangan yang memiliki <i>DelayType</i> dengan tipe 0	273 179
<i>sum1</i>	Total penerbangan yang memiliki <i>DelayType</i> dengan tipe 1	26 45
<i>sum2</i>	Total penerbangan yang memiliki <i>DelayType</i> dengan tipe 2	7 46
<i>sum3</i>	Total penerbangan yang memiliki <i>DelayType</i> dengan tipe 3	19 22

<i>sum4</i>	Total penerbangan yang memiliki <i>DelayType</i> dengan tipe 4	3 28
<i>sum5</i>	Total penerbangan yang memiliki <i>DelayType</i> dengan tipe 5	110 12
<i>Total_Flight</i>	Total penerbangan keseluruhan	626 4673
<i>Total_Delay</i>	Total penerbangan yang mengalami <i>delay</i> . <i>DelayType</i> bernilai 1-5	Menjumlahkan <i>sum1</i> sampai <i>sum5</i> . Contoh: 104, 34, 61
<i>code_delay</i>	Nilai <i>delay</i> atau tidak <i>delay</i>	0 = $sum0 > Total\_Delay$ 1 = $sum0 \leq Total\_Delay$
<i>is_deleted</i>	Nilai apakah data tersebut akan dihapus atau tidak	0 = $Total\_Flight \geq 104$ 1 = $Total\_Flight < 104$

### 3.4 Data Training

Setelah melalui tahap *data cleansing*, himpunan data penerbangan memiliki 990.321 *records* yang terdiri dari 5 maskapai, 81 kota keberangkatan, dan 90 kota tujuan. Data yang telah melewati tahap *data cleansing* telah dapat digunakan untuk melakukan *data training*. Dalam melakukan *data training*, diperlukan suatu data *frame* yang berisi *variable independent* dan *variable dependent*. *Variable independent* terdiri dari *Airlines*, *From*, *To*, *Day*, dan *DeparType*. Kemudian *variable dependent* adalah *Bdelay*.

### 3.5 Prediction Process

Langkah pertama dalam melakukan prediksi adalah dengan cara membuat model *machine learning* menggunakan *data training* yang sudah dimiliki. Pembuatan model *machine learning* ini menggunakan metode *decision tree* dengan algoritma C5.0. Fungsi yang digunakan untuk pembuatan model *decision tree* terdapat pada *library* C5.0

### 3.6 Premium Calculation

Pada proses penghitungan premi, terdapat 2 penghitungan yang dilakukan. Penghitungan pertama adalah penghitungan premi awal, yaitu premi yang dihitung menggunakan persamaan (3) sampai (6) diatas dengan menggunakan harga kompensasi awal. Penghitungan kedua adalah penghitungan premi baru, yaitu premi yang dihitung menggunakan persamaan (7) dengan menggunakan harga kompensasi baru. Dalam melakukan penghitungan premi asuransi baru, dibutuhkan himpunan data dengan atribut-atribut pada Tabel 7. Proses pengolahan data untuk menghasilkan himpunan data tersebut dilakukan dengan cara yang serupa dengan proses *data cleansing*.

Tabel 7 Atribut Himpunan Data Pada Penghitungan Premi

Nama Atribut	Deskripsi	Format dan Contoh
<i>Airlines</i>	Nama maskapai	Citilink Garuda Indonesia
<i>From</i>	Kota keberangkatan	Jakarta (HLP)

<i>To</i>	Kota tujuan	Semarang (SRG) Samarinda (SRI) Denpasar (DPS)
<i>pred0</i>	Jumlah penumpang dari penerbangan yang terprediksi tidak <i>delay</i>	87701 216045
<i>pred1</i>	Jumlah penumpang dari penerbangan yang terprediksi <i>delay</i>	44523 116086
<i>tloss</i>	Total kerugian yang harus dikeluarkan oleh maskapai	Didapatkan dengan cara menjumlahkan hasil dari total penumpang dikali nilai kompensasi baru, yang dikelompokkan berdasarkan tipe <i>delay</i> . Nilai kompensasi dapat dilihat pada Tabel 8 [6]. Contoh: 2.928.460.000

Tabel 8 Nilai Kompensasi Baru Berdasarkan Tipe Delay

Tipe Delay	Lama Delay (menit)	Kompensasi (IDR)
0	0-29	0
1	30-60	50.000
2	61-120	75.000
3	121-180	105.000
4	181-240	150.000
5	> 240	450.000

Terdapat dua nilai premi yang akan dihitung pada penghitungan premi baru, yang pertama adalah nilai premi rendah yaitu premi untuk penerbangan yang terprediksi tidak *delay* (atribut *plow*), kedua adalah premi tinggi yaitu premi untuk penerbangan yang terprediksi *delay* (atribut *phigh*). Nilai premi dapat dicari dengan persamaan (7) dibawah dengan kondisi-kondisi sebagai berikut:

- 1) Jika nilai *pred0* atau *pred1* bernilai nol, maka  $plow = phigh$ .
- 2) Jika nilai *pred0* dan *pred1* tidak nol, maka nilai *plow* dan *phigh* yang diambil adalah nilai yang memiliki selisih terkecil.

$$(pred0.plow) + (pred1.phigh) = tloss \quad (7)$$

Langkah pertama dalam mencari nilai *plow* dan *phigh* adalah dengan memberikan nilai 1000 sebagai nilai sementara untuk *plow*. Setelah itu dilakukan iterasi dengan kelipatan 1000 sampai menemukan selisih *plow* dan *phigh* yang terkecil.

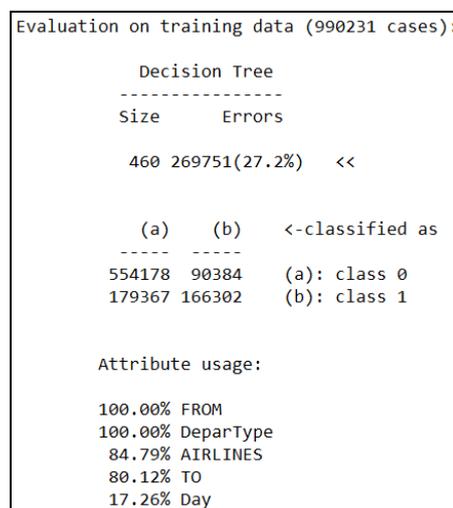
#### 4 HASIL DAN PEMBAHASAN

Pada bagian ini akan dibahas hasil penelitian yang meliputi hasil prediksi, hasil penghitungan premi dan contoh antar muka aplikasi.

##### 4.1 Hasil Prediksi

Pada Gambar 2 dibawah, *cases* menjelaskan bahwa terdapat 990.231 data yang di-*training*, *size* menunjukkan jumlah *leaf nodes decision tree* yang terbentuk, yaitu sebanyak 460 *nodes*, kemudian 269.751

pada bagian *errors* menunjukkan jumlah *records* yang salah klasifikasi dan 27,2% merupakan rasio kesalahan atau *error* dari keseluruhan *records* data. Persentase *error* yang cukup tinggi ini disebabkan oleh adanya atribut-atribut *independent variable* yang sama namun memiliki label atau *Bdelay* yang berbeda. *Confusion matrix* menunjukkan sebanyak 554.178 *records* data yang tidak *delay* diklasifikasikan tidak *delay* (*True Positive*), 90.384 *records* data yang tidak *delay* namun diklasifikasikan *delay* (*False Negative*), 179.367 *records* data yang *delay* namun diklasifikasikan tidak *delay* (*False Positive*), dan 166.302 *records* data yang *delay* diklasifikasikan *delay* (*True Negative*). Besar akurasi dapat diperoleh dengan cara membagi jumlah nilai *True Positive* dan *True Negative* atau total *records* yang terklasifikasi benar dengan total data keseluruhan. Nilai akurasi yang didapatkan dari *confusion matrix* pada Gambar 2 adalah sebesar 72,76%.



Gambar 2 Ringkasan Evaluasi Data Training

Gambar 2 juga menunjukkan besar persentase atribut-atribut yang digunakan sebagai penentu. Dapat dilihat bahwa atribut *From* dan *DepartType* digunakan 100%, diikuti dengan atribut *Airlines* dengan 84,79%, kemudian atribut *To* 80,12%, dan yang terakhir adalah atribut *Day* dengan 17,26%. Setelah model diperoleh, prediksi dilakukan. Setelah hasil prediksi dan kelas target (*Bdelay*) ditambahkan ke dalam data *frame* kelas *input* yaitu *input\_training\_set*, hasil prediksi pada data *frame* tersebut dapat dilihat pada Tabel 9.

Tabel 9 Hasil Prediksi

Airlines	From	To	Day	Depart Type	Bdelay	Prediction
Lion Air	BTH	PLM	2	Morning	0	1
Citilink	HLP	PDG	7	Afternoon	1	1
Batik Air	HLP	AMQ	2	Night	1	0
WINGS	BMU	LOP	5	Night	0	0

Pada Tabel 9 dapat dilihat bahwa penerbangan dengan maskapai *Lion Air*, dari kota Batam (BTH) menuju kota Palembang (PLM), pada hari Senin dengan kategori waktu keberangkatan pagi hari mendapatkan hasil prediksi *delay* sedangkan pada data aktual penerbangan yang diwakili dengan atribut *Bdelay* menunjukkan bahwa penerbangan tersebut tidak *delay*. Setelah mendapatkan hasil prediksi, perlu dilakukan pengecekan kembali terhadap hasil prediksi tersebut, hal ini dapat dikatakan sebagai proses validasi, proses ini dilakukan dengan cara membandingkan hasil prediksi dengan suatu atribut untuk menentukan *delay* atau tidak *delay* dengan cara membandingkan banyaknya jumlah total penerbangan yang *delay* dan total penerbangan yang tidak *delay*, atribut ini dinamakan *code\_delay* seperti yang terlihat pada Tabel 6 diatas. Atribut *code\_delay* yang diperoleh kemudian ditambahkan pada data *frame* yang sebelumnya digunakan untuk membandingkan hasil prediksi. Tabel 10 dibawah menunjukkan contoh beberapa data penerbangan yang telah ditambahkan dengan atribut *code\_delay*.

Tabel 10 Data Hasil Validasi

Air lines	From	To	Day	Depart Type	Bdelay	Predicted	code_delay
WINGS	KNO	GNS	4	Afternoon	0	1	0
Lion	BTH	PLM	2	Morning	0	1	1
Citilink	HLP	PDG	7	Afternoon	1	1	0
GA	CGK	PKU	6	Morning	1	1	1

Tabel 10 memperlihatkan penerbangan dengan maskapai *WINGS Air*, dari kota Medan (KNO) menuju kota Gunung Sitoli (GNS), pada hari Rabu, dengan kategori waktu keberangkatan siang hari, mendapatkan hasil prediksi *delay* sedangkan pada data aktual penerbangan yang diwakili dengan atribut *Bdelay* menunjukkan bahwa penerbangan tersebut tidak *delay*, akan tetapi atribut *code\_delay* menunjukkan angka 0 yang berarti tidak *delay*, maka dari itu dapat diasumsikan bahwa hasil prediksi pada penerbangan ini kurang tepat. Perbandingan antara hasil prediksi dengan *code\_delay* menunjukkan seberapa tepat hasil prediksi tersebut. Jumlah hasil prediksi yang memiliki nilai yang sama dengan *code\_delay* adalah 951.982 records dari total keseluruhan data. Hal ini membuktikan bahwa hasil prediksi memiliki kesamaan sebesar 96,14% terhadap *code\_delay*, maka dari itu dapat disimpulkan bahwa hasil prediksi yang telah didapatkan cukup *valid*.

## 4.2 Hasil Penghitungan Premi

Penghitungan premi meliputi premi awal dan premi baru. Penghitungan premi awal dilakukan dengan

menggunakan persamaan (3) sampai (6) dengan harga kompensasi awal yang dapat dilihat pada Tabel 11 (Pemerintah Indonesia, 2015). Hasil penghitungan premi awal dapat dilihat pada Tabel 12 dibawah. Penghitungan premi awal ini menjadi latar belakang mengapa penghitungan premi baru dilakukan. Terdapat total 988 penerbangan yang memiliki nilai premi yang berbeda-beda.

Tabel 11 Nilai Kompensasi Awal

Tipe Delay	Lama Delay (menit)	Kompensasi (IDR)
0	0-29	0
1	30-60	10.000
2	61-120	25.000
3	121-180	35.000
4	181-240	50.000
5	> 240	300.000

Tabel 12 Hasil Penghitungan Premi Awal

Airlines	From	To	Total Passenger per DelayType				
			0	1	2	3	4
Citilink	CGK	KNO	367850	102028	47287	4334	1472
Lion Air	DPS	CGK	326277	256313	209580	86918	29749
Sriwijaya	CGK	MLG	96959	52808	35129	8759	1392

5	Total_loss	Mean	Stdev	Premium	Total Premium (IDR)
13697	6.536.845K	12.180.43	47.323.10	59.503.52	31.933.636.014.44
14024	16.539.410K	17.921.89	37.578.44	55.500.33	51.219.090.338.91
538	1.943.870K	9.938.75	18.991.98	28.930.72	5.658.415.752.86

Tabel 12 menunjukkan bahwa biaya yang harus dikeluarkan oleh maskapai *Citilink* untuk memberikan kompensasi kepada penumpang yang mengalami *delay* pada penerbangan dari kota Jakarta (CGK) menuju kota Medan (KNO) adalah IDR 6.536.845.000, biaya ini dapat dilihat pada atribut *total\_loss*. Kemudian premi untuk penerbangan tersebut adalah IDR 59.503,52, sehingga total pemasukan yang didapatkan oleh maskapai *Batik Air* pada penerbangan tersebut adalah IDR 31.933.636.014,44. Sebagai contoh, dalam penerbangan dengan maskapai *Citilink* dari kota Jakarta (CGK) menuju kota Medan (KNO), premi asuransi yang ditawarkan kepada penumpang adalah sebesar IDR 59.503,52, tetapi apabila terjadi *delay* dengan kategori *delay* tipe 1, penumpang hanya akan mendapatkan kompensasi sebesar IDR 10.000, begitu juga dengan kategori *delay* tipe 2, penumpang hanya akan mendapatkan kompensasi sebesar IDR 25.000 padahal penumpang tersebut membeli premi asuransi dengan harga IDR 59.503,52. Penumpang baru akan mendapatkan keuntungan melalui terjadinya *delay* ketika mengalami *delay* dengan kategori *delay* tipe 5 yaitu IDR 300.000. Hal ini menjadi salah satu alasan mengapa penumpang enggan untuk membeli premi. Hasil penghitungan premi awal menunjukkan bahwa terdapat perbedaan yang sangat signifikan pada total biaya yang harus dikeluarkan untuk memberikan

kompensasi dan total pendapatan yang didapat dari hasil penjualan premi. Biaya yang harus dikeluarkan untuk memberikan kompensasi kepada penumpang yang mengalami *delay* cenderung lebih kecil ketimbang biaya yang harus dikeluarkan untuk meningkatkan layanan maskapai agar tidak terjadi *delay*. Oleh karena itu, perlu dilakukan penghitungan premi baru dengan menggunakan nilai kompensasi yang baru sehingga mendapatkan nilai *total\_loss* yang baru, kemudian nilai *total\_loss* tersebut menjadi dasar dari penghitungan premi baru yang akan dilakukan sehingga nilai *total\_loss* atau biaya yang harus dikeluarkan maskapai akan sama dengan jumlah uang yang didapatkan dari penjualan premi. Penghitungan premi baru dapat dilihat pada Tabel 13.

**Tabel 13 Hasil Penghitungan Premi Baru**

Airlines	From	To	<i>plow</i> (IDR)	<i>phigh</i> (IDR)	Status
Citilink	CGK	KNO	28.858,52	28.858,52	Dapat dijual
Lion Air	DPS	CGK	52.000,00	52.531,84	Tidak dapat dijual
Sriwijaya	CGK	MLG	33.000,00	34.462,71	Dapat dijual

Pada hasil penghitungan premi baru yang ditunjukkan pada Tabel 13, penerbangan dengan maskapai Citilink dari kota Jakarta (CGK) menuju kota Medan (KNO), memiliki nilai premi rendah dan premi tinggi yang sama yaitu sebesar IDR 28.858,52 dan memiliki status dapat dijual karena nilai premi tersebut masih di bawah besar kompensasi pada kategori *delay* tipe 1 yaitu IDR 50.000. Sedangkan pada penerbangan dengan maskapai Lion Air dari kota Denpasar (DPS) menuju kota Jakarta (CGK), nilai premi rendah yang didapatkan adalah sebesar IDR 52.000 sedangkan nilai premi tinggi yang didapatkan adalah sebesar IDR 52.531,84, dan memiliki status tidak dapat dijual dikarenakan kedua nilai premi tersebut jumlahnya lebih dari besar kompensasi yang akan diterima penumpang ketika mengalami *delay* dengan kategori *delay* tipe 1 yaitu IDR 50.000. Dengan nilai premi baru yang telah didapatkan, penumpang mendapatkan keuntungan dengan pembelian premi tersebut. Sebagai contoh, pada penerbangan dengan maskapai Sriwijaya Air dari kota Jakarta (CGK) menuju kota Malang (MLG), nilai premi yang harus dibayarkan penumpang jika penerbangan tersebut terprediksi akan mengalami *delay* adalah IDR 34.462,71. Ketika *delay* dengan kategori *delay* tipe 1 terjadi, penumpang akan mendapatkan kompensasi sebesar IDR 50.000. Hal ini diharapkan dapat membuat penumpang berminat untuk membeli premi asuransi perjalanan yang ditawarkan. Pada pihak maskapai, dengan nilai biaya yang harus dikeluarkan oleh maskapai untuk memberikan kompensasi atau *total\_loss* menjadi sebanding dengan uang yang didapatkan dari hasil penjualan premi,

diharapkan pihak maskapai dapat meningkatkan layanan agar tidak terjadi *delay* dikarenakan biaya yang dikeluarkan untuk memberikan kompensasi kepada penumpang tidaklah sedikit.

### 4.3 Antar Muka Aplikasi

Pada Gambar 3 dibawah ditampilkan antar muka aplikasi yang dibuat meminta *input* yang terdiri dari *variable Airlines, From, To, Day, dan DepartType*. Setelah *input* dimasukkan, aplikasi akan mengeluarkan *output* berupa titik kota keberangkatan dan kota tujuan; hasil prediksi yaitu *Delay* atau *On Time* beserta nilai premi yang dimiliki oleh penerbangan tersebut.



**Gambar 3** Antar Muka Aplikasi

## 5 SIMPULAN

Prediksi *delay* atau keterlambatan penerbangan menggunakan metode *decision tree* telah berhasil diimplementasikan dengan akurasi sebesar 72,76% berdasarkan *confusion matrix*. Kemudian telah dilakukan proses validasi, yaitu proses membandingkan hasil prediksi *decision tree* dengan prediksi yang didapatkan melalui data historis dengan tingkat kemiripan sebesar 96,14% sehingga hasil prediksi *decision tree* yang didapatkan memiliki hasil yang *valid*. Penghitungan premi baru dengan menggunakan metode statistik juga telah berhasil dilakukan. Hasil nilai premi baru lebih rendah dari hasil nilai premi lama sehingga dapat disimpulkan bahwa nilai premi baru merupakan nilai premi yang lebih masuk akal baik bagi pihak maskapai maupun pihak penumpang.

## KEPUSTAKAAN

- [1] Badan Pusat Statistik, *Lalu Lintas Penerbangan Dalam Negeri Indonesia Tahun 2003-2018*. (2020). <https://www.bps.go.id/statictable/2009/02/21/1402/lalu-lintas-penerbangan-dalam-negeri-indonesia-tahun-2003-2018.html>. (diakses tanggal 17 Oktober 2020).

- [2] Aviation Safety Network, *Airliner Accident Fatalities Per Year 1946-2017*. (2017). <https://aviation-safety.net/graphics/infographics/Fatal-Accidents-Per-Year-1946-2017.jpg>. (diakses tanggal 17 Oktober 2020).
- [3] Pemerintah Indonesia, *Peraturan Menteri Perhubungan Republik Indonesia Nomor PM 89 Tahun 2015 Tentang Penanganan Keterlambatan (Delay Management) Pada Badan Usaha Angkutan Udara Niaga Berjadwal di Indonesia*. Lembaran RI Tahun 2015 No. 6. Jakarta: Sekretariat Negara. (2015).
- [4] Hungelmann, J, *Insurance for dummies*, Hoboken, NJ: Wiley. (2009).
- [5] Kagan, Julia, *Insurance Premium*. (2019). <https://www.investopedia.com/terms/i/insurance-premium.asp>. (diakses tanggal 17 Oktober 2020)
- [6] Lukas, S., Margaretha, H., Stefanim D., Widjaja, P., Feng, Ben, *Insurance Premium Model For Flight Delay Using Standar Deviation, GLM And SVM*. Universitas Pelita Harapan & University of Waterloo. (2019).
- [7] Kotsiantis, S. B., *Supervised Machine Learning: A Review of Classification Techniques*. Informatica 31, 249-268. (2007).
- [8] Rokach, L.; Maimon, O., *Top-down induction of decision trees classifiers-a survey*. IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews, 35 (4): 476–487. CiteSeerX 10.1.1.458.7031. doi:10.1109/TSMCC.2004.843247. (2008).