

Penerapan Data Mining Untuk Prediksi Slot Time di Bandara Internasional di Indonesia : Algoritma J48

Renddy Wandhana Suryaman^{*1}, Gunawan Wang²

^{1,2}Magister Manajemen Sistem Informasi Universitas Bina Nusantara Jakarta, Indonesia

E-mail: ^{*1} rendddyws@gmail.com, ² gunawan.wang261@gmail.com

Abstrak

Dalam dunia penerbangan, keselamatan dan kelancaran arus lalu lintas penerbangan merupakan business core, dimana setiap pelayanan lalu lintas penerbangan diharapkan tidak terjadi delay yang disebabkan oleh terjadinya holding pergerakan pesawat baik di udara maupun di darat. Oleh karena itu slot time di bandara sangat penting untuk ketepatan pergerakan pesawat udara baik yang Departure maupun yang Arrival, hal ini dimaksudkan untuk menghindari terjadinya delay yang disebabkan penumpukan antrian pesawat yang akan berangkat dan pesawat yang akan mendarat, dengan banyaknya pergerakan pesawat udara di Bandara Internasional Soekarno Hatta, maka dibutuhkan analisa yang mendalam dengan Teknik data mining seperti algoritma J48 dan Decesion Tree serta Naïve Bayes.

Kata Kunci— Data mining, J48, Decision Tree, Naïve Bayes, Slot Time

1. PENDAHULUAN

Sebuah pelayanan lalu lintas penerbangan dikatakan baik apabila seluruh sistem di dalam pelayanan lalu lintas penerbangan itu saling terintegrasi dan berjalan dengan lancar, agar proses pelayanan lalu lintas penerbangan sebagai business core perusahaan dapat berjalan dengan baik. AirNav Indonesia merupakan perusahaan pelayanan navigasi penerbangan yang sangat mengedepankan pelayanan dan keselamatan penerbangan, oleh karena itu apabila terjadi antrian panjang untuk setiap pesawat yang akan mendarat maupun yang akan terbang di seluruh bandara di Indonesia, khususnya bandara internasional Soekarno Hatta, maka hal ini sangat mempengaruhi keselamatan penerbangan dan bisnis proses perusahaan. Dengan demikian, maka slot time di masing-masing bandara di Indonesia menjadi perhatian yang cukup serius, khususnya di jam-jam sibuk (golden time) di mana setiap airline saling berebut untuk mendapatkan slot time pada jam tersebut, apabila tidak ter-manage dengan baik, maka dapat terjadi chaos dan terdampaknya keselamatan penerbangan yang dapat mengakibatkan insiden maupun kecelakaan pesawat.

Dalam pelayanan lalu lintas penerbangan, flight plan merupakan data utama, di mana didalamnya termuat semua informasi tentang rencana keberangkatan pesawat sampaid engan tujuan dan informasi-informasi penting lainnya. Data flight plan tersebut juga merupakan data yang kemudian akan menjadi dasar untuk billing pelayanan lalu lintas penerbangan yang diberikan oleh AirNav Indonesia kepada Airline. Data flight plan tersebut akan dikirimkan kepada Air Traffic Service System (ATS System) dan akan diaktifasi oleh berita Departure apabila pesawat sudah terbang, untuk kemudian di de-aktivasi dengan berita Arrival setelah pesawat mendarat dengan selamat.

Proses pemberian slot time saat ini masih bersifat manual berdasarkan beberapa parameter yaitu :permintaan dari airline, ketersediaan parker pesawat udara di bandara, kesediaan bandara tujuan menerima, ketersediaan slot time itu sendiri baik di bandara keberangkatan maupun bandara tujuan. Apabila perusahaan memiliki analisa yang baik terhadap slot time tersebut, dengan me-manage setiap pergerakan pesawat baik itu yang sesuai jadwal maupun yang

delay atau cancel flight yang disebabkan kendala operasional atau teknis, yang bukan disebabkan karena adanya antrian, maka hal ini akan sangat efektif dan efisien yang dapat memberikan dampak langsung kepada setiap stakeholder di bandara, baik perusahaan itu sendiri maupun bandara dan airline yang berujung pada pelayanan yang prima dan tepat waktu untuk penumpang pesawat udara sebagai pelanggan utama pada bisnis proses pelayanan transportasi udara.

Bandara Internasional Soekarno Hatta sendiri memiliki pergerakan take off dan landing rata-rata setiap hari antara 800 sampai dengan 1000 pergerakan dan mencapai 1200 per hari pada waktu-waktu tertentu. Saat ini bandara internasional Soekarno Hatta memiliki kapasitas untuk menampung 86 pergerakan pesawat per jam dengan menggunakan 2 landasan pacu. Proses peningkatan kapasitas bandara dengan menambah jumlah runway atau memindahkan beberapa penerbangan ke bandara lain yang terdekat bukanlah solusi yang tepat karena membutuhkan banyak biaya dan persiapan yang sangat lama karena regulasi penerbangan yang sangat detail dan ketat, hal ini dikarenakan masalah keselamatan merupakan business core dalam bisnis transportasi udara..

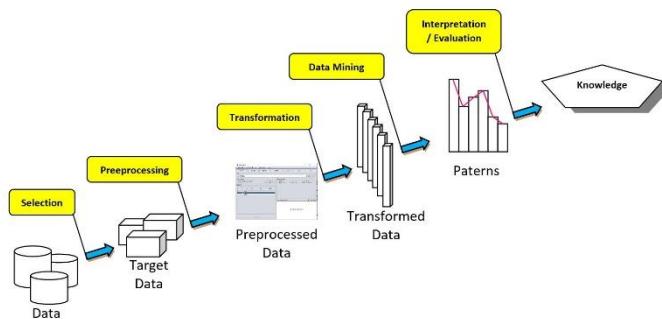
Seiring berjalaninya waktu dan perkembangan ekonomi, hal ini akan mendorong berkembangnya jalur transportasi udara yang mengakibatkan peningkatan pergerakan pesawat, sehingga perlu adanya mitigasi dengan menganalisa pergerakan pesawat untuk meningkatkan occupancy dari slot waktu yang tersedia. Setiap elemen yang merupakan bagian dari penambangan data untuk slot waktu akan dijelaskan lebih lanjut.

Oleh karena itu solusi teknik data mining dengan pemodelan klasifikasi menjadi salah satu solusi untuk memprediksi ketersediaan slot time khususnya pada bandara internasional Soekarno Hatta apakah nantinya akan available atau not-available. Karena setiap harinya akan terdapat pesawat delay atau cancel flight dan extra flight yang dapat merubah susunan slot time pada setiap jam nya, khususnya pada golden time untuk menambah occupancy yang berdampak pada kelancaran arus lalu lintas penerbangan, di mana efek terakhir yang diterima masing-masing stakeholder adalah peningkatan pelayanan terhadap konsumen, efisiensi pengeluaran (avtu rpesawat, parker pesawat, dll) serta peningkatan pendapatan perusahaan. Data mining ialah suatu proses di mana menerapkan berbagai teknik seperti statistics, math, artificial intelligence, serta learning machine. Adapun teknik-teknik tersebut dimanfaatkan sebagai identifikasi dan ekstraksi suatu informasi untuk menambah wawasan pengetahuan mengenai sumber-sumber database yang besar maupun Data Warehouse. Model prediktif merupakan proses yang menggunakan berbagai gabungan teknik data mining dan analisa data yang dapat menghasilkan informasi perkiraan yang diinginkan.

2. DATA MINING TECHNIQUE

Data mining berasal dari kata mining, yang artinya menambang jika dikembangkan dapat menjadi menggali data yang dimiliki. Data mining adalah investigasi informasi terpadu yang terdiri dari serangkaian latihan yang dijalankan berdasarkan arti dari target yang akan dipecah, dengan pemeriksaan informasi untuk pemahaman dan penilaian hasil. Koleksi penambangan informasi tidak hanya mengumpulkan informasi tetapi mencakup penyelidikan dan harapan data yang akan ditampilkan. Informasi yang terkumpul disimpan dalam suatu kumpulan data dan kemudian ditangani sehingga cenderung dimanfaatkan untuk pengambilan keputusan di dalamnya.

Untuk menggali informasi berbasis data lebih lanjut, biasanya menggunakan data mining serta Knowledge Discovery in Data Bases (KDD), di mana penerapannya dilakukan dengan bergantian. Adapun penggunaan data mining dan KDD ini dilakukan karena informasi yang akan diproses biasanya berada dalam basis data dengan ukuran yang besar, namun keduanya masih saling terkait. Berikut ialah gambaran dari proses KDD:



Gambar 1. Skema KDD

Dari tahap-tahap proses KDD ialah data mining, adapun tahapannya ialah seperti di bawah ini :

1. Data Selection (Pemilihan Data), yaitu proses pengambilan data pada lokasi terkait sesuai dengan analisis yang relevan.
2. Data Preprocessing/Cleaning (Pemrosesan/Pembersihan Data), yaitu tahap dilakukannya proses pembersihan data, sekaligus pemeriksaan apabila terdapat kekurangan data, data yang masih kosong, adanya duplikasi, serta data-data yang kurang relevan.
3. Transformation (Transformasi), yaitu tahap memilih data dari hasil data preprocessing/cleaning yang bertujuan untuk mendapatkan data seperti yang diharapkan pada data mining.
4. Data mining, yaitu cara untuk menemukan pola yang diinginkan melalui penerapan dari suatu metode dari data yang akan ditampilkan.
5. Interpretation/Evaluation (Interpretasi atau Evaluasi), adalah tahap penerjemahan dari suatu teknik atau pola dari hasil olah data mining dengan tujuan mempermudah pemahaman dari informasi yang dihasilkan.

2.1. Decision Tree

Salah satu informasi yang biasa digunakan untuk menambang metode adalah Decision Tree. Kemampuan beradaptasi dari prosedur ini membuatnya sangat memikat, terutama karena menyajikan sisi positif dari persepsi yang sangat lugas di mana bagian-bagian pohon merangkum karakterisasi. Pohon pilihan memiliki tiga gaya klasik yaitu:

1. Classification Tree (Pohon Klasifikasi), teknik ini diterapkan pada saat hasil perkiraan merupakan keanggotaan kelas, sebagai contoh Algoritma decesssion tree.
2. Regression Tree (Pohon Regresi), teknik ini diterapkan ketika hasil perkiraan merupakan bilangan asli, sebagai contoh harga bahan bakar, nilai bangunan.
3. CART (C & RT) adalah denominasi dan pohon regresi. Berkembangnya pohon keputusan telah banyak ditemukan, namun ID3 dan decision tree adalah dua teknik yang menjadi favorit untuk analisis penelitian. Kedua teknik ini memiliki kesamaan prinsip, karena Algoritma decision tree dikembangkan melalui ID3. Meskipun demikian, terdapat adanya konsep berbeda dari kedua teknik ini yaitu :
 - Pohon pilihan dapat menangani kualitas yang konsisten dan terpisah dan juga dapat menangani persiapan informasi dengan kualitas yang hilang atau informasi yang tidak valid.
 - Hasil yang diperoleh dari perhitungan pohon pilihan akan dikelola setelah penentuan sifat dilakukan dengan memanfaatkan Gain Ratio
 - Perhitungan pohon pilihan merupakan peningkatan dari ID3 menggunakan Gain Ratio Keuntungan untuk menyegarkan perolehan data menggunakan persamaan :

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)}$$

Di mana :

S : Ruang/Data Sample dimanfaatkan untuk pelatihan data

A : Atribut Gain (S,A) = information gain untuk atribut A Split Info (S,A) = split information untuk atribut A

Properti dengan Gain Ratio Keuntungan paling tinggi dipilih sebagai kualitas uji untuk hub. Pendekatan ini menerapkan standarisasi pada perolehan data dengan memanfaatkan apa yang disebut dengan parted data, dengan persamaan :

$$SplitInfo(S, A) = - \sum_{i=1}^k \frac{S_i}{S} \log_2 \frac{S_i}{S}$$

Di mana :

S : Ruang (data) sample yang dipakai untuk pelatihan.

A : Atribut.

S_i : Jumlah sample pada atribut i

Saat membangun pohon pilihan, mungkin ada keributan atau informasi kosong dalam informasi persiapan. Pemangkas pohon harus dimungkinkan untuk mengenali dan menghilangkan cabang-cabang ini sehingga pohon lebih sederhana dan lebih jelas untuk pengaturan yang lebih baik. Ada dua strategi untuk memangkas pohon pilihan, yaitu :

- Melalui Prepruning misalnya berhenti membangun dari awal subpohon sehingga tidak melangkah lebih jauh dalam pemangkas informasi persiapan. Persamaan Prepruning :

$$\theta = \frac{r + \frac{Z^2}{2n} + Z \sqrt{\frac{r}{n} - \frac{r^2}{n} + \frac{Z^2}{4n^2}}}{1 + \frac{Z^2}{n}}$$

Di mana

r : nilai perbandingan error rate

n : total sampel

z : $\Phi^{-1}(c)$

c : confidence level

- Menggunakan Postpruning misalnya mengerjakan pohon dengan menghilangkan sebagian cabang subpohon setiap kali dibuat. Teknik ini adalah bagian standar dari perhitungan pohon pilihan.

2.2. Naïve Bayes

Naive Bayes adalah strategi karakterisasi yang sering digunakan untuk mengukur nilai yang tandanya tidak jelas. Penggunaan strategi Naïve Bayes memerlukan sedikit persiapan informasi yang pasti disebut menyiapkan informasi yang digunakan untuk menentukan batas-batas yang dinilai yang diperlukan dalam siklus pesanan. Strategi Naïve Bayes juga merupakan teknik yang menggunakan perkiraan kemungkinan satu kelas dari setiap kumpulan sifat saat ini, dan menentukan kelas mana yang paling ideal. Perhitungan Naïve Bayes memiliki fase siklus yang harus diselesaikan, untuk lebih spesifiknya yaitu :

1. Hitung jumlah kelas / label.
2. Hitung Jumlah Kasus Per Kelas
3. Kalikan Semua Variable Kelas
4. Analisis Hasil Per Kelas

Di bawah ini ialah formula persamaan dari teorema Bayes :

$$P(H | X) = (P(X | H) \cdot P(H)) / (P(X))$$

Dimana :

X : Data dengan kelas yang belum diketahui

H : Hipotesis data ialah kelas yang spesifik

$P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

$P(H)$: Probabilitas hipotesis H (priorprobabilitas)

$P(X|H)$: Probabilitas X sesuai dengan kondisi pada hipotesis H

Untuk memahami teknik Naive Bayes, harus diperhatikan bahwa interaksi karakterisasi memerlukan berbagai pedoman untuk mengetahui kelas apa yang masuk akal untuk tes yang diselidiki. Dengan cara ini, teknik Naive Bayes di atas diubah sebagai berikut:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)}$$

Di mana Variabel C membahas kelas, sedangkan variabel $F_1 \dots F_n$ membahas kualitas arah yang diharapkan untuk memainkan pengaturan. Kemudian persamaan masuk akal bahwa kemungkinan masuknya contoh kualitas tertentu di kelas C (Posterior) adalah kemungkinan kehadiran kelas C (sebelum berlalunya contoh, biasa disebut sebelumnya), digandakan dengan kemungkinan kejadian atribut uji di kelas C (juga disebut probabilitas), dipisahkan dengan kemungkinan pengembangan atribut uji di seluruh dunia (juga disebut bukti). Dengan cara ini, persamaan di atas juga dapat disusun persamaan yaitu :

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

Nilai bukti (evidence) umumnya ditetapkan untuk setiap kelas dalam satu contoh. Nilai bagian belakang akan dikontraskan dan bagian belakang terbalik dari kelas yang berbeda untuk memutuskan kelas mana yang akan diklasifikasikan sebagai contoh. Elaborasi lebih lanjut dari persamaan Bayes diselesaikan dengan menguraikan $(C|F_1, \dots, F_n)$ menggunakan aturan duplikasi yang menyertainya:

$$\begin{aligned} P(C|F_1, \dots, F_n) &= P(C)P(F_1, \dots, F_n|C) \\ &= P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2) \\ &= (C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2)P(F_4, \dots, F_n|C, F_1, F_2, F_3) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \end{aligned}$$

Sangat mungkin terlihat bahwa konsekuensi dari elaborasi menyebabkan semakin banyak elemen rumit yang mempengaruhi harga kemungkinan, yang sangat sulit untuk diselidiki satu per satu. Oleh karena itu, estimasi menjadi sulit dilakukan. Ini adalah tempat di mana anggapan kebebasan yang sangat tinggi (tidak bersalah) digunakan, bahwa setiap petunjuk (F_1, F_2, \dots, F_n) adalah otonom satu sama lain. Dengan asumsi-asumsi ini, perumpamaan yang menyertainya yaitu :

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i)$$

Untuk $i \neq j$, sehingga

$$P(F_i|C, F_j) = P(F_i|C)$$

Kondisi di atas adalah model dari hipotesis Naive Bayes yang selanjutnya akan digunakan dalam sistem pengelompokan. Untuk karakterisasi dengan informasi persisten, persamaan Gauss Density digunakan :

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

Di mana :

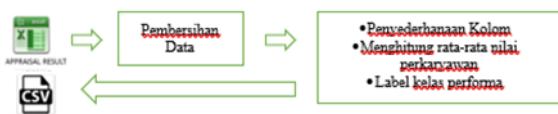
- P : Peluang
- X_i : Atribut ke i
- x_i : Nilai atribut ke i
- Y : Kelas yang dicari y_i : Sub kelas Y yang dicari
- μ : mean, yang berisi rata – rata dari seluruh atribut
- σ : Standart Deviation, yaitu berisi varian dari semua atribut.

3. DATA MINING PROCESSS

Data dianalisa menggunakan metode klasifikasi untuk memprediksi alokasi slot time pada bandara Soekarno Hatta.

3.1. Data Preperation

Kumpulan data disini berasal dari sistem webbase flight plan dan akan didapatkan data nilai appraisal dari sistem appraisal dan di export ke excel sesuai dengan template excel yang akan diolah.



Gambar 2. Ekport Dataset

3.2. Pengolahan Data Awal

Melalui informasi penerbangan yang sudah diperoleh melalui pengumpulan data, selanjutnya adalah pembersihan informasi, khususnya dengan mengeluarkan catatan-catatan yang kurang jelas dan mengulang kembali catatan-catatan tersebut, juga dapat diambil sifat-sifat yang berlebihan, misalnya nomor ciri, pendaftaran pesawat, nomor penerbangan. Hal ini dilakukan mengingat fakta bahwa karakteristik ini secara signifikan mempengaruhi penanganan informasi dalam sistem. Berikut adalah informasi yang dianggap digunakan:

- C1 : Schedule Flight
- C2 : Un-schedule Flight
- C3 : Domestic Flight
- C4 : International Flight
- C5 : PBN (performance based navigation)
- C6 : Winter Season

C7 : Summer Season

C8 : Night Flight

C9 : Day Flight

Dan masing masing dari parameter mempunyai nilai yang sebagai berikut :

1 : Tidak disarankan

2 : Kurang Disarankan

3 : Pertimbangan Operasional

4 : Disarankan

5 : Sangat Disarankan

Data set yang sudah diekspor dari sistem webbase flight plan seperti gambar dibawah :

A	B	C	D	E	F	G	H	I	J	K	L
FPL	C1	C2	C3	C4	C5	C6	C7	C8	C9	Hasil	
1	1	5	5	5	5	5	5	5	5	5 Sangat Disarankan	
2	1	1	1	2	2	2	3	2	2	2 Tidak Disarankan	
3	4	3	5	5	5	5	5	3	3	5 Disarankan	
4	5	5	5	5	5	5	5	5	5	5 Disarankan	
5	4	4	5	5	5	5	4	4	4	4 Disarankan	
6	5	5	5	5	3	4	4	4	5	5 Disarankan	
7	4	4	4	4	5	5	5	4	5	5 Disarankan	
8	4	5	4	5	4	5	4	5	4	4 Disarankan	
9	5	5	5	5	5	5	4	4	4	4 Disarankan	
10	5	5	5	5	4	4	4	5	5	4 Disarankan	
11	2	1	3	2	2	2	2	3	1	Tidak Disarankan	
12	4	4	4	4	4	4	4	5	5	5 Disarankan	
13	5	5	5	5	5	5	5	5	5	5 Sangat Disarankan	
14	4	5	4	5	4	5	4	3	3	5 Disarankan	
15	3	5	5	5	5	5	4	4	4	5 Disarankan	
16	5	5	5	5	4	4	4	4	4	4 Disarankan	
17	5	4	5	4	5	4	5	5	5	4 Disarankan	
18	4	4	5	5	4	4	5	5	5	5 Disarankan	
19	4	5	4	5	4	5	5	4	4	4 Disarankan	
20	5	3	3	5	5	5	5	5	5	5 Disarankan	
21	5	4	5	5	4	3	3	5	5	5 Disarankan	
22	2	5	5	5	5	5	5	5	5	5 Disarankan	
23	5	3	5	5	3	5	5	5	5	5 Disarankan	
24	4	5	5	4	5	4	5	4	4	4 Disarankan	
25	4	4	3	3	3	3	3	4	4	4 Disarankan	
26	4	3	3	3	3	3	3	4	4	4 Disarankan	
27	3	2	3	3	3	3	3	3	3	3 Kurang Disarankan	

Gambar 3. Dataset

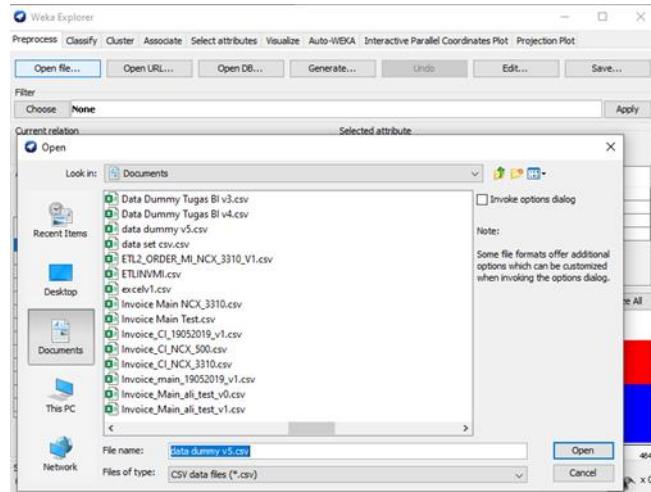
4. MINING MODELS

Berbagai algoritma dan teknik Klasifikasi, Clustering, Regresi, Kecerdasan Buatan (AI), Jaringan Saraf Tiruan, Naive Bayes, Pohon Keputusan, Algoritma Genetika, K-Neighboards dll. Klasifikasi adalah salahsatunya salah yang paling sering dipelajari oleh data peneliti pertambangan dan machine learning (ML) yang terdiri dari memprediksi nilai atribut (kategori / kelas) berdasarkan nilai-nilai attribute lainnya (atribut prediksi). Adametode klasifikasi yang berbeda. Dalam penelitian ini kami menggunakan algoritma Pohon Keputusan dan Naive Bayes.

5. APPLICATION OF DECISION TREE ON PERFORMANCE OF EMPLOYEE

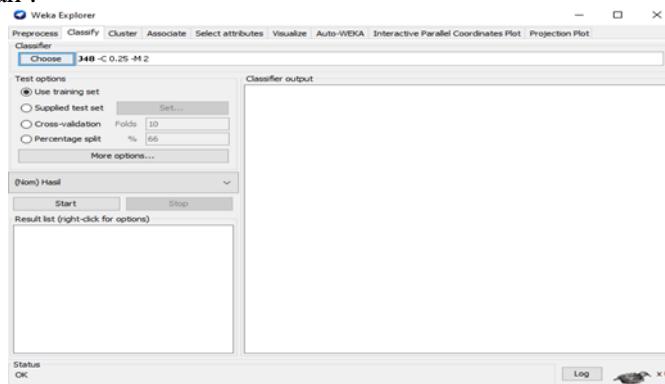
Decision Tree merupakan metode tambahan yang mendukung keputusan dengan menerapkan grafik yang berbentuk pohon atau model keputusan di mana ini akan memperlihatkan atau menjelaskan kemungkinan yang anak terjadi. Penggunaan decision tree biasa ditemukan dalam operasi penelitian, yaitu untuk membuat keputusan maupun identifikasi metode terbaik dalam pencapaian tujuan tertentu.

Pada penelitian ini pengolahan data akan menggunakan Weka. Untuk langkah pertama kita memilih file excel yang akan diolah menggunakan Weka. Langkah pertama akan seperti gambar dibawah



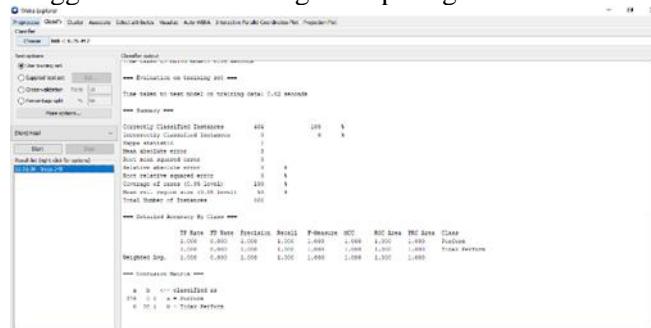
Gambar 4. Import Data Weka

Kemudian setelah melakukan import data ke Weka maka pilih algoritma yaitu J48 pada Weka seperti gambard dibawah :



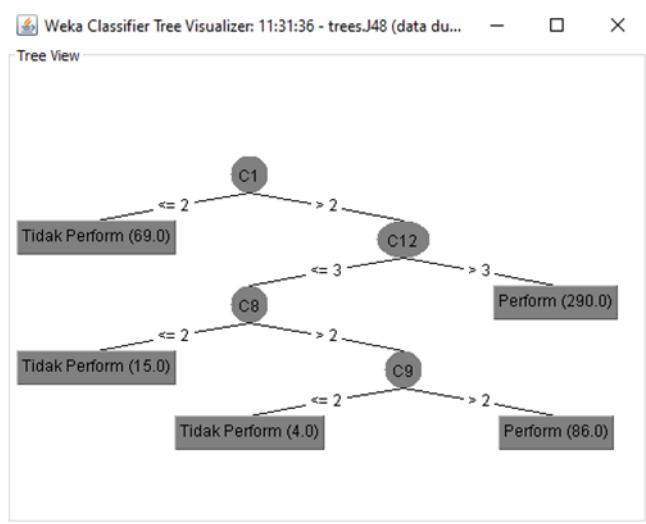
Gambar 4. Algoritma J48

Untuk pengujian disini menggunakan 2 pilihan yaitu use training set, cross validation fods 10%,. Untuk pengolahan data menggunakan use training set seperti gambar dibawah :



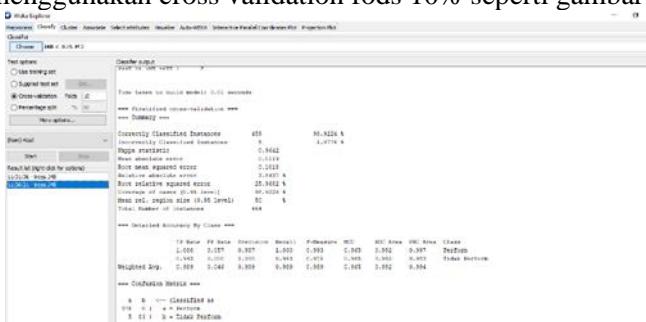
Gambar 5. Algoritma J48 Use Trainning Set

Dan Visualisasi dengan Tree seperti gambar dibawah :



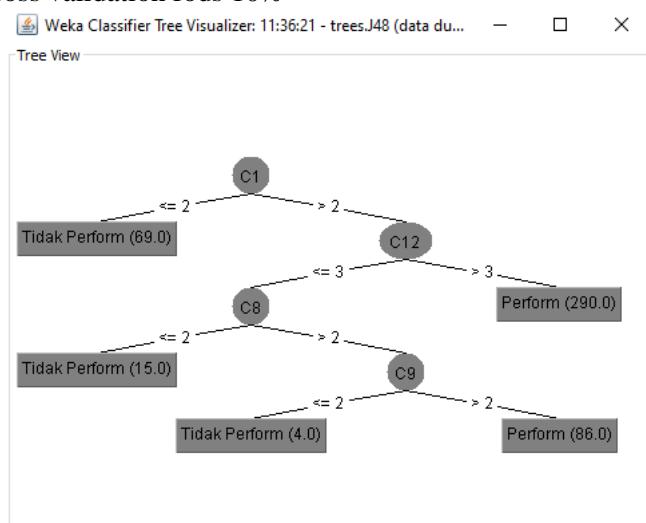
Gambar 6 . Visualisasi Tree Algoritma J48 Use Training Set

Untuk validasi dengan menggunakan cross validation fods 10% seperti gambar dibawah :



Gambar 7. Algoritma J48 Cross Validation

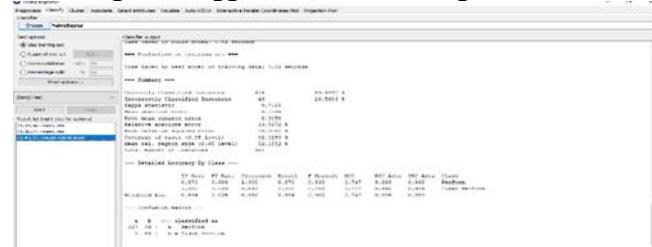
Gambar diatas adalah pengolahan data dengan menggunakan cross validation 10%. Kelebihan menggunakan cross validation dimana data yang akan ditampilkan akan lebih akurat. Berikut visualisasi tree cross validation fods 10%



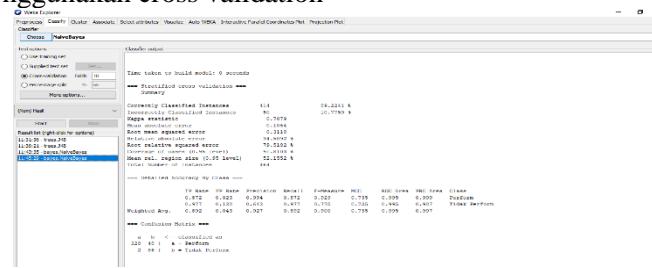
Gambar 8. Visualisasi Tree Algoritma J48 Use Cross Validation.

6. APPLICATION OF NAÏVE BAYES ON PERFORMANCE OF EMPLOYEE

Untuk Pengolahan Naïve Bayes hampir sama dengan menggunakan algoritma decision tree pada Weka dimana harus meilih file excel yang sudah ditentukan. Kemudian pilih algoritma yaitu naïve Bayes. Untuk pengolahan data sendiri sama menggunakan use training set dan cross validation fods 10%. Berikut dengan menggunakan use training set



Gambar 8. Pengolahan Data Menggunakan Naïve Bayes Use Training Set Dan gambar dibawa menggunakan cross validation



Gambar 8. Pengolahan Data Menggunakan Naive Bayes Cross Validation.

7. EVALUSI MODEL

Untuk penerapan pada algoritma decision tree dan dengan jumlah attribute setalah diolah sumbernya terdapat 9 attribute dengan sum of weight 484 dan instance 464. Dari hasil algoritma decision tree memiliki correctly classified instance sekitar 98, 9224% dan incorrectly classified instance sekitar 1.0076% dengan menggunakan cross validation dan menggunakan cross validation didapatkan correctly instance sebesar 100% dan inccoret instance 0%. ROC Area perform 0.997 dan tidak perform sekitar 0.983 dengan menggunakan cross validation dan didapatkan rule sebagai berikut :

- R1 : Jika tipe penerbangan merupakan schedule flight dan memiliki performance based navigation (PBN) maka termasuk pesawat yang sangat disarankan.
 - R2 : Jika penerbangan internasional kurang dari sama dengan 3 dan pernerbangan unschedule lebih besar dari 2 dan maka memiliki hasil disarankan
 - R3 : Jika C1 kurang dari 2 maka dapat menjadi pertimbangan operasional untuk pemberian slot time nya
 - R4 : Jika PBN kurang dari sama dengan 2 maka hasilnya adalah kurang disarankan
 - R5 : Jika C1 dan C3 kurang sama dengan 2 maka tidak hasilnya tidak disarankan

Untuk hasil dari algoritma Naive Bayes memiliki correctly classified instance 87.931% dan incorrectly classified instance 12.096%. Tingkat ROC Area perform 0.997 untuk perform dan tidak perform dengan menggunakan cross validation 10%. Menggunakan use training set didapatkan correct instance sebesar 89,4937% dan incorrect instance sebesar 10,5063% dan roc area untuk perfom sebesar 0,996 dan roc area untuk 0,996.

8. EVALUSI MODEL

Untuk penerapan pada algoritma decision tree dan dengan jumlah attribute setalah diolah sumbernya terdapat 15 attribute dengan sum of weight 484 dan instance 464. Dari hasil algoritma decision tree memiliki correctly classified instance sekitar 98, 9224% dan incorrectly classified instance sekitar 1.0076% dengan menggunakan cross validation dan menggunakan cross validation didapatkan correctly instance sebesar 100% dan incorrect instance 0%. ROC Area perform 0.997 dan tidak perform sekitar 0.983.

Dengan kesalahan yang lebih kecil maka dapat disimpulkan bahwa penerapan menggunakan algoritma decision tree lebih akurat untuk memprediksi dan melihat slot time yang tersedia secara keseluruhan dan melihat occupancy dari slot tersebut..

DAFTAR PUSTAKA

- [1] Baradwaj, B. K., & Pal, S. (2012). Mining Educational Data to Analyze Students Performance. International Journal of Advanced Computer Science and Applications, 63-69
- [2] Kotalwar, R., Chavan, R., Gandhi, S., & Parmar, V. (2014). Data Mining: Evaluating Performance of Employee's using Classification Algorithm Based on Decision Tree.
- [3] Application of Data Mining Classification in Employee Performance Prediction. (2016).
- [4] Cacho, C. d. (2010). A comparison of data mining methods for mass real estate appraisal.
- [5] Kaunang, F. J. (2019). Penerapan Algoritma J48 Decision Tree Untuk Analisa Tingkat Kemiskinan di Indonesia.
- [6] Kaur, G., & Chhabra, A. (2014). Improved J48 Classification Algorithm for the Prediction of Diabetes.
- [7] NirmalaG, & Mallikarjuna, P. B. (2014). Faculty Performance Evaluation Using Data Mining.
- [8] Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., & Rong, X. (2015). Data Mining for the Internet of Things: Literature Review and Challenges. International Journal of Distributed Sensor Networks, 2.
- [9] Dongming, L., Yan, L., Chao, Y., Chaoran, L., Huan, L., & Lijuan, Z. (2016). The Application of Decision Tree C4.5 Algorithm to Soil Quality Grade Forecasting Model. First IEEE International Conference on Computer Communication And the Internet, 552-555.
- [10] Gorunescu, F. (2011). Data Mining Concepts, Models, and Techniques. Verlag Berlin Heidelberg: Springer.
- [11] Han, J., Kamber, M., & Pei, J. (2012). Data Mining Concepts And Techniques Third Edition. Waltham: Morgan Kaufmann

- [12] Jailani, Defit, S., & Nurcahyo, G. W. (2015). Penerapan Algoritma C4.5 Pada NUPTK Untuk Menentukan Pola Sertifikasi Guru Dengan Menggunakan Metode Klasifikasi Decision Tree. Riau Journal of Computer Science Vol.1/No.1, 69-83.
- [13] Larose, D. T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. New Jersey: John Wiley & Sons, Inc.
- [14] Indonesia, AirNav (2016). Web base flight plan : web link.
- [15] Tayefi, M., Tajfard, M., Saffar, S., Hanachi, P., Amirabadizadeh, A. R., Esmaeily G. A. (2017). hs- CRP is Strongly associated with coronary heart disease (CHD): A Data mining approach using decision tree algorithm. Computer Methods and Programs in Biomedicine, 105-109..