

# Analisis *Delay* Penerbangan Akibat Cuaca di Bandara

## Ahmad Yani Semarang dengan Algoritma C4.5

( Mochamad Nur Sholikhin, Dra. Yuniarsi Rahayu )

The role of weather in a very large cost. Weather has two roles. On one side of weather information have contributed to the improvement of the efficiency and effectiveness of the activity and safety of flight, on the other hand has the potential to cause harm to death. However, it is not easy to say where the weather is dangerous, because the weather impact depends also to other factors. For that conducted research titled Analysis of Flight Delay Due to Weather Ahmad Yani Airport in Semarang with the C4.5 algorithm. This technique uses weather data and flight delay flight list in 2013 at Ahmad Yani Airport in Semarang, so the data can be used as a reference to predict whether there will be a delay flight or not. Recommendation feasibility flight is affected by wind direction, wind speed, temperature, air pressure, visibility, weather. And obtained the results that C4.5 algorithm is applied to the data set Delay flight in 2013, the data generating confusion matrix accuracy value of 94.55% and an accuracy within 0.815 AOC interval 0 seconds.

**Keyword : Flight Delay Prediction, C4.5 Algorithm, analysis, weather**

### I. PENDAHULUAN

Cuaca adalah kondisi udara di suatu tempat pada saat yang relatif singkat yang meliputi kondisi suhu, kelembaban, serta tekanan udara sebagai komponen utamanya. Faktor cuaca menjadi hal yang sulit untuk diprediksi pada penerbangan, oleh karena itu peramalan cuaca akhir-akhir ini menjadi topik yang sangat menarik untuk dibahas, karena akan sangat membantu dalam penerbangan. Hal ini membuat banyak peneliti tertarik untuk mencari metode lain untuk memprediksi cuaca [1].

Peran cuaca dalam penerbangan sangat besar. Cuaca mempunyai dua peran. Disatu sisi informasi cuaca mempunyai andil dalam peningkatan efisiensi dan efektivitas kegiatan dan keselamatan penerbangan, di sisi lain mempunyai potensi yang membahayakan sampai dapat menimbulkan kematian. Namun demikian tidak mudah untuk mengatakan cuaca yang mana yang membahayakan. Faktor cuaca sendiri dipengaruhi oleh beberapa elemen Antara lain :

1. Arah angin
2. Kecepatan Angin
3. Suhu
4. Tekanan Udara
5. Visibility

tetapi dampak cuaca bergantung pula kepada faktor lain.

Kusunya dalam penerbangan, selain kadar atau intensitas unsur cuaca, jenis pesawat, kondisi pesawat, dan posisi penerbangan juga merupakan faktor yang menentukan sensitifitasnya terhadap cuaca. Misalnya angin silang (cross wind) di landasan terbang yang berkecepatan 20 knot, mungkin dapat menimbulkan bahaya bagi pesawat kecil

yang melakukan pendaratan, tetapi tidak ada pengaruhnya bagi pesawat terbang besar dan modern. Dari posisi terbang, angin 20 knot pada paras penerbangan 30.000 kaki tidak dirasakan bagi pesawat besar yang terbang pada paras tersebut, tetapi bila terjadi pada paras rendah sangat berarti bagi pesawat terbang kecil yang terbang pada paras tersebut. Dengan demikian kriteria membahayakan bergantung juga kepada macam penerbangan. Namun demikian karena setiap pesawat terbang mempunyai tiga kegiatan yang sama, yakni tinggal landas, terbang, dan mendarat maka penggunaan arti bahaya dalam penerbangan umumnya diterapkan untuk masing-masing kegiatan tersebut [2].

Penundaan penerbangan terjadi karena cuaca yang buruk sehingga akan berakibat kepada mesin pesawat. Akibat dari penundaan penerbangan tersebut banyak penumpang yang menumpuk di bandara dan mereka kesal karena merasa tidak mendapat informasi yang memadai[3]. Untuk itu dalam rangka memenuhi kebutuhan informasi para penumpang pihak Bandara Ahmad Yani Semarang, dibuatlah analisis cuaca yang dapat memberikan informasi penundaan penerbangan. Tujuannya adalah pihak maskapai lebih mudah dan cepat mengambil keputusan untuk melakukan *delay* penerbangan. Oleh karena diperlukan data cuaca dari pihak BMKG Penerbangan Ahmad Yani Semarang untuk melakukan analisis *delay* penerbangan. Analisis tersebut dapat menggunakan metode Klasifikasi dengan algoritma C4.5. Dengan data yang sudah di olah menggunakan algoritma C4.5 diharapkan dapat mempermudah pihak Bandara maupun Maskapai Penerbangan dalam mengambil keputusan untuk melakukan *delay* penerbangan yang di akibatkan oleh cuaca yang membahayakan keselamatan penumpang.

Dari masalah yang sudah dibahas sebelumnya, maka penulis melakukan penelitian dengan judul “Analisi Delay Penerbangan Akibat Cuaca di Bandara Ahmad Yani Semarang dengan Algoritma C.4.5”.

Algoritma C 4.5 umumnya digunakan untuk melakukan klasifikasi. Ada beberapa algoritma yang pada umumnya digunakan dalam klasifikasi selain C 4.5 dan ID3 terdapat juga algoritma K-Nearest Neighbor [4]. Kelebihan Algoritma C.4.5 antara lain :

1. Daerah pengambilan keputusan yang sebelumnya kompleks dan sangat global, dapat diubah menjadi lebih simpel dan spesifik.
2. Eliminasi perhitungan-perhitungan yang tidak diperlukan, karena ketika menggunakan metode pohon keputusan maka sample diuji hanya berdasarkan kriteria atau kelas tertentu.
3. Fleksibel untuk memilih fitur dari internal node yang berbeda, fitur yang terpilih akan membedakan suatu kriteria dibandingkan kriteria yang lain dalam node yang sama. Kelembeluhan metode pohon keputusan ini meningkatkan kualitas keputusan yang dihasilkan jika dibandingkan ketika menggunakan metode penghitungan satu tahap yang lebih konvensional.
4. Mampu mengolah data nominal dan kontinyu.

## II. METODE YANG DIUSULKAN

Untuk menghindari penyimpangan dari judul dan tujuan yang sebenarnya serta keterbatasan pengetahuan yang dimiliki penulis, maka penulis membuat ruang lingkup dan batasan masalah yaitu :

1. Pada penelitian ini peneliti akan menggunakan metode klasifikasi algoritma C.4.5 untuk menentukan *delay* terhadap penerbangan.
2. Data yang digunakan untuk pengamatan adalah data cuaca tahun 2013 dari BMKG Penerbangan Bandara Ahmad Yani Semarang dan data delay penerbangan dari PT. ANGKASAPURA I (persero).
3. Menampilkan hasil prediksi *delay* penerbangan.

No	Peneliti	Tahun	Judul	Metode	Hasil
1	Aa Zezen Zaenal Abidin	2011	Implementasi Algoritma C 4.5 untuk Menentukan Tingkat Bahaya Tsunami	C4.5	Algoritma C 4.5 dapat digunakan untuk menentukan tingkat bahaya tsunami di suatu daerah pesisir pantai mengacu pada kasus tingkat bahaya tsunami yang sudah ada di wilayah pesisir pantai Kabupaten Sukabumi.
2	Vina Mandasari	2011	Analisis Kepuasan Konsumen	C4.5	Adanya hubungan sebab-akibat yang didapatkan dari rules dataset kepuasan

			Terhadap Restoran Cepat Saji Melalui Pendekatan Data Mining: Studi Kasus XYZ		konsumen memberikan informasi baru kepada manajemen restoran cepat saji bahwa atribut rasa, perilaku staf, suasana restoran dan harga berkaitan erat dalam menciptakan rasa puas untuk konsumen.
3	Risty Jayanti Yuniar	2013	Perbaikan Metode Prakiraan Cuaca Bandara Abdulrahman Saleh dengan Algoritma Neural Network Backpropagation	Algoritma Neural Network Backpropagation	Proses training menghasilkan arsitektur jaringan terbaik dengan hidden layer 5 dan nilai learning rate 0,9. Nilai MSE sebesar 0,009946 didapatkan pada epoch ke-13. Jaringan dirancang dengan menggunakan tiga variabel input yaitu suhu udara, kelembaban udara, tekanan udara dan dua variabel output yaitu kecepatan angin dan curah hujan.
4	Nur Endah Sari	2011	Prediksi Cuaca Berbasis Logika Fuzzy untuk Rekomendasi Penerbangan di Bandar Udara Haji Fisabilillah	Logik Fuzzy model prediksi Takagi-Sugeno	Untuk prediksi angin menggunakan 2 masukan yaitu suhu udara dan tekanan udara dengan keluaran berupa kecepatan angin. Untuk nilai keakuratan prediksi hujan adalah 61.73%, kecepatan angin 50.5%, jarak pandang 87.5%, angin buritan 88.6%, rekomendasi penerbangan berdasarkan jarak pandang 96.2%, rekomendasi penerbangan berdasarkan angin buritan 88.6%.
5	Liliana Swasti	2013	Penerapan Algoritma	C4.5	Algoritma Decision Tree C4.5 akurat

	na		ma C4.5 Untuk Penentu an Jurusan Mahasis wa		diterapkan untuk penentuan kesesuaian jurusan mahasiswa dengan tingkat keakuratan 93,31 % dan akurasi rekomendasi jurusan sebesar 82,64%.
6	Ilham Achmadi Yorinda		Perancangan Sistem Prediksi Cuaca Berbasis Logika Fuzzy Untuk Kebutuhan Penerbangan Di Bandara Juanda - Surabaya	logika fuzzy	nilai keakuratan prediksi curah hujan adalah 74.79%, jarak pandang 85.43%, kelayakan variabel jarak pandang 98.31%, Untuk nilai keakuratan prediksi kecepatan angin adalah 66.58%, angin buritan 95.57%, dan kelayakan variabel angin buritan 95.57%

memecahkan masalah pada latar belakang tuliskan juga literature review/ tinjauan studi untuk mendapatkan state of the art. Tuliskan pula penjelasan tahapan2 metode yang akan digunakan.

## 2.1 LANDASAN TEORI

### 2.1.1 Algoritma

Algoritma adalah urutan langkah-langkah logis penyelesaian masalah yang disusun secara sistematis dan logis. Kata logis merupakan kata kunci dalam algoritma. Langkah-langkah dalam algoritma harus logis dan harus dapat ditentukan bernilai salah atau benar

Menurut para ahli sejarah matematika menemukan asal kata tersebut yang berasal dari nama penulis buku arab yang terkenal yaitu Abu Ja'far Muhammad Ibnu Musa Al-Khuwarizmi. Al-Khuwarizmi dibaca orang barat menjadi Algorism. Al-Khuwarizmi menulis buku yang berjudul Kitab Al Jabar Wal-Muqabala yang artinya "Buku pemugaran dan pengurangan" (The book of restoration and reduction). Dari judul buku itu juga memperoleh akar kata "Aljabar" (Algebra). Perubahan kata dari Algorism menjadi algorithm muncul karena kata algorism sering dikelirukan dengan arithmetic, sehingga akhiran -sm berubah menjadi -thm.

Karena perhitungan dengan angka Arab sudah menjadi hal yang biasa, maka lambat laun kata algorithm berangsur-angsur dipakai sebagai metode perhitungan (komputasi) secara umum, sehingga kehilangan makna kata aslinya. Dalam bahasa Indonesia, kata algorithm diserap menjadi algoritma [6].

### 2.1.2 Data Mining

Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data (Larose, 2005). Beberapa faktor yang mendorong kemajuan *data mining* antara lain :

1. Pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam *data warehouse*, sehingga seluruh perusahaan memiliki akses ke dalam database yang andal.
3. Adanya peningkatan akses kedalam data melalui navigasi web dan internet.
4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Perkembangan teknologi perangkat lunak untuk *data mining* (ketersediaan teknologi).
6. Perkembangan kapasitas media penyimpanan.

Menurut Larose data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat di lakukan, salah satunya yaitu prediction.

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Contoh prediksi dalam bisnis dan penelitian adalah:

- a. Prediksi harga beras dalam tiga bulan yang akan datang.
- b. Prediksi presentase kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikan.

Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

### 2.1.3 Pengertian Pohon Keputusan (*Decision Tree*)

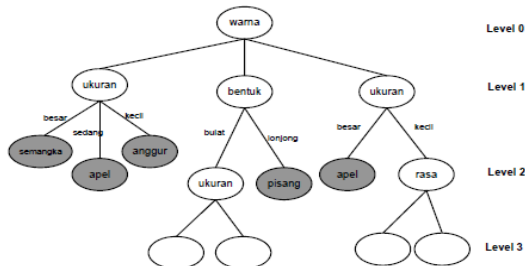
Pohon dalam analisis pemecahan masalah pengambilan keputusan adalah pemetaan mengenai alternatif-alternatif pemecahan masalah yang dapat diambil dari masalah tersebut. *Decision tree* merupakan salah satu teknik klasifikasi yang paling populer dan paling banyak digunakan dalam data mining dan machine learning . *Decision tree* terdiri dari node internal yang menggambarkan data yang diuji, cabang menggambarkan nilai keluaran dari data yang diuji, sedangkan leaf node menggambarkan distribusi kelas dari data yang digunakan. *Decision tree* digunakan untuk mengklasifikasikan suatu sampel data yang tidak dikenal

Dalam *decision tree* tidak menggunakan vector jarak untuk mengklasifikasikan obyek. Seringkali data observasi mempunyai atribut-atribut yang bernilai nominal. Seperti yang diilustrasikan pada gambar 2.6, misalkan obyeknya adalah sekumpulan buah-buahan yang bisa dibedakan berdasarkan atribut bentuk, warna, ukuran dan rasa. Bentuk, warna, ukuran dan rasa adalah besaran nominal, yaitu bersifat kategoris dan tiap nilai tidak bisa dijumlahkan atau dikurangkan. Dalam atribut warna ada beberapa nilai yang mungkin yaitu hijau, kuning, merah. Dalam atribut ukuran ada nilai besar, sedang dan kecil.

Dengan nilai-nilai atribut ini, kemudian dibuat decision tree untuk menentukan suatu obyek termasuk jenis buah apa jika nilai tiap-tiap atribut diberikan (Santoso, 2007).

Pada *decision tree* terdapat 3 jenis *node*, yaitu:

- Root Node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
- Internal Node*, merupakan *node* percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.
- Leaf node atau terminal node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.



Gambar 2.1 : Decision Tree

Ada beberapa macam algoritma *decision tree* diantaranya CART dan C4.5. Beberapa isu utama dalam *decision tree* yang menjadi perhatian yaitu seberapa detail dalam mengembangkan *decision tree*, bagaimana mengatasi atribut yang bernilai *continues*, memilih ukuran yang cocok untuk penentuan atribut, menangani data training yang mempunyai data yang atributnya tidak mempunyai nilai, memperbaiki efisiensi perhitungan (Santoso, 2007). *Decision tree* sesuai digunakan untuk kasus-kasus yang keluarannya bernilai diskrit. Walaupun banyak variasi model *decision tree* dengan tingkat kemampuan dan syarat yang berbeda, pada umumnya beberapa ciri yang cocok untuk diterapkannya *decision tree* adalah sebagai berikut:

- Data dinyatakan dengan pasangan atribut dan nilainya
- Label/keluaran data biasanya bernilai diskrit
- Data mempunyai missing value (nilai dari suatu atribut tidak diketahui)

Dengan cara ini akan mudah mengelompokkan obyek ke dalam beberapa kelompok. Untuk membuat decision tree perlu memperhatikan hal-hal berikut ini :

- Atribut mana yang akan dipilih untuk pemisahan obyek
- Urutan atribut mana yang akan dipilih terlebih dahulu
- Struktur tree
- Kriteria pemberhentian
- Pruning

### 2.1.5 Algoritma C4.5

C4.5 adalah algoritma yang sudah banyak dikenal dan digunakan untuk klasifikasi data yang memiliki atribut-atribut numerik dan kategorial. Hasil dari proses

klasifikasi yang berupa aturan-aturan dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari record yang baru. Algoritma C4.5 sendiri merupakan pengembangan dari algoritma ID3, dimana pengembangan dilakukan dalam hal : bisa mengatasi missing data, bisa mengatasi data kontinyu, pruning.

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

- Pilih atribut sebagai akar.
- Buat cabang untuk tiap-tiap nilai.
- Bagi kasus dalam cabang.
- Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Dalam algoritma C4.5 digunakan information gain untuk memilih atribut yang akan digunakan untuk pemisahan obyek. Atribut yang mempunyai information gain paling tinggi dibanding atribut yang lain relatif terhadap set y dalam suatu data, dipilih untuk melakukan pemecahan.

Pada algoritma ini, pemilihan atribut mana yang akan menempati suatu simpul dilakukan dengan melakukan perhitungan entropi informasi (information entropy) dan mencari nilai yang paling minimum. Pemilihan atribut pada algoritma ini berdasarkan pada asumsi bahwa kompleksitas yang dimiliki oleh pohon keputusan sangat berkaitan erat dengan jumlah informasi yang diberikan oleh nilai-nilai atributnya. Dengan kata lain, teknik heuristik berbasis informasi ini memilih atribut yang memberikan perolehan informasi terbesar (highest information gain) dalam menghasilkan subpohon (subtree) untuk mengklasifikasi sampel.

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \text{Entropy}(S_i) \quad (1)$$

Keterangan :

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

|S<sub>i</sub>| : jumlah kasus pada partisi ke-i

|S| : jumlah kasus dalam S

Sementara itu, penghitungan nilai entropi dapat dilihat pada persamaan berikut :

$$\text{Entropy}(S) = \sum_{i=1}^n -p_i \cdot \log_2 p_i \quad (2)$$

Keterangan :

S : himpunan Kasus

A : fitur

n : jumlah partisi S

p<sub>i</sub> : proporsi dari S<sub>i</sub> terhadap S

Langkah – langkah diatas digunakan untuk menangani atribut nominal. Perhitungan dengan metode *Entropy Based Discretization* di gunakan untuk menangani atribut yang bersifat kontinu. Metode ini menggunakan entropi sebagian dari proses pemisahan selang data

selang kontinu[10]. Untuk menemukan nilai pemisah yang terbaik maka harus dihitung nilai *split point* berikut tataranya :

1. Urutkan data subset dari yang terkecil sampai yang terbesar.
2. Hitung rata-rata per 2 data yang bersebelahan yang digunakan untuk *split point* dengan formula 2.3. Setiap nilai rata-rata merupakan titik nilai yang mungkin menjadi titik perpecahan (*split\_point*) untuk memilih titik terbaik, data akan dipecah menurut titik yang diuji.
3. Hitung nilai informasi dari kedua sample(S) dengan formula 2.5 kemudian  $T(\text{split point})$  yang dimiliki nilai informasi terkecil diambil sebagai batas node.

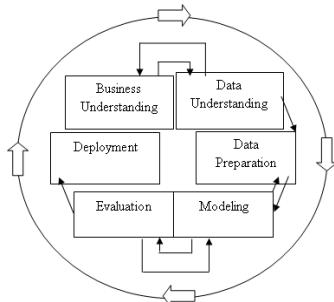
$$\text{Split\_point} = \frac{a_i + a_{i+1}}{2} \quad (3)$$

$$E(S,T) = \frac{|S_1|}{|S|} \text{Ent}(S_1) + \frac{|S_2|}{|S|} \text{Ent}(S_2) \quad (4)$$

$$\text{Nilai informasi} = \text{Gain}(A, S_\alpha) - E(S,T) \quad (5)$$

#### 2.1.6 Definisi Cross-Industry Standard Process for Data Mining (CRISP-DM)

Dalam bukunya [4], Larose menjelaskan fase siklus dari *Data Mining*. Ada 6 (enam) fase yang siklus yang berurutan yaitu:



Gambar 2.2 Proses Data Mining menurut CRISP-DM  
Dari enam fase CRISP-DM diatas menurut Larose adalah sebagai berikut:

1. Fase Pemahaman Bisnis (Business Understanding Phase)
  - a. Menentukan tujuan proyek *data mining* dan kebutuhan detail tentang manfaat untuk bisnis maupun penelitian.
  - b. Menerjemahkan tujuan dan batasan proyek dari permasalahan.
  - c. Persiapan langkah awal untuk mencapai tujuan proyek.
2. Fase Pemahaman Data (Data Understanding Phase)
  - a. Mengumpulkan data dari sumber data.
  - b. Menggunakan analisis data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal.

- c. Evaluasi kualitas dan integritas data.
  - d. Pilih sebagian kecil group data yang mungkin mengandung pola (*pattern*) dari permasalahan.
3. Fase Pengolahan Data (*Data Preparation Phase*)
    - a. Persiapan dataset yang akan digunakan. Pada tahap ini dilakukan pembersihan atribut data yang tidak diperlukan dalam fase permodelan.
    - b. Pilih kasus dan variabel yang ingin dianalisis sesuai dengan tujuan proyek.
    - c. Transformasi variabel jika dibutuhkan.
  4. Fase Permodelan (*Modeling Phase*)
    - a. Pemilihan dan penerapan teknik permodelan yang sesuai dengan kasus yang ingin dianalisis.
    - b. Kalibrasi model untuk mengoptimalkan hasil.
    - c. Perlu diperhatikan bahwa beberapa teknik mungkin untuk digunakan pada permasalahan *data mining* yang sama.
    - d. Jika diperlukan, proses dapat kembali ke fase pengolahan data untuk menjadikan data ke dalam bentuk yang sesuai dengan spesifikasi kebutuhan teknik *data mining* tertentu.
  5. Fase Evaluasi (*Evaluation Phase*)
    - a. Mengevaluasi satu atau lebih model yang digunakan dalam fase permodelan untuk mendapatkan kualitas dan efektivitas sebelum disebarkan untuk digunakan.
    - b. Menetapkan apakah terdapat model yang memenuhi tujuan pada fase awal.
    - c. Menentukan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik.
    - d. Mengambil keputusan berkaitan dengan penggunaan hasil model dari *data mining*.
  6. Fase Penyebaran (*Deployment Phase*)
    - a. menggunakan model yang dihasilkan.
    - b. Conqqqtoh sederhana penyebaran: Pembuatan laporan.

Contoh kompleks penyebaran: Penerapan proses *data mining* secara paralel pada departemen lain.

#### 2.1.7 Confusion matrix

*Confusion Matrix* merupakan sebuah metode untuk evaluasi yang menggunakan tabel matrix seperti pada tabel 1. Pada tabel 1 dapat dilihat bahwa jika dataset terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif (Bramer, 2007). Nilai *accuracy* merupakan persentase jumlah record data yang diklasifikasikan secara benar oleh sebuah algoritma dapat membuat klasifikasi setelah dilakukan pengujian pada hasil klasifikasi tersebut (Han & Kamber, 2006). Nilai *precision* atau dikenal juga dengan nama *confidence* merupakan proporsi jumlah kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. Sedangkan nilai dari *recall* atau *sensitivity*

merupakan proporsi jumlah kasus positif yang sebenarnya yang diprediksi positif secara benar (Powers, 2011).

Tabel 2.2 Model Confusion Matrix Sumber Han & Kamber (2006)

Correct classification	Classified as	
	+	-
+	True positives	False negatives
-	False positives	True negative

*True Positive* adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *false positive* adalah jumlah *record negative* yang diklasifikasikan sebagai positif, *false negative* adalah jumlah *record* positif yang diklasifikasikan sebagai negative, *true negative* adalah jumlah *record negative* yang diklasifikasikan sebagai negatif, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam *confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *sensitivity (recall)*, *Specifity*, *precision*, dan *accuracy*. *Sensitivity* digunakan untuk membandingkan jumlah *t\_pos* terhadap jumlah *record* yang positif sedangkan *Specifity*, *precision* adalah perbandingan jumlah *t\_neg* terhadap jumlah *record* yang negatif. Untuk menghitung digunakan persamaan dibawah ini [11].

$$\text{Sensitivity} = \frac{t\_pos}{t\_pos + t\_neg} \times 100\% \quad (6)$$

$$\text{Specifity} = \frac{t\_neg}{neg} \times 100\% \quad (7)$$

$$\text{Precision} = \frac{t\_pos}{t\_pos + f\_pos} \times 100\% \quad (8)$$

$$\text{Accuracy} = \frac{t\_pos + t\_neg}{t\_pos + f\_neg + f\_pos + t\_neg} \times 100\% \quad (9)$$

Keterangan :

t\_pos : jumlah *true* positif

t\_neg : jumlah *true* negatif

p : jumlah *record* positif

n : jumlah tupel negatif

f\_pos : jumlah *false* positif

f\_neg : jumlah *false* negatif

Hasil evaluasi *confusion matrix* dalam klasifikasi menunjukkan tingkat akurasi hasil klasifikasi seperti yang ditunjukkan dalam tabel 2.

### 2.1.8 Kurva ROC

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horizontal dan *treua positive* sebagai garis vertical (Vercellis, 2009).

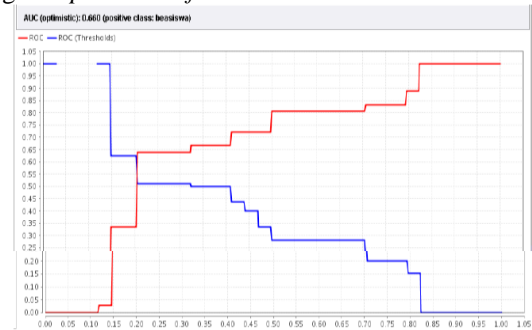
Hasil perhitungan divisualisasikan dengan kurva ROC (*Receiver Operating Characteristic*) atau AUC (*Area Under Curve*). ROC memiliki tingkat nilai diagnose yaitu :

- Akurasi bernilai 0.90 – 1.00 = *excellent classification*
- Akurasi bernilai 0.80 – 0.90 = *good classification*
- Akurasi bernilai 0.70 – 0.80 = *fair classification*

d. Akurasi bernilai 0.60 – 0.70 = *poor classification*

e. Akurasi bernilai 0.50 – 0.60 = *failure*

Hasil yang didapat dari pengolahan ROC untuk algoritma C4.5 dengan menggunakan data training sebesar 0.660 dapat dilihat pada gambar 2.3 dengan tingkat diagnose *poor classification*.



Gambar 2.3 Contoh Evaluasi dengan Kurva ROC/AUC

## III. IMPLEMENTASI

Tujuan utama dari penelitian ini adalah untuk mengetahui nilai akurasi dari algoritma C4.5 dari data delay penerbangan yang telah diperoleh dari BMKG Penerbangan Ahmad Yani Semarang dan PT. AngkasaPura I(persero). Data delay penerbangan yang diolah merupakan data tahun 2013 dan diambil record perhari, sehingga jumlah 365 data. Dari data tersebut memiliki 6 atribut yaitu :

- Arah Angin (°)
- Kecepatan Angin (knot)
- Suhu (°C)
- Tekanan Udara (mb)
- Visibility (m)
- Cuaca

Pada atribut cuaca memiliki 4 cuaca yaitu :

- RA = *Rain* di bandara
- TS = *Thunder Storm* di bandara
- VCTS = *Thunder Storm* disekitar bandara
- SN = *Sunny* di bandara

Tabel 4.1 Data set delay penerbangan di Bandara Ahmad Yani Semarang

A	B	C	D	E	F	G	H
No	Arah Angin	Kecepatan Angin	Suhu	Tekanan Udara	Visibility	Cuaca	Keterangan
1	350	10	27,6	1008,4	5000	RA VCTS	TIDAK
2	330	18	27,1	1009,4	5000	TS RA	TIDAK
3	330	13	26,9	1008,3	3000	TS RA VCTS	TIDAK
4	320	16	27,5	1006,7	3000	TS RA	DELAY
5	330	17	27,7	1006,8	5000	TS RA VCTS	TIDAK
6	270	14	27,7	1007,5	6000	TS RA	TIDAK
7	350	11	26,8	1007,3	2000	TS RA	DELAY
8	310	14	27,3	1006,9	3000	TS RA	TIDAK
9	300	10	27,7	1006,2	6000	RA	TIDAK
10	310	18	28,4	1005,7	2600	RA VCTS	DELAY
11	300	22	28,4	1007,5	4000	RA	TIDAK
12	300	23	28,6	1008,7	3000	TS RA	TIDAK
13	330	15	26,3	1008,7	3000	TS RA	TIDAK
14	300	17	26,9	1009,8	1000	TS RA	DELAY
15	320	14	25,4	1010,8	1000	RA	DELAY
16	300	16	26,1	1010,3	4000	RA	TIDAK
17	300	11	28,1	1009,6	6000	RA	TIDAK
18	320	18	27,7	1010,2	5000	RA VCTS	TIDAK
19	320	17	27,2	1010	3000	TS RA VCTS	DELAY
20	310	15	26,8	1010,2	5000	SN	TIDAK
21	320	20	27,6	1011	6000	TS RA VCTS	TIDAK
22	290	23	27,5	1010,6	5000	TS RA	TIDAK
23	310	19	27,3	1009,9	5000	RA	TIDAK
24	290	15	26,4	1010,6	2000	TS RA	DELAY
25	330	18	26,9	1010,7	5000	TS RA	TIDAK
26	310	18	27,1	1011	4000	TS	TIDAK
27	340	13	27,8	1010,9	6000	TS	TIDAK
28	330	21	28,1	1011	3000	TS RA	TIDAK
29	330	12	27,5	1011,2	6000	TS	TIDAK
30	320	16	28,5	1011,6	5000	TS VCTS	TIDAK
31	340	12	27,7	1011,4	5000	TS	TIDAK
32	330	12	28,2	1010,8	5000	TS VCTS	TIDAK

Langkah pertama cari split point, dikarenakan data cuaca berbentuk *numeric* dengan rumus berikut.

$$\text{Split\_point} = \frac{a_1 + a_2}{2}$$

$$\text{Split\_point} = \frac{270 + 300}{2} = 285$$

Pertama urutkan data Arah Angin, Kecepatan Angin, Suhu, Tekanan Udara dan Visibility dari data terkecil hingga terbesar, dan hilangkan data yang sama, kemudian gunakan rumus diatas sehingga menjadi seperti table 3.2.

Tabel 3.2 Data Hasil Split Point

Arah	270	285	300	
	300			
	310	315		
	320			
	330	340		340
	350			
Kecepatan	10	10,5	12	
	11			
	13	13,5		
	14			
	15	15,5	16,5	
	16			
	17	17,5		
	18			
	22	22,5	22,5	
	23			
Suhu	25,4	25,75	26,15	
	26,1			
	26,3			
	26,8	26,55		
	26,9			
	27,1	27	27,125	
	27,2			
	27,3	27,25		
	27,5			
	27,6	27,55	27,725	
	27,7			
	27,9	27,9		
	28,1			
	28,4	28,5	28,5	
28,6				
	1005,7	1005,95	1006,35	
	1006,2			

Tekanan Udara	1006,7	1006,75	1007,5	
	1006,8			
	1006,9	1007,1		
	1007,3			
	1007,5	1007,9		
	1008,3			
	1008,4	1008,55		1009,02 5
	1008,7			
	1009,4	1009,5		
	1009,6			
	1009,8	1009,9		1010,07 5
	1010			
	1010,2	1010,25		
	1010,3			
1010,8	1010,8	1010,8		
Visibility	1000	1500	2150	
	2000			
	2600	2800		
	3000			
	4000	4500	5250	
	5000			
	6000	6000		

### 1. Fase Permodelan (*Modeling Phase*)

Dari fase sebelumnya data preparation, setelah data diolah sesuai kebutuhan maka data tersebut bisa dimodelkan. Model yang digunakan dalam penelitian ini adalah model C4.5. Pada algoritma C4.5 harus menentukan jumlah *Entropy* total. Hasil dari perhitungan digunakan untuk menentukan  $\text{Gain}$  dari masing-masing atribut.

$$\text{Entropy}(S) = \sum_{i=1}^2 -p_i * \log_2 p_i$$

$$\text{Entropy}(\text{total}) = (-p_1 * \log_2 p_1) + (-p_2 * \log_2 p_2)$$

$$\text{Entropy}(\text{total}) = \left(-\frac{14}{20} * \log_2\left(\frac{14}{20}\right)\right) + \left(-\frac{6}{20} * \log_2\left(\frac{6}{20}\right)\right)$$

$$\text{Entropy}(\text{total}) = 0,8812909$$

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^2 \frac{|S_i|}{|S|} * \text{Entropy}(S_i)$$

$$\text{Gain}(\text{Total}, \text{Arah}) = \text{Entropy}(\text{total}) - \left(\frac{|S_1|}{|S|} * \text{Entropy}(\text{arah1}) + \frac{|S_2|}{|S|} * \text{Entropy}(\text{arah2})\right)$$

$$\text{Gain}(\text{Total}, \text{Arah}) = 0,8812909 - \left(\frac{2}{20} * 1\right) + \left(\frac{18}{20} * 0,85240528\right)$$

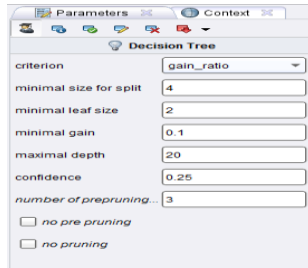
$$\text{Gain}(\text{Total}, \text{Arah}) = 0,014126238$$



#### IV. HASIL & PEMBAHASAN

##### 4.1 Validasi dan Evaluasi

Tabel 4.1 adalah data delay penerbangan dari BMKG Bandara Ahmad Yani Semarang dan PT.Angkasapural(persero). Untuk mendapatkan hasil akurasi dan model yang akan dihasilkan oleh algoritma C4.5 sesuai perhitungan sebelumnya maka digunakan RapidMiner untuk mengolah 365 data tersebut. Pada bagian seting parameter algoritma C4.5 untuk menentukan *information gain*, maka digunakan setingan seperti gambar 4.2.



Gambar 4.2 Setting parameter RapidMiner

##### 4.2 Hasil Percobaan dan Pengujian Metode

Hasil evaluasi confusion matrix dalam klasifikasi menunjukkan tingkat akurasi hasil klasifikasi seperti yang di tunjukkan dalam tabel 4.2. Dari tabel tersebut dapat diketahui tingkat akurasi hasil klasifikasi delay penerbangan sebesar 94,55. Untuk perhitungan manual dapat di gunakan persamaan 9. Dengan hasil bagi dari  $t_{pos}$  (jumlah *true positif*), dan  $t_{neg}$  (jumlah *true negatif*) dengan  $t_{post}$  (total *true positif*),  $f_{neg}$  (total *false negatif*),  $f_{pos}$  (total *false positif*),  $t_{neg}$  (total *true negatif*). Dari tabel 4.2.

$$\text{Akurasi} = \frac{t_{pos} + t_{neg}}{t_{pos} + f_{neg} + f_{pos} + t_{neg}}$$

$$\text{Akurasi} = \frac{96 + 8}{96 + 4 + 2 + 8}$$

$$\text{Akurasi} = \frac{104}{110} = 0.9455$$

Sedangkan dalam bentuk persen (%) menjadi :

$$\text{Akurasi} = 0.9455 \times 100\% = 94.55\%$$

Tabel 4.2 Hasil nilai akurasi dari data delay penerbangan yang di tampilkan oleh RapidMiner

	true TIDAK	true DELAY	class precision
pred. TIDAK	96	4	96.00%
pred. DELAY	2	8	80.00%
class recall	97.96%	66.67%	

Berdasarkan tabel 4.3 hasil evaluasi dengan confusion matrix menunjukkan nilai precision sebesar 0.8. untuk perhitungan manual dapat menggunakan persamaan 8.Precision adalah hasil bagi  $t_{pos}$  (total *true positif*), dengan jumlah  $t_{pos}$  (total *true positif*), dan  $f_{pos}$  (total *false positif*).

$$\text{Precision} = \frac{t_{pos}}{t_{pos} + f_{pos}}$$

$$\text{Precision} = \frac{8}{8 + 2}$$

$$\text{Precision} = 0.8$$

Sedangkan dalam bentuk persen(%) nilai precision menjadi.

$$\text{Precision} = 0,8 \times 100\% = 80\%$$

Tabel 4.3 Nilai precision dari data delay penerbangan yang di tampilkan oleh RapidMiner

	true TIDAK	true DELAY	class precision
pred. TIDAK	96	4	96.00%
pred. DELAY	2	8	80.00%
class recall	97.96%	66.67%	

Berdasarkan tabel 4.4 hasil recall dengan confusion matrix menunjukkan nilai recall sebesar 66.67%. Untuk perhitungan manual dapat menggunakan persamaan 6. Nilai Recall adalah hasil bagi  $t_{pos}$  (total *true positif*) dengan penjumlahan  $t_{pos}$  (total *true positif*) dan  $t_{neg}$  (total *true negatif*)total true positif

$$\text{Recall} = \frac{t_{pos}}{t_{pos} + t_{neg}}$$

$$\text{Recall} = \frac{8}{8 + 4}$$

$$\text{Recall} = 0.6667$$

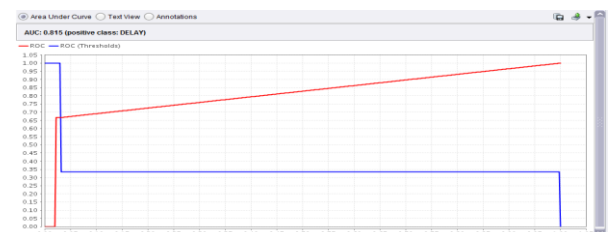
Sedangkan dalam bentuk persen(%) nilai *Recall* menjadi.

$$\text{Recall} = 0.6667 \times 100 = 66.67\%$$

Tabel 4.4 Nilai precision dari data delay penerbangan yang di tampilkan oleh RapidMiner

	true TIDAK	true DELAY	class precision
pred. TIDAK	96	4	96.00%
pred. DELAY	2	8	80.00%
class recall	97.96%	66.67%	

Hasil yang didapat dari pengolahan ROC untuk algoritma C4.5 dengan menggunakan data *training* sebesar 0.815 dapat dilihat pada gambar 4.4 dengan tingkat diagnosa good *classification*.



Gambar 4.4 Hasil ROC dari C4.5 yang ditampilkan oleh RapidMiner

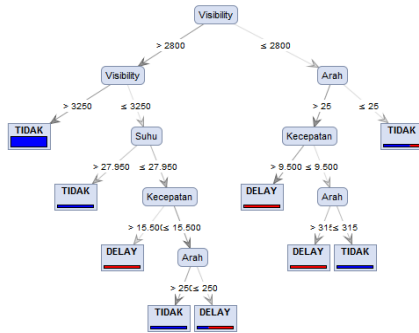
Tabel 4.1 Hasil Akurasi dan AUC dari C4.5

Percobaan	C4.5		Lama Waktu Eksekusi
	Akurasi	Performa AUC	
1	94.55%	0.815	0 S
2	94.55%	0.815	0 S



Hasil diatas menunjukkan algoritma C4.5 yang diterapkan pada *data set* Delay penerbangan tahun 2013, data menghasilkan nilai akurasi confusion matrix sebesar 94.55% dan akurasi AOC 0.815 dalam selang waktu 0 detik.

### 4.3 Hasil Pemodelan Pohon Keputusan dan Rule



Gambar 4.6 Hasil Pemodelan Tree C4.5

Rule yang tercipta dari gambar hasil pemodelan diatas adalah sebagai berikut :

1. Jika Visibility > 2800 DAN Visibility > 3250 maka TIDAK.
2. Jika Visibility > 2800 DAN Visibility ≤ 3250 DAN Suhu > 27,950 maka TIDAK.
3. Jika Visibility > 2800 DAN Visibility ≤ 3250 DAN Suhu ≤ 27,950 DAN Kecepatan Angin > 15,500 Maka DELAY.
4. Jika Visibility > 2800 DAN Visibility ≤ 3250 DAN Suhu ≤ 27,950 DAN Kecepatan Angin ≤ 15,500 DAN Arah Angin > 250 Maka Tidak.
5. Jika Visibility > 2800 DAN Visibility ≤ 3250 DAN Suhu ≤ 27,950 DAN Kecepatan Angin ≤ 15,500 DAN Arah Angin ≤ 250 Maka Delay.
6. Jika Visibility ≤ 2800 DAN Arah Angin > 25 DAN Kecepatan Angin > 9,500 Maka Delay.
7. Jika Visibility ≤ 2800 DAN Arah Angin > 25 DAN Kecepatan Angin ≤ 9,500 DAN Arah Angin > 315 MAKA DELAY.
8. Jika Visibility ≤ 2800 DAN Arah Angin > 25 DAN Kecepatan Angin ≤ 9,500 DAN Arah Angin ≤ 315 MAKA TIDAK.
9. Jika Visibility ≤ 2800 DAN Arah Angin ≤ 25 MAKA TIDAK.

### 4.4 Aplikasi Sederhana

Dari rule yang dihasilkan dari RapidMiner maka dapat di buat aplikasi berbasis website. Cara kerja aplikasi ini yaitu dengan cara menginputkan Arah Angin, Kecepatan Angin, Jarak Pandang / Visibility, Cuaca, Suhu, dan Tekanan Udara. Kemudian klik Prediksi Maka akan muncul hasil prediksi apakah penerbangan berprediksi DELAY atau TIDAK.

Aplikasi ini di tujukan untuk pihak bandara dan pihak maskapai, dikarenakan untuk informasi cuaca yang

mengetahui cara membaca dan pengolahannya hanya pihak dari bandara dan maskapai. Kemudian dari output aplikasi bisa di umumkan di papan pengumuman di bandara dengan bahasa yang mudah di mengerti oleh penumpang, sehingga para penumpang dapat mengerti kenapa terjadi delay penerbangan dari informasi yang dikeluarkan oleh pihak bandara dan maskapai penerbangan jelas kepada para penumpang apabila terjadi Delay Penerbangan. Tampilan Aplikasi Delay Penerbangan dapat di lihat pada gambar 4.7.

Gambar 4.7 Form input Delay Penerbangan

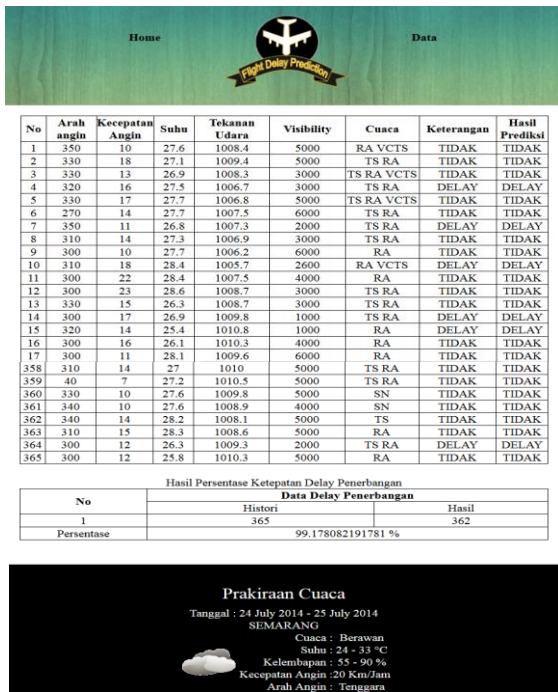
Inputkan Arah Angin(°), Kecepatan Angin(knot), Jarak Pandang / Visibility(m), Cuaca, Suhu(°C), Tekanan Udara(mb). Seperti pada gambar 4.8, kemudian klik Prediksi.

Gambar 4.8 Form input Delay Penerbangan

Setelah data diolah maka tampilah hasil prediksi dari data yang di inputkan tadi seperti pada gambar 4.9. Hasil yang keluar dari hasil inputan tadi adalah "Tidak Delay". Dengan demikian pihak bandara atau maskapai bisa memberitahukan informasi yang jelas kepada calon penumpang apabila terjadi Delay penerbangan.

Gambar 4.9 Hasil prediksi Delay Penerbangan

Hasil prediksi dibandingkan dengan data histori delay penerbangan dari 365 data tersebut ada 362 hasil prediksi yang tepat. Perbandingan data histori dengan prediksi dapat di lihat pada gambar 4.10.



Gambar 4.10 Persentase hasil delay penerbangan

## V. PENUTUP

### 5.1 Kesimpulan

Dari analisis data Delay Penerbangan menggunakan algoritma C4.5 berdasarkan literatur yang digunakan maka dapat disimpulkan bahwa dengan algoritma C4.5 dapat diterapkan untuk menentukan Delay Penerbangan yang diakibatkan oleh cuaca. Dengan menggunakan validasi model menggunakan *split-validation* dan evaluasi model menggunakan *confusion matrix* serta kurva ROC.

Hasil menunjukkan bahwa algoritma C4.5 yang diterapkan pada *data set* Delay penerbangan di tahun 2013, data menghasilkan nilai akurasi confusion matrix sebesar 94.55% dan akurasi AOC 0.815 dalam selang waktu 0 detik.

Dengan adanya penerapan Decision Tree C4.5 diharapkan mampu memberikan solusi bagi pihak bandara maupun maskapai penerbangan dalam membantu menentukan delay penerbangan akibat gangguan cuaca buruk. Sehingga pihak maskapai dapat memberikan informasi kepada penumpang bahwa ada delay penerbangan beserta alasan kenapa harus diadakan delay.

### 5.2 Saran

Proses penelitian ini mendapatkan banyak hambatan seperti terbatasnya data penelitian dan perangkat keras yang digunakan. Untuk penelitian selanjutnya terdapat beberapa saran sebagai berikut :

1. Untuk mendapatkan hasil akurasi yang lebih baik,
2. sebaiknya menambah atribut keadaan pesawat yang digunakan, serta pilot yang menerbangkannya.
3. Data yang di peroleh hanya di tahun 2013, Untuk mendapatkan hasil yang lebih akurat bisa menggunakan data set yang lebih banyak atau rentang waktu yang lebih lama.

Untuk mempercepat hasil eksekusi sebaiknya menggunakan perangkat keras dengan spesifikasi processor core i7 dengan RAM 8gb atau spesifikasi yang lebih tinggi lagi.

## REFERENCES

- [1] Nur Endah Sari, "PREDIKSI CUACA BERBASIS LOGIKA FUZZY UNTUK REKOMENDASI PENERBANGAN DI BANDAR UDARA RAJA HAJI FISABILILLAH," Universitas Gunadarma, 2011
- [2] Soejadi Wh, "MANFAAT DAN BAHAYA CUACA DALAM PENERBANGAN," 31 Agustus 2010. [online]. Available: <http://pustakacuaca.blogspot.com/2010/08/manfaat-dan-bahaya-cuaca-dalam.htm>. [Accessed 30 Januari 2014]
- [3] <http://news.detik.com/read/2012/03/01/004810/1854995/10/penunda-an-penerbangan-garuda-imbas-cuaca-buruk>. Diakses 24 Februari 2014
- [4] Aa Zezen Zaenal Abidin, "IMPLEMENTASI ALGORITMA C 4.5 UNTUK MENENTUKAN TINGKAT BAHAYA TSUNAMI", Jurusan Teknik Informatika STMIK Subang, Jawa Barat, 2 Juli 2011.
- [5] [http://usupress.usu.ac.id/files/Algoritma%20dan%20Pemrograman;%20Teori%20dan%20Praktik%20dalam%20Pascal%20Edisi%20Ke dua\\_Normal\\_bab%201.pdf](http://usupress.usu.ac.id/files/Algoritma%20dan%20Pemrograman;%20Teori%20dan%20Praktik%20dalam%20Pascal%20Edisi%20Ke dua_Normal_bab%201.pdf). Diakses 24 Februari 2014
- [6] [http://www.meteojuanda.info/index.php?option=com\\_content&view=article&id=36&Itemid=34](http://www.meteojuanda.info/index.php?option=com_content&view=article&id=36&Itemid=34). Diakses 26 Februari 2014
- [7] <http://baithanitosari.org/wp-content/uploads/2013/05/RM-GEO-VII.pdf>. Diakses 26 Februari 2014
- [8] Kusriani, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Publishing
- [9] Bramer, Max. 2007. *Principles of Data Mining*. London: Springer. ISBN-10: 1-84628-765-0, ISBN-13: 978-1-84628-765-7.
- [10] Jiawei Han, *Data Mining : Concepts and Techniques.*, 2006.
- [11] Han, J., & Kamber, M. 2006. *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman. ISBN 13: 978-1-55860-901-3.
- [12] Liliana Swastina, " Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa", *Jurnal GEMA AKTUALITA*, Vol.2 No.1, Juni 2013.
- [13] Andrea Adelheid & Khairil Nst (2012). *Buku Pintar Menguasai PHP-MySQL*. ISBN 9797943141.
- [14] MADCOMS (2009). *Menguasai XHTML, CSS, PHP, & MySQL melalui DREAMWEAFER*. ISBN 9789792909821 .