

## PREDIKSI CUACA MENGGUNAKAN METODE CASE BASED REASONING DAN ADAPTIVE NEURO FUZZY INFERENCE SYSTEM

Ria Chaniago<sup>1\*</sup>, The Houw Liong<sup>1</sup>, Ken Ratri Retno Wardani<sup>1</sup>

Departemen Teknik Informatika Institut Teknologi Harapan Bangsa

Jalan Dipatiukur No. 83–84 Bandung

Email: chaniago\_ria@yahoo.co.id, theh1007@gmail.com, ken\_ratri@ithb.ac.id

\*Korespondensi penulis

**Abstrak:** Cuaca merupakan faktor alam yang dapat mempengaruhi pengambilan keputusan yang akan dilakukan dalam berbagai aspek kehidupan. Atas dasar tersebut, penulis ingin membuat suatu aplikasi yang dapat digunakan untuk memprediksi hasil cuaca dengan tingkat akurasi yang baik. Dengan menggunakan teknologi komputer yang mengimplementasikan cabang ilmu Sistem Pakar akan dibangun sebuah sistem prediksi cuaca. Metode yang digunakan dalam penelitian adalah metode *Adaptive Neuro Fuzzy Inference System* (ANFIS) dan *Case Based Reasoning* (CBR). Kedua metode tersebut dikombinasikan untuk menghasilkan prediksi cuaca. Sistem ini juga dibangun dengan metode belajar untuk meningkatkan hasil akurasi. Metode belajar yang digunakan untuk belajar adalah *Backpropagation Error* (BPE) dan *Recursive Least Square Estimator* (RLSE). Dalam implementasinya dilakukan juga proses *clustering* dan *data cleaning* yang dibutuhkan dalam proses prediksi. *K-Means* digunakan untuk proses *clustering*. Metode *Box and Whisker plot* digunakan untuk proses *clustering*. Hasil akhir yang diharapkan dari penelitian ini adalah sebuah aplikasi yang dapat melakukan prediksi dari cuaca dengan nilai akurasi yang tinggi.

**Kata kunci:** Sistem pakar, Prediksi Cuaca, *Adaptive Neuro Fuzzy Inference System*, *Case Based Reasoning*, *Backpropagation Error*, *Recursive Least Square Estimator*, *K-Means*, *Box and Whisker plot*.

**Abstract:** Weather is one of the nature elements that can influence decision making in human's life. Based on that issue, the author wants to make an application that is able to predict weather with good accuracy. The application is a weather forecasting system, using computer technology that implements expert system. The methods used are *Adaptive Neuro Fuzzy Inference System* (ANFIS) and *Case Based Reasoning* (CBR), and a combination of both methods will applied to the system. The system also has learning methods like *Backpropagation Error* (BPE) and *Recursive Least Error* (RLSE), to increase its accuracy. *Clustering* and *data cleaning* also done inside the system, as it needed by forecasting process to achieve a good result. *K-Means* is the clustering algorithm, while *Box and Whisker Plot* is the algorithm for data cleaning. The result from this project is to create a weather forecasting system with high accuracy.

**Keywords:** Expert System, Weather Forecasting, *Adaptive Neuro Fuzzy Inference System*, *Case Based Reasoning*, *Backpropagation Error*, *Recursive Least Square Estimator*, *K-Means*, *Box and Whisker plot*.

### PENDAHULUAN

Cuaca sangat berkaitan erat dalam kehidupan manusia hal itu didasari bahwa cuaca mempengaruhi kegiatan yang akan dilakukan [4]. Cuaca mempengaruhi pengambilan keputusan yang optimal dalam aspek kehidupan dengan tujuan yang bermacam-macam contohnya pengontrolan lalu lintas udara, menentukan rute kapal dan bidang pertanian [1]. Oleh karena itu diperlukan suatu sistem dengan kemampuan yang dapat melakukan prediksi cuaca agar manusia dapat lebih waspada ketika terjadi fenomena cuaca yang dapat merugikan dan membahayakan manusia.

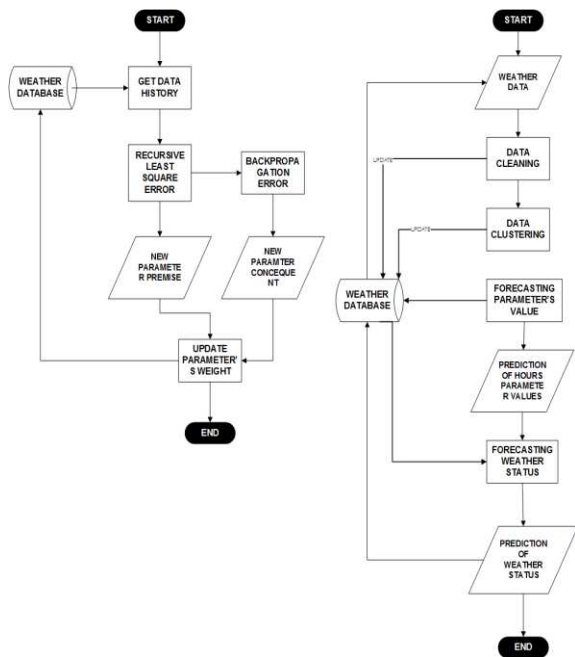
Prediksi cuaca adalah suatu proses untuk memperkirakan keadaan atmosfer bumi dengan menggunakan penerapan ilmu dan teknologi. Dalam

prosesnya ada beberapa metode yang harus digunakan untuk menghasilkan *output* prediksi cuaca. Metode pertama adalah metode yang digunakan untuk memprediksikan nilai-nilai parameter yang berhubungan dengan cuaca untuk waktu kedepan yaitu metode *time series Adaptive Neuro Fuzzy Inference System* (ANFIS). Kelebihan dari metode ANFIS yaitu ANFIS memiliki konsep penggabungan antara penalaran pengetahuan manusia dan proses pembelajaran terarah sehingga cocok di implementasikan untuk kasus prediksi *time series* [2]. Proses pembelajaran dalam penelitian ini menggunakan metode *Hybrid Learning* yaitu dengan menggabungkan metode *Backpropagation Error* dan *Recursive Least Square Error* dimana dengan menggunakan kedua metode ini diharapkan sistem dapat belajar lebih cepat [3].

Proses selanjutnya adalah menentukan pengambilan keputusan akhir dari prediksi cuaca. Pada penelitian ini metode yang digunakan adalah metode *Case Based Reasoning* (CBR). Pemilihan metode ini didasari pola cuaca yang konsisten dimana prediksi untuk kasus cuaca hari ini dapat dihasilkan dari informasi kasus cuaca sebelumnya [9]. Sesuai dengan konsep CBR dimana metode ini mengambil informasi dari kasus yang lama untuk solusi kasus yang ada. Dalam implementasinya algoritma CBR yang digunakan adalah algoritma *Fuzzy k- Nearest Neighbor* (FKNN), dimana metode ini mencari jarak terdekat suatu data uji terhadap kelompok data lampau dengan mengimplementasikan konsep *fuzzy*.

**Rancangan Sistem dan Teori Pendukung**

Berikut adalah rancangan sistem yang digambarkan dengan menggunakan *flowchart*.



Gambar 1. Flowchart Prediksi Cuaca

**Data Cuaca**

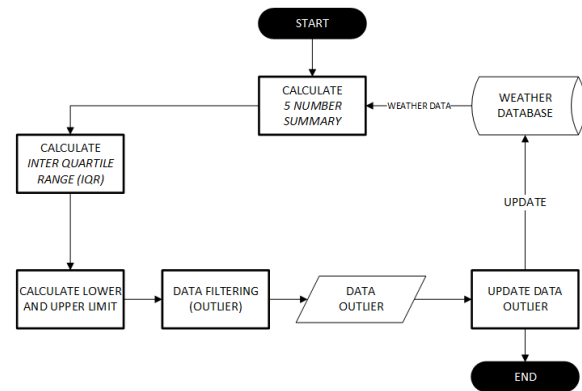
Data yang digunakan merupakan data cuaca Canada setiap jam selama 11 tahun [7]. Data dibagi menjadi 2, yaitu data belajar (1 tahun) dan data untuk CBR (10 tahun). Data cuaca yang didapatkan memiliki beberapa keterangan *field* yaitu waktu (jam, tanggal, bulan dan tahun), suhu (*temperature*) dalam satuan derajat *celcius*, titik embun (*dew point*), kelembaban (*humidity*), arah angin (*wind direction*), kecepatan angin (*wind speed*), jarak pandang (*visibility*) dan keterangan cuaca (*weather*) yang terjadi.

**Data Cleaning**

*Data cleaning* adalah proses mendeteksi dan memperbaiki data. Tujuannya menghapus atau memperbaiki data yang rusak, tidak lengkap atau ekstrim (*outlier*). Dilakukan agar data yang akan diolah menjadi relevan, akurat dan lengkap. Metode yang digunakan adalah *Box and Whisker plot*.

**Box and Whisker Plot**

Metode *box and whisker plot* (*Box plot*) adalah metode yang digunakan untuk memetakan sebaran data. Konsep utama dari metode ini adalah mencari lima dasar nilai yang disebut *five-number summary*. *Five – number summary* terdiri atas nilai minimum, quartile 1 (Q1), median (Q2), quartile 3 (Q3) dan nilai maksimum dari suatu himpunan data. Berikut adalah alur proses data *cleaning* menggunakan metode *Box plot* dalam bentuk *flowchart*.



Gambar 2. Flowchart Data Cleaning

**Data Clustering (K-Means)**

*Data Clustering* adalah proses pengelompokan data berdasarkan kecocokan suatu elemen yang telah ditentukan. Metode yang digunakan adalah metode *clustering K-Means*. *K-Means* merupakan metode *clustering partitioning* yang memisahkan data ke dalam kelompok yang berbeda secara iteratif. Secara garis besar ada empat langkah dalam penggunaan *K-Means clustering*

1. Tentukan nilai k sebagai jumlah dari kelompok *clustering* dan tentukan *center* dari *clustering*, biasanya dipilih secara acak dari data.
2. Cari jarak terdekat untuk masing-masing data terhadap keseluruhan *centroid* yang dimiliki.
3. Ketika seluruh data telah masuk ke dalam satu kelompok, maka *center* dari *cluster* akan dihitung ulang berdasarkan nilai rata-rata dari keseluruhan vektor data.
4. Kembali ke langkah 2 hingga *center* dari *clustering* tidak berubah

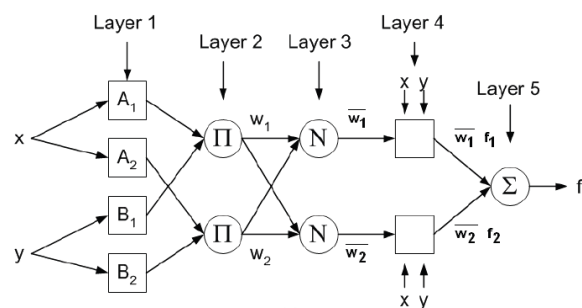
Pencarian jarak dalam kasus ini menggunakan persamaan *Euclidian Distance*.

$$d(p,q) = \sqrt{((P1 - Q1))^2 + ((P2 - Q2))^2 + \dots + ((Pn - Qn))^2}$$

- d(p,q) : Nilai jarak
- P : Jarak titik 1
- Q : Jarak titik 2
- N : Total data

**Adaptive Neuro Fuzzy Inference System (ANFIS)**

*Adaptive Neuro Fuzzy Inference System* (ANFIS) adalah penggabungan mekanisme *Artificial Neural Network* (ANN) dengan menggunakan sistem penalaran *fuzzy*. Strukturnya menggunakan konsep lapisan – lapisan seperti yang dimiliki oleh ANN kemudian setiap tingkatan lapisan akan diisi perhitungan dengan menggunakan konsep *fuzzy*.



**Gambar 3.** Arsitektur ANFIS

Terdapat dua jenis simpul yaitu simpul *adaptive* (dilandangkan kotak) dan *non-adaptive* (dilandangkan lingkaran). Memiliki 5 layer, masing memiliki fungsi yang mewakili metode *fuzzy* Takagi-Sugeno-Kang.

**Lapisan 1:** Merupakan lapisan *adaptive*, dengan fungsi keanggotaan suatu data terhadap himpunan atau kelompok yang terbentuk. Derajat keanggotaan adalah sebuah nilai fungsi keanggotaan sebuah data terhadap setiap himpunan, nilai dari derajat keanggotaan memiliki skala 0 hingga 1. Penelitian ini akan menggunakan persamaan *generalized bell* untuk menghitung derajat keanggotaan.

$$Gbell(x, a, b, c) = 1 / (1 + 3 / ((x-c)/a)^3 + 2b)$$

*Gbell* mempunyai 3 parameter yaitu parameter a, b dan c dimana setiap himpunan memiliki parameter yang berbeda. Untuk inialisasi dari ketiga parameter akan ditentukan berdasarkan data.

**Lapisan 2:** Lapisan ini merupakan lapisan *nonadaptive*, merupakan fungsi aktifasi dari aturan yang ada. Banyaknya simpul pada bagian ini akan sesuai dengan total aturan yang terbentuk dimana total

aturan aka nada sebanyak dengan jumlah kelas yang ada. Pada bagian ini seluruh derajat keanggotaan akan digabungkan dengan menggunakan persamaan T-Norm sesuai dengan himpunan masing-masing.

$$Wi = Ai(X), Bi(Y) \quad i = \text{Himpunan ke-}i$$

**Lapisan 3:** Lapisan ini merupakan lapisan yang menormalisasikan nilai derajat pengaktifan sebelumnya.

$$\bar{wi} = \frac{Wi}{\sum_{i=0}^n Wi}$$

**Lapisan 4:** Lapisan 4 adalah lapisan *nonadaptive*. yang merupakan bagian konsekuen dari sebuah aturan *fuzzy*. Pada layer ini terdapat bobot, jumlah parameter bobot akan ada sebanyak (jumlah *input* + 1), untuk inialisasi awal, nilai parameter yang diambil dari data secara acak.

$$\bar{wif_i} = Wi' (PiX + QiY + ri) \quad i = \text{Himpunan ke-}i$$

**Lapisan 5:** Lapisan ini adalah lapisan yang berfungsi untuk menjumlahkan seluruh *output* yang didapatkan dari lapisan 4. Proses ini juga dinamakan proses defuzzifikasi.

$$i = \sum_{i=0}^c \bar{wiFi}, \quad i = \text{Himpunan ke-}i$$

Dalam penelitian ini, digunakan metode *hybrid learning* untuk proses belajar. Metode *hybrid learning* tersebut adalah penggabungan antara metode *Backpropagation-error* dan *Recursive Least Square Estimator* (RLSE).

**Backpropagation Error**

*Backpropagation – Error* merupakan salah satu metode pembelajaran yang digunakan pada jaringan arsitektur syaraf tiruan. Tujuan metode pembelajaran ini adalah untuk memperkecil total kesalahan yaitu dengan cara mempropagasikan balik total kesalahan yang terdapat dalam lapisan *output* sampai pada lapisan awal, kemudian mengubah bobot dari parameter premis. Total kesalahan dapat dicari dengan melihat selisih dari hasil *output* sistem dan target asli dari data belajar yang digunakan.

$$Ep = \sum_{k=1}^{N(l)} (d_k^p - x_{l,k}^p)$$

Dengan  $d_k^p$  adalah komponen ke-k dari vektor hasil yang diinginkan, sedangkan  $x_{l,k}^p$  merupakan vektor keluaran aktual. Kemudian dengan mendefinisikan sinyal kesalahan pada sebuah lapisan

sebagai nilai *ordered derivative* maka sebuah sinyal kesalahan pada lapisan ke-1, simpul ke-i dapat dicari dengan menurunkan fungsi terhadap hasil.

$$\varepsilon_{l,i} = \frac{\partial Ep}{\partial x_{l,i}}$$

Maka sinyal kesalahan pada persamaan adalah:

$$\varepsilon_{l,i} = -2(d_i^p - x_{l,i}^p)$$

Untuk mendapatkan kesalahan pada lapisan – lapisan berikutnya, sinyal kesalahan pada lapisan ke – 1 akan dipropagasikan kepada lapisan ke – (1 +1). Proses ini akan terus diulang hingga mencapai lapisan masukan atau lapisan paling awal. Untuk mempropagasi balik sinyal kesalahan ke lapisan

berikutnya digunakan aturan rantai. Jika  $\varepsilon_{l,i} = \frac{\partial Ep}{\partial x_{l,i}}$  merupakan sinyal kesalahan pada lapisan ke l, maka lapisan (l+1) adalah  $\varepsilon_{l,i} = \frac{\partial Ep}{\partial x_{l+1,i}}$ . Kemudian

untuk mencari bobot parameter premis akan dapat dilakukan ketika kesalahan pada lapisan yang terdapat penggunaan parameter premis tersebut ditemukan. Berikut merupakan persamaan yang digunakan untuk mencari bobot baru untuk parameter premis.

$$\frac{\partial Ep}{\partial \alpha} = \varepsilon_{l,i} \frac{\partial F_{l,i}}{\partial \alpha}$$

Ketika setiap kesalahan lapisan telah dapat didefinisikan, parameter premis dapat diubah dengan menggunakan persamaan *gradient simple steepest decent*, dimana hasil dari persamaan tersebut akan ditambahkan dengan bobot parameter premis sebelumnya. Berikut adalah persamaan *gradient simple steepest decent*. dengan  $\eta$  sebagai *learning rate*.

$$\Delta \alpha = -\eta \frac{\partial Ep}{\partial \alpha}$$

**Recursive Least Square Estimator (RLSE)**

Metode *Recursive Least Square Estimator* merupakan metode yang digunakan untuk mencari ketepatan bobot parameter konsekuen. Metode ini mengubah nilai linear menjadi bentuk matriks. Parameter konsekuen terdapat dalam lapisan 4 dan jika dijabarkan maka akan membentuk sebuah persamaan linear.

$$\begin{aligned} f &= w_1' f_1 + w_2' f_2 \\ &= w_1' (p_1 x + q_1 y + r_1) + w_2' (p_2 x + q_2 y + r_2) \\ &= [(w_1') x] p_1 + [(w_1') y] q_1 + [(w_1') r_1] + \\ &\quad [(w_2') x] p_2 + [(w_2') y] q_2 + [(w_2') r_2] \end{aligned}$$

Jika diberikan sejumlah data belajar maka fungsi akan dapat dibentuk menjadi sebuah matriks. Berikut adalah contoh penjabaran dari fungsi linear menjadi bentuk matriks.

$$\begin{aligned} &[(w_1') x] p_1 + [(w_1') y] q_1 + [(w_1') r_1] + \\ &[(w_2') x] p_2 + [(w_2') y] q_2 + [(w_2') r_2] = y_1 \\ &\vdots \\ &\vdots \\ &[(w_1') x] p_1 + [(w_1') y] q_1 + [(w_1') r_1] + \\ &[(w_2') x] p_2 + [(w_2') y] q_2 + [(w_2') r_2] = y_n \end{aligned}$$

Menjadi  
(x)<sub>2</sub>)

Dari penjabaran diatas dapat dilihat pola matriks yang terbentuk dari sejumlah data, maka penjabaran tersebut akan dapat dinyatakan sebagai persamaan berikut.  $A \theta = y$

Dengan A merupakan matriks yang berisi penjabaran fungsi dalam lapisan 4 dan  $\theta$  merupakan matriks yang berisikan parameter konsekuen. Dimensi matriks A akan sebesar n-data dikali (total kelas \* (total input+1)) sedangkan untuk dimensi matriks  $\theta$  dan y adalah 1 dikali (total kelas \* (total input+1)). Setelah mendapatkan persamaan matriks, maka persamaan untuk mencari nilai dari  $\theta$  dengan menggunakan metode *Least Square Error*.  
 $\theta = (A^T A)^{-1} A^T y$

**Case Based Reasoning (CBR)**

*Case Based Reasoning* (CBR) adalah sebuah metode yang memiliki konsep mengambil solusi untuk sebuah kasus berdasarkan pengalaman dari kasus dan solusi yang pernah terjadi sebelumnya. Secara garis besar CBR memiliki empat proses utama yaitu *retrieve, reuse, revise* dan *retain*.

1. *Retrieve*  
*Retrieve* adalah proses mencari kasus yang serupa dengan kasus yang tersimpan dalam basis data, serta memunculkan solusi yang relevan dengan kasus tersebut.
2. *Reuse*  
Proses pemilihan informasi apa saja dari kasus sebelumnya yang dapat digunakan. sekarang.
3. *Revise*  
*Revise* adalah proses evaluasi dilakukan, mengkalkulasi dan mempertimbangkan solusi kasus yang telah didapatkan dari proses sebelumnya, jika solusi dapat berhasil dijalankan maka proses akan berlanjut ke dalam proses *retain*.
4. *Retain*  
*Retain* adalah proses penyaringan sebuah informasi dari kasus dan solusi yang baru yang akan disimpan ke dalam basis data.

Dalam menggunakan metode CBR ini hal yang harus diperhatikan adalah perhitungan untuk matrik serupa (*similarity matrices*) yang akan digunakan untuk proses pengambilan (*retrieve*). Dalam penelitian ini perhitungan matriks serupa yang digunakan adalah metode *k-Nearest Neighbor*.

**Fuzzy K-Nearest Neighbor**

Metode *fuzzy k Near Neighbor* (FKNN) adalah sebuah metode yang digunakan untuk pengklasifikasian ketepatan suatu data uji dengan mencari perbedaan jarak terdekat antara satu objek dengan objek lainnya. Metode ini adalah penggabungan algoritma *k-Nearest Neighbor* klasik dengan konsep *fuzzy*.

Jarak data uji terhadap tiap kelompok yang terbentuk akan dihitung, kemudian data uji akan memiliki derajat keanggotaan pada setiap kelompok, semakin kecil jarak data uji terhadap suatu kelompok maka derajat keanggotaan data uji terhadap kelompok akan semakin besar. Berikut adalah langkah – langkah proses *fuzzy k-Nearest Neighbor*:

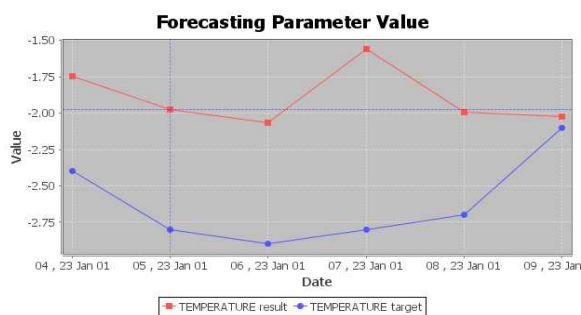
1. Tentukan data terdekat sebanyak K dari tiap kelompok terbentuk. Untuk mencari data terdekat dapat menggunakan rumus *Manhattan Distance* atau *Euclidian Distance*.
2. Akumulasikan k data terdekat yang terpilih dari setiap kelompok.
3. Hitung keseluruhan nilai akumulasi jarak tiap kelompok.
4. Tentukan nilai derajat keanggotaan. Pilih kelompok dengan nilai derajat keanggotaan paling tinggi.

**PENGUJIAN DAN HASIL PENGUJIAN**

Telah dilakukan pengujian terhadap hasil prediksi menggunakan ANFIS dan CBR. Serta dilakukan pengujian dari pengaruh proses *data cleaning*. Berikut adalah hasil dari beberapa hasil pengujian.

**Pengujian terhadap ANFIS**

Berikut adalah tampilan dari aplikasi prediksi yang telah dibangun.



**Gambar 3.** Grafik pengujian ANFIS dalam aplikasi

Berikut adalah detail dari pengujian yang dilakukan beberapa kali dengan jumlah data belajar yang bervariasi.

**Tabel 1.** Tabel Hasil Pengujian ANFIS

NO	RMSE		AKURASI		Data Belajar
	Tanpa Belajar	Belajar	Tanpa Belajar	Belajar	
1	17.6531	1.1760	82.3469	98.8239	500
2	16.2762	1.1760	83.7238	98.8239	
	Rata - rata		83.03535	98.8239	
1	10.8094	1.1005	89.1906	98.8994	1000
2	8.0912	1.1005	91.9088	98.89947	
	Rata - rata		90.5497	98.89947	
1	10.2467	1.2565	89.7533	98.7435	8000
2	8.974	1.2565	91.026	98.7435	
	Rata - rata		90.38965	98.7435	

**Pengujian terhadap CBR**

Dilakukan pengujian terhadap hasil prediksi status cuaca menggunakan CBR. Berikut adalah pengujian yang dilakukan dengan pengelompokan data (*clustering*) secara langsung menggunakan 10 tahun data. Banyaknya total *clustering* yang paling sesuai untuk sistem ini adalah 3.

Ketika dilakukan *clustering* dengan total *cluster* sebanyak 5, 8, 10 dan 17 hasil bobot yang didapatkan dari RLSE yang digunakan dalam sistem ANFIS tidak lebih akurat dari proses dengan total *clustering* sebanyak 3. Dilakukan pengujian terhadap pengelompokan terhadap keseluruhan data sekaligus dan pengelompokan setiap musim.

**Tabel 2.** Tabel Hasil Pengujian CBR

	Non – Musim	Musim	Target Cuaca
Date	2001-12-31 09:00:00.0		
probability 1	Mostly Cloudy	Snow	Snow Showers
probability 2	Cloudy	Cloudy	
Date	2001-12-31 08:00:00.0		
Probability 1	Mostly Cloudy	Snow	Snow Showers
probability 2	Cloudy	Cloudy	
Date	2001-12-31 07:00:00.0		
probability 1	Mostly Cloudy	Snow	Snow Showers
probability 2	Cloudy	Cloudy	
Date	2001-12-31 04:00:00.0		
probability 1	Mostly Cloudy	Mostly Cloudy	Mostly Cloudy
probability 2	Cloudy	Mainly Clear	

Dari pengujian non-musim, dapat terlihat rata – rata hasil prediksi ketika cuaca *clear* ataupun *cloudy* cukup tepat, namun ketika cuaca lain yang muncul hasil akan menjadi kurang tepat. Ini disebabkan cuaca dominan adalah *cloudy*, *clear* dan sebagainya yang mendekati *cloudy* dan *clear*, sehingga pada saat proses *clustering* status cuaca ini juga yang sangat banyak terdapat dalam sebuah *cluster*.

Untuk pengujian permusim, hasilnya telah dapat terbentuk kelompok yang memiliki probabilitas cuaca yang sesuai prioritas cuaca dalam musim itu walaupun status tersebut bukan merupakan cuaca dominan secara keseluruhan pada 10 tahun data. Status cuaca dominan dalam sebuah *cluster* akan memiliki probabilitas tertinggi sesuai dengan masing – masing musim. Namun ketika masih terdapat hasil yang minor dari cuaca, hasil dari prediksi masih tidak terlalu akurat.

## KESIMPULAN DAN SARAN

Berikut adalah beberapa kesimpulan yang didapatkan dari hasil penelitian ini:

1. Banyaknya data ketika proses belajar mempengaruhi keakuratan hasil prediksi.
2. Rata-rata akurasi dari hasil prediksi parameter cuaca menggunakan pembelajaran *Hybrid* dengan metode *Backpropagation Error* dan *Recursive Least Square Estimator* terhadap skala data 12 bulan (8760 baris) dengan parameter *wind speed* 95.720%, *temperature* 98.82%, *dew point* 97.514%, *humidity* 95.720%, *visibility* 95.319% dan *wind direction* 98.262%.
3. Proses data cleaning sebelum melakukan proses pembelajaran berpengaruh terhadap hasil akhir RMSE ketika proses belajar. Data yang sudah bersih dari nilai ekstrim memiliki total RMSE yang lebih kecil.
4. Semakin banyak total data ekstrim (*outlier*) yang terdapat dalam data belajar maka total RMSE semakin besar.
5. Dalam menangani nilai ekstrim/*outlier* nilai ditarik kedalam nilai batas yang paling mendekati nilai *outlier* tersebut, sehingga tidak terjadi perubahan nilai yang tidak terlalu besar.
6. Dengan menggunakan sekaligus data selama 10 tahun dalam proses *clustering* rata-rata hasil prediksi status cuaca menggunakan CBR kurang akurat ketika target cuaca merupakan cuaca minor dari keseluruhan data belajar. Dengan membagi data per musim saat proses *clustering*, terbentuk kelompok – kelompok yang tidak harus

cuaca dominan dari keseluruhan data, sehingga hasil akan lebih akurat.

7. Hasil prediksi status cuaca tidak akurat untuk hasil minor dalam data histori.

Berikut adalah saran yang perlu dipertimbangkan untuk pengembangan penelitian kedepan:

1. Dilakukan prediksi tersendiri terhadap nilai data kosong atau *outlier*.
2. Menggunakan metode yang lebih kompleks dan akurat untuk penerapan CBR agar hasil yang didapatkan lebih bagus seperti ANN.
3. Menggunakan data cuaca dari Indonesia.

## DAFTAR PUSTAKA

- [1] Fraley, Chris, dkk. *The R Journal*, 3(1), Juni 2011.
- [2] Jang, S. R. ANFIS Adaptive Neuro Fuzzy Inference System. *IEEE on Transaction and System, Man, and Cybernetics*, 23(3), May/June 1993.
- [3] Jang, S.R dkk. *Neuro-Fuzzy and Soft Computing*. United State of America: Prentice-Hall, Inc. 1997.
- [4] Kiskac, Bulent dan Yardimci, Harun. *Weather Prediction Expert System Approaches (Ceng-568 Literature Survey)*. Ankara, Turkey: Middle East Technical University, 2004.
- [5] Kusumadewi, Sri. *Artificial Intelligence (Teknik dan aplikasinya)*. Yogyakarta: Graha Ilmu, 2003.
- [6] National Climate Data and Informat Archive Canada. 2012. (online). ([www.climate.weather-office.gc.ca](http://www.climate.weather-office.gc.ca) diakses tanggal 4 November 2012)
- [7] NOAA's National Weather Service: Aviation Weather Center. 2012. (online). (<http://aviation-weather.gov/products/gairmet/info> diakses tanggal 23 Mei 2013).
- [8] R. E. Huschke. *Glossary of Meteorology*, American Meteorological Society, Boston, Massachusetts, USA, 1959. Pp. 106, 419.
- [9] Riordan, Denis dan Hansen Bjarne K. September, *A Fuzzy case-based system for weather prediction*, 10(3), 2002.