

# DECISION SUPPORT SYSTEMS PRAKIRAAN CUACA HARIAN BERBASIS SEMI-SUPERVISED LEARNING MENGGUNAKAN RECURSIVE K-MEANS DI BANDAR UDARA JUANDA SURABAYA

Fanny Ratna Sari<sup>1</sup>, Lilik Anifah<sup>1</sup>

<sup>1</sup>Program Studi Teknik Informatika, Universitas Negeri Surabaya

Fannysari.ratna@gmail.com

lilikanifah@unesa.ac.id

**Abstract**— Weather is an air condition in the atmosphere layer somewhere. There are many human activities that depend on weather conditions. Based on this, weather forecasting is able to solve problems that occur around. One of the human activities that is very influential on climate change is flight activities at Juanda Airport in Surabaya. Where all activities that occur in aircraft flight are determined based on weather conditions observing surface air observations which are then sent to the analysis and weather forecasting section. The process of determining weather forecasts with manual observation methods often occur errors. Based on this, research is conducted using learning methods based on weather history data in the hope that it can provide weather forecasts that are more efficient and have a higher level of accuracy. The method used is semi supervised learning using recursive k-means. Semi supervised learning uses recursive k-means is a method of development of the k-means clustering method. This study uses five weather variables: Minimum temperature, maximum temperature, average temperature, humidity and duration of solar radiation. the best learning process is done with a threshold value of ten which results in an accuracy rate of 100%. And to make it easier for users to use this system, this research also creates an interface system that can be used by users.

**Keywords**— Decission Support System, Weather Forecasting, Recursive K-means, Semi-Supervised, Daily Weather

**Abstrak**— Cuaca merupakan keadaan udara pada lapisan atmosfer di suatu tempat. Terdapat banyak kegiatan manusia yang bergantung pada keadaan cuaca sekitarnya. Berdasarkan hal tersebutlah prakiraan cuaca mampu menyelesaikan permasalahan yang terjadi disekitar. Salah satu kegiatan manusia yang sangat berpengaruh terhadap perubahan cuaca adalah kegiatan penerbangan di Bandara Udara Juanda Surabaya. Dimana segala kegiatan yang terjadi dalam penerbangan pesawat ditentukan berdasarkan keadaan cuaca pengamatan observasi udara permukaan yang kemudian dikirimkan ke bagian analisa dan prakiraan cuaca. Proses penentuan prakiraan cuaca dengan metode pengamatan manual sering terjadi error. Berdasarkan hal tersebutlah dilakukan penelitian menggunakan metode pembelajaran berdasarkan data history cuaca dengan harapan dapat memberikan prakiraan cuaca yang lebih efisien dan memiliki tingkat akurasi yang lebih tinggi. Adapun metode yang digunakan adalah semi supervised learning menggunakan recursive k-means. Semi supervised learning menggunakan recursive k-means merupakan metode pengembangan dari metode k-means clustering. Penelitian ini menggunakan lima variabel cuaca, antara lain: Suhu minimum, suhu maksimum, suhu rata-rata, kelembapan dan lama penyinaran matahari. Berdasarkan proses learning yang dilakukan dengan nilai threshold terbaik adalah sepuluh menghasilkan tingkat akurasi sebesar 100%. Dan untuk memudahkan pengguna dalam menggunakan sistem ini, maka penelitian ini juga membuat sistem antarmuka yang dapat digunakan oleh pengguna.

**Kata Kunci**— Decission Support System, Prakiraan Cuaca, Recursive K-means, Semi-Supervised, Cuaca Harian

## I. PENDAHULUAN

Cuaca merupakan keadaan udara pada lapisan atmosfer di suatu tempat. Terdapat banyak kegiatan manusia yang bergantung pada keadaan cuaca sekitarnya. Berdasarkan hal tersebutlah prakiraan cuaca mampu menyelesaikan permasalahan yang terjadi disekitar. Salah satu kegiatan manusia yang sangat berpengaruh terhadap perubahan cuaca adalah kegiatan penerbangan di Bandara Udara Juanda Surabaya. Dimana segala kegiatan yang terjadi dalam penerbangan pesawat ditentukan berdasarkan keadaan cuaca pengamatan observasi udara permukaan yang kemudian dikirimkan ke bagian analisa dan prakiraan cuaca.

“Prakiraan cuaca adalah proses identifikasi dan prediksi untuk akurasi kondisi iklim tertentu menggunakan beberapa teknologi”[1]. Maka dengan Prakiraan cuaca dapat menyelesaikan atau setidaknya mengurangi permasalahan yang disebabkan oleh dampak cuaca yang tidak terkendali. Adapun faktor-faktor yang mempengaruhi kondisi cuaca seperti suhu, kelembapan, arah dan kecepatan angin, jarak pandang, tekanan atmosfer dan lain sebagainya. Untuk mencapai prakiraan cuaca yang akurat selalu menjadi tantangan besar bagi para ahli meteorologi. Adapun metode yang mereka gunakan saat ini bergantung berdasarkan pengamatan serta analisis ahli meteorologi yang menggunakan statistik dan prinsip mekanika dan fisika [2].

Proses penentuan prakiraan cuaca dengan metode pengamatan manual sering terjadi error. Berdasarkan hal tersebutlah dilakukan penelitian menggunakan metode pembelajaran berdasarkan data history cuaca dengan harapan dapat memberikan prakiraan cuaca yang lebih efisien dan memiliki tingkat akurasi yang lebih tinggi. Adapun metode yang digunakan adalah semi supervised learning menggunakan recursive k-means. Semi supervised learning menggunakan recursive k-means merupakan metode pengembangan dari metode k-means clustering. Metode semi supervised learning menggunakan recursive k-means adalah metode yang menggabungkan pendekatan supervised dan unsupervised dalam mengelompokkan data set yang memiliki label kelas maupun tidak memiliki label kelas kedalam beberapa klaster [3].

Karena variabel curah hujan yang terdapat pada data cuaca harian yang tersedia di Stasiun Meteorologi BMKG Juanda Surabaya ada beberapa hari yang tidak memiliki nilai di variabel tersebut. Maka dengan menggunakan semi supervised k-means clustering tidak membuang banyak data yang mengalami missing value. Hasil penelitian yang dilakukan oleh (Hanmin, Hao, & Qianting, 2016) menyatakan bahwa dalam menghasilkan pusat klaster yang lebih baik dilakukan oleh metode semi supervised k-means clustering dibandingkan dengan metode k-means tradisional [4]. Pada Stasiun Meteorologi Bandar Udara Juanda Surabaya tingkat akurasi dari prakiraan cuaca hasil pengamatan forecaster hanya mencapai 70%, maka salah satu upaya untuk meningkat akurasi kebenaran dalam memprakirakan cuaca adalah dengan melakukan “Decision Support Systems Prakiraan Cuaca Harian Berbasis Semi Supervised Learning menggunakan Recursive K-Means di Bandar Udara Juanda Surabaya” yang akan dibuat berdasarkan data history keadaan cuaca di sekitar Bandar Udara Juanda.

## II. METODOLOGI PENELITIAN

Penelitian mengenai prakiraan cuaca harian ini menggunakan metode semi-supervised recursive k-means dalam menentukan status prakiraan cuaca di hari berikutnya yang berdasarkan masukan data hari ini. Penelitian ini berdasarkan data history atau data masa lalu keadaan cuaca harian yang terjadi di daerah sekitar Bandar Udara Juanda Surabaya. Adapun unsur cuaca atau variabel yang digunakan dalam penelitian ini adalah suhu minimum, suhu maksimum, suhu rata-rata, kelembapan, penyinaran matahari dan intensitas hujan.

Dengan menggunakan metode semi-supervised recursive k-means, maka kekurangan metode k-means tradisional dalam menentukan cluster awal dapat diatasi dan target data yang ada pada unsur cuaca intensitas hujan tidak terbuang sia-sia. Adapun tahapan penelitian ini sebagai berikut:

### A. Preprocessing Data

Tahap preprocessing data ini dilakukan untuk menghilangkan atribut-atribut atau variabel yang tidak digunakan dalam penelitian ini. Selain itu juga merubah

nilai intensitas hujan ke dalam bentuk linguistik. Nilai dari variabel intensitas curah hujan berada di rentang 0 – diatas seratus, namun ada yang bernilai tak terukur (ttu). Berdasarkan hasil wawancara dengan ahli meteorology, nilai “ttu” termasuk kedalam kategori “Hujan Ringan” karena kondisi hujan namun tidak terukur atau tidak tertampung di alat pengukur intensitas curah hujan.

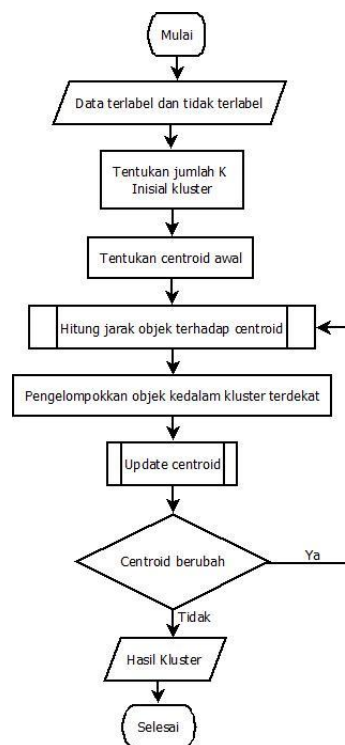
TABEL I  
KATEGORI STATUS CUACA BERDASARKAN NILAI INTENSITAS CURAH HUJAN

No.	Kategori	Intensitas Curah Hujan (mm/hari)
1.	Tidak Hujan	-
2.	Hujan Ringan	0.1 – 19,9 dan ttu
3.	Hujan Sedang	20,0 – 49,9
4.	Hujan Lebat	50,0 - 100
5.	Hujan Sangat Lebat	> 100,0

Sehingga tahap preprocessing ini digunakan untuk penghapusan atribut yang tidak digunakan, merubah nilai intensitas hujan ke bentuk linguistik dan normalisasi data untuk mengubah nilai masing-masing unsur cuaca menjadi nilai dengan rentang 0 – 1. Adapun normalisasi didapatkan dengan persamaan sebagai berikut:

$$\text{Nilai Normalisasi} = \frac{\text{Nilai awal} - \text{Nilai minimal}}{\text{Nilai maksimal} - \text{Nilai minimal}}$$

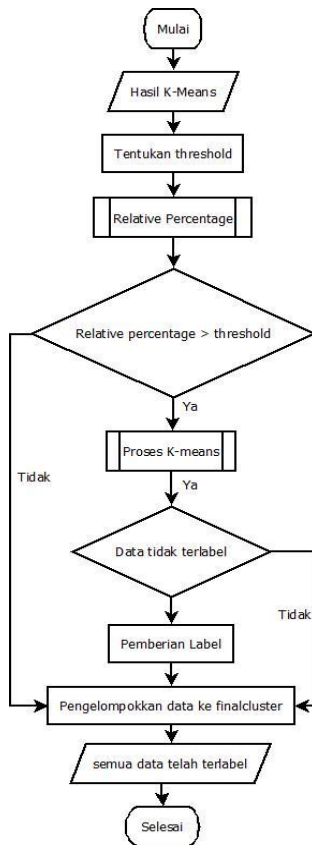
### B. Proses K-Means



Gbr. 1 Flowchart Algoritma K-Means

### C. Proses Recursive K-Means

Proses recursive k-means dimulai saat proses k-means telah mengeluarkan hasil berupa data yang sudah terpartisi. Recursive k-means ini berfungsi untuk mengurangi tingkat kesalahan pengelompokan yang terjadi di proses k-means. Adapun alur dari algoritma recursive k-means sebagai berikut:



Gbr.2 Flowchart Algoritma Recursive K-Means

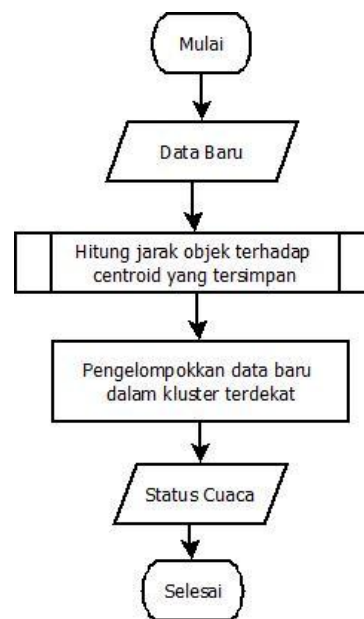
- 1) Hasil pengelompokan data dari proses k-means merupakan pengolahan data awal dari algoritma recursive k-means.
- 2) Menentukan threshold, threshold merupakan batas toleransi kesalahan suatu sistem. Penelitian yang berjudul Semi-Supervised Text Categorization using Recursive K-means clustering menyatakan bahwa threshold yang terbaik adalah threshold yang bernilai kecil namun tidak menyebabkan infinite recursive atau error yang terjadi karena looping yang berlebihan. Sehingga dalam menentukan threshold terbaik perlu dilakukan berbagai percobaan nilai threshold [5].
- 3) Menghitung nilai relative percentage untuk masing-masing partisi. Adapun formula dari relative percentage sebagai berikut:

$$\text{Relative Percentage} = \frac{\text{Jumlah data ke } [i]}{\text{Jumlah data label terbanyak}} * 100 \%$$

- 4) Jika nilai dari relative percentage lebih besar dibandingkan nilai *threshold* maka data yang terdapat di partisi tersebut dilakukan pengulangan proses, yaitu mulai dari proses awal (k-means) hingga memperoleh *relative percentage* yang lebih kecil atau sama dengan nilai *threshold*.
- 5) Pada proses *recursive k-means* ini, memungkinkan untuk terciptanya partisi baru diluar partisi yang ditentukan di awal. Namun label dari partisi baru merupakan kluster yang telah ditentukan di awal.
- 6) Ketika semua data sudah berada di partisinya masing-masing dengan nilai *relative percentage* tidak lebih besar dari *threshold* maka partisi-partisi tersebut merupakan kluster akhir (final clusters) dari proses *recursive k-means*.

### D. Proses Learning dan Testing

Proses *learning* merupakan tahapan pembelajaran yang dilakukan oleh algoritma *recursive k-means* terhadap data cuaca yang dimasukkan. Perhitungan akurasi dilakukan setelah proses *learning* yang disebut dengan proses *testing*.



Gbr.3 Proses testing data baru

Setelah proses testing dilakukan, maka perhitungan tingkat akurasi terhadap data testing dilakukan. Adapun formula dari perhitungan akurasi sebagai berikut:

$$\text{Akurasi} = \frac{\text{Total data benar}}{\text{Jumlah data}} * 100 \%$$

## III. HASIL DAN PEMBAHASAN

### A. Pengolahan Data

Tahap pertama dalam melakukan penelitian ini adalah kesiapan data cuaca dari BMKG Juanda Surabaya. Penelitian

ini juga membandingkan jumlah data, yaitu menggunakan seluruh data yang ada dan menggunakan sebagian data. Dalam analisa pertama, data cuaca yang digunakan merupakan data cuaca harian yang didapatkan dari BMKG Juanda Surabaya mulai tahun 2015 hingga 2017. Data yang digunakan sebagai data learning adalah data cuaca harian pada tahun 2015-2016 yang berjumlah 732 data. Sedangkan yang digunakan sebagai data training adalah data cuaca tahun 2017 yang berjumlah 364 data. Adapun rincian jumlah data sebagai berikut:

TABEL III  
RINCIAN JUMLAH DATA ANALISA SATU

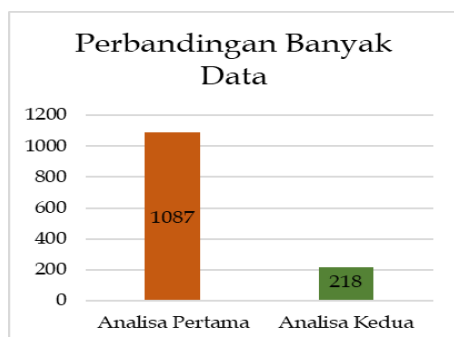
Kategori	Jumlah Data
Tidak Hujan	379
Hujan Ringan	254
Hujan Sedang	63
Hujan Lebat	26
Hujan Sangat Lebat	1
Total Data Learning	723
Total Data Testing	364

Sedangkan pada analisa kedua, data cuaca yang digunakan merupakan gabungan dari data learning dan data testing pada analisa pertama dengan pengambilan data secara acak. Dan pada analisa kedua ini, data learning dan data testing berasal dari data yang sama. Adapun jumlah data pada analisa kedua sebanyak 2018 dengan rincian sebagai berikut:

TABEL IIIII  
RINCIAN JUMLAH DATA ANALISA DUA

Kategori	Jumlah Data
Tidak Hujan	60
Hujan Ringan	60
Hujan Sedang	60
Hujan Lebat	37
Hujan Sangat Lebat	1
Total Data Learning	218
Total Data Testing	218

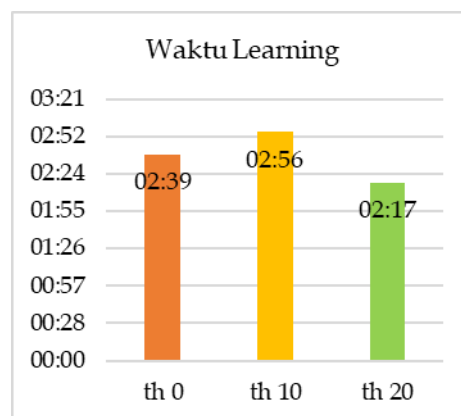
Perbandingan jumlah data dari Analisa pertama dan Analisa kedua sebagai berikut:



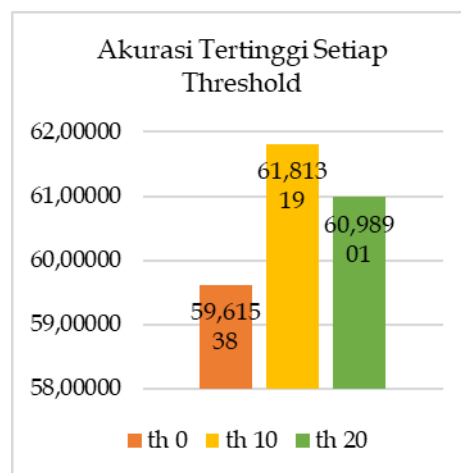
Gbr.4 Perbandingan Jumlah Data Analisa 1 dan Analisa 2

## B. Tahap Learning

Proses learning atau proses pembelajaran data merupakan proses pengolahan data yang berguna memberikan pengalaman terhadap sistem berdasarkan banyak data yang digunakan. Pada proses learning dibutuhkan nilai threshold. Untuk menemukan nilai threshold terbaik maka dilakukan percobaan terhadap tiga nilai threshold yaitu: 0, 10 dan 20 dengan hasil sebagai berikut:



Gbr.5 Perbandingan Lama Waktu Learning tiap Threshold



Gbr.6 Perbandingan Tingkat Akurasi tiap Threshold

Berdasarkan hasil tersebut, maka penelitian ini melanjutkan proses *learning* menggunakan nilai *threshold* 10.

## C. Tahap Testing

Tahap testing merupakan tahap validasi sistem dimana pada tahap ini dapat diketahui seberapa sistem mampu melakukan prediksi terhadap suatu data berdasarkan tahap learning yang sudah dilalui. Adapun perbandingan Analisa pertama dan Analisa kedua disajikan pada diagram berikut:

#### IV. KESIMPULAN

Kesimpulan yang diperoleh dari seluruh proses dan hasil pembahasan penelitian yang dilakukan dapat menghasilkan sebuah sistem perangkat lunak prakiraan cuaca berdasarkan algoritma semi-supervised menggunakan recursive k-means yang memiliki tampilan antarmuka. Dengan memperhitungkan jangka waktu proses learning maka nilai threshold yang digunakan adalah sepuluh dengan memakan waktu empat puluh lima hari menggunakan data Analisa kedua.

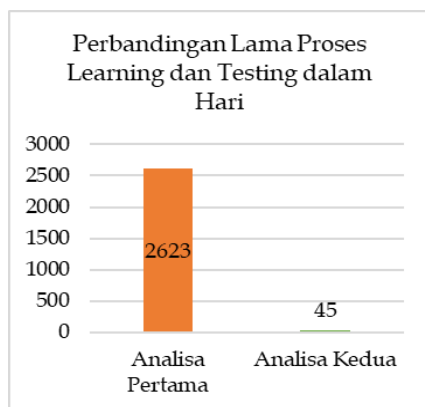
Dalam percobaan yang dilakukan, pada Analisa kedua didapatkan tingkat akurasi tertinggi 61,8131868% sedangkan Analisa kedua didapatkan tingkat akurasi tertinggi 100%.

#### REFERENSI

- [1] Narvekar, M., & Fargose, P. (2015, July). Daily Weather Forecasting using Artificial. *International Journal of Computer Applications*, 121. Retrieved November 25, 2018
- [2] Sobrevilla, K. L., Quinones, A. G., Lopez, K. V., & Azana, V. T. (2016). Daily Weather Forecast in Tiwi, Albay, Philippines. doi:10.1109/TENCON.2016.7848592
- [3] Paramartha, G., Ratnawati, D., & Widodo, A. (2017, Juni 9). Analisis Perbandingan Metedo K-Means Dengan Improved Semi-Supervised K-Means Pada Data Indeks Pembangunan Manusia (IPM). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1, 813-824. Retrieved Desember 10, 2018
- [4] Hanmin, Y., Hao, L., & Qianting, S. (2016). An Improved Semi-Supervised K-means Clustering Algorithm. *ITNEC*. doi:10.1109/ITNEC.2016.7560315
- [5] Gowda, H., Suhil, M., Guru, D., Raju, & Lavanya Narayana. (2017). Semi-Supervised Text Categorization using Recursive K-means clustering. *Communications in Computer and Information Science*, (pp. 217-227).



Gbr. 7 Perbandingan Tingkat Akurasi Pertama dan Analisa Kedua



Gbr.8 Perbandingan Lama Waktu Proses Analisa Pertama dan Kedua