Model Gaussian Mixture pada Distribusi Kecepatan Angin dengan Algoritma EM

1st Naufal Farras Deofanny
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
naufaltheodst@students.telkomuniversity.ac.id

2nd Aniq Atiqi Rohmawati
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
aniqatiqi@telkomuniversity.ac.id

3rd Indwiarti
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
indwiarti@telkomuniversity.ac.id

Abstrak

Dalam pemanfaatan sumber daya yang terbarukan, energi angin merupakan salah satu indikator penting dalam membangun fasilitas yang menggunakan energi angin seperti pembangkit listrik tenaga angin. Selain itu, kecepatan angin memiliki kaitan yang erat dengan indikator cuaca dan fenomena alam yang lain. Dengan kata lain, variabel kecepatan angin menjadi salah satu faktor yang penting dalam melakukan pemodelan yang berhubungan dengan fenomena alam. Berdasarkan uraian diatas, maka pada penelitian ini dilakukan pemodelan Gaussian Mixture pada distribusi kecepatan angin di daerah Padalarang, Jawa Barat. Hasil pengujian dilakukan dengan melakukan iterasi pemodelan Gaussian Mixture pada tiap komponen Gaussian dari satu sampai dua puluh, dari hasil pemodelan tersebut, maka dipilihlah Gaussian dengan 4 komponen sebagai model Gaussian Mixture terbaik yang mana diperoleh dari skor BIC yang terkecil yaitu sebesar 33814.149 dengan model Gaussian yang diperoleh adalah $p(x) = (0.17)P(X|\hat{\theta}_1) + (0.30)P(X|\hat{\theta}_2) + (0.31)P(X|\hat{\theta}_3) + (0.22)P(X|\hat{\theta}_4)$. Dalam hal ini, pada penilaian BIC tersebut didapat dari melihat kompleksitas pada parameter Gaussian yang diperoleh dan seberapa besar log-likelihood yang didapat.

Kata kunci : algoritma EM, fungsi peluang, kecepatan angin, *log-likelihood*, model gaussian mixture

Abstract

When it comes to the use of renewable resources, wind energy is one of the key indicators of buildings that use wind energy, such as wind turbines. In addition, wind speed is closely related to meteorological indicators and other natural phenomena. In other words, wind speed variables are one of the key factors in modeling natural phenomena. Based on the description above, Gaussian Mixture modeling was carried out on the distribution of wind speed in Padalarang, West Java. The test results are carried out by iterating Gaussian Mixture modeling on each Gaussian component up to twenty, from the modeling results, the Gaussian model with 4 components is chosen as the best Gaussian Mixture model obtained from the BIC score which is 33814.149 with the Gaussian model obtained is $p(x) = (0.17)P(X|\hat{\theta}_1) + (0.30)P(X|\hat{\theta}_2) + (0.31)P(X|\hat{\theta}_3) + (0.22)P(X|\hat{\theta}_4)$. In this case, the BIC assessment is obtained by observing the complexity of Gaussian's parameters and how much log-likelihood is obtained.

Keywords: EM Algorithm, log-likelihood, gaussian mixture model, probability density function, wind speed

I. PENDAHULUAN

Dalam perencanaan pembangunan Indonesia berjangka panjang, pemanfaatan sumber daya alam terbaharui menjadi prioritas, salah satunya adalah energi angin untuk membangun faslitas yang berhubungan dengan energi angin seperti pembangkit listrik tenaga angin. Selain itu, kecepatan angin berkaitan erat dengan indikator-indikator cuaca dan fenomena alam. Dalam hal ini, variabel kecepatan angin menjadi predictor yang signifikan dalam pemodelan fenomena alam. Beberapa metode distribusi probabilistik seperti Rayleigh, Weibull dan Gaussian telah banyak digunakan dalam penelitian power engineering untuk melakukan permodelan pada kecepatan angin [1]. Salah satu contoh penelitian yang berhubungan dengan energi angin antara lain, memprediksi tenaga yang dihasilkan pada turbin angin [2], lalu ada juga penelitian tentang penganalisaan potensi energi yang dihasilkan oleh angin di daerah Karnataka, India [1][3].

Pemodelan Gaussian Mixture (GMM) merupakan model statistik yang melibatkan fungsi distribusi Gaussian multivariate dari data (observasi) untuk menggambarkan sebaran frekuensi (magnitude) dari data. Pada penyajian data yang memiliki banyak *record*, biasanya akan lebih mudah dipahami apabila data tersebut dapat divisualisasikan kedalam suatu kategori atau kelas-kelas tertentu bergantung pada frekuensi pada data sehingga ketika dioberservasi akan menghasilkan nilai yang objektif pada suatu data dan untuk menghindari terjadinya bias atau penilaian yang hanya berfokus pada satu kategori saja. Untuk itu, pemodelan Gaussian Mixture merupakan salah satu pemodelan pendistribusian data yang dapat digunakan .Model tersebut merupakan metode permodelan yang popular

digunakan dan banyak diimplementasikan di bidang analisis pada data statistik, pengenalan pola dan *machine learning.* Segala permasalahan yang berhubungan dengan analisis data dan *clustering* dapat diselesaikan dengan Gaussian dengan pertimbangan semua titik datanya diambil dari distribusi Gaussian yang bersifat campuran dengan parameter yang tidak diketahui [4]. Gaussian Mixture Models secara umum dapat mengetahui distribusi datanya tanpa perlu mengetahui label dari data tersebut. Oleh karena itu GMM merupakan *unsupervised learning*.

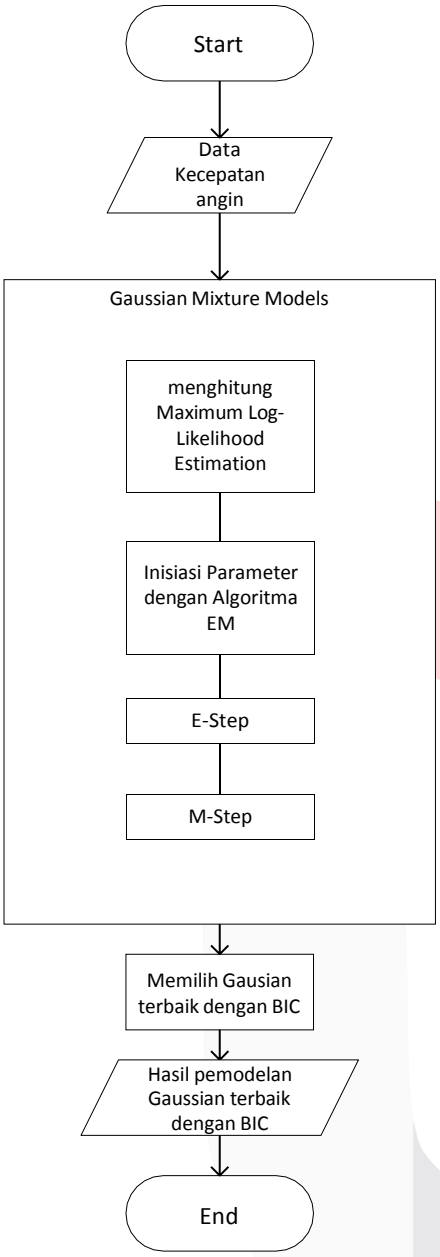
Oleh karena itu pada penelitian ini, akan dilakukan pemodelan Kecepatan Angin dengan menggunakan Gaussian Mixture Model pada data kecepatan angin di Indonesia yang mana data yang digunakan diambil di daerah Padalarang, Jawa Barat pada tahun 2015. Tujuan pada penelitian ini adalah untuk mengetahui bagaimana memodelkan distribusi kecepatan angin di Indonesia dengan menggunakan *Gaussian Mixture Model*, dan bagaimana mengevaluasi performansi dari pemodelan tersebut pada tiap *Gaussian*. Untuk pengimplementasiannya penulis juga akan mengevaluasi metode Log-Likelihood untuk mengukur parameter pada GMM dan untuk mengetahui performansinya peneliti menggunakan *Bayesian Information Criteria* untuk menghitung skor pada tiap *Gaussian*. Diharapkan pada hasil penelitian ini dapat membantu pengambilan keputusan dalam aktivitas sehari-hari yang berhubungan dengan kecepatan angin di Indonesia. pada penelitian ini membahas tentang studi terkait tentang penelitian yang menggunakan topik yang terkait pada bagian 2, pada bagian 3 juga dibahas mengenai sistem yang telah dibangun beserta penjelasan teorinya, pada bagian 4 akan dijelaskan mengenai eksekusi skenario dan hasil pengujian untuk dianalisis, untuk terakhir ada bagian 5 yang akan menjelaskan tentang kesimpulan dari hasil penelitain dan saran untuk penelitian mendatang

II. KAJIAN TEORI

Terdapat beberapa penelitian terkait yang menggunakan Gaussian Mixture Models yang digunakan untuk melakukan *learning* pada dsitribusi data. Contohnya, pada penelitian yang ditulis oleh R. Sanchez dan Reillo yaitu “Pengenalan Pola Geometri pada Tangan Menggunakan Gaussian Mixture” yang mana penelitian tersebut bertujuan untuk menyediakan kontrol akses sekuritas yang aman dengan menggunakan Biometrik pada telapak tangan. Hasil penelitian tersebut menunjukkan bahwa pengenalan menggunakan GMM memiliki akurasi yang memuaskan dengan success rate sebesar 96% dan memiliki Equal Error Rate (EER) sebanyak 4.6%. Dari hasil tersebut, terbukti GMM memiliki hasil pengenalan pola geometri pada tangan yang lebih bagus daripada menggunakan Hamming Distance System yang memiliki successs rate sebesar 87% dan memiliki Equal Error Rate (EER) sebanyak 8.3% [5]. Pada paper yang ditulis oleh Shi Qiu, Keyang Cheng, Lei Cui, Dongmei Zhou, dan Qiang Guo berjudul ”Algoritma pelacakan kendaraan bergerak menggunakan *deep learning”* melakukan penelitian dengan menggunakan dataset berupa kendaraan yang bergerak di jalan raya. Pendeteksian kendaraan bergerak sangatlah krusial dalam berbagai aplikasi khususnya pada bidang komputer visi. Penelitian dilakukan dengan menghitung seberapa besar error yang dihasilkan metode GMM dengan metode lain. Hasil penelitian menunjukkan bahwa dengan menggunakan GMM, pendeteksian gambar dapat dilakukan dengan lebih baik dengan error rate sebesar 0.79 pada dataset dibanding metode *frame difference method (FD)* dengan error rate 0.76. Hal ini menunjukkan GMM lebih baik dalam melakukan pendeteksian kendaraan yang bergerak [6]. Selain itu, pada penelitian [7] yang ditulis oleh Jose G. Dias dan Michel Wedel yaitu merupakan penelitian perbandingan empirikal terhadap performa dari algoritma EM dan MCMC terhadap pemodelan Gaussian. Tujuan utama dari penelitian ini adalah untuk melakukan perbandingan antara performa pada algoritma EM (Expectation-Maximization) dan MCMC (Markov Chain Monte Charlo). Dataset yang digunakan pada penelitian kali ini menggunakan data pendapatan perkapita pada 174 negara pada tahun 1998 yang dikonversikan ke dalam bentuk dolar. Hasil penelitian tersebut dilakukan melakukan pemodelan beberapa iterasi pada dataset. Menunjukkan bahwa hasil pada algoritma MCMC pada beberapa kondisi menghasilkan log-likelihood yang sedikit lebih baik daripada algoritma EM contoh pada satu kasus iterasi yaitu sebanyak -249.46 dibandingkan dengan algoritma EM yang sebanyak -249.45. Namun, pada hasil log-likelihood yang konvergen, algoritma EM menghasilkan nilai yang lebih stabil daripada MCMC yaitu sebanyak 75.6 % dari iterasi yang menghasilkan nilai yang konvergen daripada algoritma MCMC sebanyak 73.2 % dari iterasi yang menghasilkan nilai yang konvergen. Selain itu, pada paper tersebut juga dijelaskan terdapat beberapa masalah pada MCMC jika terjadi perubahan label.

III. METODE

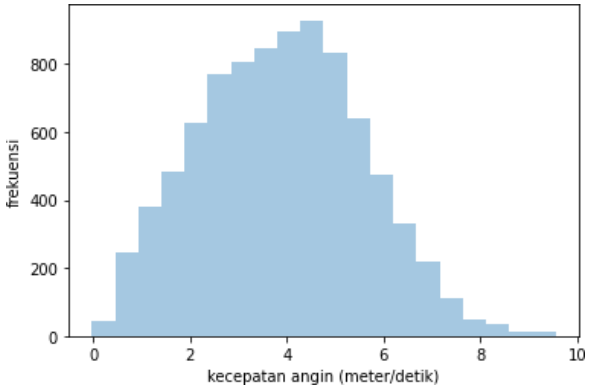
Pada penelitian ini, Peneliti akan membangun sistem untuk memodelkan distribusi kecepatan angin dengan model Gaussian Mixture melibatkan algoritma EM dengan flowchart sistem berikut ini:



GAMBAR 1. Flowchart Sistem

A. Data Kecepatan Angin

Data yang digunakan dalam pemodelan distribusi dengan Gaussian Mixture Model akan diperoleh dari penelitian yang dilakukan dengan menghitung kecepatan angin di daerah Padalarang di Jawa Barat. Sumber pada data tersebut diambil dari website http://dataonline.bmkg.go.id/ yang mana data angin tersebut diukur dengan menggunakan alat anemometer. Data tersebut berisi dua data yang mengandung data angin dan data ombak. Untuk penelitian kali ini peneliti akan menggunakan data angin untuk pemodelan distribusi dengan Gaussian Mixture Models. Rentang waktu data tersebut diambil pada tahun 2015. Berikut adalah data kecepatan angin yang disajikan dalam bentuk diagram batang:



GAMBAR 2. Data kecepatan angin di Indonesia

B. Gaussian Mixture Model (GMM)

GMM didefinisikan sebagai fungsi kepadatan probabiltas parametrik yang dapat direpresentasikan sebagai bobot/weight jumlah dari komponen K dari Gaussian [4]. Setiap komponen memiliki ciri yang sederhana dalam bentuk parametrik. Formula GMM dituliskan sebagai berikut.

$$p(x) = \sum_{j=1}^{K} w_j \, P(X|\mu_j, \Sigma_j) \quad (1)$$

Dimana :

K : Banyaknya distribusi Gaussian yang terlibat

w_j : peluang mixing ke-j

X : data kecepatan angin

μ_j : mean Gaussian ke-j

Σ_j : variansi Gaussian ke-j

w_j disebut sebagai peluang mixing parameter yang harus memenuhi $0 \leq p_j \leq 1$. Maka X adalah data set dimana $X = (X_1, \ldots, X_N)$ N pada rumus tersebut merujuk berapa banyak dimensi yang ada pada data. Fungsi probability density (pdf) pada distribusi Gaussian akan dituliskan sebagai berikut:

$$P(X|\mu, \Sigma) = \frac{1}{\sqrt{2\pi\Sigma}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sqrt{\Sigma}}\right)^2} \quad (2)$$

Dimana :

X : data kecepatan angin

μ : mean

Σ : variansi

e : exponen

π : 3.141 atau phi

C. Maximum Log-Likelihood Estimation

Dalam memperkirakan parameter GMM sejauh ini, metode yang paling banyak digunakan adalah Maximum Log-Likehood Estimation(MLE). Tujuan dari metode tersebut adalah untuk menemukan model parameter yang mana dapat memaksimalkan likelihood(kemungkinan) dari GMM. Dalam permodelan, parameter adalah angka-angka yang membentuk nilai distribusi yang aktual. Maximum Log-Likelihood Estimation ditulis pada rumus sebagai berikut:

$$\log[p(X|\Theta)] = \sum_{i=1}^{N} \log\left(\sum_{j=1}^{K} w_j P(X|\mu_j, \Sigma_j)\right) \quad (3)$$

Dimana:

Θ : banyaknya parameter
N : banyaknya data
K : Banyaknya distribusi Gaussian yang terlibat
w_j : peluang mixing ke-j
X : data kecepatan angin
μ_j : mean Gaussian ke-j
Σ_j : variansi Gaussian ke-j

Kekurangan pada metode log-likelihood adalah metode tersebut merupakan fungsi yang tidak linear dan memerlukan inisiasi parameter agar metode tersebut dapat berjalan. Untuk mengatasi hal tersebut, algoritma Expectation-maximization (EM) dapat digunakan untuk mengatasi kekurangan pada metode Maximum Log-Likelihood Estimation [8][9].

D. Algoritma EM

Seringkali terjadi masalah dalam mengestimasi parameter ketika melakukan fungsi likelihood yang mengakibatkan kesulitan dalam menghitung nilai maximization. Untuk itu, Algoritma Expectation-Maximization (EM) adalah pendekatan yang dapat digunakan untuk melakukan untuk melakukan memaksimalkan perkiraan pada metode MLE. Yang mana dapat berguna pada pada permasalahan data yang tidak lengkap. Algoritma tersebut bekerja dengan melakukan estimasi pada data yang hilang, dimana parameter akan diestimasi setelah mengisi nilai awal pada data yang hilang, dan akan terus terulang hingga hasilnya menjadi konvergen. Pada tiap iterasi pada algoritma EM, terdapat dua Langkah dalam melakukan prosesnya, yaitu langkah Expectation (E-Step), yang mana rumusnya ditulis sebagai berikut:

$$P(j|X_i) = \frac{w_j P(X_i|\mu_j, \Sigma_j)}{\sum_{m=1}^{K} p_m P(X_i|\mu_m, \Sigma_m)} \quad (4)$$

dan Langkah Maximization (M-Step) dengan rumus yang ditulis sebagai berikut:

$$\hat{w} = \frac{1}{N}\sum_{i=1}^{N} P(j|X_i) \quad (5)$$

$$\hat{\mu}_j = \frac{\sum_{i=1}^{N} P(j|X_i)X_i}{\sum_{i=1}^{N} P(j|X_i)} \quad (6)$$

$$\hat{\Sigma}_j = \frac{\sum_{i=1}^{N} P(j|X_i)(X_i - \mu_j)^2}{\sum_{i=1}^{N} P(j|X_i)} \quad (7)$$

Dimana:

w_j : peluang mixing ke-j
X_i : data kecepatan angin
μ_j : mean Gaussian ke-j
Σ_j : variansi Gaussian ke-j
N : banyaknya data

E. Cara Kerja Maximum Log-Likelihood Estimation dengan menggunakan algoritma EM

Dalam mengevaluasi metode Log-Likelihood, Algoritma EM adalah salah satu metode yang dapat diterapkan. Algoritma tersebut digunakan untuk memaksimalkan fungsi dari log-likehood dengan mengikuti parameter (mean, variansi dan iterasi pada probabilitas) pada model. Algoritma EM akan terus dilakukan secara berulang-ulang dan nantinya proses pengulangan akan dibagi menjadi 2 proses, yaitu: proses expectation dan proses maximization. Nantinya kedua proses tersebut akan terus dilakukan hingga maximum likehood estimationnya didefinisikan seperti pada rumus (3).
Inisiasi akan dimulai dengan parameter (Θ), pada proses expectation, dilakukan penghitungan mengikuti iterasi yang ada pada probabilitias dengan menggunakan rumus (4). Selanjutnya pada proses maximization, parameter pada model akan terus diperbarui berdasarkan nilai penghitungan yang dilakukan pada rumus (5, 6, dan 7). Pada tahap akhir, dilakukan evaluasi pada log-likehood dan melakukan pengecekan kriteria rumus berikut ini:

$$(||\log[p(X|\Theta_{t+1})] - \log[p(X|\Theta_t)]|| < \epsilon) \quad (8)$$

Dimana:

$\log[p(X|\Theta_{t+1})]$: log-likelihood pada iterasi t+1
$\log[p(X|\Theta_t)]$: log-likelihood pada iterasi t
ϵ : batas *error* prediksi

Jika kondisi tersebut belum terpenuhi maka akan dilakukan pengulangan kembali mulai dari proses expectation. dimana ϵ adalah batas *error* prediksi [4].

F. Bayesian Information Criteria

Bayesian Information Criteria (BIC) adalah kriteria simtotik yang konsisten jika distribusi pada data bersifat regular. Seperti yang sudah ditunjukkan pada penelitian [10], BIC adalah metode pengukuran yang konsisten dalam memilih berapa banyak K komponen pada GMM dengan mengukur seberapa kompleksitas parameter pada Gaussian Mixture Models [11]. Pengukuran dengan menggunakan BIC ditulis pada rumus berikut ini:

$$BIC = -2 \log p(X|\Theta) + \mathrm{m} \log(\mathrm{N}) \quad (9)$$

Dimana :

$\log p(X|\Theta)$: log-likelihood

m : banyaknya parameter model GMM

N : banyaknya data

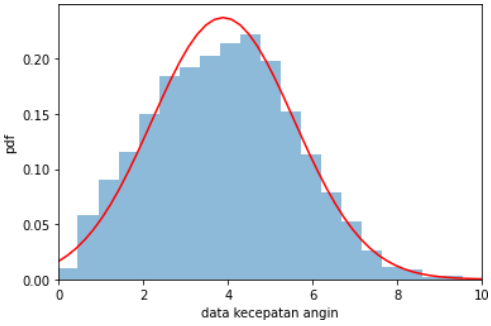
Nantinya, hasil akhir penilaian BIC pada tiap Gaussian K, yang terbaik yang diambil adalah BIC dengan nilai terkecil. Hal ini dapat dilihat karena BIC dilakukan dengan mengukur kompleksitas parameter pada GMM

IV. HASIL DAN PEMBAHASAN

A. Skenario Pengujian

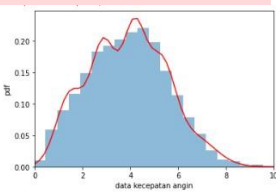
Dalam menentukan hasil dari pemodelan, karena distribusi datanya tidak diketahui pengelompokannya. Untuk itu, akan dilakukan skenario untuk memastikan hasil dari penelitian agar mengeluarkan hasil yang terbaik. Skenarionya adalah dilakukan pemodelan pada tiap Gaussian K hingga iterasinya mencapai angka 20. Dan selanjutnya akan dilakukan perbandingan mengenai hasil yang didapat, supaya prosesnya dapat diamati pemodelannya, peneliti akan menampilkan visualisasi pemodelan dari tiap iterasi ke 5 pemodelan.

B. Analisis Hasil Pengujian

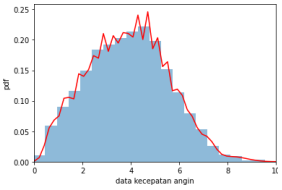


GAMBAR 3. Pemodelan Univariate Gaussian

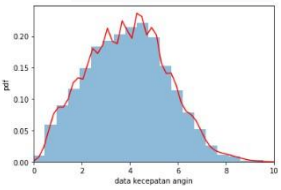
Pada gambar 3 ditunjukkan merupakan pemodelan Univariate Gaussian yang mana kurvanya berbentuk lonceng, disini karena hanya memiliki satu puncak maka dari keseluruhan data tersebut kurvanya hanya memiliki satu mean(μ), satu variansi(Σ), dan satu peluang mixing(w) yang memiliki nilai probabilitas bernilai satu karena hanya memiliki satu distribusi yang berarti data yang diamati hanya memiliki satu variabel yang dependen (variabel yang tidak bebas). Karena hanya memiliki satu variabel yang dependen saja, maka Gaussian Univariate tidak terlalu menggambarkan kurvanya secara presisi pada diagram batang, dapat terlihat jelas pada puncak pada kurva dan puncak pada diagram batang yang terlihat tidak terlalu fit ketika dimodelkan. Hal ini dikarenakan Gaussian Univariate hanya terbatas pada satu puncak saja.



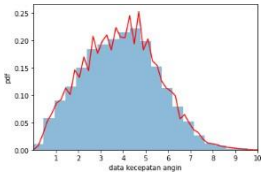
GAMBAR 4. Model Gaussian Mixture dengan K = 5



GAMBAR 6. Model Gaussian Mixture dengan K = 15



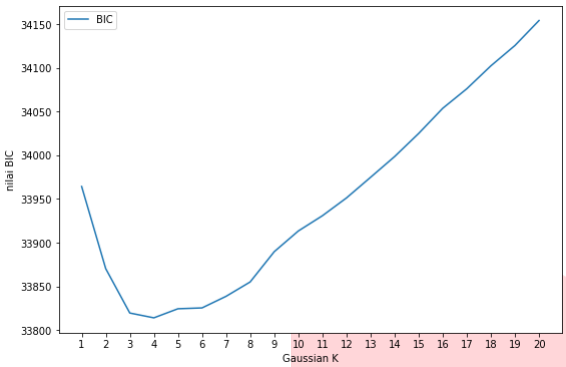
GAMBAR 5. Model Gaussian Mixture dengan K = 10



GAMBAR 7. Model Gaussian Mixture dengan K = 20

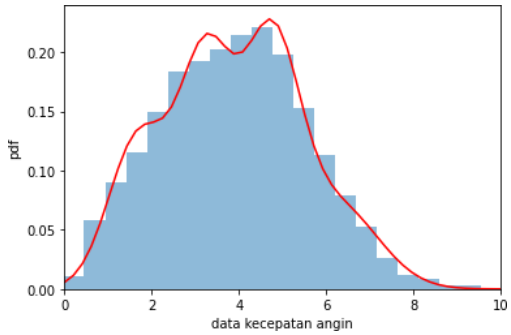
Sedangkan pada Gambar 4, 5, 6 dan 7 pada pemodelan menunjukkan pemodelan sesuai dengan banyaknya K yang ada pada Gaussian. Berbeda dengan Gaussian Univariat yang sebelumnya, pemodelan tersebut memiliki beberapa puncak tergantung dari seberapa banyaknya K Gaussian yang terlibat yang berarti pemodelan pada gambar tersebut memiliki lebih dari satu distribusi. Maka dari itu pemodelan pada gambar 4, 5, 6 dan 7 disebut Pemodelan Gaussian Mixture, karena pada tiap K memiliki 3 parameter yang dapat membangun Gaussian Mixture yaitu peluang mixing(w), mean(μ) dan variansi(Σ). Ketiga parameter tersebut merupakan variabel pembentuk Gaussian Mixture pada rumus (1). Dalam hal ini, kurva pada Gambar 4, 5, 6, dan 7 dibentuk dengan menjumlahkan tiap

Gaussian K yang terlibat. Pada tiap iterasi juga terlihat semakin besar Gaussian K terlihat maka kurva yang ditampilkan semakin presisi pada diagram batang. Namun dengan banyaknya Gaussian yang dilibatkan maka akan semakin banyak pula parameter yang akan dilibatkan dalam membentuk pemodelan dengan hasil nilai dari BIC akan semakin membesar sesuai dengan nilai log-likelihood yang didapat.



GAMBAR 8. Skor Penilaian BIC pada tiap n Gaussian K

Bisa dilihat pada Gambar 8 di atas menunjukkan kurva penilaian dengan menggunakan BIC yang dihitung dengan rumus (9) untuk menilai seberapa baik Gaussian K yang diinput, pada skor penilaian BIC, semakin kecil nilainya maka nilai pemodelan akan semakin baik. Seperti yang terlihat, bahwa Gaussian dengan 4 komponen merupakan Gaussian K yang terbaik dalam pemodelan, pada Gaussian univariate sampai Gaussian dengan 4 komponen, terlihat bahwa kurva yang turun menunjukkan terjadi peningkatan kualitas pada skor BIC, kurva terus turun sampai pada Gaussian dengan 4 komponen yang merupakan Gaussian K terbaik. Setelah menuju Gaussian dengan 5 komponen dan seterusnya, kurvanya terus naik hingga sampai pada Gaussian dengan 20 komponen yang telah ditujukan pada tabel merupakan Gaussian K dengan nilai BIC tertinggi. Dari data yang disajikan, maka peneliti akan memodelkan pemodelan GMM pada komponen yang terbaik yang mana adalah Gaussian dengan 4 komponen yang memiliki nilai BIC sebesar 33814.149. Hasil pemodelannya ditunjukkan pada gambar di bawah ini:



GAMBAR 9. Pemodelan Gaussian Mixture terbaik, Gaussian K = 4

Pada Gambar 9 terlihat pada Gaussian K ke-4 memiliki 4 puncak distribusi yang mempresentasikan diagram batang yang dihasilkan dari data asli. Pada koordinat Y ditunjukkan seberapa besar probability density function (pdf) yang dihasilkan pada koordinat X yang mana mempresentasikan data kecepatan angin. Dengan kata lain, sumbu Y adalah nilai probabilitas yang ada dari setiap kejadian X yang mana penghitungannya dijelaskan pada rumus (2), karena pdf menyatakan nilai fungsi kepadatan probabilitas, maka nilainya antara 0 hingga 1. Seperti contoh pada kecepatan angin 5 m/s diperoleh pdfnya adalah 0.23, artinya kecepatan angin tersebut intensitas kejadiannya adalah 0.23. Dari permodelan diatas, Model Gaussian Mixture yang diperoleh adalah $p(x) = (0.17)P(X|\hat{\Theta}_1) + (0.30)P(X|\hat{\Theta}_2) + (0.31)P(X|\hat{\Theta}_3) + (0.22)P(X|\hat{\Theta}_4)$, dengan $P(X|\Theta_i)$ bersesuaian dengan formula (2) dan $\hat{\Theta}$ adalah hasil estimasi parameter Gaussian Mixture, untuk lebih lengkapnya dapat dilihat pada Tabel 1:

TABEL 1 Parameter Gaussian Mixture Model pada Gaussian K = 4

K	Estimasi Parameter ($\hat{\Theta}$)		
	Peluang mixing (w_j)	Mean (μ_j)	Variansi (Σ_j)
1	0.169	6.235	0.927
2	0.3	3.242	0.386
3	0.314	4.733	0.408
4	0.216	1.689	0.460

Telah disajikan pada Tabel 1 yang merupakan parameter Gaussian Mixture data angin dengan komponen K ke=4 yang mana ketiga parameter tersebut mempunyai peran penting dalam merepresentasikan distribusi data sesuai pada rumus (1) dan rumus (2) dalam mencari nilai *pdf*. Untuk memenuhi rumus (1) itu sendiri, dilakukan dengan memaksimalisasikan log-likelihood yang dilakukan dengan rumus (3), di dalam rumus maximization log-likelihood itu sendiri parameternya diinisasikan dengan menggunakan Algoritma EM yang mana prosesnya dimulai dari E-step yang dihitung dengan rumus (4). Selanjutnya ada proses M-step yang penghitungannya dilakukan dengan rumus (5, 6, dan 7), kedua proses tersebut (Algoritma EM) akan terus berulang sampai menemukan hasil yang konvergen sesuai pada rumus (8). Sehingga menghasilkan ketiga parameter dari tiap Gaussian. Hasil dari GMM ini dapat digunakan untuk memetakan intensitas kecepatan angin di Padalarang, Jawa Barat

V. KESIMPULAN

Dari penelitian ini dapat ditarik kesimpulan bahwa dalam mencari distribusi kecepatan angin di

Padalarang, Jawa Barat. Model Gaussian Mixture terbaik model yang melibatkan empat Gaussian dengan skor BIC terkecil yaitu 33814.149 , dengan model yang diperoleh adalah $p(x) = (0.17)P(X|\theta_1) + (0.30)P(X|\theta_2) + (0.31)P(X|\theta_3) + (0.22)P(X|\theta_4)$. Menentukan Gaussian K berpengaruh secara signifikan terhadap pemodelan Gaussian Mixture, dengan menentukan Gaussian K yang tepat, kita dapat menentukan model distribusi kecepatan angin. Model Gaussian yang tepat nantinya dapat digunakan untuk *clustering* atau prediksi kecepatan angin dimana hasil prediksi tersebut diperoleh dengan melibatkan fungsi kepadatan probabilitas atau yang disebut dengan *probability density function* (pdf).

REFERENSI

[1] M. Wahbah, O. Alhussein, T. H. M. El-Fouly, B. Zahawi, and S. Muhaidat, "Evaluation of Parametric Statistical Models for Wind Speed Probability Density Estimation," *2018 IEEE Electr. Power Energy Conf. EPEC 2018*, pp. 1–6, 2018, doi: 10.1109/EPEC.2018.8598283.

[2] K. Sunderland, T. Woolmington, J. Blackledge, and M. Conlon, "Small wind turbines in turbulent (urban) environments: A consideration of normal and Weibull distributions for power prediction," *J. Wind Eng. Ind. Aerodyn.*, vol. 121, pp. 70–81, 2013, doi: 10.1016/j.jweia.2013.08.001.

[3] A. S. A. Awad, T. H. M. EL-Fouly, and M. M. A. Salama, "Optimal ESS Allocation for Benefit Maximization in Distribution Networks," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1668–1678, 2017, doi: 10.1109/TSG.2015.2499264.

[4] F. Najar, S. Bourouis, N. Bouguila, and S. Belghith, "A comparison between different Gaussian-based mixture models," *Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl. AICCSA*, vol. 2017-Octob, pp. 704–708, 2018, doi: 10.1109/AICCSA.2017.108.

[5] R. Sanchez-Reillo, "Hand geometry pattern recognition through Gaussian Mixture Modelling," *Proc. - Int. Conf. Pattern Recognit.*, vol. 15, no. 2, pp. 937–940, 2000, doi: 10.1109/icpr.2000.906228.

[6] S. Qiu, K. Cheng, L. Cui, D. Zhou, and Q. Guo, "A moving vehicle tracking algorithm based on deep learning," *J. Ambient Intell. Humaniz. Comput.*, no. 2017, 2020, doi: 10.1007/s12652-020-02352-w.

[7] J. G. Dias and M. Wedel, "An empirical comparison of EM, SEM and MCMC performance for problematic Gaussian mixture likelihoods," *Stat. Comput.*, vol. 14, no. 4, pp. 323–332, 2004, doi: 10.1023/B:STCO.0000039481.32211.5a.

[8] C. Based, "Gabor Jets," *Encycl. Biometrics*, pp. 627–627, 2009, doi: 10.1007/978-0-387-73003-5_334.

[9] M. Allen, "Maximum Likelihood Estimation," *SAGE Encycl. Commun. Res. Methods*, pp. 1–5, 2017, doi: 10.4135/9781483381411.n322.

[10] C. Keribin, "Consistent estimate of the order of mixture models," *Comptes Rendus l'Academie des Sci. - Ser. I Math.*, vol. 326, no. 2, pp. 243–248, 1998, doi: 10.1016/s0764-4442(97)89479-7.

[11] A. Mehrjou, R. Hosseini, and B. Nadjar Araabi, "Improved Bayesian information criterion for mixture model selection," *Pattern Recognit. Lett.*, vol. 69, pp. 22–27, 2016, doi: 10.1016/j.patrec.2015.10.004.