IMPLEMENTASI *POISSON HIDDEN MARKOV MODELS* DAN *EXPECTATION MAXIMIZATION ALGORITHM* DALAM PENENTUAN MODEL TERBAIK

(Studi Kasus : Data Banyaknya Kejadian Angin Ribut di Jawa**)**

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana
Program Studi Statistika



Irina Hidayati

14 611 002

JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2018

HALAMAN PERSETUJUAN PEMBIMBING

TUGAS AKHIR

Judul	:	Implementasi *Poisson Hidden Markov Models* dan *Expectation Maxmization Algorithm* dalam Penentuan Model Terbaik (Studi Kasus : Data Banyaknya Kejadian Angin Ribut di Jawa)
Nama Mahasiswa	:	Irina Hidayati
Nomor Mahasiswa	:	14 611 002

TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK DIUJIKAN

Yogyakarta, April 2018

Pembimbing

(Dr. Edy Widodo, S.Si., M.Si.)

HALAMAN PENGESAHAN

TUGAS AKHIR

IMPLEMENTASI *POISSON HIDDEN MARKOV MODELS* DAN *EXPECTATION MAXIMIZATION ALGORITHM* DALAM PENENTUAN MODEL TERBAIK

(Studi Kasus : Data Banyaknya Kejadian Angin Ribut di Jawa)

Nama Mahasiswa : Irina Hidayati

Nomor Mahasiswa : 14 611 002

TUGAS AKHIR INI TELAH DIUJIKAN

PADA TANGGAL APRIL 2018

Nama Penguji	**Tanda tangan**
1 Ir. Ali Parkhan, M.T	.....
2 M. Hasan Sidiq Kurniawan, M.Sc	.....
3 Dr. Edy Widodo, S.Si., M.Si	.....

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

(Drs. Allwar, M.Sc., Ph.D.)

KATA PENGANTAR



Assalamu'alaikum Warahmatullaahi Wabarakaatuh

Alhamdulillaahirabbil'aalamiin, puji syukur kehadirat Allah SWT, yang telah melimpahkan rahmat dan hidayah-Nya, serta shalawat dan salam peneliti haturkan kepada nabi Muhammad SAW sehingga peneliti dapat menyelesaikan tugas akhir yang berjudul "**IMPLEMENTASI *POISSON HIDDEN MARKOV MODELS* DAN *EXPECTATION MAXIMIZATION ALGORITHM* DALAM PENENTUAN MODEL TERBAIK (Studi Kasus : Data Banyaknya Kejadian Angin Ribut di Jawa)**" dengan baik.

Tugas akhir ini disusun sebagai salah satu persyaratan untuk memperoleh gelar sarjana di Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia. Peneliti menyadari bahwa selama penyusunan tugas akhir ini tidak lepas dari saran, kritik, bimbingan, maupun dorongan dari berbagai pihak. Oleh karena itu, perkenankanlah peneliti menyampaikan ucapan terimakasih kepada:

1. Bapak Drs. Allwar, M.Sc., Ph.D selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia, Yogyakarta.
2. Bapak Dr. Raden Bagus Fajriya Hakim, S.Si., M.Si., selaku Kepala Program Studi Statistika FMIPA UII beserta jajarannya.
3. Bapak Dr. Edy Widodo, S.Si., M.Si., selaku dosen pembimbing dalam pelaksanaan tugas akhir yang telah memberi bimbingan dan arahannya.
4. Seluruh dosen prodi Statistika, atas ilmu yang telah diberikan selama proses belajar, semoga menjadi amal kebaikan Bapak/Ibu sekalian.
5. Orang tua tersayang, Bapak M. Yasin dan Ibu Netti Aslamiah yang telah mencurahkan kasih sayangnya, memberikan doa, dukungan serta dorongan moril maupun materil, serta kakak satu-satunya, Muhammad Yuhandi yang selalu memberikan motivasi dan semangatnya.

6. Sendhyka Cakra Pradana, terima kasih selalu ada dan menjadi penyemangat.
7. Nilam Novita Sari dan Oktavia, sahabat tersayang yang selalu sabar dan selalu ada dalam keadaan apapun. *You're my best friends*.
8. Noor Asyiah, Rima Juridar, dan Sulhaerati, sahabat seperjuangan yang selalu membagi ceritanya, selalu menemani, dan memberikan motivasinya.
9. Venanda, Suci, Widya, Syahriani, dan Delpiana, teman satu *kost*-an yang sudah seperti keluarga, selalu menghibur, menemani waktu begadang. Terima kasih atas kebersamaan kita yang hampir *so sweet*.
10. Atiqah Febry Nurlyta, sahabat paling sabar, paling pengertian, penampung curhatan dan kegalauan. *You make me happy in a way no one else can*.
11. Ziane, Yani, Calu, Indana, Putri, dan Anis, sahabat tersayang yang selalu memberikan keceriaan, dan hiburannya.
12. Renita, Rinanti, Anilda, dan Ainun, terima kasih selalu menghibur ketika pulang ke rumah.
13. Teman-teman Statistika UII angkatan 2014 serta seluruh keluarga besar Statistika UII.
14. Semua pihak yang telah mendukung dan membantu peneliti.

Semoga bimbingan, dukungan dan bantuan dari semua pihak senantiasa diberikan rahmat dan anugerah oleh Allah SWT. Peneliti menyadari bahwa tugas akhir ini masih jauh dari sempurna. Oleh karena itu, peneliti menerima kritik dan saran yang membangun demi kesempurnaan tugas akhir ini. Semoga dapat bermanfaat bagi semua pihak.

Wassalamu'alaikum Warahmatullaahi Wabarakaatuh

Yogyakarta, April 2018

Irina Hidayati

DAFTAR ISI

HALAMAN JUDUL i
HALAMAN PERSETUJUAN PEMBIMBING ii
HALAMAN PENGESAHAN iii
KATA PENGANTAR iv
DAFTAR ISI vi
DAFTAR TABEL ix
DAFTAR GAMBAR x
DAFTAR LAMPIRAN xi
PERNYATAAN xii
INTISARI xiii
ABSTRACT xiv
DAFTAR SINGKATAN xv
BAB I PENDAHULUAN 1
1.1. Latar Belakang Masalah 1
1.2. Rumusan Masalah 5
1.3. Batasan Masalah 5
1.4. Tujuan Penelitian 5
1.5. Manfaat Penelitian 6
BAB II TINJAUAN PUSTAKA 7
2.1 Angin Ribut 7
2.2 *Hidden Markov Models* dan *Expectation Maximization Algorithm* 9
BAB III LANDASAN TEORI 12
3.1 Konsep Dasar Peluang 12
3.1.1 Ekspektasi 13
3.1.2 Kejadian-kejadian 14
3.1.3 Ruang Sampel Diskrit 14
3.2 Variabel Acak 15
3.2.1 Konsep Variabel Acak Diskrit 15

3.2.2 Fungsi Massa Probabilitas Variabel Acak Diskrit 16
3.2.3 Fungsi Distribusi Kumulatif Variabel Acak Diskrit 16
3.2.4 Momen Variabel Acak Diskrit 17
3.3 Model Fungsi Variabel Acak Diskrit 18
3.3.1 Model *Poisson* 18
3.4 Proses *Poisson* 19
3.5 Overdispersi 19
3.6. *Independent Mixture Models* 20
3.7. Pengantar Stokastik 21
3.8. Proses Rantai *Markov* 22
3.9. Rantai Markov Diskrit 22
3.10. Prediksi Peluang dengan Pendekatan Matriks Transisi 25
3.11. *Poisson Hidden Markov Models* (*PHMM*) 26
3.11.1 *Hidden Markov Models* (*HMM*) 26
3.12 Distribusi Marginal dan Moment *HMM* 27
3.13. Penskalaan *Likelihood* 27
3.14. Estimasi *Expectation Maximization Algorithm* (EM)..... 27
3.15. Pemilihan Model berdasarkan *Akaike Information Criterion* (*AIC*) 29
3.16. Definisi Angin 29
3.16.1 Jenis-jenis Angin 30
3.16.2 Skala Angin *Beaufort* 31
BAB IV METODOLOGI PENELITIAN 32
4.1. Data Penelitian 32
4.2. Sampel dan Teknik Sampling 32
4.3. Variabel dan Definisi Operasional Variabel 32
4.4. Metode Analisis Data 33
4.5. Proses Analisis Data 33
BAB V ANALISIS DAN PEMBAHASAN 36
5.1. Deskripsi Data 36
5.2. Pengecekan *Overdispersi* 38

5.3. Pemodelan Menggunakan PHMM (*Poisson Hidden Markov Models*) 39
5.3.1 Penentuan Parameter *Input* 40
5.3.2 Penaksiran Parameter-parameter *PHMM* dengan Algoritma *EM* 43
5.4. Prediksi dengan Pendekatan Matriks 44
BAB VI PENUTUP 45
6.1. Kesimpulan 45
6.2. Saran 46
DAFTAR PUSTAKA 47
LAMPIRAN 50

DAFTAR TABEL

Tabel 2.1	Penelitian terdahulu tentang angin ribut	7
Tabel 2.2	Penelitian terdahulu tentang *HMM* dan Algoritma *EM*	9
Tabel 3.1	Peluang Keadaan Cuaca.....	24
Tabel 3.2	Skala Angin *Beaufort*.....	31
Tabel 4.1	Definisi Operasional Variabel	32
Tabel 5.1	Data Kecepatan Angin Wilayah Jawa Berdasarkan Stasiun Meteorologi.....	36
Tabel 5.2	Banyaknya Kejadian Angin ribut di wilayah Jawa Januari 2014-Desember 2017	38
Tabel 5.3	Summary Banyaknya Kejadian Angin ribut di wilayah Jawa Januari 2014-Desember 2017	39
Tabel 5.4	Hasil Perhitungan Parameter λ dan δ pada m = 2.....	41
Tabel 5.5	Peluang pada Kasus 2 Keadaan Tersembunyi	42
Tabel 5.6	Parameter λ, δ, dan ⌈ pada masing-masing *PHMM*	42
Tabel 5.7	Hasil Estimasi Parameter *PHMM* dengan Algoritma *EM* pada masing-masing Keadaan Tersembunyi	43

DAFTAR GAMBAR

Gambar 1.1 Grafik Jumlah Kejadian Bencana Angin Ribut Tahun 1990-2011 3

Gambar 3.1 Sebuah himpunan dalam sebuah diagram *Venn M* 14

Gambar 3.2 Graf Dasar *HMM* 26

Gambar 4.1 Flowchart Penelitian 33

Gambar 5.1 Peta Potensi Angin Wilayah Jawa Tahun 2014-2017 37

DAFTAR LAMPIRAN

Lampiran 1 *Script R untuk Peluang Forward dan Backward pada PHMM* 51

Lampiran 2 *Script R untuk Estimasi Algoritma EM*..... 52

Lampiran 3 *Banyaknya Kejadian Angin Ribut di Wilayah Jawa pada Kecepatan Angin ≥ 34 Knot Januari 2014 – Desember 2017* 53

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, April 2018



Hidayati

IMPLEMENTASI *POISSON HIDDEN MARKOV MODELS* DAN *EXPECTATION MAXIMIZATION ALGORITHM* DALAM PENENTUAN MODEL TERBAIK

(Studi Kasus : Data Banyaknya Kejadian Angin Ribut di Jawa**)**

Irina Hidayati

Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Islam Indonesia

INTISARI

Kondisi geografis di Indonesia yang merupakan daerah tropis sangat mendukung terjadinya angin ribut karena dilihat dari letak lintang dan bujurnya. Angin ribut adalah suatu fenomena meteorologis dan dampaknya dapat menimbulkan kerusakan serta kerugian yang cukup besar. Banyaknya kejadian angin ribut dalam kurun waktu tertentu dapat diprediksi dan hasilnya dapat dijadikan dasar untuk penanganan kasus bencana angin ribut. Jika banyaknya kejadian angin ribut merupakan proses observasi dan mengalami overdispersi, serta penyebab kejadiannya tidak diamati secara langsung dan membentuk suatu rantai Markov, maka proses observasi dan penyebabnya dapat dimodelkan dengan Poisson Hidden Markov. Model tersebut dicirikan oleh parameternya yang dapat diduga menggunakan Algoritma EM. Dari model yang diperoleh kemudian dipilih model terbaik berdasarkan nilai AIC. Penelitian dengan PHMM ini diperoleh model 2 state sebagai model terbaik dengan nilai AIC terkecil, yaitu 132,2833.

Kata kunci : *Algoritma EM, Angin ribut, Hidden Markov Model, Poisson*

IMPLEMENTATION OF POISSON HIDDEN MARKOV MODELS AND EXPECTATION MAXIMIZATION ALGORITHM IN DETERMINING THE BEST MODEL

(Case Study : Amount of Wind Cases in Jawa)

Irina Hidayati

Statictics Department, Faculty of Mathematics and Natural Sciences

Islamic University of Indonesia

ABSTRACT

Geographical conditions in Indonesia which is a tropical area strongly supports the occurrence of a hurricane because of the latitude and longitude location . Hurricane is a meteorological phenomenon and its impact can cause considerable damage and losses. The number of occurrences of hurricanes within a certain time can be predicted and prediction results can be used as the basis for handling the case of a hurricane disaster. If the number of occurrences of hurricanes is an observation process and overdispersed, also the cause of the event is not observed directly and form a Markov chain, then the process of observation and its causes can be modeled with Poisson Hidden Markov. The model is characterized by its predictable parameters using the EM Algorithm. From the model obtained then selected the best model based on the AIC value. Research with PHMM obtained 2 model state as the best model with the smallest AIC value, that is 132,2833.

Keywords : *EM Algorithm, Hurricane, Hidden Markov Model, Poisson*

DAFTAR SINGKATAN

CDF	: *Cumulative Distribution Function*
CMAX	: *Colume Max*
dBz	: *Decibels of z*
EM	: *Expectation Maximization*
PHMM	: *Poisson Hidden Markov Model*
PMF	: *Probability Mass Function*
WRF	: *Weather Research and Forecasting*

BAB I
PENDAHULUAN

1.1. Latar Belakang Masalah

Angin memiliki peranan penting dalam menentukan suatu kondisi iklim dan cuaca di suatu tempat. Secara umum, angin adalah udara yang bergerak yang disebabkan oleh perbedaan tekanan udara dengan arah angin dari tempat yang bertekanan tinggi ke tempat yang bertekanan rendah. Arah angin itu sendiri dapat dilihat dari arah mana angin itu datang, misalnya dari barat maka disebut angin barat (Kartasapoetra, 2004).

Data kecepatan dan arah angin secara akurat dapat diukur dengan suatu alat yang dinamakan dengan anemometer. Anemometer umumnya digunakan pada stasiun pengamatan cuaca. Dalam pengukurannya, anemometer jenis *cup* menerapkan metode mekanik. Instrumen ini dengan tiga atau empat lubang kecil yang sudah ditetapkan, sehingga mampu menangkap angin dan berputar dengan dibantu batang vertikal (Azlina, 2013).

Angin juga memiliki peranan yang besar bagi kehidupan manusia. Sebagai contoh, pada masa lampau, bangsa-bangsa Barat yang datang ke Indonesia menggerakkan kapalnya menggunakan tenaga angin. Selain itu, ketersediaan angin yang tidak ada habisnya di alam ini dapat dimanfaatkan sebagai dasar pembangunan pembangkit listrik bertenaga angin. Sebagai sarana olahraga dan rekreasi, angin juga dapat dimanfaatkan oleh manusia untuk bermain layang-layang, terjun payung, dan sebagainya. Namun, sebaliknya, kecepatan angin dalam kisaran tertentu yang melebihi batas maksimum kondisi normal juga dapat merugikan manusia.

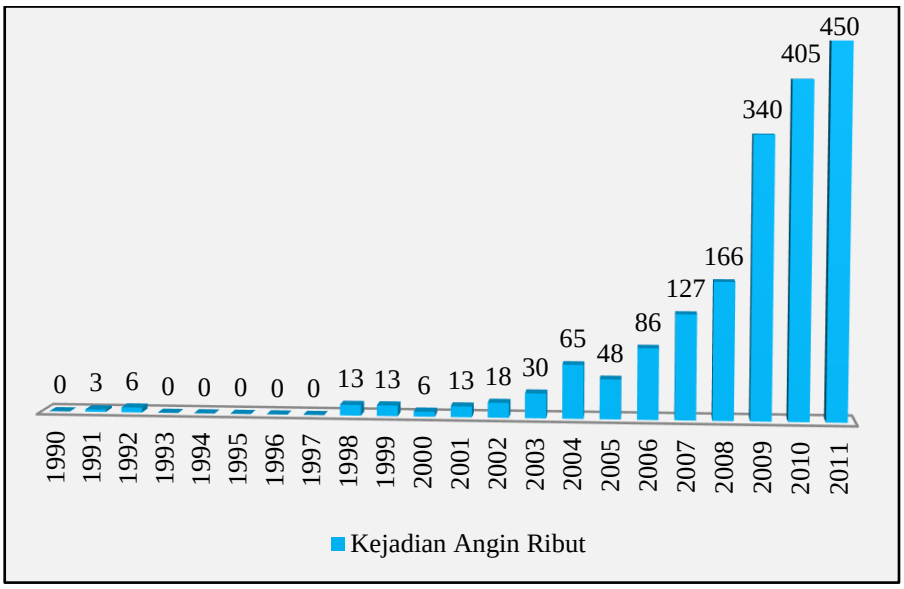
Angin berkecepatan tinggi terjadi karena adanya perbedaan tekanan yang sangat besar antara dua lokasi yang berdekatan. Angin ini disebut pula dengan angin ribut karena kecepatannya dapat menimbulkan daya rusak terhadap berbagai media yang dilaluinya. Terjadinya angin ribut dipengaruhi oleh pembentukan awan *cumulonimbus* dari konveksi maupun orografi massa udara yang tidak stabil atau dari adveksi massa udara relatif dingin dengan massa udara yang relatif panas

dalam frontal massa udara (Nikaryanto, 1979). Keadaan tersebut yang menyebabkan posisi konvergensi dan divergensi mempunyai perbedaan tekanan udara yang sangat besar pada posisi yang berdekatan sehingga memicu terjadinya angin ribut. Angin tersebut juga mempunyai arah yang tidak beraturan, sehingga berpotensi menyebabkan kerusakan dalam segala bentuk budi daya yang ada di permukaan bumi, seperti rusaknya pemukiman, sektor sosial ekonomi, serta sarana dan prasarana umum.

Kondisi geografis di Indonesia yang merupakan daerah tropis sangat mendukung terjadinya angin ribut karena dilihat dari letak lintang dan bujurnya. Indonesia merupakan daerah tropis yang memiliki kelembaban di atas 75% yang menyebabkan ketidakstabilan massa udara. Letak Indonesia yang berbatasan langsung dengan Samudera Hindia dan Samudera Pasifik serta dekat dengan Benua Asia dan Australia juga mempengaruhi terjadinya angin ribut. Adanya Angin Monsun Barat dan Angin Monsun Timur memicu terjadinya angin ribut di daratan. Angin ribut banyak terjadi pada musim penghujan, namun tidak menutup kemungkinan angin ribut terjadi di saat musim kemarau.

Dalam skala angin *Beaufort*, yang disebut angin ribut adalah angin mulai skala 8 dimana angin ini memiliki kecepatan 34-40 knot atau 62-74 km/jam (Tjasyono, 2007). Sedangkan, menurut peraturan kepala BMKG (Nomor 009 Tahun 2010), angin yang memiliki kecepatan lebih dari 34 knot dinamakan dengan angin puting beliung dimana angin ini berputar kencang yang berasal dari awan *cumulonimbus*. Angin ini kerap terjadi pada musim pancaroba pada siang atau sore hari.

Jumlah kejadian bencana angin ribut yang tercatat, terus mengalami peningkatan dari tahun ke tahun seperti yang tercatat pada tahun 1990-2011. Pada tahun 1990-1997 sebagian besar data bernilai nol kejadian karena pada tahun-tahun tersebut belum banyak obyek-obyek bencana (*element at risk*) yang mengalami kerusakan sehingga kejadian bencana angin ribut belum begitu diperhatikan oleh masyarakat dan pihak terkait. Jumlah kejadian angin ribut disajikan pada grafik berikut.



Gambar 1.1 *Grafik Jumlah Kejadian Bencana Angin Ribut Tahun 1990-2011*
(Sumber: Pengolahan Data Sekunder dari BNPB)

Berdasarkan **Gambar 1.1**, pada tahun 1997-2007 hanya sedikit saja kejadian bencana yang dicatat namun jumlahnya sudah terus meningkat dari tahun ke tahun. Peningkatan ini menunjukan bahwa sudah mulai ada perhatian dari pemerintah terkait dengan kejadian bencana angin ribut. Tahun 2008-2009 mengalami kenaikan jumlah kejadian yaitu dari 166 kejadian menjadi 340 kejadian. Hal ini karena sistem pendataan bencana sudah semakin baik dan juga *monitoring* cuaca dengan satelit terus dimanfaatkan. Kemudian dari tahun 2010-2011 mengalami kenaikan menjadi 450 kejadian. Konsentrasi bencana terbanyak berada di wilayah Jawa Tengah. Beberapa media juga memberitakan kejadian angin ribut yang terjadi di wilayah Jawa Tengah. Media *online* Republika.co.id memberitakan bahwa sebanyak 17 rumah mengalami kerusakan di Kabupaten Temanggung setelah diterjang angin ribut. Kerusakan terjadi di beberapa lokasi, diantaranya Dusun Potrokasan, Bringinsari, Kauman, Carikan, Potrogalan dan Desa Kedu. Selain itu, angin ribut juga melanda wilayah Banjarnegara yang dilansir dari media *online* Detik News. Akibat dari kejadian tersebut, satu orang dilaporkan meninggal dunia dan satu korban lainnya dilarikan ke rumah sakit untuk perawat lebih lanjut. Angin ribut tersebut terjadi di Kecamatan Batur, Kecamatan Karangkobar, dan Kecamatan Pagentan.

Berdasarkan grafik jumlah kejadian bencana angin ribut tahun 1990-2011, dapat diketahui bahwa beberapa provinsi di Pulau Jawa memiliki jumlah kejadian yang relatif lebih banyak dibanding di luar Jawa. Hal ini karena Pulau Jawa dikelilingi oleh lautan seperti Lautan Hindia, Laut Jawa, Laut Natuna dan Selat Malaka dan juga adanya dataran yang letaknya berdekatan dengan Jawa yaitu Pulau Sumatera dan Benua Australia. Sehingga arus angin yang berasal dari wilayah-wilayah tersebut mempunyai perbedaan suhu, tekanan dan kelembaban yang relatif tinggi.

Angin ribut banyak terjadi pada Bulan Oktober hingga Maret. Bulan-bulan yang mengalami banyak kejadian bencana angin ribut tersebut merupakan bulan- bulan saat terjadinya musim hujan di Indonesia yaitu pada Bulan Oktober-April. Dimana biasanya kejadian angin ribut disertai dengan hujan lebat dan petir yang menjadi ciri khas fenomena di musim hujan. Angin ribut juga banyak terjadi pada musim Pancaroba (Emilya, Arum dan Febriyan, 2013). Pada musim pancaroba posisi zona konvergensi antar tropik berkisar diatas ekuator sehingga mengalami variabilitas suhu tertinggi yang memicu terjadinya pergerakan angin yang kencang (Nirkaryanto, 1979). Pada wilayah darat menyebabkan kerusakan-kerusakan pada bangunan, sementara jika angin kencang terus-menerus bertiup di atas lautan maka dapat menimbulkan gelombang besar. Jika terjadi gelombang besar maka nelayan di pesisir akan sukar untuk melaut mencari ikan.

Informasi mengenai dampak kejadian angin ribut perlu diperhatikan untuk mengetahui seberapa besar kerugian yang terjadi di daerah kejadian. Dampak kerugian tersebut meliputi kerugian secara material dan non material yang ditanggung oleh individu dan pemerintah. Secara umum, karakteristik kejadian banyak fenomena angin ribut dalam periode waktu tertentu didekati oleh distribusi *Poisson* yang memiliki nilai variansi dan nilai *mean* sama. Dalam pengaplikasian distribusi *Poisson* ini relatif terjadi overdispersi. Overdispersi disebabkan adanya pengelompokan data pada distribusi tersebut dimana untuk tiap-tiap kelompok memiliki parameter yang berbeda-beda. Karena *Hidden Markov Models* adalah perluasan dari rantai Markov, maka hasil pengamatannya bergantung pada keadaan proses Markov yang *unobserved*. Pada setiap kasus kebencanaan,

penyebab bencana yang berbeda akan menyebabkan jumlah bencana yang terjadi pula. Oleh karena itu, maka digunakan model *Poisson* campuran atau dikenal dengan istilah *Poisson Hidden Markov Models* dimana setiap observasinya dihasilkan oleh salah satu dari *m* keadaan tersembunyi (*hidden state*) (Nurhasanah, 2016). Pengaplikasian dengan *Poisson Hidden Markov Models* bertujuan untuk mengidentifikasi pola-pola barisan keadaan yang *unobserved* yang mendasari barisan observasi. Dengan menggunakan metode *Poisson Hidden Markov Models*, peneliti akan membangun sebuah model dari banyaknya kejadian angin ribut, serta menggunakan *Expectation Maximization Algorithm* untuk mendapatkan estimator dari parameternya.

1.2. Rumusan Masalah

Berdasarkan latar belakang permasalahan di atas, maka rumusan masalah yang dapat diidentifikasi sebagai berikut :

1. Bagaimana sebaran potensi angin menurut skala *Beaufort* di wilayah Jawa pada tahun 2014-2017?
2. Bagaimana hasil estimasi parameter dari model banyaknya kejadian angin ribut menggunakan *Poisson Hidden Markov Models* dan *Expectation Maximization Algorithm*?
3. Bagaimana prediksi banyaknya kejadian angin ribut dari model yang terbentuk?

1.3. Batasan Masalah

Agar penelitian tugas akhir ini tidak menyimpang dari tujuan yang semula direncanakan sehingga mempermudah mendapatkan informasi, maka peneliti menetapkan batasan-batasan masalah sebagai berikut :

1. Penelitian ini menggunakan data banyaknya kejadian angin ribut.
2. Peneliti menggunakan bantuan *R 3.4.0* untuk melakukan proses analisis data.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

1. Mengetahui sebaran potensi angin berdasarkan skala *Beaufort* di wilayah Jawa tahun 2014-2017.

2. Memperoleh hasil estimasi parameter dari model banyaknya kejadian angin ribut menggunakan *Poisson Hidden Markov Models* dan *Expectation Maximization Algorithm*.
3. Mengetahui prediksi banyaknya kejadian angin ribut dari model yang terbentuk.

1.5. Manfaat Penelitian

Adapun manfaat yang diharapkan dari hasil penelitian ini yaitu:

1. Sebaran potensi angin dapat mengetahui wilayah yang memiliki kategori kecepatan angin minimum, menengah, dan maksimum .
2. Dapat diketahui hasil estimasi rata-rata banyaknya kejadian angin ribut di wilayah Jawa.
3. Hasil prediksi dapat mengetahui peluang kedepan terjadinya bencana angin ribut.

BAB II
TINJAUAN PUSTAKA

Hal terpenting dalam membuat suatu penelitian adalah dengan adanya penelitian terdahulu, dimana bertujuan untuk mengetahui hubungan antara penelitian sebelumnya dengan penelitian yang akan dilakukan. Selain menghindari adanya tindakan duplikasi, dengan menggunakan penelitian terdahulu, penelitian yang dilakukan peneliti ini mempunyai arti penting sehingga memberikan kotribusi pada perkembangan ilmu pengetahuan. Berikut beberapa penelitian-penelitian terdahulu yang berkenaan dengan data dan metode yang digunakan yang dijadikan acuan oleh peneliti sebagai berikut.

2.1 Angin Ribut

Pada **Tabel 2.1** menunjukkan penelitian-penelitian terdahulu yang berkaitan dengan kecepatan angin.

Tabel 2.1 *Penelitian terdahulu tentang angin ribut*

No	Nama Peneliti	Variabel Penelitian	Metode Penelitian	Hasil Penelitian
1.	Emilya Nurjani, Arum Rahayu, dan Febriyan Rachmawati (2013)	- Frekuensi kejadian angin ribut - Data digital penggunaan lahan	Statistik deskriptif dan analisis spasial	Angin ribut banyak terjadi di Pulau Jawa dimana provinsi Jawa Tengah merupakan daerah dengan kejadian angin ribut terbanyak. Sedangkan, provinsi Bengkulu dan Papua Barat tidak terjadi angin ribut. Upaya mitigasi bencana diutamakan pada daerah yang rawan bencana.
2.	Linda Fitrotul Muzayanah (2015)	- Frekuensi kejadian puting beliung - Data radar cuaca sebelum dan sesudah kejadian	Analisis reflektifitas dengan produk *CMAX*	Hasil *CMAX*, diketahui bahwa terdapat dua kejadian pola *bow echo* di Bangakalan dan Pasuruan. Nilai reflektifitas >50 dBz yang diindikasikan sebagai awan cumulonimbus sebagai pemicu puting beliung.

Tabel 2.1 *Penelitian terdahulu tentang angin ribut (lanjutan)*

No	Nama Peneliti	Variabel Penelitian	Metode Penelitian	Hasil Penelitian
3.	Erlya Afriani dan Suyatim (2015)	- Kejadian angin kencang - Angin gradien	Analisis Model *WRF*	Berdasarkan hasil *running* model *WRF* dengan data *Final Analysis* dan *Global Forecasting System* kurang cukup baik dalam menangkap kejadian angin kencang >30 knots, namun menginterpretasikan dengan cukup baik kondisi meteorologis yang mendukung pertumbuhan awan-awan konvektif.
4.	Ray Jeremy Sanger, Charitas Fibriani, dan Yessica Nataliani (2012)	- Kecepatan angin	Analisis *Hardware* Inframerah	Kepekaan aplikasi *hardware* inframerah dalam memantau kecepatan angin menghasilkan nilai pemantauan yang detail pada tiap detik dibandingkan alat pemantauan milik BMKG Wilayah II Ciputat. Pengujian menghasilkan 7908 data rekam kecepatan angin.

Penelitian terdahulu memiliki kesamaan dan perbedaan terhadap penelitian yang dilakukan saat ini. Penelitian terdahulu oleh Emilya Nurjani, Arum Rahayu, dan Febriyan Rachmawati (2013) memiliki kesamaan menggunakan variabel frekuensi kejadian angin ribut, namun pada penelitian saat ini tidak menggunakan data digital penggunaan lahan dan metode yang digunakan juga memiliki perbedaan. Analisis yang sama juga menggunakan data frekuensi kejadian oleh peneliti Linda Fitrotul Muzayanah (2015), akan tetapi metode yang digunakan berbeda dengan penelitian yang dilakukan. Penelitian tersebut menggunakan metode *CMAX*, sedangkan penelitian yang dilakukan saat ini menggunakan *PHMM*.

Penelitian sebelumnya oleh Erlya Afriani dan Suyatim (2015)

memiliki kesamaan menggunakan variabel kejadian angin kencang, perbedaannya terletak pada metode yang digunakan, serta pada penelitian saat ini tidak menggunakan angin gradien. Analisis kecepatan angin oleh Ray Jeremy Sanger, Charitas Fibriani, dan Yessica Nataliani (2012) mempunyai perbedaan pada penelitian saat ini. Pada analisis tersebut menggunakan metode *Hardware* inframerah dalam memantau kecepatan angin.

2.2 *Hidden Markov Models* dan *Expectation Maximization Algorithm*

Pada **Tabel 2.2** menunjukkan penelitian-penelitian terdahulu yang berkaitan dengan metode *Hidden Markov Models* (*HMM*) dan *Algoritma EM* (*Expectation Maximization*).

Tabel 2.2 *Penelitian terdahulu tentang HMM dan Algoritma EM*

No	Nama Peneliti	Variabel Penelitian	Metode Penelitian	Hasil Penelitian
1.	Sri Mulyana, Afiahayati, dan Wijaya Adhi Surya (2008)	- Data *sequence* protein globin.	*Hidden Markov Models* dan *Baum-Welch*	*Clustering sequence* protein globin belum optimal, disebabkan arsitektur *HMM* belum optimal, belum memperhatikan dan mengatasi masalah pada prosedur *training Baum-Welch*, seperti *local* minima dan *overfitting*.
2.	Hendra Gustra (2014)	- Data kedatangan klaim perhari	*PHMM* dan Algoritma *EM*	Diperoleh model terbaik *Poisson Hidden Markov Models* 3 *state* menurut *AIC* dengan dugaan rata-rata klaim yang datang perhari sebanyak 3.236 dan *Poisson Hidden Markov Models* 2 *state* menurut *BIC* dengan
3.	Aprilia Fitriatul Aisyah (2015)	- Data rekam suara penderita tuna laring (data.wav)	*Hidden Markov Models*	Hasil akhir yang diperoleh berupa gelombang sinus hubungan antara gain yang bersatuan *decibel* (dB) dan frekuensi (Hz). Selain itu, juga diperoleh nilai *MSE* dan *SNR*, sebesar 8,2357 dan 36,0305dB.

Tabel 2.2 *Penelitian terdahulu tentang HMM dan Algoritma EM (lanjutan)*

No	Nama Peneliti	Variabel Penelitian	Metode Penelitian	Hasil Penelitian
4.	Agung Setiyoaji, Lailil Muflikhah, dan M. Ali Fauzi (2017)	- nama, tempat, zat, dan kegunaan dari teks tanaman obat	*Hidden Markov Models* dan Algoritma Viterbi	Secara keseluruhan pengenalan entitas nilai yang paling rendah pada skenario ke tiga pada entitas nama dengan *f-measure* 0.4100, dan yang paling tinggi pada skenario pertama pada entitas guna dengan *f- measure* 0.7255.
5.	Rosyid Suryandaru (2015)	- Data banyaknya gempa bumi	*PHMM* dan Algoritma *EM*	Diperoleh 3 model *PHMM*, yaitu model dengan keadaan tersembunyi m = (2,3,4). Model dengan 3 keadaan tersembunyi merupakan model estimasi banyaknya gempa bumi terbaik dengan nilai *AIC* sebesar 537,429.
6.	Sri Wahyuni Mamonto, Yohanes A. R. Langi, dan Altien J. Rindengan (2016)	- Data harga saham harian dengan periode satu minggu, satu bulan, dan satu tahun	*Hidden Markov Models* dan Algoritma Viterbi	*HMM* menunjukkan peluang kenaikan harga saham yang paling tinggi untuk satu minggu dan satu bulan, yaitu PT. Bank Mandiri. Pada Algoritma Viterbi, kemungkinan harga saham dari bank PT. Bank BNI Tbk, PT. Bank BRI Tbk, PT. Bank BTN Tbk, dan PT. Bank Mandiri Tbk akan lebih banyak turun.

Penelitian sebelumnya memiliki kesamaan dan perbedaan terhadap penelitian yang dilakukan saat ini. Analisis menggunakan metode *Hidden Markov Model* dengan estimasi *Baum-Welch* dilakukan oleh Sri Mulyana, Afiahayati, dan Wijaya Adhi Surya (2008) dengan menggunakan data *sequence* protein globin. Penelitian oleh Hendra Gustra (2014) memiliki kesamaan pada metode yang digunakan, yaitu *PHMM* dan Algoritma *EM*, namun pada penelitian tersebut menggunakan *software Mathematica* dan objek yang digunakan juga berbeda.

Analisis menggunakan metode *HMM* juga dilakukan oleh Aprilia Fitriatul Aisyah (2015) dengan objek yang digunakan menggunakan data rekam dalam bentuk *wav*. Penelitian sebelumnya oleh Agung Setiyoaji, Lailil Muflikhah, dan M. Ali Fauzi (2017) memiliki perbedaan pada metode estimasi. Penelitiannya menggunakan estimasi Algoritma Viterbi, sedangkan penelitian yang dilakukan saat ini menggunakan estimasi Algoritma *EM*.

Rosyid Suryandaru (2015) memiliki perbedaan yang terletak pada variabel penelitian. Pada penelitian tersebut mencari model estimasi banyaknya gempa bumi. Analisis menggunakan *HMM* dengan estimasi Algoritma Viterbi pada data harga saham menunjukkan peluang kenaikan harga saham yang paling tinggi untuk satu minggu dan satu bulan. (Sri Wahyuni Mamonto, Yohanes A. R. Langi, dan Altien J. Rindengan (2016).

BAB III
LANDASAN TEORI

3.1. Konsep Dasar Peluang

Statistik merupakan suatu ilmu yang berhubungan dengan analisis data dan proses pengambilan keputusan mengenai sistem dari data yang diperoleh. Cabang utama statistik adalah probabilitas (peluang) dan statistik induktif (*inferential statistics*). Tugas statistika baru dianggap selesai jika berhasil membuat kesimpulan dengan baik dan dapat dipertanggungjawabkan tentang sifat atau karakteristik populasi. Jelas bahwa kesimpulan yang dibuat, kebenarannya tidaklah pasti secara absolute, sehingga timbul persoalan bagaimana keyakinan untuk mempercayai kebenaran kesimpulan yang dibuat. Untuk ini diperlukan teori yang disebut teori peluang. (Sudjana, 1982)

Mengundi dengan sebuah dadu, menghitung banyak barang cacat yang diproduksi tiap hari, mencatat banyak kendaraan yang masuk ke parkiran setiap jam merupakan eksperimen yang dapat diulangi. Dari eksperimen demikian semua hasil yang mungkin terjadi bisa dicatat. Segala bagian yang mungkin didapat dari hasil ini dinamakan dengan kejadian. Dalam menyatakan kejadian digunakan huruf-huruf besar, A, B, C, … baik disertai indeks maupun tidak. Sebagai contoh, mata uang logam mempunyai dua muka yang berlainan, misal muka A dan muka B. Kemudian, dilakukan undian dengan sebuah mata uang, sehingga muncul muka-muka yang nampak. Maka kejadian-kejadian muka A dan muka B yang Nampak merupakan dua kejadian yang saling eksklusif. Dari contoh tersebut maka definisi peluang adalah dua kejadian atau lebih dinamakan saling eksklusif jika terjadinya kejadian yang satu mencegah terjadinya yang lain.

Definisi klasik untuk peluang, misalkan sebuah kejadian E dapat terjadi sebanyak n kali di antara N kejadian yang saling eksklusif dan masing-masing terjadi dengan kesempatan yang sama. Maka peluang kejadian E terjadi adalah $\frac{n}{N}$ dan disingkat dengan:

$$P(E) = \frac{n}{N} \tag{3.1}$$

dengan :

P(E) = Peluang munculnya suatu kejadian E

n = Banyaknya anggota dalam kejadian E

N = Banyaknya anggotaseluruh kejadian

Dari definisi, jelas bahwa paling kecil n = 0, yakni dalam hal kejadian E tidak ada, dan paling besar n = N, yakni dalam hal semua yang terjadi merupakan kejadian E, sehingga didapat:

$$0 \leq P(E) \leq 1 \tag{3.2}$$

Jika P(E) = 0, maka diartikan kejadian E pasti tidak terjadi, sedangkan jika P(E) = 1 diartikan kejadian E pasti terjadi. Selanjutnya, dari definisi (3.1), jika $\overline{E}$ menyatakan bukan kejadian E, maka didapat:

$$P(\overline{E}) = 1 - P(E) \tag{3.3}$$

dengan :

$P(\overline{E})$ = Peluang bukan kejadian E

Kejadian E dan $\overline{E}$ juga merupakan dua kejadian yang saling eksklusif. (Sudjana, 1982)

3.1.1 Ekpektasi

Sebuah eksperimen yang menghasilkan k buah kejadian dapat terjadi, dimana peluang terjadinya tiap kejadian masing-masing p_1, p_2, …, p_k dan setiap kejadian dengan peluang tersebut terdapat satuan-satuan d_1, $d_{2, \ldots,}$ d_k. Satuan-satuan tersebut bisa bernilai nol, positif maupun negatif. Maka ekspektasi didefinisikan:

$$E = p_1d_1 + p_2d_2 + \ldots + p_kd_k \tag{3.4}$$

dengan :

E = Ekspektasi

p_i = Peluang tiap kejadian (i = 1, 2, …, k)

d_i = Satuan (Sudjana, 1982).

Contoh 1.1

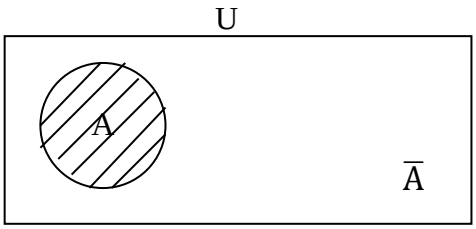
Produksi semacam barang cacat sebesar 6%. Diambil sebuah sampel acak yang

terdiri atas 50 barang, maka setiap sampel diharapkan rata-rata berisi:
$0.06 \times 50 = 3$ barang cacat.

3.1.2 Kejadian-kejadian

Menurut William dan Douglas, sebuah kejadian misalkan A, dihubungkan dengan ruang sampel percobaan. Jika ruang sampel dianggap menjadi himpunan yang universal, maka kejadian A menjadi himpunan bagian dari ruang sampel. Sebagai ketentuan umum, huruf besar akan menunjukkan sebuah kejadian. Berikut ini kejadian-kejadian yang dapat dijelaskan untuk masing-masing kasus, antara lain: koin dilemparkan ke atas menghasilkan sebuah gambar {G}, jumlah produk yang rusak dilas tidak melebihi 5 {0, 1, 2, 3, 4, 5}, jumlah permukaan yang "muncul" adalah tujuh {(1,6), (2,5), (3,4), (4,3), (6,1), (5,2)}.

Diagram Venn dapat digunakan untuk menjelaskan operasi himpunan tertentu. Penggambaran dengan empat persegi panjang ini untuk menunjukkan himpunan universal (U).



Gambar 3.1 *Sebuah himpunan dalam sebuah diagram Venn*

Pada Gambar 3.1 menjelaskan bahwa A himpunan bagian dari U yang ditunjukkan oleh daerah dalam lingkaran yang digambar di dalam empat persegi panjang. Sedangkan, $\overline{A}$ ditunjukkan oleh daerah di dalam empat persegi panjang sebelah luar dalam lingkaran.

3.1.3 Ruang Sampel Diskrit

Percobaan seringkali dijumpai dimana ruang sampelnya tidak terbatas tetapi berisikan hasil-hasil yang dapat dihitung. Persamaan 3.3 yang digunakan dalam menetapkan probabilitas pada ruang sampel terbatas, harus dimodifikasi untuk diperluas bahwa probabilitas menunjukkan hasil-hasilnya yaitu:

1. $p_i \geq 0$ untuk semua i (i = 1, 2, ….)
2. $\sum_{i=1}^{\infty} p_i = 1$

Jadi, ruang sampel diskrit adalah suatu ruang sampel yang mengandung titik yang berhingga banyaknya atau sederetan anggota yang banyaknya sebanyak bilangan bulat. (Ronald dan Raymond, 1972)

3.2 Variabel Acak

Jika dalam ruang sampel mengandung titik sampel yang berhingga banyaknya atau banyaknya sama dengan banyaknya bilangan bulat, maka variabel acak yang dapat mendefinisikannya pada ruang sampel tersebut adalah variabel acak diskrit. Peubah acak atau variabel acak merupakan suatu fungsi yang mengaitkan suatu bilangan *real* pada setiap unsur dalam ruang sampel. Variabel acak dinyatakan dengan huruf besar, sementara nilainya dinyatakan dengan huruf kecil.

Contoh 1.2

Bila tiga sukucadang elektronik diuji dan ruang sampel yang menunjukkan hasilnya adalah sebagai berikut :

S = {BBB, BBC, BCB, CBB, BCC, CBC, CCB, CCC}

B menyatakan sukucadang elektronik baik dan C menyatakan sukucadang elektronik cacat Jika ingin mengetahui berapa banyaknya sukucadang yang cacat, maka setiap titik pada ruang sampel akan dikaitkan dengan suatu bilangan 0, 1, atau 2. Bilangan tersebut tentunya besaran acak yang ditentukan oleh hasil eksperimen. Bilangan ini juga dipandang sebagai nilai yang dicapai oleh sebuah peubah acak, X, banyaknya barang elektronik yang cacat. (R Ronald dan Raymond, 1972)

3.2.1 Konsep Variabel Acak Diskrit

Notasi yang digunakan untuk variabel acak adalah huruf kapital, misalnya X. Himpunan seluruh nilai yang mungkin dalam X merupakan kisaran (*range*) X. *Range* dari variabel acak dinotasikan dengan huruf S dengan *subscript* yang merupakan nama dari variabel acak. Sebagai contoh, S_x adalah *range* dari variabel acak X. Penggunaan S_x untuk menotasikan *range* X disebabkan karena himpunan seluruh nilai yang mungkin dari X adalah analog dengan S, yaitu himpunan dari seluruh *outcome* yang mungkin dalam eksperimen. (Agustinah, 2014)

X disebut variabel acak diskrit jika *range* dari X adalah himpunan yang dapat dihitung, dengan ruang sampel $S_x = \{x_1, x_2, x_3, \ldots\}$. Sebaliknya, variabel acak Y yang dapat dinyatakan pada setiap bilangan *real* dalam interval $a \leq y \leq b$ disebuk dengan variabel acak kontinu.

Bila *range* dari variabel acak X adalah terbatas, $S_x = \{x_1, x_2, x_3, \ldots, x_n\}$ disebut variabel acak diskrit terbatas. Variabel acak diskrit biasanya memuat nilai-nilai *integer*, meskipun dalam beberapa kasus tertentu dapat bernilai bukan *integer*.

3.2.2 Fungsi Massa Probabilitas Variabel Acak Diskrit

Model probabilitas yang menugaskan bilangan antara 0 dan 1 untuk tiap *outcome* bernilai diskrit dari eskperimen. Model probabilitas untuk variabel acak diskrit ini dideskripsikan sebagai fungsi massa probabilitas dalam *range* seluruh bilangan *real*.

Fungsi massa probabilitas (*probability mass function*-PMF) didefinisikan sebagai $P_X(x) = P(X = x)$. Pada variabel acak, huruf besar (X) menyatakan nama variabel dan huruf kecil (x) digunakan untuk nilai yang mungkin dalam variabel tersebut.

PMF berisi seluruh informasi tentang variabel acak X. Karena $P_X(x)$ adalah probabilitas dari *event* $\{X = x\}$, maka $P_X(x)$ mempunyai beberapa sifat penting yang diturunkan dari aksioma probabilitas untuk variabel acak diskrit. Sifat-sifat PMF sebagai berikut:

1. $P_X(x) \geq 0, \forall x$

   PMF variabel acak diskrit selalu bernilai tak negatif.

2. $\sum_{x \in S_X} P_X(x) = 1$

   Jumlah PMF dari variabel acak X sama dengan 1. (Agustinah, 2014)

3.2.3 Fungsi Distribusi Kumulatif Variabel Acak Diskrit

Deskripsi model probabilitas untuk variabel acak diskrit dapat ditunjukkan melalui fungsi distribusi kumulatif. Fungsi ini merupakan penjumlahan probabilitas massa dari tiap nilai dalam variabel acak tersebut. Fungsi distribusi

kumulatif (*cumulative distribution function*-CDF) variabel acak X didefinisikan sebagai probabilitas *event* $\{X \leq x\}$:

$$F_X(x) = P(X \leq x) \tag{3.5}$$

dengan :

$F_X(x)$ = Fungsi distribusi kumulatif dari variabel x

$P(X \leq x)$ = Probabilitas *event*

Event $\{X \leq x\}$ dan probabilitasnya bervariasi sesuai dengan nilai x, karenanya $F_X(x)$ merupakan fungsi dari variabel x.

Bila nilai tertentu dalam variabel acak diskrit X dinotasikan sebagai x_i, maka fungsi distribusi kumulatif $F_X(x)$ dapat ditulis sebagai :

$$F_X(x) = \sum_{i=1}^{N} P_X(x_i)u(x - x_i) \tag{3.6}$$

dengan :

$P_X(x_i)$ = Fungsi massa probabilitas (PMF)

u = Fungsi tangga satuan

Sifat-sifat CDF sebagai berikut:

1. $F_X(-\infty) = 0$ dan $F_X(\infty) = 1$

   $F_X(x)$ dimulai dari nol dan berakhir pada nilai satu.

2. Untuk semua $x' \geq x$, $F_X(x') \geq F_X(x)$

   CDF tidak pernah turun, dari kiri ke kanan.

3.2.4 Momen Variabel Acak Diskrit

Selain dinyatakan dengan fungsi probabilitas, variabel acak diskrit dinyatakan juga dalam *moment-moment*nya. Dari *moment* terhadap origin dan *moment* sentral dapat dikembangkan pengukuran karakteristik variabel acak dalam bentuk nilai. Nilai-nilai tersebut adalah *mean* dan varians.

Nilai ekspektasi variabel acak X didefinisikan :

$$E[X] = \sum_{x \in S_X} xP_X(x) \tag{3.7}$$

dengan :

$E[X]$ = Ekspektasi dari variabel acak X

Nilai ekspektasi disebut juga sebagai nilai *mean* dan dinotasikan dengan μ_X. Sedangkan, varians dari variabel acak :

$$\sigma^2{}_X = E[(X - \mu_X)^2] \tag{3.8}$$

dengan :

$\sigma^2{}_X$ = Varians dari variabel acak X

$(X - \mu_X)^2$ = Fungsi X

Akar dari varians (σ_X) disebut standar deviasi dari X. Nilai ini adalah ukuran sebaran variabel acak X dalam fungsi kepadatan terhadap nilai *mean*.

$$\sigma_X = \sqrt{var[X]} \tag{3.9}$$

dengan :

σ_X = Standar deviasi dari X

3.3 Model Fungsi Variabel Acak Diskrit

Model fungsi probabilitas untuk variabel acak diskrit salah satunya adalah model *Poisson*. Model *Poisson* banyak digunakan dalam aplikasi perhitungan (*counting*), misalnya berapa jumlah unit cacat dalam produksi elektronik pada *shift* pertama, jumlah panggilan telepon dalam tiap jam, dan sebagainya.

3.3.1 Model *Poisson*

Fungsi massa probabilitas (PMF) menyatakan probabilitas terjadinya X sebanyak k dalam selang waktu tertentu didefinisikan dengan:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \; k = 0, 1, 2, \ldots \tag{3.10}$$

dengan :

λ = rata-rata banyaknya kejadian tiap satuan waktu

k = banyaknya kejadian per unit

e = 2.71828…

dengan λ >0 merupakan *rate* banyaknya kejadian tiap satu satuan waktu. Fungsi distribusi CDF dari X didefinisikan sebagai :

$$F_X(x) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} u\,(x - k) \tag{3.11}$$

Model *Poisson* untuk variabel acak X memiliki *mean* :

$$E[X] = \lambda \tag{3.12}$$

dan varians :

$$var[X] = \lambda \tag{3.13}$$

3.4 Proses *Poisson*

Menurut Ronald dan Raymond (1972), percobaan *Poisson* adalah suatu percobaan yang menghasilkan peubah atau variabel X yang memiliki nilai numerik, yang merupakan banyaknya hasil dalam selang waktu tertentu. Selang waktu tersebut memiliki panjang waktu semenit, sejam, sehari, seminggu, sebulan, ataupun setahun. Sebagai contoh, pengamatan untuk peubah acak X menyatakan bahwa banyaknya suatu panggilan telepon perjam yang diterima suatu kantor, atau banyaknya pertandingan sepakbola yang diundur karena hujan selama musim hujan.

Proses *Poisson* memiliki beberapa sifat seperti berikut :

1. Proses *Poisson* dapat dikatakan tak punya ingatan, maksudnya adalah banyaknya hasil yang terjadi dalam selang waktu tertentu tidak terpengaruh pada apa yang terjadi dalam selang waktu lain yang terpisah.
2. Peluang untuk terjadinya hasil dalam selang waktu yang singkat sebanding dengan panjang selang waktu dan tidak tergantung dengan banyaknya hasil yang terjadi di luar selang waktu tersebut.
3. Peluang untuk lebih dari satu hasil yang terjadi dalam selang waktu yang singkat tersebut dapat diabaikan.

Banyaknya hasil pada peubah acak X dalam suatu percobaan *Poisson* disebut dengan peubah acak *Poisson* dan distribusi dari peluangnya adalah distribusi *Poisson*. Distribusi Poisson juga disebut distribusi yang melibatkan jumlah percobaan (n) yang besar dan peluang sukses (p) yang kecil. Rataan (μ) dari banyaknya hasil adalah λ_t, dengan t menyatakan waktu.

3.5 Overdispersi

Pada distribusi *Poisson* terdapat asumsi yang harus dipenuhi, yaitu asumsi kesamaan rataan dan varians dari variabel acak atau disebut dengan ekuidispersi.

Sedangkan, dalam analisis data diskrit sering dijumpai nilai varians dari variabel acak lebih besar dari nilai rataan. Keadaan ini disebut dalam keadaan overdispersi.

Ada beberapa hal yang menyebabkan terjadinya overdispersi, diantaranya adanya pengamatan yang hilang pada variabel prediktor, adanya pencilan pada data, adanya sumber keragaman yang tidak teramati yang menyebabkan rataan dalam suatu populasi berbeda-beda antar observasi. Hal ini menunjukkan bahwa populasi tersebut bersifat heterogen dan bergantung pada parameter μ_i. Pelanggaran terhadap asumsi tersebut mengakibatkan taksiran parameter yang diperoleh dari distribusi *Poisson* menjadi tidak efisien walaupun tetap konsisten, dan memberikan taksiran yang tidak sesuai. Data yang mengalami overdispersi dapat diminimalisir atau diatasi menggunakan beberapa model salah satunya yaitu model campuran (*Mixture Model*).

3.6 *Independent Mixture Models*

Distribusi *Poisson* yang memiliki fungsi massa peluang seperti pada persamaan 3.10 mempunyai ciri khas bahwa bahwa nilai rataan sama dengan nilai variansi. Namun, ketika nilai variansi sampel (σ^2) lebih besar dari rata-rata sampelnya ($\bar{x}$) maka akan terjadi overdispersi relatif pada distribusi *Poisson* dan ketidaktepatan distribusi sebagai model. Dalam hal ini, diperlukan suatu metode untuk mengatasi pengamatan overdispersi, yaitu dengan menggunakan *Mixture Model* (model campuran).

Mixture Models merupakan suatu metode untuk mengakomodasi heterogenitas yang tidak teramati (*unobserved*) pada populasi, dimana populasi dimungkinkan terdiri dari kelompok-kelompok yang tidak teramati. Sebagai contoh, misalkan Z adalah distribusi banyaknya bungkus rokok yang dibeli oleh pelanggan di supermarket. Para pelanggan tersebut dapat dibagi menjadi beberapa kelompok, misalkan dibagi menjadi dua kelompok yaitu perokok pasif dan perokok aktif. Jika banyaknya bungkus rokok yang dibeli oleh masing-masing kelompok pelanggan berdistribusi *Poisson*, maka distribusi Z belum tentu berdistribusi *Poisson* dan memungkinkan terjadinya overdispersi relatif pada distribusi tersebut. Anggaplah masing-masing kelompok berdistibusi *Poisson*

memiliki rata-rata λ_1 dengan perluang δ_1 dan λ_2 dengan peluang $\delta_2 = 1 - \delta_1$, dimana dalam menentukan nilai rata-rata tersebut didasarkan pada beberapa mekanisme acak lain yang disebut "proses parameter". Jika proses parameter dimana serangkaian variabel acaknya saling bebas (*independent*) dan banyaknya juga saling bebas maka proses tersebut disebut "*mixture* independen".

Secara umum, distribusi campuran yang saling bebas (*independent mixture distribution*) terdiri dari sejumlah bilangan/komponen yang berhingga (m komponen) dan distribusi pencampuran yang dipilih dari komponen tersebut. Misalkan δ_1, …, δ_m adalah peluang pada masing-masing m komponen dan misalkan p(1), …, p(m) adalah peluang fungsi densitas masing-masing komponen tersebut. Misalkan X menyatakan variabel acak yang memiliki disribusi campuran. Peluang fungsi densitas X didefinisikan sebagai :

$$p(x) = \sum_{i=1}^{m} \delta_i \, p_i \, (x) \tag{3.14}$$

dengan :

δ = Peluang pada tiap-tiap komponen

p = Peluang fungsi densitas

Sehingga untuk kasus diskrit , peluang fungsi untuk denstitas X adalah :

$$\Pr (X=x) = \sum_{i=1}^{m} \Pr(X=x \mid C = i) \; \Pr (C=i) \tag{3.15}$$

dengan :

C = Himpunan variabel acak

3.7 Pengantar Stokastik

Konsep proses stokastik didasarkan pada perluasan konsep variabel stokastik dengan memasukkan waktu. Karena variabel stokastik X berdasarkan definisinya merupakan fungsi dari *outcome* yang mungkin s dari eksperimen, maka proses stokastik menjadi fungsi s dan waktu. Dengan kata lain, fungsi waktu $x(t, s)$ untuk setiap *outcome* s. Keluarga dari seluruh fungsi ini dinotasikan $X(t, s)$ disebut proses stokastik. Proses stokastik $X(t, s)$ mempresentasikan suatu ansambel dari fungsi waktu bila t dan s variabel. Setiap anggota fungsi waktu disebut fungsi sampel atau seringkali disebut dengan realisasi dari proses. Jadi,

proses stokastik juga mempresentasikan fungsi sampel bila t adalah variabel dan s tetap pada nilai tertentu (*outcome*).

Menurut Praptono (1986), proses stokastik didefinisikan sebagai himpunan variabel acak yang merupakan fungsi waktu atau disebut *random process*. Menurut Taylor dan Karlin (1998), kumpulan peubah acak X(t) dengan t menunjukkan elemen-elemen pada himpunan T yang berkoresponden terhadap waktu.

Kumpulan peubah acak yang memiliki indeks (X_t) pada suatu himpunan T dimana anggota-anggotanya berkoresponden terhadap nilai waktu (Paris, 2011).

Proses stokastik juga mempresentasikan variabel stokastik bila t adalah tetap dan s variabel. Sebagai contoh, variabel stokastik $X(t_1,s) = X(t_1)$ diperoleh dari proses bila waktu dipertahankan pada nilai t_1. Seringkali digunakan notasi X_1untuk menotasikan variabel stokastik yang dihubungkan dengan proses $X(t)$ pada waktu t_1. X_1berhubungan dengan irisan secara vertikal dari seluruh ansambel pada waktu t_1. Sifat-sifat statistik dari $X_1 = X(t_1)$ mendeskripsikan sifat-sifat statistik dari proses stokastik pada waktu t_1. Nilai ekspektasi dari X_1 ini disebut rata-rata ansambel atau nilai *mean* dari proses stokastik (pada waktu t_1).

3.8 Proses Rantai *Markov*

Menurut Papoulis (1992), proses rantai *markov* merupakan proses stokastik dimana kejadian masa datang hanya dipengaruhi oleh masa sekarang dan tidak dipengaruhi oleh masa lalu. Secara sistematis rantai markov dinyatakan sebagai berikut:

$$P(X_{n+1} = j \mid X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n) = P(X_{n+1} = j \mid X_n = i_n) \qquad (3.16)$$

dengan :

X_{n+1} = Keadaan yang akan datang

X_n = Keadaan sekarang

Persamaan 3.16 merupakan *Markovian Property* atau sifat utama rantai *markov*.

3.9 Rantai Markov Diskrit

Analisis rantai Markov (*Markov Chain Analysis*) merupakan suatu teknik peluang yang menganalisis pergerakan peluang dari suatu keadaan ke keadaan

lainnya. Proses stokastik $\{C_t : t \in N\}$ dikatakan sebuah rantai Markov dengan waktu diskrit jika untuk setiap $t \in N$ berlaku :

$$Pr(C_{t+1} \mid C_t, C_{t-1}, \ldots, C_1) = Pr\,(C_{t+1} \mid C_t) \quad (3.17)$$

dengan :

C = Kejadian

Definisi tersebut dapat diinterpretasikan bahwa untuk suatu rantai Markov, sebaran bersyarat dari sembarang keadaan yang akan datang $\{C_{t+1}\}$ dengan syarat keadaan pada masa lalu adalah $\{C_{t-1}, \ldots, C_1\}$ dan keadaan sekarang adalah $\{C_t\}$, adalah bebas terhadap semua keadaan yang lalu dan hanya bergantung dari keadaan sekarang (kita definisikan $C^{(t)}$ sebagai kejadian $\{C_1, C_2, \ldots, C_t\}$. Persamaan (3.17) dapat ditulis sebagai berikut:

$$Pr(C_{t+1} \mid C^{(t)}) = Pr\,(C_{t+1} \mid C_t) \quad (3.18)$$

Kuantitas terpenting yang terkait dengan rantai Markov adalah peluang bersyarat yang disebut dengan peluang transisi (*transition probabilities*) dimana pada kasus rantai Markov homogen persamaannya adalah sebagai berikut:

$$\gamma_{ij} = Pr\,(C_{t+1} = j \mid C_n = i) \quad (3.19)$$

untuk semua i, j, n = {1,2,…}. Nilai y_{ij} merupakan elemen dari ⌈ (1), yaitu elemen dari matriks peluang transisi satu langkah (*one-step transition probabilities matrix*). Kita definisikan ⌈ (1)= ⌈. Nilai ⌈ menyatakan peluang bahwa jika proses rantai Markov berada pada keadaan i, maka keadaan berikutnya ke keadaan j setelah satu langkah. Misal, didefinisikan ⌈ (1) = ⌈. Nilai γ_{ij} menunjukkan peluang bahwa ketika proses rantai *Markov* berada pada keadaan i, maka keadaan berikutnya memenuhi syarat-syarat berikut:

1. $\gamma_{ij} \geq 0$, untuk semua i, j = {1, 2, …, m}
2. $\sum_{j=1}^{m} \gamma_{ij} = 1$, untuk semua i, = {1, 2, …, m}

   dimana m menyatakan banyaknya keadaan pada rantai *Markov* dengan jumlah masing-masing baris pada matriks ⌈ = 1.

Pada peluang transisi t langkah, yaitu peluang bahwa suatu proses yang pada awalnya berada pada keadaan i akan berpindah pada keadaan j setelah t langkah yang didefinisikan sebagai berikut:

$$\gamma_{ij}\ (t)\ =\ \Pr(C_{s+t} = j \mid C_s = i) \tag{3.20}$$

untuk semua i, j, s, t ∈ N. Nilai $\gamma_{ij}\ (t)$ merupakan elemen matriks ⌈ (t) (*t-step transition probabilities matrices*). Berdasarkan persamaan *Chapman Kolmogorov* matriks peluang transisi t langkah dapat ditentukan dengan mengalikan matriks ⌈ sebanyak t kali dengan persamaan sebagai berikut:

$$⌈ (t + u) = ⌈ (t) . ⌈ (u) \text{ , untuk semua t, u} \in N \tag{3.21}$$

dengan :

⌈ = Matriks peluang transisi

Persamaan *Chapman Kolmogorov* mengimpilkasikan bahwa

$$⌈ (t) = ⌈ (1)^t, \text{ untuk semua t} \in N \tag{3.22}$$

Sejauh ini, semua peluang yang ditentukan merupakan peluang bersyarat. Misalnya ⌈ (t) adalah peluang bahwa pada waktu ke t proses berada pada keadaan j dengan syarat keadaan mula-mula adalah i. Pada sebaran tidak bersyarat rantai Markov, nilai peluang sebaranya ditentukan oleh sebaran peluang pada keadaan awal (*initial state*). Peluang tidak bersyarat (*unconditional probabilities*) $\Pr(C_t = j)$ dari rantai Markov berada pada keadaan tertentu pada waktu ke t didefinisakan sebagai vektor baris sebagai berikut:

$$u(t) = \Pr(C_t = 1), \ldots, \Pr(C_t = m)), \ t \in N \tag{3.23}$$

Sebagai contoh, misalkan peluang besok akan terjadi hujan bergantung pada keadaan cuaca hari ini, serta tidak bergantung pada keadaan cuaca sebelumnya (sifat rantai Markov). Misalkan diberikan peluang transisi satu langkah rantai Markov dengan dua keadaan, dimana keadaan jika cuaca hujan = 1 dan jika cuaca cerah = 2 dengan peluang transisinya ditunjukkan pada tabel berikut:

Tabel 3.1 *Peluang Keadaan Cuaca*

Hari ke-t	t+1	
	Hujan	Cerah
Hujan	0.8	0.2
Cerah	0.4	0.6

Tabel di atas menjelaskan bahwa peluang cuaca besok akan hujan sebesar 0,8 dengan syarat jika cuaca hari ini hujan dan sebesar 0,2 jika cuaca hari ini adalah cerah. Matriks peluang transisi ⌈ untuk memodelkan kondisi di atas adalah:

$$\lceil = \begin{Vmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{Vmatrix}$$

Jika cuaca hari ini (t = 1) adalah cerah, maka peluang tidak bersyarat cuaca hari ini adalah

$$u\,(1) = \Pr(C_1 = 1), \Pr(C_1 = 2)) = (0\ ,\ 1)$$

dan peluang cuaca untuk t = {2,3, … }, yaitu peluang cuaca pada hari-hari berikutnya adalah

$$u\,(2) = \Pr(C_2 = 1), \Pr(C_2 = 2)) = u\,(1)\ .\ \lceil = (\ 0.4\ ,\ 0.6)$$

Pada rantai *Markov*, setelah proses *Markov* berjalan dalam jangka yang panjang (t $\rightarrow \infty$), peluang yang dihasilkan akan bernilai tetap, dengan kata lain akan menuju stasioner. Hal ini berarti, proses *Markov* memiliki peluang yang konvergen ke suatu nilai yang sama, dimana proses tersebut tidak lagi bergantung pada keadaan awal (u (1)). Rantai Markov dengan matriks peluang transisi ⌈, dikatakan memiliki distribusi yang stasioner δ (vektor baris dengan elemen tak negatif) jika:

$$\delta\lceil = \delta \text{ dan } \sum_{i=1}^{m} \delta_i = 1 \qquad (3.24)$$

3.10 Prediksi Peluang dengan Pendekatan Matriks Transisi

Keadaan transisi adalah perubahan dari suatu keadaan (status) ke keadaan (status) lainnya pada periode berikutnya. Keadaan transisi dinyatakan dalam bentuk probabilitas, probabilitas ini dikenal sebagai probabilitas transisi. Probabilitas dapat digunakan untuk menentukan probabilitas keadaan atau periode berikutnya.

Probabilitas diasumsikan dengan 1 dan 0, kemudian disunsun ke dalam vektor baris sehingga terbentuk $(1 \quad 0)$. Adapun rumus untuk menghitung probabilitas periode berikutnya (i+1) adalah :

$$(\mathrm{N}(\mathrm{i}+1) \quad \mathrm{M}(\mathrm{i}+1)) = (\mathrm{N}(\mathrm{i}) \quad \mathrm{M}(\mathrm{i})) \times \text{Matriks Probabilitas Transisi} \qquad (3.25)$$

dengan :

N = Suatu keadaan/status

M = Keadaan/status lainnya

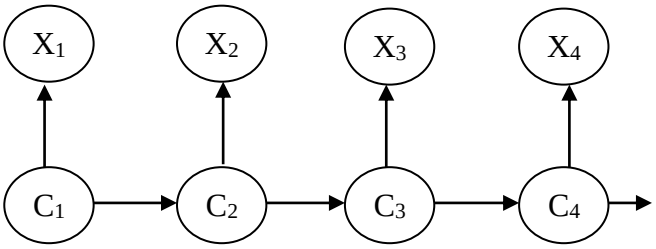
i = 1, 2, ….

3.11 *Poisson Hidden Markov Models* (*PHMM*)

Poisson Hidden Markov Models (*PHMM*) yang merupakan perluasan dari *Hidden Markov Models* ini, untuk setiap observasinya dihasilkan oleh salah satu dari m keadaan tersembunyi (*hidden state*) yang berdistribusi *Poisson*. Dalam memahami *PHMM*, diperlukan penjelasan mengenai dasar teori *Hidden Markov Model* (*HMM*), distribusi marginal dan *moment HMM*, serta *likelihood HMM*.

3.11.1 *Hidden Markov Models* (*HMM*)

Hidden Markov Model merupakan sebuah proses stokastik yang terdiri dari bagian yang tidak teramati (*unobserved*) $\{ C_t : t \in N \}$ dimana memenuhi sifat Markov (*parameter process*), serta bagian yang teramati/terobservasi $\{X_t : t \in N\}$, atau disebut sebagai "proses keadaan bergantung" (*state-dependent process*). Pada bagian yang terobservasi, distribusi X_t hanya bergantung pada kondisi saat C_t dan tidak bergantung pada keadaan terobservasi sebelumnya. *Hidden Markov Model* dapat direpresentasikan pada **Gambar 3.2**.



Gambar 3.2 *Graf Dasar HMM*

Hidden Markov Model $\{X_t : t \in N\}$ merupakan distribusi campuran yang bergantung, dengan $X^{(t)}$ dan $C^{(t)}$ mewakili kejadian masa lalu dari waktu 1 hingga waktu t, yang disimpulkan model sederhana tersebut dengan persamaan :

$$\Pr(C_t \mid C^{(t-1)}) = \Pr(C_t \mid C_{t-1}),\ t = 2, 3, \ldots \quad (3.26)$$

$$\Pr(X_t \mid X^{(t-1)}, C^{(t)}) = \Pr(X_t \mid C_t),\ t \in N \quad (3.27)$$

Jika rantai *Markov* $\{C_t\}$ mempunyai m keadaaan tersembunyi, kita katakan $\{X_t\}$ adalah *HMM* dengan m keadaaan.

Pada kasus pengamatan diskrit, jika rantai Markov pada waktu t berada pada keadaan i maka fungsi masa peluang (*pmf*) X_t didefinisikan sebagai[5]:

$$pi(x) = \Pr(X_t = x \mid C_t = i),\ i = 1, 2, \ldots, m \quad (3.28)$$

Distribusi p_i dengan m keadaan tersembunyi dapat dikatakan sebagai distribusi keadaan bergantung (*state-dependent distributions*).

3.12 Distribusi Marginal dan Moment *HMM*

Pada kasus stasioner, nilai harapan bergantung (*state-dependent*) $E(g(X_t))$ dan $E(g(X_t, X_{t+k}))$ untuk setiap fungsi g adalah sebagai berikut :

$$E(g(X_t)) = \sum_{i=1}^{m} \delta_i E\,(g(X_t)|C_t = i) \tag{3.29}$$

dan

$$E(g(X_t, X_{t+k})) = \sum_{i,j=1}^{m} E(g(X_t, X_{t+k}|C_t = i, C_{t+k} = j))\delta_i \gamma_{ij}(k)$$

$$= \sum_{i,j=1}^{m} E(g_1(X_t|C_t = i))\, E(g_2(X_{t+k})|C_{t+k} = j)\delta_i \gamma_{ij}(k) \tag{3.30}$$

Persamaan (3.28) dan Persamaan (3.29) berguna untuk mendapatkan nilai ekspektasi pada *HMM*. Misal, jika terdapat dua keadaan *HMM* dimana rantai markov berdistribusi *Poisson* dan stasioner, maka:

$$E(X_t) = \delta_1 \lambda_1 + \delta_2 \lambda_2 \tag{3.31}$$

3.13 Penskalaan *Likelihood*

Pada kasus distribusi *state-dependent* diskrit, hasil peluang elemen α_t akan semakin kecil nilainya saat t meningkat, dan pada akhirnya dibulatkan menjadi 0, atau biasa disebut *underflow*. Salah satu cara dalam mengatasi masalah *underflow* digunakan penskalaan likelihood, yaitu menghitung nilai log dari likelihood, atau bisa disebut dengan penskalaan vektor peluang *forward* α_t . Didefinisikan sebuah vektor, untuk t = 0,1, … , T adalah:

$\emptyset_t = \alpha_t / w_t$ (3.32)

dengan: $w_t = \sum_i \alpha_t$ (i).

3.14 Estimasi *Expectation Maximization Algorithm* (EM)

Salah satu metode umum yang digunakan untuk mengestimasi parameter-parameter pada *HMM* adalah dengan menggunakan metode Algoritma *EM* (*Estimation Maximization Algorithm*). Dalam konteks *HMM*, algoritma *EM* dikenal sebagai algoritma Baum-Welch, dimana rantai *Markov* pada *HMM* adalah homogen dan tidak diharuskan stasioner. Parameter *HMM* yang diestimasi dengan Algoritma *EM* adalah distribusi keadaan bergantung p_i, matriks peluang transisi,

dan distribusi inisial δ. Dalam penerapannya, Algoritma *EM* membutuhkan perangkat yaitu peluang *forward* dan peluang *backward*, dimana kedua peluang tersebut dapat digunakan untuk prediksi *state*.

1. Peluang *Forward*

   Peluang *forward* α_t untuk t = 1, 2, … , T didefinisikan sebagai vektor baris :

   $$\alpha_t = \delta P(x_1) \lceil P(x_2) \dots \lceil P(x_t) = \delta P(x_1) \prod_{s=2}^{t} \lceil P(x_s) \qquad (3.33)$$

   dengan δ adalah distribusi inisial rantai Markov. Berdasarkan definisi peluang *forward* di atas, untuk t = 1, 2, … , T −1.

2. Peluang *Backward*

   Peluang *backward* β_t untuk t = 1, 2, … , T didefinisikan sebagai vektor baris :

   $$\beta'_t = \lceil P(x_{+1}) \lceil P(x_{+2}) \dots \lceil P(x_T)\, 1' = (\prod_{s=t+1}^{t} \lceil P(x_s)\, 1' \qquad (3.34)$$

   dimana untuk t = T, $\beta_T = 1$. Berdasarkan definisi peluang *backward* di atas, untuk t = 1, 2, … , T −1.

1. Peluang *Forward* dan Peluang *Backward*

   Gabungan antara peluang *forward* dan *backward* α_t dan β_t dapat digunakan untuk menghitung $Pr(X^{(T)} = x^{(T)}, C_t = i)$. Gabungan antar peluang tersebut sangat dibutuhkan dalam *Expectation Maximization Algorithm* pada *HMM*. Pada pengaplikasiannya, dibutuhkan dua sifat diantaranya :

   a. t = 1, 2, …, T,

   $$Pr(C_t = j \mid X^{(T)} = x^{(T)}) = \alpha_t(j)\, \beta_t(j)/L_T \qquad (3.35)$$

   b. t = 2, …, T.

   $$Pr(C_{t-1} = j, C_t = k \mid (X^{(T)} = x^{(T)}) = \alpha_{t-1}(j)\, \gamma_{jk} p_k\, (x_t)\, \beta_t(k)/L_T \qquad (3.36)$$

Barisan keadaan rantai *Markov* yang *unobserved* memungkinkan terdapat data yang hilang, sehingga berakibat pada data yang tidak lengkap. Algoritma *EM* sebagai metode iteratif berfungsi dalam menghitung estimasi maksimum *likelihood* untuk data yang tidak lengkap tersebut. Pada setiap iterasi Algoritma *EM* terdapat E (Ekspektasi) dan tahap M (Maksimisasi).

1. Tahap Ekspektasi

   Mengganti semua nilai v_{jk} dan u_j (t) dengan ekspektasi bersyaratnya jika diberikan observasi $x^{(T)}$.

$$\hat{u}_j\ (t) = Pr(C_t = j \mid x^{(T)}) = \alpha_t(j)\ \beta_t(j)/L_T \quad (3.37)$$

serta

$$\hat{v}_{jk}\ (t) = Pr(C_{t-1} = j, C_t = k \mid x^{(T)}) = \alpha_{t-1}(j)\ \gamma_{jk} p_k\ (x_t)\ \beta_t(k)/L_T \quad (3.38)$$

2. Tahap Maksimisasi

Setelah dilakukan penggantian nilai v_{jk} dan u_j (t) dengan $\hat{u}_j$ (t) dan $\hat{v}_{jk}$, maka dilakukan pemaksimalan *CDLL* dimana berkenaan dengan tiga parameter, yaitu parameter distribusi *state*-dependen ($\lambda_1, \lambda_2, \ldots, \lambda_3$), distribusi inisial δ, dan matriks peluang transisi. *Log-likelihood CDLL* terdapat barisan observasi $x_1, x_2, \ldots, x_T$ serta data hilang $c_1, c_2, \ldots, c_T$ ditunjukkan pada persamaan berikut:

$$\log(Pr(X^{(T)}, c^{(T)}) = \sum_{j=1}^{m} u_j(1)\log \delta_j + \sum_{j=1}^{m}\sum_{k=1}^{m}\left(\sum_{t=2}^{T} v_{jk}\ (t)\right)\log\sum_{j=1}^{m} u_j + \sum_{j=1}^{m}\sum_{t=1}^{T} u_j(t)\log p_j(x_t) \quad (3.39)$$

3.15 Pemilihan Model berdasarkan *Akaike Information Criterion* (*AIC*)

Akaike Information Criterion (*AIC*) diperkenalkan pertama kali oleh Akaike (1974) yang digunakan untuk mengidentifikasikan model dari suatu *dataset*. Penerapan metode ini m pendekatan *Maximum Likelihood*. Persamaan *AIC* dalam melakukan pemilihan model yang dikemukakan oleh Akaike adalah

$$AIC = -2 \log L + 2p \quad (3.40)$$

dengan :

log L = *log-likelihood* pada model

p = jumlah parameter pada model

Misal, jika model diberikan 2 keadaan tersembunyi m = 2, maka model tersebut memiliki parameter sebanyak $p = m^2 + (m - 1) = 5$. Hal ini dikarenakan ⌈ memiliki parameter sebanyak m^2, δ sebanyak (m – 1), dan λ sebanyak m.

dengan :

m = banyaknya *state*

3.16 Definisi Angin

Menurut Buys Ballot, ahli ilmu cuaca dari Perancis, angin adalah massa udara yang bergerak dari daerah bertekanan maksimum ke daerah bertekanan

minimum. Gerakan massa udara yang arahnya horizontal dikenal dengan istilah angin. Anemometer mangkok adalah alat yang digunakan untuk mengukur kecepatan angin. Satuan yang biasa digunakan dalam menentukan kecepatan angin adalah km/jam atau knot (1 knot = 0,5148 m/det = 1,854 km/jam). Sisteman penamaan angin biasanya dihubungkan dengan arah datangnya massa udara tersebut.

3.16.1 Jenis-jenis Angin

Beberapa jenis angin antara lain:

1. Angin Tetap

   Angin tetap adalah angin dengan arah hembusnya yang tetap di sepanjang tahun. Angin tetap ini terdiri atas :

   a. Angin timur, yaitu angin yang arah hembusnya dari daerah kutub.
   b. Angin barat, yaitu angin yang berhembus dari daerah subtropik.
   c. Angin pasat, yaitu angin yang berhembus dari subtropik selatan dan utara ke daerah khatulistiwa.
   d. Angin anti pasat, yaitu angin yang bertiup dari daerah khatulistiwa ke daerah subtropik selatan dan utara.

2. Angin Periodik

   Angin periodik disebut angin tidak tetap karena angin ini terjadi pada waktu-waktu tertentu dan biasanya relatif tidak lama. Angin periodik terdiri atas :

   a. Angin darat, yaitu angin yang terjadi pada waktu malam hari dengan arah hembusnya dari darat menuju laut.
   b. Angin laut, yaitu angin yang terjadi pada waktu siang hari dengan arah hembusnya dari laut menuju darat.
   c. Angin muson, yaitu angin yang bertiup secara periodik biasanya setengah tahun sekali dan selalu berganti arah secara berlawanan pada setiap setengah tahunnya.
   d. Angin lembah, yaitu angin yang terjadi pada waktu siang hari dengan arah hembusnya dari lembah menuju puncak gunung.
   e. Angin gunung, yaitu angin yang terjadi pada waktu malam hari dengan arah hembusnya dari lereng gunung menuju lembah.

3. Angin Lokal

Angin lokal adalah angin yang bertiup di daerah-daerah tertentu. Angin lokal terdiri atas :

a. Angin fohn, yaitu angin yang bersifat kering dan panas dengan arah tiupannya dari pegunungan. Sebagai contoh, angin brubu di Makassar.
b. Angin siklon, yaitu angin yang berhembus di daerah yang mempunyai minimum serta dikelilingi barometris maksimum.
c. Angin antisiklon, yaitu angin yang bertiup berlawanan dengan angin siklon. (Wikipedia)

3.16.2 Skala Angin *Beaufort*

Pengkategorian kecepatan angin berdasarkan besar nilai kecepatan angin terbagi atas 13 kategori menurut skala *Beaufort*. Semakin besar skala *Beaufort*, maka angin yang bertiup semakin kencang dan bisa semakin merusak. Skala Beaufort dimulai dari angka 0 (nol) dengan tiupan angin yang sangat tenang hingga angka 12 yang menimbulkan kerusakan. Skala *Beaufort* ini dikemukakan oleh armada *Beaufort*, yaitu Francis Beaufort pada tahun 1805. Hingga saat ini skala *Beaufort* tetap berguna dan dipakai sampai sekarang (Tjasyono, 2007). Pengkategorian tersebut disajikan pada **Tabel 3.2.**

Tabel 3.2 *Skala Angin Beaufort*

Skala *Beaufort*	Kategori	Gejala yang diamati	Kecepatan Angin	
			knot	km/jam
0	Tenang	Asap naik vertikal	< 1	< 1
1	Udara Ringan	Arah angin dapat dilihat dari arah asap	1 - 3	1 – 3
2	Sepoi Lemah	Kincir angin bergerak	4 - 6	4 – 7
3	Sepoi Lembut	Daun dan ranting bergerak	7 - 10	12 – 19
4	Sepoi Sedang	Debu dan kertas naik ke atas	11 - 16	20 – 28
5	Sepoi Segar	Pohon kecil mulai bergoyang	17 - 21	29 – 38
6	Sepoi Kuat	Sulit memakai payung	22 - 27	39 - 49
7	Angin Ribut Lemah	Seluruh pohon bergerak	28 - 33	50 - 61
8	Angin Ribut	Ranting pohon patah	34 - 40	62 - 74
9	Angin Ribut Kuat	Bangunan rusak ringan	41 - 47	75 - 88
10	Badai	Menumbangkan pohon	48 - 55	89 - 102
11	Badai Amuk	Kerusakan yang luas	56 - 63	103 - 117
12	Siklon	Kerusakan hebat	≥ 64	≥ 118

BAB IV
METODOLOGI PENELITIAN

4.1. Data Penelitian

Data dalam penelitian ini menggunakan data sekunder yang diperoleh dari *website* BMKG. Data tersebut berupa data kecepatan angin (knot) untuk wilayah pulau Jawa dengan masing-masing Stasiun Meteorologi. Dari data tersebut diperoleh 1.739 data yang dimulai dari Januari 2014 hingga Desember 2017.

4.2. Sampel dan Teknik Sampling

Sampel yang dimaksud adalah data yang akan digunakan sebagai data untuk pemodelan *PHMM* guna memperoleh model terbaik menggunakan estimasi *Algoritma EM*. Sampel tersebut diambil dari populasi dengan melakukan penyaringan data yang didasarkan pada besarnya resiko dan bahaya yang diakibatkan oleh angin ribut tersebut, khususnya bagi masyarakat di sepanjang wilayah yang diamati. Besarnya kecepatan angin tersebut ≥ 34 knot.

Berdasarkan hasil penyaringan data, diperoleh jumlah sampel data secara keseluruhan yang digunakan sebagai data untuk pemodelan *PHMM* sebanyak 144 data.

4.3. Variabel dan Definisi Operasional Variabel

Variabel yang akan digunakan pada penelitian ini ditunjukkan dalam **Tabel 4.1** tentang penjelasan dan definisi operasional penelitian :

Tabel 4.1 *Definisi Operasional Variabel*

Variabel	Definisi Operasional Variabel	Satuan
Stasiun Meteorologi	Stasiun yang terdiri atas bangunan, peralatan, dan perlengkapan lain untuk keperluan pengamatan meteorologi	-
Kecepatan Angin	Kecepatan udara yang bergerak dari daerah yang bertekanan udara tinggi (maksimum) ke rendah (minimum)	knot

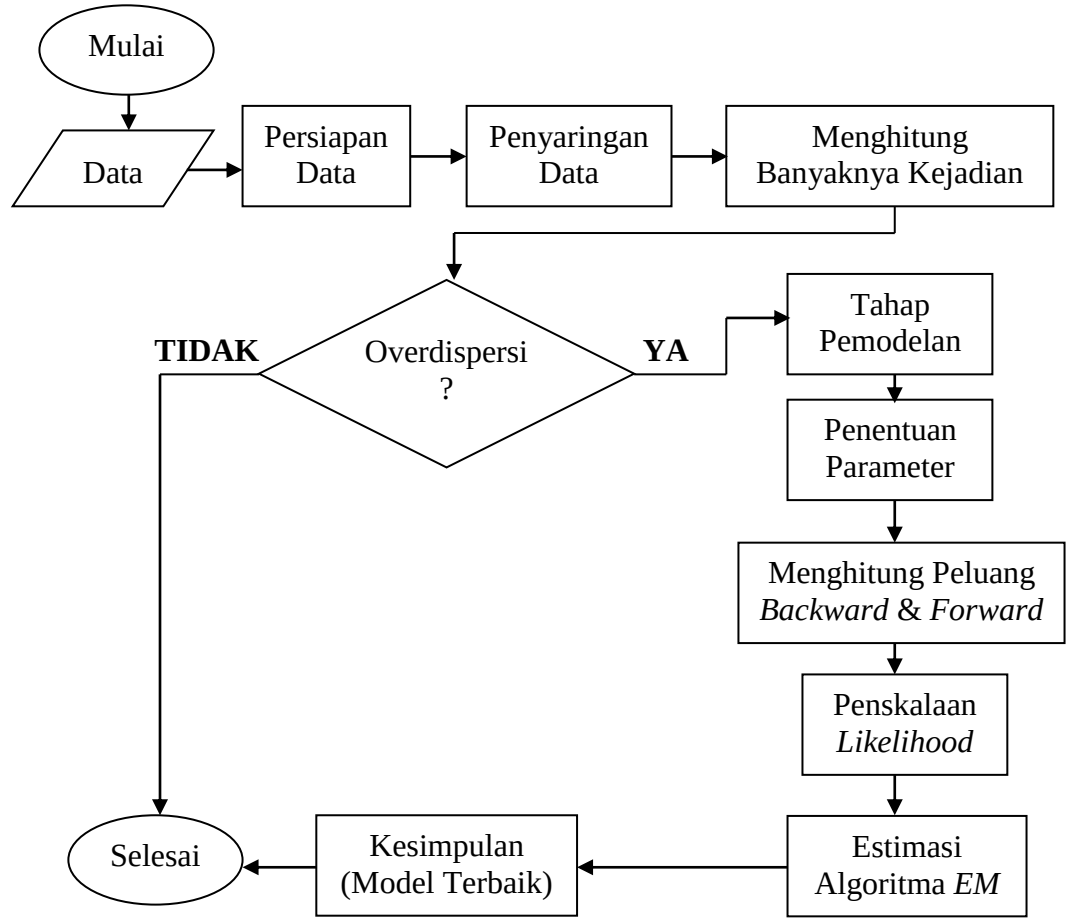
4.4 Metode Analisis Data

Proses analisis dalam penelitian ini menggunakan *software R 3.4.0*. Adapun metode analisis data yang digunakan dalam penelitian ini sebagai berikut:

1. Analisis deskriptif, digunakan untuk mendeskripsikan arah dan kecepatan angin yang bergerak di wilayah Indonesia dengan masing-masing Stasiun Meteorologi, khususnya wilayah konsentrasi bencana angin ribut.
2. *PHMM*, digunakan untuk mendapatkan parameter-parameter *input* yang selanjutnya dilakukan estimasi dari parameter tersebut.
3. Estimasi Algoritma *EM*, digunakan untuk memperoleh model terbaik dari keadaan tersembunyi yang diberikan dengan melihat nilai *AIC*.

4.5 Proses Analisis Data

Adapun langkah-langkah penelitian ini divisualisasikan dalam diagram yang disajikan pada **Gambar 4.1** berikut ini :



Gambar 4.1 *Flowchart Penelitian*

Adapun tahapan-tahapan penelitian sebagai berikut :

1. Pada tahap persiapan data, peneliti menyiapkan data kecepatan angin pada masing-masing Stasiun Meteorologi di wilayah konsentrasi bencana angin ribut. Kemudian, dari data tersebut dilakukan penyaringan data dengan menggunakan kriteria yang didasarkan pada besarnya resiko dan bahaya yang diakibatkan oleh angin ribut, khususnya bagi masyarakat di sepanjang wilayah yang diamati. Besarnya kecepatan angin tersebut ≥ 34 knot. Data hasil penyaringan tersebut kemudian dihitung banyaknya kejadian yang terjadi berdasarkan ≥ 34 knot dalam kurun waktu tertentu. Proses pengambilan data tersebut mengikuti pengambilan data berdistrubusi *Poisson*, sehingga tahapan berikutnya adalah pengecekan overdispersi.
2. Pengecekan overdispersi tersebut dilakukan dengan membandingkan nilai variansi dan nilai *mean* dari data banyaknya kejadian yang terjadi. Jika dari pengecekan overdispersi tersebut nilai variansi lebih besar dibandingkan nilai *mean*, maka dikatakan data terjadi overdispersi terhadap distribusi *Poisson*.
3. Pada tahap pemodelan, peneliti menentukan parameter-parameter *input* pada *PHMM*. Pada tahapan ini merupakan tahapan dalam mencari nilai parameter awal untuk masing-masing model. Parameter-paremeter tersebut diantaranya parameter rata-rata banyaknya kejadian $\lambda_i = (\lambda_1, \ldots, \lambda_m)$ dimana pada setiap λ_i memiliki kriteria berdistribusi *Poisson* dengan peluang awal kejadian adalah δ, dan matriks peluang transisi keadaan tersembunyi (⌈). Setelah diperoleh tiga parameter tersebut, kemudian parameter untuk setiap model di*input* ke *software R* untuk memperoleh model estimasi parameter tersebut.
4. Metode estimasi dalam *Hidden Markov Models* ini salah satunya dapat menggunakan metode Algoritma *EM*. Perangkat yang diperlukan dalam metode ini adalah menghitung nilai peluang *Forward* dan *Backward* yang ditunjukkan pada persamaan 3.33 dan persamaan 3.34. Masing-masing dari hasil kedua peluang tersebut kemudian dilakukan penskalaan *Likelihood* untuk mengatasi kasus *underflow*. Pada tahap estimasi Algoritma *EM,* terdiri atas dua langkah, yaitu tahap E (*E Step*) dan tahap M (*M Step*). Tahap E adalah menghitung ekspektasi bersyarat dari data yang hilang (*missing data*)

yang kemudian dari hasil *E Step* tersebut mengganti data pengamatan yang hilang. Selanjutnya, dimaksimalkan pada tahap M.

5. Pemilihan model estimasi terbaik dengan menggunakan nilai *AIC*. Pada penelitian ini akan dicari 3 model estimasi dengan menggunakan metode Algoritma *EM* dengan m = (2,3,4). Model estimasi yang terbaik adalah model yang memiliki nilai *AIC* terkecil.

BAB V
ANALISIS DAN PEMBAHASAN

5.1. Deskripsi Data

Deskripsi data dalam penelitian ini digunakan untuk melihat gambaran umum mengenai kejadian angin ribut di wilayah Jawa. Data tersebut berupa kecepatan angin yang diambil 4 tahun terakhir dimulai dari tanggal Januari 2014 hingga Desember 2017, yang kemudian disaring. Kriteria penyaringan data tersebut didasarkan pada besarnya resiko dan bahaya yang diakibatkan oleh angin ribut, khususnya bagi masyarakat di sepanjang wilayah yang diamati dengan besar kecepatan angin ≥ 34 knot.

Berikut ini data kecepatan angin untuk wilayah Jawa berdasarkan Stasiun Meteorologi:

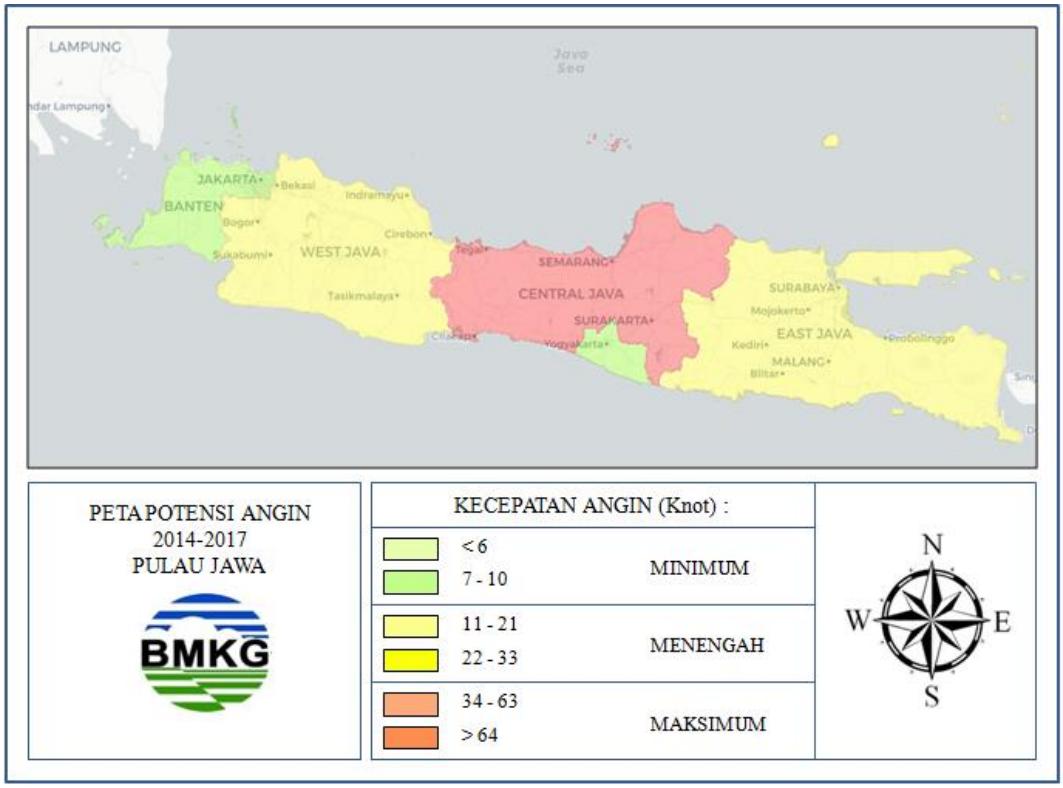
Tabel 5.1 *Data Kecepatan Angin Wilayah Jawa Berdasarkan Stasiun Meteorologi*

Nama Stasiun Meteorologi	Tanggal	Arah Angin (deg)	Kecepatan Angin (knot)
Maritim Perak II	01/01/2014	NW	7
Jatiwangi	02/01/2014	N	6
Maritim Perak II	03/01/2014	S	7
Tangerang Selatan	04/01/2014	W	6
⋮	⋮	⋮	⋮
Semarang	28/12/2017	NW	6

Tabel 5.1 di atas menginformasikan bahwa kecepatan angin untuk tanggal 1 Januari sampai dengan 4 Januari 2014 berada < 10 knot, yang berarti termasuk kategori angin sepoi lembut.

Peta sebaran kecepatan angin dapat dilihat pada **Gambar 5.1**. Peta tersebut menginfomasikan bahwa wilayah Jawa yang memiliki kecepatan angin minimum adalah provinsi DKI Jakarta, Banten, dan DIY. Kecepatan angin tersebut berkisar antara 0-10 knot. Sementara, wilayah yang memiliki kecepatan angin menengah adalah provinsi Jawa Barat dan Jawa Timur dengan kecepatan

angin berkisar antara 11-33 knot. Wilayah Jawa Tengah memiliki kecepatan angin maksimum, yaitu >34 knot.



Gambar 5.1 *Peta Potensi Angin Wilayah Jawa Tahun 2014-2017*

Pulau Jawa dikelilingi oleh lautan seperti Lautan Hindia, Laut Jawa, Laut Natuna dan Selat Malaka dan juga adanya dataran yang letaknya berdekatan dengan Jawa yaitu Pulau Sumatera dan Benua Australia. Sehingga arus angin yang berasal dari wilayah-wilayah tersebut mempunyai perbedaan suhu, tekanan dan kelembapan yang relatif tinggi. Kondisi topografi Pulau Jawa yang banyak terdapat perbukitan dan pegunungan memungkinkan terjadinya perbedaan suhu yang tinggi yang akan mempengaruhi gerakan angin sehingga menyebabkan terjadinya pertemuan antar jenis angin.

Pergerakan angin dipengaruhi oleh kondisi kekasaran permukaan suatu wilayah terkait dengan gaya gesekan. Gaya gesekan yang besar menyebabkan pergerakan angin melemah. Pada daerah yang permukaannya datar dan halus memiliki gaya gesek lemah, maka kecepatan angin akan besar. Oleh karena itu angin kencang sering terjadi pada daerah dataran dan pesisir seperti topografi pada

Provinsi Jawa Tengah bagian utara dan selatan yang memiliki topografi dengan permukaan yang relatif lebih halus dibanding bagian tengah.

5.2 Pengecekan *Overdispersi*

Setelah dilakukan pendeskripsian data, selanjutnya adalah mengecek overdispersi pada data. Pengecekan tersebut dilakukan pada data banyaknya kejadian angin ribut ≥34 knot dalam periode 10 hari. Pengecekan dapat dilakukan dengan cara menghitung banyaknya kejadian yang berkecepatan ≥34 knot pada Tabel 5.1. dalam periode 10 hari, kemudian membandingkan nilai *mean* dengan nilai variansi dari data tersebut. Dalam periode 10 hari, diperoleh data banyaknya kejadian sebanyak 144 data yang dapat dilihat pada Tabel 5.2.

Tabel 5.2 *Banyaknya Kejadian Angin ribut di wilayah Jawa Januari 2014-Desember 2017*

10 Hari Ke-	Banyaknya Kejadian Angin ribut
1	0
2	1
3	0
4	0
5	0
6	0
⋮	⋮
143	0
144	0

Banyaknya kejadian angin ribut di wilayah Jawa dilihat dari kecepatan angin yang bersumber dari *website* BMKG. **Tabel 5.2** di atas menginformasikan bahwa pada periode 10 hari ke-1 terdapat 0 kejadian. Pada 10 hari ke-2 terjadi kejadian sebanyak 1 kejadian, hingga pada 10 hari ke-144 terjadi 0 kejadian. Karakteristik pengambilan data tersebut mengikuti pengambilan data yang berdistribusi *Poisson*. Penghitungan nilai *mean* dan variansi dapat dilihat pada *output software R* berikut:

Tabel 5.3 *Summary Banyaknya Kejadian Angin ribut di wilayah Jawa Januari 2014-Desember 2017*

N	$\bar{x}$	s^2	Maksimum	Minimum
144	0.229	1.563	9	0

Summary dari **Tabel 5.3**. tersebut memberikan informasi bahwa jumlah banyaknya kejadian yang terjadi sebanyak 144 data dengan nilai *mean* ($\bar{x}$) sebesar 0.229, dan nilai variansi (s^2) sebesar 1.563. Sedangkan, untuk nilai maksimum dari banyaknya kejadian adalah 9 kejadian. Oleh karena nilai variansi (s^2) yang diperoleh lebih besar dibandingkan nilai *mean* ($\bar{x}$), maka dapat dikatakan data banyaknya kejadian angin ribut pada Tabel 5.2 terjadi overdispersi terhadap distribusi *Poisson*.

5.3 Pemodelan Menggunakan *PHMM* (*Poisson Hidden Markov Models*)

Setelah dilakukan pengecekan overdispersi, selanjutnya adalah mencari model estimasi banyaknya kejadian angin ribut terbaik dari tiga model. Ketiga model tersebut merupakan model keadaan tersembunyi, yaitu m = (2, 3, 4). Oleh karena dalam penelitian ini data tidak memadai, maka peneliti mencari model estimasi hingga keadaan tersembunyi m = 4. Pemodelan menggunakan *PHMM* dengan estimasi Algoritma *EM* terlebih dahulu menentukan parameter *input*, kemudian menaksir parameter-parameter *PHMM* dengan Algoritma *EM* menggunakan *software R*.

Barisan data banyaknya kejadian angin ribut pada wilayah Jawa diasumsikan bahwa dibangkitkan oleh *Poisson Hidden Markov*, dimana penyebab kejadian tersebut membentuk rantai *markov* yang *unobserved*. Faktor yang menyebabkan kejadian diasumsikan dengan *state* dari suatu rantai *markov* $\{C_t\}_{t\in N}$. Sementara, data banyaknya kejadian merupakan barisan observasi per-10 hari dimana t menunjukkan periode 10 hari ke-t, Y_t menunjukkan banyaknya kejadian angin ribut pada 10 hari ke-t, maka Y_t bersifat diskrit. C_t merupakan rantai markov yang *unobserved*, atau faktor penyebab pada kejadian 10 hari ke-t. Dengan kata lain, Y_t bergantung pada C_t.

5.3.1 Penentuan Parameter *Input*

Penentuan parameter *input* dengan metode *PHMM* adalah membuat tabel distribusi frekuensi dari data banyaknya kejadian angin ribut (Tabel 5.2), dimana jumlah kelas pada tabel frekuensi tersebut ditentukan oleh banyaknya keadaan tersembunyi yang ditentukan. Parameter-parameter tersebut diantaranya, parameter rata-rata banyaknya kejadian angin ribut ≥34 knot dalam per-10 hari atau nilai $\lambda = (\lambda_1, \ldots, \lambda_m)$, peluang awal kejadian atau nilai $\delta = (\delta_1, \ldots, \delta_m)$, dan matriks peluang transisi keadaan tersembunyi yang berukuran m×m (⌈).

Selanjutnya, peneliti menentukan *range* untuk pembagian interval masing-masing kelas secara seragam. Dalam hal ini, peneliti mengambil *range* 0 sampai 11. Misalkan, peneliti memberikan 2 keadaan tersembunyi (m = 2) dengan parameter rata-rata banyaknya kejadian adalah $\lambda = (\lambda_1, \lambda_2)$, maka terdapat 2 kelas dengan masing-masing kelasnya memiliki panjang interval c = 6. Hal ini berarti ruang sampel banyaknya kejadian angin ribut pada keadaan tersembunyi 1 adalah {0,1,2,3,4,5}, sementara sisanya adalah ruang sampel untuk keadaan tersembunyi 2.

Pada 3 keadaan tersembunyi (m = 3) dengan parameter rata-rata banyaknya kejadian adalah $\lambda = (\lambda_1, \lambda_2, \lambda_3)$, maka terdapat 3 kelas dengan masing-masing kelasnya memiliki panjang interval c = 4. Hal ini berarti ruang sampel banyaknya kejadian angin ribut pada keadaan tersembunyi 1 adalah {0,1,2,3}, untuk keadaan tersembunyi 2 adalah {4,5,6,7}, sementara sisanya adalah ruang sampel untuk keadaan tersembunyi 3, yaitu {8,9,10,11}.

Pada 4 keadaan tersembunyi (m = 4) dengan parameter rata-rata banyaknya kejadian adalah $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, maka terdapat 4 kelas dengan masing-masing kelasnya memiliki panjang interval c = 3. Hal ini berarti ruang sampel banyaknya kejadian angin ribut pada keadaan tersembunyi 1 adalah {0,1,2}, untuk keadaan tersembunyi 2 ruang sampelnya adalah {3,4,5}, untuk keadaan tersembuni 3 ruang sampelnya adalah {6,7,8}, sementara sisanya adalah ruang sampel untuk keadaan tersembunyi 4.

Selanjutnya, menghitung nilai λ dengan cara memasukkan data pada Tabel 5.2 ke masing-masing kelompok keadaan tersembunyi berdasarkan ruang sampel

yang telah diperoleh. Sementara, untuk parameter δ yang merupakan peluang awal pada keadaan tersembunyi, yang diperoleh dengan menghitung jumlah frekuensi pada masing-masing kelompok keadaan tersembunyi, lalu dari frekuensi tersebut dibagi dengan frekuensi dari keseluruhan keadaan tersembunyi. Berikut ini hasil perhitungan untuk parameter λ dan δ pada kasus 2 keadaan tersembunyi (m = 2) yang disajikan pada tabel berikut:

Tabel 5.4 *Hasil Perhitungan Parameter λ dan δ pada m = 2*

15 Hari Ke-	**Banyaknya Kejadian Angin ribut**	**Keadaan Tersembunyi**	**Banyaknya Kejadian Keadaan Tersembunyi 1**	**Banyaknya Kejadian Keadaan Tersembunyi 2**
1	0	1	0	-
2	1	1	1	-
3	0	1	0	-
4	0	1	0	-
5	0	1	0	-
6	0	1	0	-
⋮	⋮	⋮	⋮	⋮
143	0	1	0	-
144	0	1	0	-
	f	144	141	3
	λ		0.057	8.333
	δ		0.979	0.021

Dari **Tabel 5.4** diperoleh hasil perhitungan untuk parameter λ = (0.057, 8.333) dan parameter δ = (0.979, 0.021), artinya adalah dalam waktu 10 hari, rata-rata banyaknya kejadian angin ribut pada keadaan tersembunyi 1 sebesar 0.057 kejadian dengan peluang awal kejadiannya sebesar 0.979, dan rata-rata banyaknya kejadian angin ribut pada keadaan tersembunyi 2 sebesar 8.333 kejadian dengan peluang awal kejadiannya sebesar 0.021.

Perhitungan untuk parameter matriks peluang transisi (Γ) diperoleh dengan menghitung frekuensi pada masing-masing kemungkinan perpindahan keadaan tersembunyi, lalu dibagi dengan total masing-masing baris keadaan tersembunyi.

Tabel 5.5 *Peluang pada Kasus 2 Keadaan Tersembunyi*

λ_i	1	2
1	139/140	1/140
2	1/3	2/3

Banyaknya kejadian angin ribut pada kasus 2 keadaan tersembunyi terdapat empat kemungkinan perpindahan keadaan tersembunyi yang disajikan pada **Tabel 5.5**, sehingga diperoleh matriks peluang transisi pada kasus 2 keadaan tersembunyi seperti berikut :

$$\Gamma = \begin{Vmatrix} 0.993 & 0.007 \\ 0.333 & 0.667 \end{Vmatrix}$$

dari matriks peluang transisi tersebut, dapat diinterpretasikan bahwa jika pada periode ini berada pada keadaan tersembunyi 1, maka peluang pada 10 hari mendatang berada pada keadaan tersembunyi 1 sebesar 0.993. Sementara, jika pada periode ini berada pada keadaan tersembunyi 1, maka peluang untuk 10 hari kedepan berada pada keadaan tersembunyi 2 sebesar 0.007, dan seterusnya.

Tabel 5.6 *Parameter λ, δ, dan Γ pada masing-masing PHMM*

Model	**i**	$\boldsymbol{\lambda}$	$\boldsymbol{\delta}$	$\boldsymbol{\Gamma}$			
				1	**2**	**3**	**4**
m = 2	1	0.057	0.979	0.993	0.007	-	-
	2	8.333	0.021	0.333	0.667	-	-
m = 3	1	0.029	0.972	0.993	0.007	0	-
	2	5.5	0.014	0.5	0	0.5	-
	3	9	0.014	0	0.5	0.5	-
m = 4	1	0.029	0.972	0.993	0	0.007	0
	2	4	0.007	1	0	0	0
	3	7	0.007	0	0	0	1
	4	9	0.014	0	0.5	0	0.5

Tabel 5.6 menampilkan hasil lengkap dari perhitungan parameter λ, δ, dan Γ untuk keadaan tersembunyi m = (2,3,4). Pada kasus 3 dan 4 keadaan tersembunyi, dilakukan dengan cara yang sama untuk memperoleh parameter λ, δ, dan Γ. Tabel tersebut menjelaskan bahwa model dengan 2 keadaan tersembunyi (m = 2), mempunyai parameter rata-rata banyaknya kejadian λ dalam per-10 hari atau λ = (0.057, 8.333) dengan peluang awal kejadian δ = (0.979, 0.021) dan matriks peluang transisi $\Gamma = \begin{Vmatrix} 0.993 & 0.007 \\ 0.333 & 0.667 \end{Vmatrix}$. Pada model dengan 3 keadaan

tersembunyi (m = 3), mempunyai parameter λ = (0.029, 5.5, 9) dengan peluang awal kejadian δ = (0.972, 0.014, 0.014) dan matriks peluang transisi ⌈ = $\begin{Vmatrix} 0.993 & 0.077 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0.5 \end{Vmatrix}$. Sedangkan, untuk model dengan 4 keadaan tersembunyi mempunyai parameter λ = (0.029, 4, 7, 9) dengan peluang awal kejadian δ = (0.972, 0.007, 0.007, 0.014) dan matriks peluang transisi ⌈ = $\begin{Vmatrix} 0.993 & 0 & 0.007 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0.5 & 0 & 0.5 \end{Vmatrix}$.

5.3.2 Penaksiran Parameter-parameter *PHMM* dengan Algoritma *EM*

Setelah menentukan parameter *PHMM*, langkah selanjutnya adalah menghitung nilai estimasi dari parameter-parameter tersebut, yaitu $\hat{\lambda}$, $\hat{\delta}$, dan $\hat{⌈}$ untuk masing-masing model dengan menggunakan Algoritma *EM*. Perhitungan nilai estimasi Algoritma *EM* dikerjakan *software R* dengan nilai toleransi yang digunakan adalah 1e-06, dimana nilai toleransi ini sebagai kriteria peneliti.

Tabel 5.7 *Hasil Estimasi Parameter PHMM dengan Algoritma EM pada masing-masing Keadaan Tersembunyi*

Model	i	AIC	$\hat{\lambda}$	$\hat{\delta}$	$\hat{⌈}$			
					1	2	3	4
m = 2	1	132.2833	0.044	0.999	0.978	0.022	-	-
	2		7.870	0.001	0.375	0.625	-	-
m = 3	1	133.2268	0.03	1	0.985	0.015	0	-
	2		5.639	0	0.405	0	0.595	-
	3		9.134	0	0	0.736	0.264	-
m = 4	1	138.3017	0.078	1	0.993	0	0.007	0
	2		2.999	0	1	0	0	0
	3		5.999	0	0	0	0	1
	4		5.999	0	0	1	0	0

Tabel 5.7 menunjukkan hasil estimasi parameter *PHMM* dengan Algoritma *EM* pada masing-masing keadaan tersembunyi. Setelah diperoleh hasil estimasi parameter, langkah berikutnya adalah menentukan model estimasi terbaik dari banyaknya kejadian angin ribut dengan membandingkan nilai *AIC*, dimana nilai *AIC* terkecil adalah model estimasi paling baik. Berdasarkan Tabel tersebut, terlihat bahwa nilai *AIC* terkecil berada pada saat diberikan 2 keadaan

tersembunyi yaitu sebesar 132.2833, maka dapat dikatakan bahwa model dengan 2 keadaan tersembunyi (m = 2) merupakan model terbaik dibandingkan dengan m=3 dan m=4. Berikut ini hasil estimasi parameter terbaik *PHMM*, yaitu model dengan 2 keadaan tersembunyi:

$$\hat{\lambda} = (0.044, 7.870);$$

$$\hat{\delta} = (0.999, 0.001);$$

$$\hat{\Gamma} = \begin{Vmatrix} 0.978 & 0.022 \\ 0.375 & 0.625 \end{Vmatrix}.$$

5.4 Prediksi dengan Pendekatan Matriks

Prediksi peluang pada periode yang sangat besar dapat dihitung menggunakan metode pendekatan matriks probabilitas. Jika diasumsikan bahwa angin ribut yang disebabkan oleh *state* 1 maka berlaku probabilitas dalam vektor baris sebagai berikut :

$$(1 \quad 0)$$

dengan menggunakan persamaan (3.25), diperoleh peluang periode berikutnya atau periode setelah 10 hari ke-144 adalah

$((\mathrm{N}(145) \quad \mathrm{M}(145)) = (0.965 \quad 0.035)$

Dari peluang tersebut, dapat diprediksi bahwa kejadian angin ribut yang disebabkan oleh *state* 1 pada periode ke-145 setelah beberapa periode di masa depan juga disebabkan oleh *state* 1, maka peluangnya sebesar 0.965 dan peluang disebabkan oleh *state* 2 sebesar 0.035. Sementara, untuk periode berikutnya memiliki peluang sebagai berikut :

$((\mathrm{N}(146) \quad \mathrm{M}(146)) = (0.957 \quad 0.043)$

$((\mathrm{N}(147) \quad \mathrm{M}(147)) = (0.952 \quad 0.048)$

$((\mathrm{N}(148) \quad \mathrm{M}(148)) = (0.949 \quad 0.051)$

BAB VI
PENUTUP

6.1. Kesimpulan

Berdasarkan hasil analisis dam pembahasan pada bab sebelumnya, peneliti dapat menarik beberapa kesimpulan sebagai berikut :

1. Berdasarkan skala *Beaufort*, kecepatan angin minimum yang berpotensi terjadinya angin sepoi lembut terpantau di provinsi DKI Jakarta, Banten, dan DIY. Kecepatan angin menengah yang berpotensi angin sepoi kuat terpantau di provinsi Jawa Barat dan Jawa Timur. Sementara, kecepatan angin maksimum yang berpotensi terjadinya angin ribut terpantau di provinsi Jawa Tengah.
2. Berdasarkan nilai *AIC* terkecil, model *Poisson Hidden Markov* dengan 2 *state* atau m = 2 merupakan nilai m yang optimum untuk pendugaan parameter model. Hasil estimasi parameter model menggunakan *Expectation Maximization Algorithm* sebagai berikut :
   a. Matriks peluang transisi ($\hat{\Gamma}$) merupakan peluang penyebab terjadinya angin ribut. Terlihat bahwa setelah terjadi angin ribut yang disebabkan penyebab pada *state* 1 akan terjadi kembali dengan peluang 0.978 bahwa angin ribut tersebut akan disebabkan oleh penyebab pada *state* 1. Terdapat peluang 0.022 bahwa akan terjadi angin ribut karena penyebab pada *state* 2 setelah terjadinya penyebab pada *state* 1. Jika terjadi angin ribut karena sebab pada *state* 2 maka terdapat peluang 0.375 bahwa akan terjadi angin ribut karena sebab pada *state* 1 setelahnya, dan terdapat peluang 0.625 bahwa setelah terjadi angin ribut karena sebab pada *state* 2 akan terjadi kembali angin ribut karena sebab yang sama.
   b. Dari matriks peluang transisi diperoleh vektor peluang awal ($\hat{\delta}$), dimana peluang bahwa angin ribut pertama terjadi disebabkan oleh *state* 1 adalah 0.999, dan peluang bahwa angin ribut pertama terjadi disebabkan oleh *state* 2 adalah 0.001.

c. Rata-rata banyaknya kejadian angin ribut ($\hat{\lambda}$) karena sebab pada *state* 1 dan *state* 2 adalah 0.044 dan 7.870 per 10 hari.

3. Dari model yang terpilih, dapat diprediksi bahwa banyaknya kejadian angin ribut untuk periode 10 hari ke-145 adalah $(0.965 \quad 0.035)$ yang artinya jika pada awalnya kejadian angin ribut disebabkan oleh *state* 1, setelah beberapa periode di masa depan juga disebabkan oleh *state* 1 maka peluangnya sebesar 0.965 dan peluang disebabkan oleh *state* 2 sebesar 0.035.

6.2. Saran

Berdasarkan hasil analisis dan kesimpulan, dapat diberikan saran sebagai berikut:

1. Untuk pihak pemerintah, diharapkan adanya publikasi kejadian bencana dan upaya mitigasi yang tepat pada daerah-daerah yang rawan. Selain itu, perlu adanya kegiatan peningkatan pemahaman dan peredaman risiko bencana kepada masyarakat serta simulasi bencana agar masyarakat menjadi tangguh dalam menghadapi bencana dan mampu mengelola risiko bencana dengan mandiri.
2. Untuk masyarakat, perlu mengenali dengan betul tempat tinggal yang akan dihuni dengan mencari tahu keadaan atau kondisi lingkungan sehingga dapat dikenali dampak positif serta negatif dari hunian yang akan ditinggali. Selain itu, perlu adanya kesadaran untuk bergotong-royong menanam pohon atau melakukan penghijauan dilingkungan tempat tinggal untuk mengurangi atau meredam kecepatan angin.
3. Untuk peneliti selanjutnya, estimasi parameter kejadian angin ribut dapat dilakukan dengan mempertimbangkan pembagian lokasi pengamatan berdasarkan sebaran angin ribut yang sering terjadi. Selain itu, juga dapat menggunakan kriteria lain sebagai penentuan model terbaik.

DAFTAR PUSTAKA

Afriani, Erlya, dan Suyatim. 2015. *Prediksi Cuaca Ekstrim dengan Menggunakan Model WRF-ARW di Palangkaraya*. Jurnal Meteorologi Klimatologi dan Geofisika, Volume 2, Nomor 2, Desember 2015.

Aisyah, Aprilia Fitriatul. 2015. *Implementasi Hidden Markov Models (HMM) Sebagai Filter Untuk Mereduksi Noise pada Esophageal Speech*. Jurnal Inovasi Fisika Indonesia, Volume 4, Nomor 3, hlm. 7-14.

Agustinah, Trihastuti. 2014. *Diktat Kuliah Probabilitas dan Proses Stokastik*. Jurusan Teknik Elektro ITS, Surabaya.

Azlina, Maya. 2013. *Pembuatan Alat Ukur Kecepatan Angin dan Penunjuk Arah Angin Berbasis Mikrokontroler AT-Mega 8535*. Tugas Akhir. Program Studi Fisika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sumatera Utara Medan.

Badan Nasional Penanggulangan Bencana. 2012. *Data: Jumlah Kejadian Bencana*. http://dibi.bnpb.go.id diakses pada tanggal 23 Januari 2018 pukul 06:32 WIB.

DetikNews. 2017. *Angin Ribut Mengamuk 4 Desa di Banjarnegara, 1 Warga Tewas*. https://news.detik.com/berita-jawa-tengah/d-3750560/angin-ribut-mengamuk-4-desa-di-banjarnegara-1-warga-tewas. diakses pada tanggal 28 Maret 2018 pukul 22:10 WIB.

Gustra, Hendra. 2014. *Pemodelan Klaim Asuransi Kerugian Menggunakan Poisson Hidden Markov Untuk Data Overdispersi*. Tugas Akhir. Program Studi Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Pertanian Bogor.

Hines, William W dan Montgomery, Douglas C. 1972. *Probability and Statistics in Engineering and Management Science*. Edisi ke 2. Diterjemahkan oleh: Rudiansyah. Jakarta: UI Press.

Indonesia, *Peraturan Kepala Badan Meteorologi, Klimatologi, dan Geofisika Tentang Prosedur Standar Operasional Pelaksanaan Peringatan Dini,*

Pelaporan, dan Diseminasi Informasi Cuaca Ekstrim, KEP No.009 Tahun 2010.

Kartasapoetra, Ance Gunarsih. 2004. *Klimatologi: Pengaruh Iklim Terhadap Tanah dan Tanaman*. Jakarta: PT Bumi Aksara.

Mamonto, Sri Wahyuni, Langi, Yohanes A. R. dan Rindengan, Altien J. 2016. *Penerapan Hidden Markov Model pada Harga Saham*. Jurnal de Cartesian, Volume 5, Nomor 1, Maret 2016.

Mulyana, Sri, Afiahayati dan Surya, Wijaya Adhi. 2008. Prosiding dari Seminar Nasional Teknoin Bidang Teknik Informatika di Yogyakarta. *Penerapan Hidden Markov Model dalam Clustering Sequence Protein Globin*. Hal: 117-122.

Muzayanah, Linda Fitrotul. 2015. *Interpretasi Radar Cuaca Sebagai Kajian Puting Beliung dan Angin Kencang Wilayah Jawa Timur*. Tugas Akhir. Program Studi Terapan Meteorologi Sekolah Tinggi Meteorologi dan Geofisika Tangerang Selatan.

Nikaryanto. 1979. *Angin Ribut di Jawa.* Jakarta: Fipia Universitas Indonesia.

Nurhasanah. 2016. *Pemodelan Poisson Hidden Markov Untuk Prediksi Banyaknya Kecelakaan di Jalan Tol Jakarta-Cikampek.* Journal of Mathematical Analysis, Volume 15, Nomor 1, Juli 2016, hlm. 55-64.

Nurjani, Emilya, Rahayu, Arum dan Rachmawati, Febriyan. 2013. *Kajian Bencana Angin Ribut di Indonedia Periode 1990-2011: Upaya Mitigasi bencana*. Jurnal Geomedia, Volume 11, Nomor 2, November 2013.

Papoulis, Athanasios. 1992. *Probabilitas, Variabel Random, dan Proses Stokastik*. Yogyakarta: Gadjah Mada University Press.

Paris, Carmen Molina. 2011. *Mathematical Models and Immune Cell Biology*. New York: Springer.

Praptono. 1986. *Materi Pokok Pengantar Proses Stokastik I*. Jakarta: Karunika.

Republika.co.id. 2018. *Angin Ribut Rusak 17 Rumah di Temanggung*. http://nasional.republika.co.id/berita/nasional/daerah/18/03/30/p6eld8284-angin-ribut-rusak-17-rumah-di-temanggung diakses pada tanggal 28 Maret 2018 pukul 22:30 WIB.

Sanger, Ray Jeremy, Fibriani, Charitas dan Nataliani, Yessica. 2012. *Perancangan Aplikasi Sistem Informasi Pemantauan Kecepatan Angin Beserta Pengkategorian Jenis Angin dengan Hardware Inframerah Sebagai Media Kalibrasi*. Jurnal Teknologi Informasi-Aiti, Volume 9, Nomor 2, Agustus 2012, hlm. 101-200.

Setyoaji, Agung, Muflikhah, Lailil dan Fauzi, Mochammad Ali. 2017. *Named Entity Recognition Menggunakan Hidden Markov Model dan Algoritma Viterbi pada Teks Tanaman Obat*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Volume 1, Nomor 12, Desember 2017, hlm. 1858-1864.

Sudjana. 1982. *Metoda Statistika*. Bandung: Tarsito.

Suryandaru, Rosyid. 2015. *Estimasi Model Terbaik Banyaknya Gempa Bumi Menggunakan Poisson Hidden Markov Models dan Algoritma EM*. Tugas Akhir. Program Studi Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah Jakarta.

Taylor, Howard M. dan Samuel, Karlin. 1998. *An Introduction to Stochastic Modeling 3rd Edition*. San Diego: Academic Press.

Tjasyono, Bayong. 2007. *Sistem Angin*. Makalah dalam "Workshop Turbin Angin Kecepatan Rendah dan Peta Potensi Angin Resolusi Tinggi" diselenggarakan oleh FITB, Institut Teknologi Bandung. Bandung 21 – 22 Agustus.

Walpole, Ronald E. dan Myers, Raymond H. 1972. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan Edisi ke-4*. Jakarta: Gramedia Pustaka Utama.

Wikipedia. *Jenis-jenis Angin*, (online), (https://id.wikipedia.org/wiki/Angin, diakses pada tanggal 10 Maret 2018).

LAMPIRAN

Lampiran 1 *Script R untuk Peluang Forward dan Backward pada PHMM*

```
pois.HMM.lalphabeta<-function(x,m,lambda,gamma,delta=NULL)
{
if(is.null(delta))delta<-solve(t(diag(m)-gamma+1),rep(1,m))
n            <- length(x)
lalpha      <- lbeta<-matrix(NA,m,n)
allprobs    <- outer(x,lambda,dpois)
foo          <- delta*allprobs[1,]
sumfoo      <- sum(foo)
lscale      <- log(sumfoo)
foo          <- foo/sumfoo
lalpha[,1] <- log(foo)+lscale
for (i in 2:n)
{
foo          <- foo%*%gamma*allprobs[i,]
sumfoo      <- sum(foo)
lscale      <- lscale+log(sumfoo)
foo          <- foo/sumfoo
lalpha[,i] <- log(foo)+lscale
}
lbeta[,n]  <- rep(0,m)
foo          <- rep(1/m,m)
lscale      <- log(m)
for (i in (n-1):1)
{
foo          <- gamma%*%(allprobs[i+1,]*foo)
lbeta[,i]  <- log(foo)+lscale
sumfoo      <- sum(foo)
foo          <- foo/sumfoo
lscale      <- lscale+log(sumfoo)
}
list(la=lalpha,lb=lbeta)
}
```

Lampiran 2 *Script R untuk Estimasi Algoritma EM*

```
pois.HMM.EM <- function(x,m,lambda,gamma,delta,maxiter=300,tol=1e-6,...)
{
lambda.next     <- lambda
gamma.next      <- gamma
delta.next      <- delta
for (iter in 1:maxiter)
{
lallprobs     <- outer(x,lambda,dpois,log=TRUE)
fb   <-  pois.HMM.lalphabeta(x,m,lambda,gamma,delta=delta)
la   <-  fb$la
lb   <-  fb$lb
c    <-  max(la[,n])
llk <- c+log(sum(exp(la[,n]-c)))
for (j in 1:m)
{
for (k in 1:m)
{
gamma.next[j,k] <- gamma[j,k]*sum(exp(la[j,1:(n-
1)]+lallprobs[2:n,k]+lb[k,2:n]-llk))
}
lambda.next[j] <- sum(exp(la[j,]+lb[j,]-llk)*x)/sum(exp(la[j,]+lb[j,]-llk))
}
gamma.next <- gamma.next/apply(gamma.next,1,sum)
delta.next <- exp(la[,1]+lb[,1]-llk)
delta.next <- delta.next/sum(delta.next)
crit         <- sum(abs(lambda-lambda.next)) +sum(abs(gamma-gamma.next))
+sum(abs(delta-delta.next))
if(crit<tol)
{
np      <- m*m+m-1
AIC     <- -2*(llk-np)
return(list(lambda=lambda,gamma=gamma,delta=delta,mllk=-
llk,AIC=AIC,BIC=BIC))
}
lambda      <- lambda.next
gamma       <- gamma.next
delta       <- delta.next
}
print(paste("No convergence after",maxiter,"iterations"))
NA
}
```

Lampiran 3 *Banyaknya Kejadian Angin Ribut di Wilayah Jawa pada Kecepatan Angin ≥ 34 Knot Januari 2014 – Desember 2017*

10 hari ke-	Banyak AR >=34
1	0
2	1
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	1
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0
26	0
27	0
28	0
29	0
30	0
31	0
32	0
33	0
34	0
35	0
36	0

10 hari ke-	Banyak AR >=34
37	0
38	0
39	0
40	0
41	0
42	0
43	0
44	0
45	0
46	0
47	0
48	0
49	0
50	0
51	0
52	0
53	0
54	0
55	0
56	0
57	0
58	0
59	0
60	0
61	0
62	0
63	0
64	0
65	0
66	0
67	0
68	0
69	0
⋮	⋮
143	0
144	0