



TESIS - KS185411

**PEMODELAN PADA DATA STATUS PENERIMAAN CALON  
MAHASISWA BARU MELALUI JALUR UJIAN TULIS BIDANG  
SAINTEK TAHUN 2018 MENGGUNAKAN *BERNOULLI  
MIXTURE MODEL***

**SHIELA NOVELIA DHARMA PRATIWI  
06211750010003**

Dosen Pembimbing  
Dr. Ismaini Zain, M.Si  
Prof. Drs. Nur Iriawan, M.Ikom, Ph.D.

Departemen Statistika  
Fakultas Sains dan Analitika Data  
Institut Teknologi Sepuluh Nopember  
2020



TESIS - KS185411

**PEMODELAN PADA DATA STATUS PENERIMAAN CALON  
MAHASISWA BARU MELALUI JALUR UJIAN TULIS BIDANG  
SAINTEK TAHUN 2018 MENGGUNAKAN *BERNOULLI  
MIXTURE MODEL***

**SHIELA NOVELIA DHARMA PRATIWI  
06211750010003**

**Dosen Pembimbing  
Dr. Ismaini Zain, M.Si  
Prof. Drs. Nur Iriawan, M.Ikom, Ph.D**

**Departemen Statistika  
Fakultas Sains dan Analitika Data  
Institut Teknologi Sepuluh Nopember  
2020**



THESIS - KS185411

**ON THE MODELING OF THE NEW STUDENT  
ACCEPTANCE STATUS THROUGH SCIENCE AND  
TECHNOLOGY WRITTEN TEST USING BERNOULLI  
MIXTURE MODEL**

**SHIELA NOVELIA DHARMA PRATIWI  
06211750010003**

**Supervisor  
Dr. Ismaini Zain, M.Si  
Prof. Drs. Nur Iriawan, M.Ikom, Ph.D**

**Department of Statistics  
Faculty of Science and Data Analytics  
Institut Teknologi Sepuluh Nopember  
2020**

LEMBAR PENGESAHAN TESIS

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar  
Magister Statistika (M.Stat.)

di

Institut Teknologi Sepuluh Nopember

Oleh:

SHIELA NOVELIA DHARMA PRATIWI

NRP : 062117 5001 0003

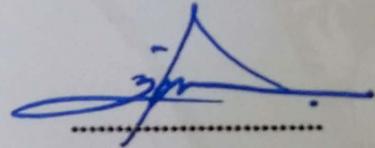
Tanggal Ujian : 20 Januari 2020

Periode Wisuda : Maret 2020

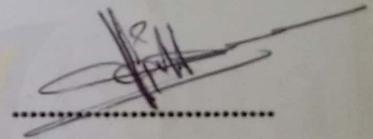
Disetujui oleh :

Pembimbing :

1. Dr. Dra. Ismaini Zain, M.Si  
NIP. 19600525 198803 2 001

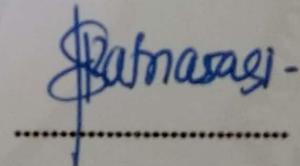


2. Prof. Drs. Nur Iriawan, M.Ikom, Ph.D  
NIP. 19621015 198803 1 002

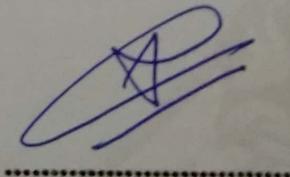


Penguji :

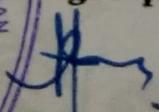
1. Dr. Vita Ratnasari, M.Si  
NIP. 19700910 199702 2 001



2. Dr. Sutikno, M.Si  
NIP. 19710313 199702 1 001



Kepala Departemen  
Fakultas Sains dan Analitika Data  
Institut Teknologi Sepuluh Nopember

  
Dr. Kartika Fithriasari, M.Si.  
NIP. 19691212 199303 2 002

**PEMODELAN PADA DATA STATUS PENERIMAAN CALON  
MAHASISWA BARU MELALUI JALUR UJIAN TULIS BIDANG  
SAINTEK TAHUN 2018 MENGGUNAKAN *BERNOULLI MIXTURE  
MODEL***

Nama Mahasiswa : Shiela Novelia Dharma Pratiwi  
NRP : 06211750010003  
Dosen Pembimbing : Dr. Ismaini Zain, M.Si.  
Co-Pembimbing : Prof. Drs. Nur Iriawan, M.Ikom, Ph.D.

**ABSTRAK**

Ujian tulis untuk seleksi penerimaan mahasiswa baru di Institut Teknologi 10 Nopember dilakukan dengan mengujikan 10 mata ujian yang dianggap mampu memberikan gambaran dari kemampuan setiap calon mahasiswa, meliputi tes kemampuan Potensi Akademik (TKPA) dan Tes Kemampuan Dasar Sains dan Teknologi (TKD SAINTEK). Dari tahun ketahun sistem penilaian pada ujian tulis hanya menitik beratkan pada nilai total sebagai tolak ukur penerimaan mahasiswa tanpa mempertimbangkan nilai kompetensi dan prodi pilihan calon mahasiswa baru. Untuk calon mahasiswa baru yang memilih prodi ilmu pasti seperti Fisika, Kimia, Biologi, Matematika, Aktuaria dan Statistika, kompetensi pada kemampuan dasar yang berkorespondensi dengan prodi pilihan perlu untuk dievaluasi, mengingat prodi tersebut hanya berfokus pada kemampuan khusus yang ditunjukkan oleh nilai TKD SAINTEK. Secara khusus, penelitian ini bertujuan untuk memodelkan status penerimaan calon mahasiswa baru dengan menggunakan *mixture* yaitu kesesuaian antara nilai TKD SAINTEK dengan prodi pilihan. Dengan harapan, hasil ketepatan klasifikasi dari pemodelan dapat digunakan sebagai pertimbangan untuk sistem penilaian pada tahun-tahun yang akan datang, sehingga sistem penilaian tidak hanya berfokus pada nilai total, tetapi juga pada kompetensi calon mahasiswa baru. Analisis pada penelitian ini menggunakan konsep *data driven*, sehingga data berpola multimodal menggunakan *Bernoulli Mixture Model* (BMM). Metode *Markov Chain Monte Carlo* (MCMC) akan digunakan untuk mengurai proses estimasi kompleks pada metode *Bayesian*. Pada penelitian ini, pemeriksaan konvergensi akan dilakukan secara inferensia dengan menggunakan tiga metode yaitu *Gelman-Rubin*, *Geweke* dan *Raftery-Lewis*. Pemodelan yang didapatkan pada penelitian ini memberikan ketepatan klasifikasi untuk prodi Fisika sebesar 73.02%, prodi Kimia sebesar 73.91%, prodi Biologi sebesar 68.13%, Aktuaria sebesar 71.18%, Matematika sebesar 63.54% dan Statistika sebesar 50%. Metode pemeriksaan konvergensi yang tepat digunakan pada penelitian ini adalah metode *Geweke* dan *Raftery-Lewis*.

**Kata kunci** : *Bernoulli Mixture*, Ujian Tulis, MCMC, *Data Driven*.

*(Halaman ini sengaja dikosongkan)*

***ON THE MODELING OF THE NEW STUDENT ACCEPTANCE STATUS  
THROUGH SCIENCE AND TECHNOLOGY WRITTEN TEST USING  
BERNOULLI MIXTURE MODEL***

Name of Student : Shiela Novelia Dharma Pratiwi  
NRP : 06211750010003  
Supervisors : Dr. Ismaini Zain, M.Si.  
Co- Supervisors : Prof. Nur Iriawan, Ph. D.

**ABSTRACT**

The written test to select registrants to be accepted as students of Sepuluh Nopember Institute of Technology is conducted by testing them in ten subjects considered providing an overview of their abilities. The ten subjects are included in the Academic Potential Ability Test (TKPA) and the Basic Ability Test of Science and Technology (TKD SAINTEK). From year to year, the scoring system of the written test only emphasises on the total score as a benchmark for the eligibility of the prospective students to be accepted, without considering the values of their competencies and the study programs they have chosen. For those who have chosen certain study programs of science such as Physics, Chemistry, Biology, Mathematics, Actuarial Science, and Statistics, their competencies in the Basic Abilities, which are directly corresponding to the study programs they chose, need to be evaluated, bearing in mind that the study programs only focus on the specific abilities shown by the scores of TKD SAINTEK. Specifically, this study aimed to model the status of the acceptance of prospective students by using a mixture, namely the suitability of the TKD SAINTEK scores with the chosen study program. Hopefully, the results of the model classification accuracy can be used as a consideration in the scoring system in the next years, so that it will not only focus on the total score, but also on the competency of each test participant. The data analysis in this study used the data-driven concept so that the multimodal-patterned data used the Bernoulli Mixture Model (BMM). The Monte Carlo Markov Chain (MCMC) method was employed to parse the complex estimation process in the Bayesian method. The convergence test was carried out inferentially using three methods, namely of Gelman-Rubin, Geweke, and Raftery-Lewis. The model obtained in this study provides the classification accuracy levels for Physics, Chemistry, Biology, Actuarial Science, Mathematics, and Statistics of 73.02%, 73.91%, 68.13%, 71.18%, 63.54%, and 50%, respectively. The exact convergence testing methods used in this study were those of Geweke and Raftery-Lewis.

**Keywords:** Bernoulli Mixture, Writing Exam, MCMC, Data Driven.

*(Halaman ini sengaja dikosongkan)*

## KATA PENGANTAR

Allahuakbar, Allahuakbar, Allahuakbar, Segala puji dan syukur penulis panjatkan atas kehadiran Allah SWT yang telah memberikan rahmat, hidayah dan kekuatannya serta shalawat dan salam selalu tercurah pada Nabi Muhammad SAW atas suri tauladannya dalam kehidupan ini. Alhamdulillah, setelah drama panjang dalam kehidupan, penulis dapat menyelesaikan Tesis yang berjudul **“Pemodelan Pada Data Status Penerimaan Calon Mahasiswa Baru Melalui Jalur Bernoulli Mixture Model Ujian Tulis Bidang Saintek Tahun 2018”**. terselesaikannya Tesis ini tidak lepas dari bantuan berbagai pihak, dan secara personal penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Dr. Dra. Ismaini Zain, M.Si dan Prof. Nur Iriawan, M.Ikom., Ph.D selaku dosen pembimbing yang telah banyak meluangkan waktunya untuk memberikan ilmu, bimbingan, arahan dan motivasi yang tiada henti kepada penulis dalam penyelesaian Tesis ini.
2. Dr. Vita Ratnasari, M.Si dan Dr. Sutikno, M.Si selaku dosen penguji yang telah meluangkan waktu serta memberi saran dan kritiknya demi kesempurnaan Tesis ini.
3. Dr. Kartika Fithriasari selaku Ketua Departemen Statistika FSAD-ITS.
4. Dr. rer. pol. Dedy D. Prastyo, M.Si selaku Ketua Program Magister Departemen Statistika FSAD-ITS.
5. Dr. Dra. Ismaini Zain, M.Si selaku dosen wali yang telah memberikan semangat, motivasi dan nasihat-nasihat yang selalu menenangkan hati.
6. Segenap dosen pengajar dan para staf Departemen Statistika FSAD-ITS yang memberikan bekal ilmu, memfasilitasi dan membantu penulis selama masa perkuliahan.
7. Mama, Papa, Adek, Ibu dan Bapak yang selalu memberikan semangat, doa dan motivasi untuk lulus. Terimakasih atas doa dan pemaklumannya selama ini.
8. Suami tercinta, Ary Budi Prakoso yang siap siaga membantu dalam penyusunan *draft* tesis, supply semangat, makanan dan cinta yang gak pernah putus, *i love you to the moon and back!*

9. Elly pusporani sebagai teman *sharing*, dosen darurat dan tempat curhat. Ardianto, Raisya yang ga pernah berhenti *support* dan jadi *reminder* supaya selalu inget sama Tesis.
10. Teman-teman seperjuangan #Wisuda121, Anjar, Lidya, Merin dan kawan-kawan lain yang tidak bisa saya sebutkan satupersatu dalam Tesis, terimakasih atas segala bantuan dan semangatnya. Semoga jalan apapun yang kita ambil setelah ini, selalu di rahmati Allah SWT.
11. Seluruh pihak yang telah banyak membantu penulis dan tidak dapat disebutkan satu per satu.

Dalam penulisan ini, penulis menyadari masih banyak kekurangan dalam penyusunan Tesis ini. Oleh karena itu, penulis sangat berharap pembaca berkenan untuk mengkritik atau memberikan saran yang membangun. Akhir kata, semoga Tesis ini bermanfaat bagi pembaca. Sukses selalu untuk kita semua.

Surabaya, Januari 2020

Penulis



2.5.1	<i>Diagnostic plot</i> .....	19
2.5.2	Metode Inferensia .....	21
2.6	<i>Credible Interval</i> .....	23
2.7	Evaluasi Performansi Untuk <i>Bernoulli Mixture Model</i> .....	23
2.8	Seleksi Penerimaan Mahasiswa Baru Melalui Ujian Tulis.....	24
2.8.1	Ujian Tulis Tahun 2017 .....	24
2.8.2	Ujian Tulis Tahun 2018 .....	26
2.9	OpenBUGS .....	26
<b>BAB 3 METODOLOGI PENELITIAN .....</b>		<b>29</b>
3.1	Sumber Data.....	29
3.2	Variabel Penelitian.....	29
3.2.1	Variabel Respon.....	30
3.2.2	Variabel Prediktor .....	30
3.3	Struktur Data Penelitian.....	32
3.4	Tahapan Analisis Data .....	33
3.5	Diagram Alir .....	35
<b>BAB 4 ANALISIS DAN PEMBAHASAN .....</b>		<b>37</b>
4.1	Kajian Algoritma Bernoulli Mixture Model dan Pemeriksaan Konvergensi .. .....	37
4.1.1	Algoritma <i>Bernoulli Mixture Model</i> pada OpenBUGS .....	37
4.1.2	Pemeriksaan Konvergensi.....	39
4.2	Karakteristik data dan Pemodelan dengan Bernoulli Mixture Model .....	53
4.2.1	Karakteristik Calon Mahasiswa Baru di FSAINS dan FMKSD ....	53
4.2.2	Proporsi dan Keketatan Ujian Tulis di FSAINS dan FMKSD.....	54
4.2.3	Karakteristik Calon Mahasiswa Baru dengan <i>Boxplot</i> dan <i>Interval Plot</i> .....	55
4.2.4	Pemodelan Bernoulli Mixture.....	61

4.2.5	Hasil Pemodelan dengan Metode BMM Pada Fakultas Sains (Fisika, Biologi, Kimia).....	64
4.2.6	Hasil Pemodelan dengan Metode BMM Pada Fakultas Matematika, Komputasi dan Sains Data (Aktuaria, Matematika, Statistika) .....	77
<b>BAB 5 KESIMPULAN DAN SARAN .....</b>		<b>89</b>
5.1	Kesimpulan.....	89
5.2	Saran .....	90
<b>DAFTAR PUSTAKA .....</b>		<b>93</b>
<b>LAMPIRAN.....</b>		<b>99</b>

*(Halaman ini sengaja dikosongkan)*

## DAFTAR TABEL

Tabel 2.1 Analogi Penerimaan calon Mahasiswa Tahun 2018 .....	13
Tabel 2.2 Identifikasi <i>Mixture Bernoulli</i> .....	13
Tabel 2.3 <i>Confusion Matrix Table</i> .....	24
Tabel 3.1 Rincian Jumlah Pengamatan di tiap Program Studi .....	29
Tabel 3.2 Rincian Jumlah Pengamatan di tiap Program Studi .....	30
Tabel 3.3 Struktur Data Penelitian .....	32
Tabel 3.4 Identifikasi <i>Mixture</i> Sesuai dengan Kriteria Penerimaan .....	34
Tabel 4.1 <i>Syntax</i> dari <i>Nodes Bernoulli Mixture Model</i> .....	38
Tabel 4.2 <i>Pseudo Prior</i> untuk Proses MCMC Tahap Pertama Prodi Fisika .....	40
Tabel 4.3 Hasil Uji <i>Geweke</i> Tahap Pertama Prodi Fisika .....	43
Tabel 4.4 Hasil Uji <i>Raftery-Lewis</i> Tahap Pertama Prodi Fisika .....	43
Tabel 4.5 <i>Prior</i> dan <i>Initial</i> untuk Proses MCMC Tahap Kedua Prodi Fisika .....	44
Tabel 4.6 Hasil Uji <i>Geweke</i> Tahap Kedua Prodi Fisika .....	47
Tabel 4.7 Hasil Uji <i>Raftery-Lewis</i> Tahap Kedua Prodi Fisika .....	47
Tabel 4.8 <i>Prior</i> dan <i>Initial</i> Tahap Ketiga Prodi Fisika .....	48
Tabel 4.9 Hasil Uji <i>Geweke</i> Tahap Ketiga Prodi Fisika .....	50
Tabel 4.10 Hasil Uji <i>Raftery-Lewis</i> Tahap Ketiga Prodi Fisika .....	51
Tabel 4.11 Perbandingan Hasil Konvergensi .....	52
Tabel 4.12 Rincian Jumlah Pendaftar dan Daya Tampung Tiap Program Studi ..	55
Tabel 4.13 Nilai TKD dan Prodi yang Berkoresponden .....	61
Tabel 4.14 Deskripsi Data Penerimaan Calon Mahasiswa Baru Berdasar Kriteria P .....	62
Tabel 4.15 Pembagian Data <i>Training</i> dan <i>Testing</i> pada FMKSD .....	63
Tabel 4.16 Pembagian Data <i>Training</i> dan <i>Testing</i> pada FSAINS .....	64
Tabel 4.17 Pemeriksaan Konvergensi dengan <i>Raftery-Lewis</i> Prodi Fisika .....	67
Tabel 4.18 Hasil Estimasi Parameter Untuk Prodi Fisika .....	67
Tabel 4.19 Hasil Klasifikasi Pada Tiap <i>Mixture</i> Prodi Fisika .....	69
Tabel 4.20 Rincian Prediksi Status Penerimaan Pada <i>Mixture 1</i> Prodi Fisika .....	69
Tabel 4.21 Rincian Prediksi Status Penerimaan Pada <i>Mixture 2</i> Prodi Fisika .....	70
Tabel 4.22 Pemeriksaan Konvergensi dengan <i>Raftery-Lewis</i> Prodi Kimia .....	70
Tabel 4.23 Hasil Estimasi Parameter Untuk Prodi Kimia .....	71
Tabel 4.24 Hasil Klasifikasi Pada Tiap <i>Mixture</i> Prodi Kimia .....	72
Tabel 4.25 Rincian Prediksi Status Penerimaan Pada <i>Mixture 1</i> Prodi Kimia .....	72
Tabel 4.26 Rincian Prediksi Status Penerimaan Pada <i>Mixture 2</i> Prodi Kimia .....	73
Tabel 4.27 Pemeriksaan Konvergensi dengan <i>Raftery-Lewis</i> Prodi Biologi .....	74
Tabel 4.28 Hasil Estimasi Parameter untuk Prodi Biologi .....	74
Tabel 4.29 Hasil Klasifikasi Pada Tiap <i>Mixture</i> Prodi Biologi .....	75
Tabel 4.30 Rincian Prediksi Status Penerimaan Pada <i>Mixture 1</i> Prodi Biologi ...	76
Tabel 4.31 Rincian Prediksi Status Penerimaan Pada <i>Mixture 2</i> Prodi Biologi ...	76
Tabel 4.32 Variabel yang Signifikan pada Tiap <i>Mixture</i> .....	77

Tabel 4.33 Hasil Uji <i>Raftery-Lewis</i> Prodi Aktuaria .....	78
Tabel 4.34 Hasil Estimasi Parameter Prodi Aktuaria .....	78
Tabel 4.35 Hasil Klasifikasi Pada Tiap <i>Mixture</i> Prodi Aktuaria.....	79
Tabel 4.36 Rincian Prediksi Status Penerimaan Pada <i>Mixture</i> 1 Prodi Aktuaria..	79
Tabel 4.37 Rincian Prediksi Status Penerimaan Pada <i>Mixture</i> 2 Prodi Aktuaria..	80
Tabel 4.38 Hasil Uji <i>Raftery-Lewis</i> Prodi Matematika .....	80
Tabel 4.39 Hasil Estimasi Parameter untuk Prodi Matematika.....	81
Tabel 4.40 Hasil Klasifikasi Pada Tiap <i>Mixture</i> Prodi Matematika.....	82
Tabel 4.41 Rincian Prediksi Status Penerimaan Pada <i>Mixture</i> 1 Prodi Matematika .....	83
Tabel 4.42 Rincian Prediksi Status Penerimaan Pada <i>Mixture</i> 2 Prodi Matematika .....	83
Tabel 4.43 Pemeriksaan Konvergensi dengan <i>Raftery-Lewis</i> Prodi Statistika.....	84
Tabel 4.44 Hasil Estimasi Parameter untuk Prodi Statistika .....	84
Tabel 4.45 Hasil Klasifikasi Pada Tiap <i>Mixture</i> Prodi Statistika.....	85
Tabel 4.46 Rincian Prediksi Status Penerimaan Pada <i>Mixture</i> 1 Prodi Statistika.	86
Tabel 4.47 Rincian Prediksi Status Penerimaan Pada <i>Mixture</i> 2 Prodi Statistika.	86
Tabel 4.48 Variabel yang Signifikan pada Tiap Prodi di FMKSD .....	87

## DAFTAR GAMBAR

Gambar 2.1 Visualisasi Plot ACF, (a) Belum Memenuhi Syarat; (b) Memenuhi Syarat.....	19
Gambar 2.2 Visualisasi <i>Density Plot</i> ; (a) belum memenuhi syarat; (b) Memenuhi Syarat.....	20
Gambar 2.3 Visualisasi <i>Serial plot</i> ; (a) belum memenuhi syarat; (b) Memenuhi Syarat.....	20
Gambar 3.1 Kerangka Konseptual antar Variabel .....	32
Gambar 3.2 Alur Pembagian Data <i>Training</i> dan <i>Testing</i> .....	34
Gambar 3.3 Diagram Alir Penelitian .....	35
Gambar 3.4 Tahapan Preprocessing Data .....	35
Gambar 3.5 Tahapan dalam Pemodelan dengan BMM .....	36
Gambar 4.1 Struktur DAG Bernoulli <i>Mixture Model</i> .....	37
Gambar 4.2 <i>Density Plot</i> Tahap Pertama Prodi Fisika .....	41
Gambar 4.3 <i>Autocorrelation Plot</i> Tahap Pertama Prodi Fisika .....	42
Gambar 4.4 <i>Serial Plot</i> Tahap Pertama Prodi Fisika .....	42
Gambar 4.5 <i>Density plot</i> Tahap Kedua Prodi Fisika.....	45
Gambar 4.6 <i>Autocorrelation Plot</i> Tahap Kedua Prodi Fisika.....	46
Gambar 4.7 <i>Serial Plot</i> Tahap Kedua Prodi Fisika.....	46
Gambar 4.8 <i>Density plot</i> Tahap Ketiga Prodi Fisika .....	49
Gambar 4.9 <i>Autocorrelation Plot</i> Tahap Ketiga Prodi Fisika.....	49
Gambar 4.10 <i>Serial Plot</i> Tahap Ketiga Prodi Fisika.....	50
Gambar 4.11 Proporsi Penerimaan Calon Mahasiswa di ITS Secara Keseluruhan .....	53
Gambar 4.12 Proporsi Penerimaan Calon Mahasiswa Di FSAINS Dan FMKSD	54
Gambar 4.13 <i>Interval Plot</i> pada Tiap Prodi di FSAINS .....	56
Gambar 4.14 <i>Interval Plot</i> pada Tiap Prodi di FMKSD .....	57
Gambar 4.15 <i>Boxplot</i> pada Tiap Prodi di FSAINS .....	58
Gambar 4.16 <i>Box Plot</i> pada Tiap Prodi di FMKSD.....	60
Gambar 4.17 <i>Serial Plot</i> dari Hasil Estimasi Parameter BMM Prodi Fisika.....	65
Gambar 4.18 <i>Autocorrelation Plot</i> Prodi Fisika .....	65
Gambar 4.19 <i>Density plot</i> dari Hasil Estimasi Parameter BMM Prodi Fisika.....	66

*(Halaman ini sengaja dikosongkan)*

## DAFTAR SINGKATAN

BMM	:	<i>Bayesian Mixture Model</i>
MCMC	:	Markov Chain Monte Carlo
UTBC	:	Ujian Tulis Berbasis Cetak
UTBK	:	Ujian Tulis Berbasis Komputer
TKPA	:	Tes Kemampuan dan Potensi Akademik
TKD	:	Tes Kemampuan Dasar
TPA	:	Tes Potensi Akademik
IPA	:	Ilmu Pengetahuan Alam
Saintek	:	Sains dan Teknologi
FSAINS	:	Fakultas Sains
FMKSD	:	Fakultas Matematika, Komputasi dan Sains Data
SMA	:	Sekolah Menengah Atas
MA/MI	:	Madrasah Aliyah/Madrasah Ibtidaiyah
ITS	:	Institut Teknologi 10 Nopember
TPS	:	Tes Potensi Skolastik

*(Halaman ini sengaja dikosongkan)*

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Untuk mendapatkan calon mahasiswa yang berkualitas, pemerintah membuat beberapa sistem seleksi masuk perguruan tinggi. Berdasar proses seleksi, seleksi penerimaan mahasiswa baru dibagi menjadi dua, yaitu ujian tulis dan jalur prestasi. Untuk jalur prestasi, proses seleksi dilakukan dengan melihat hasil belajar siswa selama di bangku SMA/SMK/MA yang ditunjukkan dengan nilai rapor. Sedangkan untuk ujian tulis, proses seleksi dilakukan dengan mengadakan ujian bersama secara serentak yang dapat diikuti oleh siswa SMA/SMK/MA lulusan tahun 2016, 2017 dan 2018. Ujian tulis dibedakan menjadi dua berdasar penggunaan alat tes, yaitu Ujian Tulis berbasis Cetak (UTBC) dan Ujian Tulis berbasis Komputer (UTBK).

Ujian tulis dilakukan dengan cara mengukur kompetensi dari calon mahasiswa baru, yang mana kompetensi ini akan diukur dengan mengujikan 10 mata uji yang terbagi ke dalam dua kelompok tes. Kelompok tes pertama yaitu tes kemampuan potensi akademik (TKPA) yang terdiri dari 6 mata ujian meliputi : TPA Verbal, TPA Figural, TPA Numerik, Tes Matematika Dasar, Tes Bahasa Inggris dan Tes Bahasa Indonesia. Kelompok tes kedua yaitu tes kemampuan dasar (TKD) yang dibagi menjadi dua kategori yaitu Tes Kemampuan Dasar Sains dan Teknologi (TKD Saintek) yang terdiri dari 4 mata uji meliputi : Matematika IPA, Kimia, Fisika dan Biologi.

Keseluruhan mata uji yang diujikan pada ujian tulis ini dirancang sedemikian rupa untuk mengukur kemampuan TKPA dan TKD dari calon mahasiswa baru secara menyeluruh. Sistem penilaian yang digunakan sebagai acuan dalam penerimaan calon mahasiswa baru adalah total nilai dari seluruh mata uji. Dari tahun ke tahun sistem penilaian hanya berfokus pada total nilai hasil ujian, yang kemudian untuk keputusan penerimaan akan diurutkan dari nilai tertinggi dengan batas *cut-off* penerimaan adalah daya tampung pada setiap program studi (prodi). Secara singkat, dapat dikatakan bahwa sistem penilaian pada proses seleksi calon mahasiswa baru hanya berdasar pada total nilai secara

keseluruhan, tanpa mempertimbangkan kompetensi dari calon mahasiswa baru dan prodi pilihan.

Penilaian tanpa mempertimbangkan nilai permata uji dan program studi yang dituju oleh calon mahasiswa tentunya akan memberikan efek yang perlu untuk di evaluasi, karena tidak ada jaminan bahwa calon mahasiswa yang mendapat nilai total lebih kecil akan memiliki kemampuan yang lebih rendah dibanding mahasiswa yang memiliki nilai total lebih tinggi. Sebagai contoh : mahasiswa A, B dan C memilih matematika sebagai program studi pilihan, kemudian mengikuti tes. Dari hasil perhitungan nilai total, calon mahasiswa A dan C diterima pada prodi matematika, sedangkan calon mahasiswa B tidak diterima. Saat dikaji lebih lanjut nilai yang berkoresponden dengan prodi pilihan, yaitu nilai matematika, dari calon mahasiswa A dan C mendapatkan nilai jauh lebih kecil dibanding calon mahasiswa B, padahal secara logika mahasiswa yang masuk jurusan matematika seharusnya lebih kompeten di bidang matematika. Hal inilah yang menjadi *concern* peneliti untuk melakukan penelitian dengan harapan hasil dari penelitian ini dapat memberikan pertimbangan dalam penentuan kebijakan sistem penilaian dalam lingkup seleksi penerimaan mahasiswa baru pada tahun-tahun yang akan datang secara nasional.

Data yang digunakan pada penelitian ini merupakan data status penerimaan calon mahasiswa baru melalui jalur ujian tulis bidang saintek di Institut Teknologi Sepuluh Nopember (ITS), khususnya pada Fakultas Matematika, Komputasi dan Sains Data (FMKSD) dan Fakultas Sains (FSAINS). Respon pada penelitian ini merupakan data biner dengan keterangan nol (0) jika calon mahasiswa tidak diterima dan 1 jika calon mahasiswa diterima. Penelitian ini berfokus untuk mengevaluasi status penerimaan calon mahasiswa baru dengan mengaitkan antara nilai total dengan kriteria penerimaan. Kriteria penerimaan yang akan dibentuk pada penelitian ini berasal dari perbandingan nilai TKD. Nilai TKD akan digunakan sebagai kriteria penerimaan karena nilai TKD pada FMKSD dan dianggap sebagai nilai mata uji dasar yang harus dikuasai oleh calon mahasiswa baru dan juga sebagai identitas tolak ukur kompetensi untuk masuk pada program studi yang dituju. Penentuan kriteria, ditentukan berdasarkan pada pertimbangan peneliti, karena data biner tidak dapat diplotkan ataupun dibuat

histogram. Kriteria penerimaan akan dikatakan benar jika nilai TKD yang berkoresponden dengan prodi memiliki nilai yang lebih tinggi dibandingkan dengan tiga mata ujian TKD lainnya dan sebaliknya, dikatakan salah jika nilai TKD yang berkoresponden dengan prodi memiliki nilai yang lebih rendah dibandingkan dengan tiga mata ujian TKD lainnya.

Konsep analisis data *driven* berperan dalam pembentukannya kriteria penerimaan dalam penelitian ini, karena sesuai dari pengertian dari konsep data *driven* yaitu analisis data akan mempertahankan macam dan sifat data sesuai dengan data aslinya (Iriawan, 2012). Analisis dengan menggunakan data *driven* menyebabkan pola data yang multimodal. Data multimodal ini dapat diselesaikan dengan bantuan *mixture* dalam analisisnya, sehingga metode Bernoulli *Mixture Model* (BMM) akan digunakan pada penelitian ini. *Mixture* pada penelitian ini ditunjukkan dengan kriteria penerimaan. Untuk penelitian dengan data biner pembentukan *mixture* ditentukan berdasar pada pertimbangan penulis, karena data biner tidak dapat diplotkan. Sebelum menggunakan metode BMM, peneliti mengusulkan untuk menggunakan metode analisis regresi logistik multilevel, dengan level satu adalah nilai TKPA dan TKD dari tiap calon mahasiswa atau individu dan level dua adalah daya tampung pada tiap program studi.

Pemodelan dengan menggunakan analisis regresi logistik multilevel terbatas pada kriteria yang diinginkan, yaitu kompetensi dengan memperhatikan prodi tujuan. Metode BMM tidak memberi batasan jumlah *mixture* yang digunakan atau bisa dikatakan lebih fleksibel, bahkan penambahan *mixture* secara hierarki juga dapat dilakukan. Fleksibilitas dalam penentuan kriteria inilah yang menjadi salah satu faktor metode BMM dirasa akan lebih tepat untuk diaplikasikan pada data ini. Penggunaan model *mixture* atau campuran pada data menunjukkan keunggulan dibandingkan dengan metode statistik lainnya (Stephens, 1997). Secara umum pemodelan dengan analisis *mixture* yaitu dengan cara melakukan pengelompokan data, kemudian melakukan estimasi parameter di tiap kelompok data yang terbentuk.

Analisis dengan menggunakan metode *mixture* menghasilkan estimasi yang tidak *closed form*, sehingga untuk mengestimasi model yang kompleks dapat dilakukan dengan menggunakan pendekatan Bayesian dengan bantuan rantai

markov dengan iterasi monte carlo atau yang populer disebut metode *Markov Chain Monte Carlo* (MCMC). Penggunaan metode MCMC diterapkan untuk mengurai proses estimasi sehingga lebih mudah dilakukan (Ye, *et.al.*, 2018). Kombinasi Bayesian MCMC dengan *mixture* dapat menghasilkan parameter yang lebih akurat, karena parameter yang diestimasi memiliki distribusi marginal posterior tertentu (Dapiah, 2018).

Penelitian dengan menggunakan metode BMM pernah dilakukan dalam rangka memodelkan penerimaan beasiswa bidikmisi serta membandingkan ketepatan klasifikasi antara metode BMM dengan *Bayesian Binary Logistic Regression*. Hasil yang didapat dari penelitian ini adalah metode BMM mampu mengklasifikasikan penerimaan beasiswa bidikmisi lebih baik, dibanding dengan metode *Bayesian Binary Logistic Regression*. (Suryaningtyas, *et. al.*, 2018) Penelitian lain yang berhubungan dengan BMM adalah penelitian tentang pengkategorian kata, dengan tujuan untuk mengkategorisasikan kata dalam rangka memproses pengenalan kata, identifikasi topic dan translasi kata. Peneliti menyebutkan bahwa untuk kategorisasi kata memiliki problem yang kompleks utamanya yaitu dalam memilih kumpulan model data yang sesuai, pemodelan dengan menggunakan *Bernoulli mixture* mampu menyelesaikan permasalahan tersebut dan memberikan hasil pengelompokkan kata yang relevan (Gonzales, *et. al.*, 2001).

Penelitian yang dilakukan oleh Neal (2015) dengan tujuan mengklusterkan data biner dengan metode BMM, merupakan salah satu jurnal yang mendasari penelitian ini. Hasil dari penelitian menyebutkan bahwa melakukan klaster, dengan komponen K yang tidak diketahui, dengan menggunakan metode *Bernoulli Mixture Model* (BMM) dengan bantuan metode MCMC membantu mempermudah dan mempercepat estimasi. MCMC memberikan hasil yang lebih akurat, dikarenakan hasil estimasi parameter dari metode BMM mendekati hasil yang sebenarnya. Penelitian selanjutnya yaitu *penelitian* penggunaan BMM untuk *machine learning* yang bertujuan membandingkan hasil klasifikasi *machine learning* dengan metode BMM dengan metode *machine learning* lainnya (Saeed, *et. al.*, 2013). Proses MCMC dua tahap, yaitu penggunaan hasil estimasi parameter yang belum konvergen sebagai *prior*

pada proses MCMC selanjutnya hingga mencapai konvergen, juga membantu dalam mempercepat proses estimasi parameter (Efendiev, *et. al.*, 2013).

Hasil estimasi parameter dengan menggunakan MCMC algoritma *gibbs sampling* dianggap mampu memberikan hasil yang akurat dan mendekati hasil yang sebenarnya (Susanto, *et al.*, 2018). Tetapi, pada kenyatannya penentuan hasil estimasi yang baik dan sesuai cukup sulit dilakukan mengingat syarat utama untuk mendapatkan hasil estimasi yang mendekati sebenarnya adalah melihat konvergenitas dari rantai markov. Selama ini, pemeriksaan konvergensi hanya diukur berdasar pada *diagnostic plot* yang didapatkan dari hasil estimasi parameter dengan memeriksa *density plot*, *serial plot* dan *autocorrelation plot*. Tetapi pada penerapannya, penggunaan *diagnostic plot* bukan satu-satunya cara untuk memeriksa konvergensi, ada beberapa cara lain yang dapat digunakan untuk pemeriksaan konvergensi dengan lebih tepat yaitu secara inferensia dengan menggunakan 3 metode pemeriksaan konvergensi yaitu *Geweke*, *Raftery-Lewis* dan *Gelman-Rubin* (Cowles & Carlin, 1996).

Metode Inferensia dalam pemeriksaan konvergensi ini menjadi perlu untuk dilakukan, mengingat penggunaan *diagnostic plot* pada pemeriksaan konvergensi saja belum cukup sehingga pada penelitian ini akan dikaji lebih dalam tentang metode inferensia untuk pemeriksaan konvergensi, sehingga hasil estimasi parameter yang dihasilkan benar-benar hasil yang tepat dan baik dalam menggambarkan hasil pemodelan data. Dari latar belakang yang telah dijabarkan, maka penelitian ini akan mengadopsi penggunaan *Bernoulli Mixture Model* yang mana proses estimasi parameternya akan dilakukan dengan MCMC algoritma *gibbs sampling*. Untuk mempercepat proses estimasi parameter, maka penerapan proses MCMC dua tahap akan dilakukan dan pemeriksaan konvergensi akan dilakukan dengan membandingkan metode inferensia dengan metode *diagnostic plot*.

## **1.2 Rumusan Masalah**

Dari uraian latar belakang yang telah dijelaskan akan disusun rumusan masalah sebagai berikut.

1. Bagaimana pembentukan algoritma *Bernoulli Mixture Model* dan pemeriksaan konvergensi secara inferensia ?
2. Bagaimana karakteristik data dan pemodelan pada data status penerimaan calon mahasiswa baru melalui jalur ujian tulis bidang saintek dengan menggunakan metode *Bernoulli Mixture Model*?

### **1.3 Tujuan Penelitian**

Berdasarkan rumusan masalah, diperoleh tujuan penelitian sebagai berikut.

1. Mengkaji tentang pembentukan algoritma *Bernoulli Mixture Model* dan pemeriksaan konvergensi dengan metode inferensia
2. Mendeskripsikan karakteristik data dan mendapatkan pemodelan pada data status penerimaan calon mahasiswa baru melalui jalur ujian tulis bidang saintek dengan menggunakan metode *Bernoulli Mixture Model*

### **1.4 Manfaat Penelitian**

Manfaat yang diharapkan dapat diperoleh dari penelitian ini adalah sebagai berikut.

1. Manfaat untuk penulis, yaitu penulis dapat mengembangkan wawasan dan mendapat ilmu baru tentang *Bernoulli Mixture Model*
2. Manfaat yang diperoleh pembaca, yaitu pembaca dapat mengembangkan atau menyempurnakan metode *Bernoulli Mixture Model*
3. Manfaat secara luas, utamanya untuk panitia ujian tulis, yaitu dapat menggunakan penelitian ini sebagai bahan evaluasi serta pertimbangan dalam penentuan atau kebijakan penerimaan calon mahasiswa baru.

### **1.5 Batasan Penelitian**

Batasan yang digunakan dalam penelitian ini adalah sebagai berikut.

1. Sumber data yang digunakan adalah data calon mahasiswa baru dari SMA atau MA yang mendaftar di ITS melalui jalur ujian tulis tahun 2018 dengan pilihan uji yaitu kelompok ujian bidang saintek.
2. Prodi yang digunakan dalam penelitian ini hanya berada dibawah naungan FMKSD dan FSAINS.

## BAB 2 TINJAUAN PUSTAKA

### 2.1 Pengertian Distribusi *Bernoulli*

Distribusi *Bernoulli* merupakan distribusi acak diskrit yang menggambarkan suatu fenomena dengan dua hasil, dengan symbol  $p$  adalah probabilitas sukses dan  $q = 1 - p$  adalah probabilitas gagal. Penamaan distribusi *Bernoulli* diambil dari seorang matematikawan berkebangsaan *Swiss* yang bernama *Jacob Bernoulli* (Uspensky, 1937). Distribusi *Bernoulli* merupakan kasus khusus dari distribusi binomial dengan  $n = 1$  jadi hanya dilakukan uji tunggal dengan probabilitas  $p$  jika hasilnya benar atau sukses dan  $q$  jika hasilnya salah atau gagal dengan nilai  $0 < p < 1$  (McCullagh & Nelder, 1989). Distribusi *Bernoulli* memenuhi persamaan (2.1).

$$P(y|p) = p^y(1 - p)^{1-y} \quad (2.1)$$

Nilai  $y$  pada persamaan (2.1) adalah variabel *Bernoulli* yang bernilai nol (0) atau satu (1), yang mana  $p$  menunjukkan probabilitas sukses dan  $1 - p$  menunjukkan probabilitas gagal. Jika nilai  $y$  merupakan variabel random yang berdistribusi *Bernoulli* dengan parameter  $p$ , maka nilai  $y$  dapat dinotasikan seperti pada persamaan (2.2) (Bain & Engelhardt, 1992).

$$y \sim \text{Bern}(p) \quad (2.2)$$

Dengan nilai *mean* dan *variansi* dari distribusi *Bernoulli* dengan parameter  $p$  masing-masing adalah  $\mu = p$  dan  $\sigma^2 = p(1 - p)$

### 2.2 Pemodelan pada Data Mixture

Model Mixture merupakan model probabilistik untuk memodelkan serta mengelompokkan data yang belum terdefinisi menjadi kelompok-kelompok baru yang terdefinisi. model *mixture* merupakan suatu model khusus yang dapat memodelkan data multimodal yang mana data tersebut terdiri atas subpopulasi sebagai komponen penyusun dan memiliki proporsi yang bervariasi. (McLachlan & Basford, 1988) Konsep utama dari *mixture* model adalah menggabungkan beberapa data, baik data yang memiliki distribusi sama atau berbeda, kemudian

menganalisisnya secara bersamaan sehingga akan didapat hasil analisa yang menggambarkan sifat-sifat data dengan lebih teliti lagi (Suryaningtyas, *et. al.*, 2018). Model *mixture* biasanya digunakan untuk menyelesaikan data *univariate* multimodal, yaitu data yang apabila dilakukan uji kebaikan model akan selalu menolak hipotesis nol ( $H_0$ ) (Iriawan, 2001). Fungsi densitas dari model *mixture finite* dapat dilihat pada persamaan (2.3) (McLachlan & Peel, 2000).

$$P(\mathbf{y}|\pi, \theta) = \sum_{k=1}^K \pi_k \theta_k(\mathbf{y}|\theta_k), \quad (2.3)$$

yang mana  $\pi_k$  adalah proporsi komponen *mixture* yang memenuhi persamaan (2.4) untuk  $0 \leq \pi_k \leq 1$ ,  $k = 1, 2, \dots, K$ .

$$\sum_{k=1}^K \pi_k = 1, \quad (2.4)$$

dengan :

$P(\mathbf{y}|\pi, \theta)$  = Fungsi densitas dari model *mixture*

$p_k(\mathbf{y}|\theta_k)$  = Fungsi densitas ke- $k$  dari total  $K$  komponen penyusun model *mixture*

$K$  = Banyak komponen *mixture*

Model *mixture* yang dinyatakan pada persamaan (2.3) berlaku jika  $P(\mathbf{y}|\pi, \theta)$  merupakan kumpulan populasi yang memiliki beberapa model penyusun.

### **2.3 Bernoulli Mixture Model**

Metode *finite mixture models* semakin banyak digunakan dalam pemodelan distribusi pada berbagai macam variasi fenomena random dataset. Model *mixture* adalah salah satu model khusus yang mampu mengakomodir karakteristik multimodal. Data karakteristik multimodal adalah data yang terdiri dari beberapa subpopulasi yang mana tiap subpopulasi merupakan komponen populasi yang tersusun secara *mixture* dengan proporsi yang bervariasi. *Finite mixture models* menjadi salah satu metode pemodelan yang populer karena keflexibelan metode dalam pemodelan dan pengklasteran data (McLachlan & Peel, 2003). Penggunaan model *mixture* pertama kali digunakan untuk

menganalisis data species kepiting dengan menggunakan *mixture* dari dua data yang berdistribusi *mixture* normal (Pearson, 1894).

Pada umumnya penelitian dengan menggunakan *mixture* berfokus pada data kontinu sedangkan penerapannya banyak digunakan untuk analisa pada *machine learning*, yang pemodelannya berdasar pada *gaussian mixture model*. sedangkan pemodelan *mixture* dengan menggunakan data biner merupakan perluasan dari pemodelan *mixture* yang kemudian disebut dengan *Bernoulli Mixture Model* (BMM) (Celeux & Govaert, 1991). Penggunaan BMM banyak digunakan untuk menganalisis *supervised* dan *unsupervised learning* data yang berhubungan dengan *text mining* atau pengkatagorisasian kata (Juan & Vidal, 2002). Selain itu metode BMM juga sangat tepat digunakan untuk mengklasifikasikan gambar pixel dan juga dalam beberapa bidang lain seperti biomedik dan robotika (Saeed, *et. al.*, 2013).

Untuk Sebuah vector random  $= y_1, y_2, \dots, y_N$ , dengan respon biner yang hanya memiliki dua kemungkinan, sukses atau gagal berdistribusi *Bernoulli* univariat hanya memiliki satu parameter, sesuai pada persamaan (2.1). Sedangkan pada data multivariat dengan komponen sebanyak  $K$  dengan  $p_k(\mathbf{y}|p_k)$  berdsitribusi *Bernoulli* akan didapat densitas *Bernoulli mixture* dengan mengaplikasikan persamaan (2.3). Persamaan yang ditunjukkan pada persamaan (2.4) (Saeed, *et. al.*, 2013).

$$P(\mathbf{y}) = \sum_{k=1}^K \pi_k p_k(\mathbf{y}|p_k), \quad (2.4)$$

Tujuan utama dari *Bernoulli mixture* adalah mempartisi  $\mathbf{y}$  ke dalam kelompok  $K$ , nilai  $K$  tidak diketahui tetapi terbatas. Pada penelitian ini, kelompok  $K$  merupakan kondisi penerimaan benar dan kondisi penerimaan salah yang mana masing-masing parameternya berdistribusi *Bernoulli* yang memiliki sebuah dimensi  $d$ , jumlah parameter yang akan dimiliki adalah sebanyak  $K + K \times j = K \times (1 + j)$ . Dengan demikian akan didapatkan probabilitas random vector  $\mathbf{y} = y_1, y_2, \dots, y_N$  dengan respon biner yang komponennya dianggap independent, pada sebuah  $y_{nj} \in \{0,1\}^j$  akan ditunjukkan pada persamaan (2.5)

$$P(\mathbf{y}|\pi, p) = \sum_{k=1}^K \pi_k \prod_{j=1}^J (p_{kj})^{y_{kj}} (1 - p_{kj})^{1-y_{kj}}, \quad (2.5)$$

Dari persamaan (2.5) akan diperoleh fungsi *log likelihood* dari metode BMM yang ditunjukkan pada persamaan (2.6) (Grantham, 2015).

$$L_m = \ln(\mathbf{y}|\pi, p) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k \prod_{j=1}^J (p_{kj})^{y_{kj}} (1 - p_{kj})^{1-y_{kj}} \right\} \quad (2.6)$$

solusi maksimum *likelihood* pada persamaan (2.6) memiliki bentuk yang tidak *closed form*.

### 2.3.1 Bernoulli Mixture Regression Model

Pada model linier sederhana kita mengetahui bahwa sebuah variabel dependen berdistribusi normal  $Y \sim N(\mu, \sigma^2)$  memiliki hubungan yang linier dengan *covariate*  $X_1, X_2, \dots, X_N$ . Pada kenyataannya kondisi dilapangan, data tidak selalu atau bahkan tidak pernah berdistribusi normal. Pada penelitian ini data berdistribusi Bernoulli dengan rata-rata  $p$  dan varian  $q = p(1 - p)$ , sehingga pemodelan dan estimasi parameter pada penelitian ini tidak menggunakan model linier sederhana melainkan dengan menggunakan *Generalize Linier Model* (GLMs). GLMs merupakan salah satu perluasan dari regresi linier sederhana yang mana persamaan pada GLMs dapat mengakomodir selain data normal. Ada tiga komponen utama dalam GLMs, meliputi : (McCullagh & Nelder, 1989)

1. Komponen acak atau *random component*, yaitu komponen  $Y$  yang bebas dan fungsi sebaran peluang  $Y$  termasuk dalam sebaran keluarga eksponensial dengan  $E(Y_n) = \mu_n$
2. Komponen sistematis atau *systematic component*, yaitu kombinasi linear dari  $\mathbf{x} = x_1, x_2, \dots, x_n$  untuk  $n = 1, 2, 3, \dots, N$  dengan parameter  $\boldsymbol{\beta}_n = \beta_1, \beta_2, \dots, \beta_n$  yang menghasilkan penduga linear  $\eta_n$ , dengan  $\eta_n = \beta_0 + \beta_1 x_1 + \dots + \beta_j x_n$
3. fungsi penghubung (link function) disimbolkan dengan  $g(\cdot)$  yang menggambarkan hubungan antara penduga linear  $\eta_n$  dengan nilai tengah  $\mu_n$  atau dapat dituliskan sesuai dengan persamaan 2.7

$$\eta_n = \beta_0 + \beta_1 x_1 + \dots + \beta_j x_n, \quad (2.7)$$

Untuk distribusi Bernoulli dengan *mean*  $p$ , merupakan distribusi acak diskrit yang hanya memiliki hasil biner yaitu 0 dan 1 sehingga untuk menuliskan  $p$  sebagai kombinasi linier dari persamaan (2.8) tidak akan konsisten dalam hukum probabilitas. Dengan demikian untuk dapat disesuaikan dengan hukum probabilitas, maka akan dilakukan transformasi terhadap  $p$ , kemudian disebut  $g(p_n)$ , dengan memetakan interval  $(0, 1)$  ke  $(-\infty, \infty)$ . Adanya hubungan linear ini menyebabkan distribusi dari variabel dependen dapat diubah bentuknya ke dalam distribusi lain seperti *gaussian*, *bernoulli*, *poison* dan *gamma*, sehingga persamaan dapat dituliskan seperti pada persamaan (2.8)

$$g(p_n) = \eta_n = \beta_0 + \beta_1 x_1 + \dots + \beta_j x_n = \beta_0 + \mathbf{x}_n^T \boldsymbol{\beta}, \quad (2.8)$$

dengan link function yang digunakan yaitu memenuhi persamaan (2.9) (Ntzoufras, 2009).

$$g(p_n) = \text{logit}(p_n) = \ln\left(\frac{p_n}{1 - p_n}\right), \quad (2.9)$$

sehingga probabilitas dari  $p$  akan dapat diketahui dari hasil transformasi *link function*  $g(p_n)$

$$\ln\left(\frac{p_n}{1 - p_n}\right) = \mathbf{x}_n^T \boldsymbol{\beta}, \quad (2.10)$$

$$\frac{p_n}{1 - p_n} = \exp(\mathbf{x}_n^T \boldsymbol{\beta}) \quad (2.11)$$

$$p_n = \frac{\exp(\mathbf{x}_n^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_n^T \boldsymbol{\beta})}, \quad (2.12)$$

Metode BMM sendiri merupakan pengembangan dari regresi linier sehingga dengan menggunakan *link function* akan didapatkan distribusi bernoulli yang didefinisikan sebagai fungsi logit seperti pada persamaan (2.9). dengan demikian, distribusi dari densitas komponen *mixture* yang berdistribusi *Bernoulli* pada persamaan (2.4) dapat didefinisikan ulang dengan *link function* dari GLMs ditunjukkan pada persamaan (2.13) (Iriawan, *et. al.*, 2018).

$$P_{mix}(\mathbf{y}|\boldsymbol{\pi}, \boldsymbol{\beta}) = \sum_{k=1}^K \pi_k p_k(\mathbf{y}|\boldsymbol{\beta}_k), \quad (2.13)$$

dengan nilai  $p_k(\mathbf{y}|\mathbf{x}, \boldsymbol{\beta}_k) \sim \text{Bern}(\text{logit}_k(p_n))$ ,  $p_n = (p_1, p_2, \dots, p_K)'$ ,  $\mathbf{x}' = x_1, x_2, \dots, x_n$  dan  $\boldsymbol{\beta}_k = (\beta_1, \beta_2, \dots, \beta_K)'$  Sehingga dari model (2.13) akan dituliskan ulang sesuai dengan persamaan

$$P_{mix}(\mathbf{y}|\mathbf{x}, \boldsymbol{\beta}) = \sum_{k=1}^K \pi_k \text{bern}(p_k), \quad (2.14)$$

dengan mensubstitusi persamaan (2.14) maka akan didapat persamaan BMM sebagai berikut (2.15).

$$p_{mix}(\mathbf{y}|\mathbf{x}, \boldsymbol{\beta}) = \sum_{k=1}^K \pi_k \frac{\exp(x_n^T \boldsymbol{\beta})}{1 + \exp(x_n^T \boldsymbol{\beta})} \quad (2.15)$$

### 2.3.2 Identifikasi Bernoulli Mixture

Pemodelan dengan menggunakan BMM dianggap penting dilakukan karena bantuan *mixture* pada penelitian ini dapat mengevaluasi hasil klasifikasi lebih dalam berdasar kondisi yang diinginkan atau ditetapkan. Penentuan *mixture* penelitian dilakukan dengan memeriksa distribusi dari variable respon. Pemeriksaan distribusi dapat dilakukan dengan dua cara yaitu dengan menggunakan histogram dan *goodness of fit test*. Dalam penelitian ini, data respon merupakan variabel dengan respon biner, sehingga pemeriksaan distribusi tidak dapat dilakukan dengan histogram ataupun dengan *goodness of fit test*. Oleh sebab itu, penentuan *mixture* untuk data respon biner akan didasarkan pada pertimbangan peneliti (Suryaningtyas, et., al., 2018)

*Mixture* yang dibentuk dalam penelitian bertujuan untuk mengevaluasi sistem penilaian yang mempertimbangkan kompetensi calon mahasiswa baru dan prodi pilihan, sehingga memberikan sistem penilaian yang lebih *fair* dan terarah. Terarah dalam artian kemampuan calon mahasiswa pada tiap mata uji dapat terukur. Sebagai contoh sekelompok calon mahasiswa A, B dan C mengikuti ujian tulis bidang saintek memilih biologi sebagai program studi pilihan. Kemudian hasil penilaian menunjukkan calon mahasiswa A mendapat nilai total 610 kemudian nilai calon mahasiswa B adalah 607 dan calon mahasiswa C adalah 616.

Dari ketiga calon mahasiswa jika sistem menggunakan penilaian klasik yang berfokus pada total nilai maka mahasiswa A dan C yang akan lebih direkomendasikan untuk masuk prodi Biologi.

Jika sistem penilaian menggunakan system yang diajukan pada penelitian ini, maka fokus nilai tidak hanya pada nilai total, melainkan juga nilai kompetensi bidang biologi. Hal ini akan lebih logis, mengingat prodi Biologi hanya menginginkan siswanya kompeten di bidang biologi bukan bidang lain. Persoalan lain yang dikhawatirkan adalah jika calon mahasiswa masuk pada prodi Biologi, tetapi kompetensi Biologi tidak baik. Tabel 2.1 merupakan analogi penerimaan calon mahasiswa baru.

**Tabel 2.1** Analogi Penerimaan calon Mahasiswa Tahun 2018

Siswa	TKPA	TKD				Total	Status Penerimaan
		Mat IPA	Kimia	Fisika	Biologi		
A	93	80	85	85	40	610	Diterima
B	90	78	83	63	85	607	Tidak Diterima
C	89	90	88	63	85	616	Diterima

Dari Tabel 2.1 dapat dilihat bahwa siswa A dengan nilai total 610 diterima di jurusan biologi, padahal nilai biologi tidak lebih baik dibanding nilai mata uji lainnya. Oleh sebab itu, sistem penilaian akan di evaluasi dengan metode BMM dengan menggunakan identifikasi mixture yaitu 4 nilai TKD dari hasil ujian tulis bidang saintek pada tahun 2018 yang mana *mixture 1* merupakan kondisi penerimaan benar dan *mixture 2* adalah kondisi penerimaan salah.

**Tabel 2.2** Identifikasi *Mixture Bernoulli*

<i>Mixture</i>	<i>y</i>	<b>P</b>	<b>Komponen</b>
$\pi_1$	0	0	Kondisi Penerimaan benar
	1	1	
$\pi_2$	0	1	Kondisi Penerimaan salah
	1	0	

Kondisi penerimaan dikatakan benar apabila calon mahasiswa berstatus diterima dengan nilai TKD yang sesuai dengan pilihan program studi mendapat nilai lebih tinggi dari nilai TKD lainnya dan calon mahasiswa berstatus tidak diterima dengan nilai TKD yang sesuai dengan pilihan program studi mendapat

nilai lebih rendah dari nilai TKD lainnya dan sebaliknya. Untuk lebih memahami identifikasi *mixture*, dapat dilihat pada Tabel 2.2 Berdasar pada Tabel 2.2, akan didapatkan dua komponen *mixture* bernoulli yaitu kondisi penerimaan salah yang akan di lambangkan dengan  $\pi_2$  dan kondisi penerimaan benar yang akan di lambangkan dengan  $\pi_1$  (Juan & Vidal, 2002).

## 2.4 Analisis Bayesian

Analisis bayesian merupakan salah satu metode statistika yang mulai populer dalam penerapan statistika, hal ini dikarenakan metode bayesian menawarkan cara yang berbeda dalam menguji hipotesis. Hal ini sesuai dengan teorema bayes yaitu model yang terbentuk berdasarkan pada data, bukan dari penelitian sebelumnya. Dalam analisis bayesian semua parameter yang tidak diketahui dianggap sebagai variabel random dan memiliki distribusi peluang. Berada dengan teori statistika klasik (*frequensi*), statistika klasik memperlakukan parameter yang tidak diketahui sebagai variabel tetap (*fixed*).

Teorema bayesian memiliki tiga *point* penting yang pertama adalah pengetahuan yang mendalam terhadap parameter dalam model yang akan diujikan. *Point* ini disebut dengan *prior distribution*, yang mana jika semakin kecil variansi distribusi *prior* maka presisinya akan semakin baik. Poin ke dua adalah informasi dari data itu sendiri, maksudnya adalah karena teorema bayesian berdasar pada pembelajaran dari data, maka akan di cek kembali *likelihood function* dari data tersebut. Poin ketiga yaitu distribusi posterior yang merupakan hasil kombinasi antara distribusi *prior* dengan fungsi *likelihood* dari data pengamatan, persamaan distribusi posterior ditunjukkan oleh persamaan (2.16) (Box & Tiao, 1973); (Ntzoufras, 2009).

$$f(\theta|y) \propto f(y|\theta)f(\theta), \quad (2.16)$$

$f(\theta|y)$  adalah probabilitas *posterior* pada data  $y$ ,  $f(y|\theta)$  adalah *likelihood* dari data dan  $f(\theta)$  merupakan distribusi *prior*. Menurut Iriawan (2001) dalam Pallo (2016), jika analisis bayesian diterapkan pada model *mixture* maka parameter  $p$  dapat disinkronkan sebagai parameter yang memuat semua parameter model penyusun *mixture*. Distribusi *prior* merupakan distribusi awal yang

memberikan informasi mengenai parameter yang mana harus ditentukan terlebih dahulu sebelum menentukan distribusi posteriornya (Box & Tiao, 1973). Penentuan distribusi *prior* dapat diperoleh dari informasi terdahulu tentang parameter. Penentuan distribusi *prior* adalah subjektif, tergantung pada masing-masing peneliti. Ada beberapa jenis distribusi *prior*, untuk lebih jelasnya akan ditunjukkan pada poin berikut (Carlin & Chib, 1995)

1. *Conjugate prior* atau *non conjugate prior*, penentuan penggunaan distribusi ini berdasar pada pola *likelihood* dari data pengamatan.
2. *Proper prior* atau *improper prior*, penelitian penggunaan distribusi ini berdasar pada pemberian bobot pada tiap titik di sepanjang domain parameter berdistribusi *uniform*
3. *Informative prior* atau *non informative prior*, penentuan penggunaan distribusi ini berdasarkan pada ketersediaan informasi pada penelitian sebelumnya pada atau frekuensi distribusi pada data.
4. *Pseudo prior*, penentuan penggunaan distribusi ini berdasar pada nilai yang disertakan dengan hasil elaborasi cara *frequentist*.

Penelitian ini menggunakan *pseudo prior* karena penentuan *prior* pada analisa ini berdasar pada nilai yang disertakan dengan hasil elaborasi cara *frequentist*. Dengan demikian, Dapat dikatakan bahwa pada distribusi *posterior* terkandung informasi dari distribusi *prior* dan data hasil pengamatan (Fungsi *likelihood*). Pada penelitian ini, fungsi *log likelihood* dari *Bernoulli mixture model* ditunjukkan pada persamaan (2.17).

$$L(\mathbf{x}, z | \beta, \pi) = \sum_{n=1}^N \sum_{k=1}^K z_{nk} \left\{ \ln \pi_k + \prod_{j=1}^J \{x_{nj} \ln \beta_{kj} + (1 - x_{nj}) \ln (1 - \beta_{kj})\} \right\} \quad (2.17)$$

sedangkan *prior* pada penelitian ini terdiri dari dua prior yaitu *prior*( $\beta$ ) dan *prior*( $\pi$ ) ditunjukkan secara berurutan pada persamaan (2.18) dan (2.19).

$$p(\beta) = \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2} \left( \frac{\beta_{kj} - \mu}{\sigma} \right)^2 \right] \right\} \quad (2.18)$$

$$p(\pi) = \frac{\Gamma(\sum_{n=1}^N \alpha_{\pi})}{\prod_{n=1}^N \Gamma(\alpha_{\pi})} \prod_{n=1}^N y_n^{\alpha_{\pi}-1} \quad (2.19)$$

Mengacu pada persamaan (2.16), perkalian dari *prior* dan *likelihood* pada persamaan (2.17) maka didapat *joint posterior* yang ditunjukkan pada persamaan (2.20).

$$p(\pi, \boldsymbol{\beta} | y) \propto p(y | x, \boldsymbol{\beta}) p(\boldsymbol{\beta}) \quad (2.20)$$

Sehingga diperoleh persamaan sebagai berikut.

$$\begin{aligned} p(\pi, \boldsymbol{\beta} | y) = & \sum_{n=1}^N \sum_{k=1}^K z_{nk} \left\{ \ln \pi_k + \prod_{j=1}^J \left\{ x_{nj} \ln \beta_{kj} + (1 - x_{nj}) \ln(1 - \beta_{kj}) \right\} \right\} \\ & \times \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2} \left( \frac{\beta_{kj} - \mu_\beta}{\sigma_\beta} \right)^2 \right] \right\} \\ & \times \frac{\Gamma(\sum_{n=1}^N \alpha_\pi)}{\prod_{n=1}^N \Gamma(\alpha_\pi)} \prod_{n=1}^N y_n^{\alpha_\pi - 1} \end{aligned} \quad (2.21)$$

Dengan demikian, perkalian *log likelihood* dari data yang ditunjukkan pada persamaan (2.21) akan menghasilkan perkalian yang tidak *closed form* dan sulit untuk di estimasi. Sehingga akan diselesaikan dengan menggunakan MCMC algoritma *gibbs sampling* (Iriawan, 2003).

#### 2.4.1 Markov Chain Monte Carlo (MCMC) Algoritma Gibbs Sampling

Dalam kurun waktu yang singkat metode *Markov Chain Monte Carlo* (MCMC) menjadi alat statistika populer untuk mengestimasi model yang kompleks. Ide dasar dari MCMC ini adalah membangkitkan data sampel dari distribusi *joint posterior* sesuai proses *markov chain* dengan iterasi *monte carlo* secara iteratif hingga diperoleh kondisi yang konvergen (Ntzoufras, 2009). Penggunaan metode MCMC ini banyak digunakan dalam analisis bayesian karena pada saat proses untuk mendapatkan distribusi posterior, analisis bayesian melibatkan proses integral yang kompleks dan rumit, sehingga metode MCMC digunakan untuk membantu mengurai kekompleksan untuk mendapat distribusi posteriornya (Cowles & Carlin, 1996). Menurut Iriawan (2003) dalam Pallo (2016), penguraian kekompleksan dapat dilakukan dengan menyederhanakan model berdimensi tinggi (multi-dimensi) menjadi uni-variabel dan uni-dimensi, sehingga akan lebih mudah dalam menyelesaikan persamaan.

Konsep dari MCMC adalah membangkitkan kumpulan rantai markov yang konvergen terhadap distribusi target, yaitu distribusi *posterior* dari parameter yang diestimasi. Membangkitkan rantai markov pada proses MCMC dilakukan dengan algoritma *gibbs sampling* dengan membuat *full conditional posterior* untuk setiap parameter *posterior*. Algoritma *gibbs sampling* merupakan suatu teknik simulasi untuk membangkitkan random variabel dari distribusi tertentu tanpa menghitung fungsi kepadatan dari distribusi data (Casella & George, 1992). Pada penelitian ini, *full conditional posterior* yang didapat sesuai pada persamaan (2.21). Sehingga dengan metode *Bayesian* proses estimasi parameter akan dilakukan dengan cara mengintegrasikan terhadap parameternya. Proses integral pada persamaan multi-dimensi akan sangat sulit dilakukan, sehingga proses estimasi parameternya diselesaikan dengan MCMC algoritma *gibbs sampling* sesuai pada algoritma berikut.

Algoritma *Gibbs Sampler* untuk *full conditional posterior*

1. Penentuan nilai awal untuk parameter

$$\beta = \beta_{10}^0, \beta_{11}^0, \beta_{12}^0, \dots, \beta_{1p}^0, \beta_{20}^0, \beta_{21}^0, \beta_{22}^0, \dots, \beta_{2p}^0$$

2. Membangkitkan parameter untuk iterasi sebanyak  $t$  dengan iterasi  $t = 1, 2, \dots, T$  sehingga

Tahap 1  $\beta_{10}^1$  dari  $f(\beta_{10} | \beta_{11}^0, \beta_{12}^0, \dots, \beta_{1p}^0, \beta_{20}^0, \beta_{21}^0, \beta_{22}^0, \dots, \beta_{2p}^0, x, y)$

Tahap 2  $\beta_{11}^1$  dari  $f(\beta_{11} | \beta_{10}^1, \beta_{12}^0, \dots, \beta_{1p}^0, \beta_{20}^0, \beta_{21}^0, \beta_{22}^0, \dots, \beta_{2p}^0, x, y)$

Tahap 3  $\beta_{12}^1$  dari  $f(\beta_{12} | \beta_{10}^1, \beta_{11}^1, \dots, \beta_{1p}^0, \beta_{20}^0, \beta_{21}^0, \beta_{22}^0, \dots, \beta_{2p}^0, x, y)$

⋮

Tahap d  $\beta_{1(p-1)}^1$  dari  $f(\beta_{1(p-1)} | \beta_{10}^1, \beta_{11}^1, \dots, \beta_{1p-1}^0, \beta_{20}^0, \beta_{21}^0, \beta_{22}^0, \dots, \beta_{2p}^0, x, y)$

Sehingga akan diperoleh nilai data bangkitan pada iterasi pertama yaitu  $\beta^1 = \beta_{10}^1, \beta_{11}^1, \dots, \beta_{1p-1}^0, \beta_{20}^0, \beta_{21}^0, \beta_{22}^0, \dots, \beta_{2p}^0, x, y$

3. Setelah mendapatkan estimasi pada iterasi pertama, simpan nilai estimasi kemudian gunakan nilai parameter tersebut untuk membangkitkan nilai parameter pada iterasi  $t + 1$ . Proses ini akan dilakukan hingga iterasi ke- $T$ , hingga mendapatkan hasil yang konvergen.

4. Dari hasil membangkitkan, akan diperoleh data sebanyak  $T$ , kemudian akan dilakukan estimasi karakteristik pada setiap fungsi kepadatan marginal  $f(\beta_{10}), f(\beta_{11}), f(\beta_{12}), \dots, f(\beta_{20}), f(\beta_{21}), \dots, f(\beta_{2p})$ .

Dari algoritma *gibbs sampling* dapat diketahui jika, proses pembangkitan data oleh *gibbs sampling* memiliki konsep yang sama dengan MCMC, yaitu membangkitkan data random pada tiap iterasi ke- $N$  sangat bergantung pada nilai random pada iterasi sebelumnya (ke- $(N-1)$ ) (Iriawan, 2003). Pada penelitian ini, proses MCMC dua tahap akan digunakan pada penelitian ini dengan cara mengambil hasil estimasi parameter yang belum konvergen pada estimasi sebelumnya untuk dijadikan *prior* dan initial pada penelitian yang berikutnya (Minkoff, *et.,al.*, 2019). Proses ini dilakukan dengan tujuan mempercepat proses estimasi untuk mencapai target distribusi.

## 2.5 Pemeriksaan Konvergensi

Kumpulan rantai markov yang dibangkitkan pada proses MCMC harus konvergen terhadap distribusi *posterior*. Rantai markov yang konvergen akan bersifat ergodik apabila memenuhi tiga kriteria (Boldstad, 2007), yaitu tak tereduksi (*irreducible*) yaitu kondisi dimana rantai markov hanya memiliki satu kelas *state* atau keadaan sehingga tidak memungkinkan lagi untuk di reduksi, *aperiodic* yaitu suatu *state* atau keadaan yang memiliki periode sama dengan satu dan berulang positif (*positive reccurent*) yaitu suatu keadaan dimana ketika ada suatu proses pada *state i* maka nilai harapan dari awal mulai hingga proses tersebut kembali ke *state i* adalah bilangan *finite* atau berhingga.

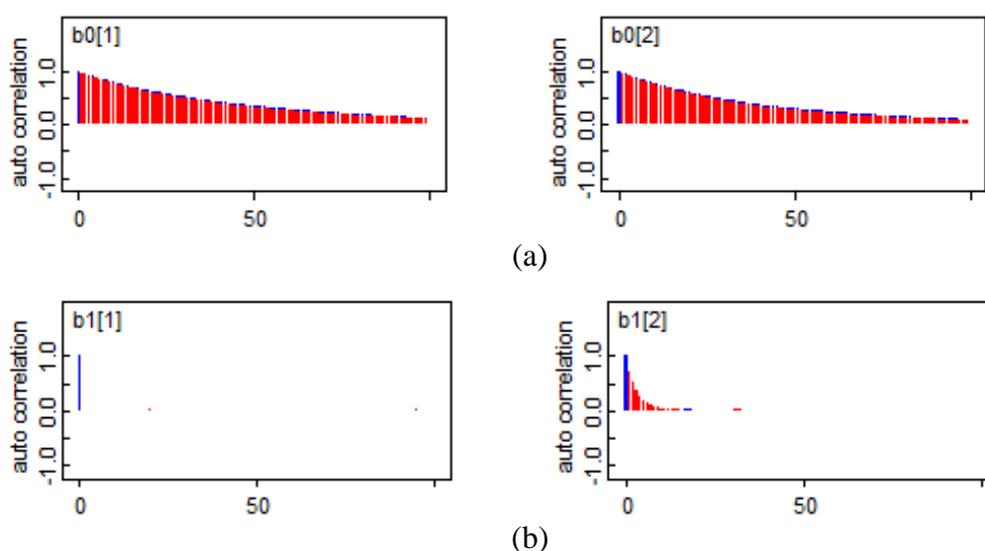
Sifat-sifat ini dapat diperiksa melalui *diagnostic plot* yang didapatkan dari proses estimasi MCMC dengan algoritma *gibbs sampling*. Selama ini, pemeriksaan konvergensi hanya diperiksa berdasar pada *diagnostic plot* yang didapatkan dari hasil estimasi parameter dengan memeriksa *density plot*, *serial plot* dan *autocorrelation plot*. Tetapi pada penerapannya, penggunaan *diagnostic plot* bukan satu-satunya cara untuk memeriksa konvergensi, ada beberapa cara lain yang dapat digunakan untuk pemeriksaan konvergensi dengan lebih tepat yaitu secara inferensia dengan menggunakan 3 metode pemeriksaan konvergensi yaitu *Geweke*, *Raftery-Lewis* dan *Gelman-Rubin* (Cowles & Carlin, 1996).

Penggunaan metode inferensia perlu untuk dilakukan mengingat metode inferensia, memeriksa secara kuantitatif dan memberikan pemeriksaan yang lebih terukur dengan menggunakan ketiga metode tersebut. Pada subbab selanjutnya, akan dibahas lebih rinci tentang pemeriksaan konvergensi.

### 2.5.1 Diagnostic plot

Pemeriksaan konvergensi dapat dilakukan dengan berbagai macam *plot* beberapa diantaranya adalah *autocorrelation plot*, *density plot* dan *serial plot*.

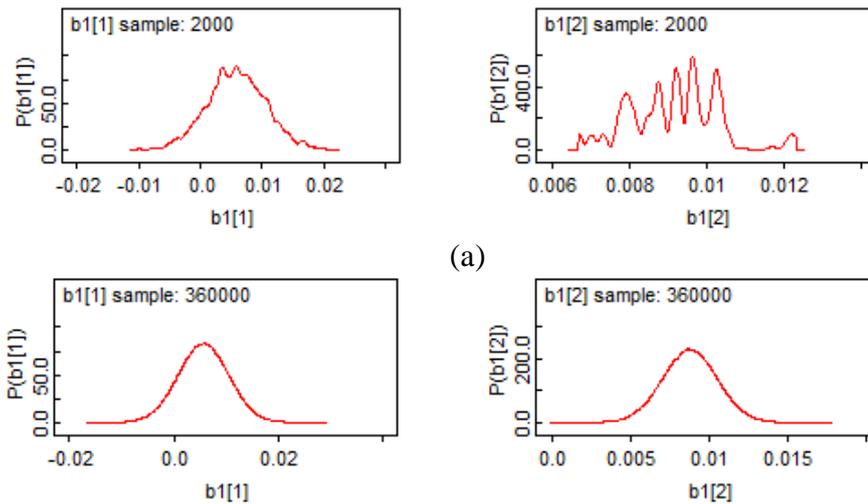
1. *Autocorrelation plot* atau plot yang menggambarkan nilai *Autocorrelations Functions* (ACF). Plot ACF akan dikatakan memenuhi kriteria konvergen jika *lag 0* bernilai satu dan *lag* selanjutnya bernilai nol atau mendekati nol. sebagai ilustrasi Gambar 2.1 (a) dan (b) menunjukkan visualisasi plot ACF yang belum dan sudah memenuhi syarat *ergodic*.



**Gambar 2.1** Visualisasi Plot ACF, (a) Belun Memenuhi Syarat; (b) Memenuhi Syarat

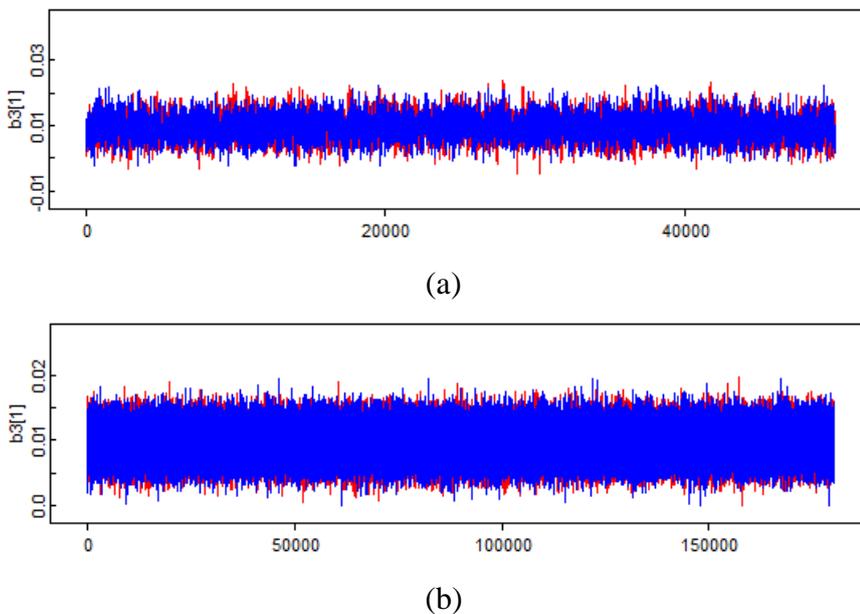
2. *Kernel Density plots* yang merupakan plot dari distribusi pada tiap-tiap parameter. *Prior* pada penelitian ini berdistribusi normal, dikarenakan pembentukan *prior* pada penelitian ini didapat dari hasil estimasi parameter dengan menggunakan teori *link function* dengan fungsi Logit, sehingga estimasi parameter pada penelitian ini juga akan berdistribusi normal. Puncak pada *density plot* merupakan hasil estimasi parameter yang sudah konvergen. Dengan demikian proses estimasi parameter akan

dikatakan konvergen jika hasil visualisasi *density plot* pada tiap parameter telah memiliki satu titik puncak. Sebagai ilustrasi Gambar 2.2 (a) dan (b) menunjukkan visualisasi *density plot* yang belum dan sudah memenuhi syarat konvergen.



**Gambar 2.2** Visualisasi *Density Plot* ; (a) belum memenuhi syarat; (b) Memenuhi Syarat

Berbeda dengan *density plot*, *serial plot* juga memeriksa iterasi pada sampel rantai markov. Perbedaan yang mendasar adalah, jika *density plot* memeriksa sampel pada tiap iterasi maka *serial plot* memeriksa seluruh sampel pada rantai markov. Sebagai ilustrasi Gambar 2.3 (a) dan (b) menunjukkan visualisasi *serial plot* yang belum dan sudah memenuhi syarat konvergen.



**Gambar 2.3** Visualisasi *Serial plot* ; (a) belum memenuhi syarat; (b) Memenuhi Syarat

## 2.5.2 Metode Inferensia

Pengukuran secara inferensia perlu dilakukan mengingat jika hanya dengan menggunakan visual bisa saja terjadi eror dan ada kesulitan tersendiri dalam penentuan kapan dan dalam keadaan seperti apa sebuah rantai markov dikatakan konvergen (Cowles dan Carlin, 1996). Penelitian ini akan mengkaji dan membandingkan *diagnostic plot* dan beberapa metode inferensia untuk pemeriksaan konvergensi antara lain metode Gelman-Rubin (1992), Raftery-Lewis (1992) dan Geweke (1992). Pemeriksaan dengan tiga metode tersebut dapat dilakukan dengan menggunakan CODA *diagnostic*. Disamping itu, *monitoring* nilai *MC Errors* juga dapat dilakukan untuk meyakinkan bahwa hasil estimasi yang diberikan mendekati target distribusinya. Jika nilai *MC Errors* lebih kecil daripada nilai standar deviasi, maka hasil estimasi yang diberikan oleh model mendekati target distribusinya (Ntzoufras, 2009).

### A. Metode Raftery-Lewis

Salah satunya yaitu metode *Raftery-Lewis*. Metode *Raftery-Lewis* digunakan untuk rantai tunggal dengan tujuan mendeteksi ketidak konvergenan ke distribusi stasioner. Dengan menggunakan metode Raftery-Lewis akan didapat jumlah iterasi minimal yang harus dilakukan ( $N_{\min}$ ) serta nilai M yaitu jumlah iterasi yang harus dibuang sebagai *burn-in* (Pratola, 2017). Untuk menggunakan metode Raftery-Lewis ada 4 point utama yang harus dipahami yaitu : Q, R, S, A yang mana

- Q adalah kuantil yang akan diestimasi
- R adalah nilai presisi yang kita inginkan
- S adalah interval probabilitas yang berasosiasi dengan presisi
- A adalah toleransi konvergensi yang digunakan untuk menghitung berapa banyak *burn-in* yang harus dibuang.

Persamaan (2.22) merupakan formula yang digunakan untuk menghitung  $N_{\min}$  (Susanto, *et., al.*, 2017)

$$N_{\min} = \left\{ F^{-1} \left( \frac{S+1}{2} \right) \right\}^2 \frac{Q(1-Q)}{R^2}, \quad (2.22)$$

Suatu rantai markov akan diindikasikan konvergen jika nilai  $I = N/N_{min} < 5$

#### B. Metode Gelman-Rubin (1992)

Metode Gelman-Rubin merupakan salah satu dari dua metode yang paling populer yang digunakan oleh komunitas statistik karena ke praktisan dalam penggunaan program statistiknya (Cowles dan Carlin, 1996). Metode Gelman-Rubin dilakukan dengan cara mendefinisikan rantai markov dengan nilai awal yang berbeda kemudian membandingkan variansi antar rantai markov dengan variansi pada tiap rantai markov. Konvergensi dimonitor dengan cara mengestimasi *potential scale reduction factor* (PSRF) yang memenuhi persamaan (2.19) (Susanto *et. al.*, 2017).

$$\hat{R} = \sqrt{\frac{\widehat{Var}(\Theta)}{avg}}, \quad (2.19)$$

dengan m adalah banyak jumlah rantai markov, kemudian n adalah jumlah iterasi  $n = 1, 2, \dots, N$ , nilai  $\widehat{Var}(\Theta)$  adalah yaitu estimasi variansi dari  $\Theta$  yang memenuhi persamaan  $\widehat{V}(\Theta) = (1 - \frac{1}{n})avg + \frac{1}{n}B$ ,  $avg = \frac{1}{m} \sum_{i=1}^m s_i^2$ , kemudian  $s_i^2 = \frac{1}{n-1} \sum_{n=1}^N (\Theta_i - \bar{\Theta}_i)^2$ , nilai B memenuhi persamaan  $B = \frac{n}{m-1} \sum_{i=1}^m (\bar{\Theta}_i - \bar{\bar{\Theta}})^2$  dan nilai  $\bar{\bar{\Theta}} = \frac{1}{m} \sum_{i=1}^m \bar{\Theta}_i^2$ . Jika nilai PSRF mendekati 1, maka rantai markov dianggap sudah konvergen.

#### C. Metode Geweke

Ide dari metode Geweke adalah melihat rantai markov sebagai *time-series* dalam rangka pengecekan stasioneritas. Metode Geweke mengusulkan diagnosa konvergensi rantai markov berdasar pada uji Z untuk melihat apakah subsampel dari total *output* MCMC adalah sama. Maksud subsampel dalam penelitian ini adalah observasi dari rata-rata dengan cara membagi *generated* rantai markov menjadi  $P_1\%$  dilambangkan dengan A dan  $P_2\%$  dilambangkan B (secara umum  $P_1\%$  sebesar 10% dan  $P_2\%$  sebesar 50% dari total iterasi) yang memiliki nilai rata-rata yang *equal* atau sama. Pada metode Geweke jika nilai  $|Z| < 2$ , maka dapat dikatakan bahwa rantai markov sudah konvergen dan memiliki

nilai rata-rata antara  $P_1\%$  dan  $P_2\%$  yang sama. Persamaan untuk menghitung  $|Z|$  pada metode Geweke dapat dilihat pada persamaan (2.21) (Susanto *et. al.*, 2017).

$$Z = \frac{\bar{\theta}^B - \bar{\theta}^A}{\sigma_{(\bar{\theta}^B - \bar{\theta}^A)}}, \quad (2.21)$$

dengan  $\bar{\theta}^A$  dan  $\bar{\theta}^B$  adalah rata-rata sampel dari dua subsampel dan  $\sigma_{(\bar{\theta}^B - \bar{\theta}^A)}$  adalah standar deviasi dari  $\bar{\theta}^B - \bar{\theta}^A$ .

## 2.6 *pCredible Interval*

Perbedaan mendasar antara pendekatan klasik dan inferensia statistik adalah dalam pembentukan selang kepercayaan. Pada pendekatan klasik, selang kepercayaan dibentuk berdasar pada sebaran dari estimasi parameter. Sebaran dari estimasi parameter kemudian disebut dengan *Confidence Interval*. Berbeda dengan pendekatan analisis *Bayesian*, pendekatan selang kepercayaan metode *Bayesian* didasarkan pada Highest Posterior Density (HPD). Konsep HPD mengacu pada konsep densitas seimbang yang mana pembentukan *interval* dilakukan dengan menggunakan lokasi modus sebagai pusat acuan data.

Misalkan  $Y$  merupakan variabel random dengan PDF  $f(y|p)$  dan  $p$  adalah variabel yang akan diestimasi. Maka  $100(1 - \alpha)\%$  *credible interval* untuk  $p$  akan memenuhi persamaan (2.23) (Marchand, *et. al.*, 2012).

$$P(p \in [a, b]|y) = \int_a^b f(y|p)dp = 1 - \alpha, \quad (2.23)$$

Untuk nilai  $0 < \alpha < 1$ . Uji Hipotesis untuk  $p$ , menggunakan *credible interval*  $H_0: p = 0$  dan  $H_1: p \neq 0$ . Nilai  $H_0$  akan ditolak apabila nilai *credible interval* tidak mengandung nol (0) (Koop, 2003).

## 2.7 Evaluasi Performansi Untuk *Bernoulli Mixture Model*

Evaluasi performansi pada metode BMM digunakan untuk melihat seberapa pengelompokan data di lakukan. Proses evaluasi dilakukan dengan membuat tabel *confusion matrix* yang berisikan informasi tentang data aktual, direpresentasikan pada baris matriks dan data hasil prediksi, direpresentasikan pada

kolom matriks (Han & Kamber, 2006). Tabel *confusion matrix* ditunjukkan pada Tabel 2.3.

**Tabel 2.3** *Confusion Matrix Table*

		Prediksi	
Aktual	Negatif	Positif	
Negatif	TN ( <i>True Negative</i> )	FN ( <i>False Negative</i> )	
	FP ( <i>False Positive</i> )	TP ( <i>True Positive</i> )	

TP atau *True Positive* adalah data aktual kelas positif yang tepat diprediksi sebagai kelas positif. FP atau *False Positive* adalah proporsi data yang salah diklasifikasikan yaitu data aktual kelas positif yang diprediksi sebagai kelas negatif. TN atau *True Negative* adalah data aktual kelas negatif yang tepat diprediksi sebagai kelas negatif. FN atau *False Negative* adalah proporsi data yang salah diklasifikasikan yaitu data aktual kelas negatif yang diprediksi sebagai kelas positif (Salim, 2016). Tabel *confusion matrix* dapat diukur dengan menggunakan nilai akurasi sesuai dengan persamaan 2.24.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.24)$$

## 2.8 Seleksi Penerimaan Mahasiswa Baru Melalui Ujian Tulis

Seleksi penerimaan mahasiswa baru selalu mengalami pembaharuan baik dari segi sistem pembuatan soal maupun sistem penilaian, berikut adalah kajian pustaka tentang perbedaan ujian tulis pada rentang tahun 2017 – 2019.

### 2.8.1 Ujian Tulis Tahun 2017

Tahun 2017 merupakan tahun kedua penyelenggaraan seleksi calon mahasiswa yang langsung diselenggarakan oleh Kemenristekdikti. Pada tahun 2017 ujian tulis dibagi menjadi dua berdasar alat yang digunakan yaitu *Paper Based Test* (PBT) dan *Computer Based Test* (CBT). Untuk materi yang diujikan meliputi :

1. Tes kemampuan dan Potensi Akademik (TKPA)

TKPA terdiri dari 6 mata ujian yaitu matematika dasar, bahasa indonesia, bahasa inggris, TPA verbal, TPA numerikal dan TPA figural.

2. Tes kemampuan Dasar Sains dan Teknologi (TKD Saintek)

TKD Saintek terdiri atas tiga mata ujian pokok yaitu matematika, IPA, kimia, fisika, dan biologi.

3. Tes kemampuan Dasar Sosial dan Humaniora (TKD Soshum)

TKD Soshum terdiri atas empat mata ujian yaitu sosiologi, sejarah, geografi dan ekonomi.

Tidak ada persyaratan khusus yang harus dipersiapkan untuk mengikuti ujian tulis pada tahun 2017. Hanya saja, panitia ujian tulis mensyaratkan peserta ujian memilih satu PTN di dekat daerah domisili dan satu PTN di luar daerah domisilinya. Untuk materi yang diujikan dikelompokkan menjadi 3 yaitu :

1. Kelompok ujian sains teknologi dengan materi ujian utama adalah TKPA dan TKD saintek.
2. Kelompok ujian sosial humaniora dengan materi ujian utama adalah TKPA dan TKD soshum.
3. Kelompok ujian campuran dengan materi ujiann TKPA, TKD Saintek dan TKD Soshum.

Untuk sistem penilaian masih menggunakan sistem klasik +4 -1 sehingga nilai masih berfokus pada nilai total, meskipun soal yang diberikan telah mengadaptasi sistem HOTS. Penggunaan sistem penilaian klasik +4 -1 yang berfokus pada nilai total dirasa kurang adil, dikarenakan penilaian sistem klasik membuat calon mahasiswa hanya akan berfokus pada mata ujian yang dikuasai dan mengesampingkan mata ujian lain yang harusnya juga dikerjakan. Karena tujuan utama dilakukannya penilaian ini adalah mengukur kemampuan calon mahasiswa. Adanya kesenjangan antar mata uji ini menunjukkan bahwa ada ketidakmampuan peserta pada mata ujian tertentu. Hal ini akan terlihat terlihat sangat kontras jika peserta tidak menguasai mata uji A tetapi mendaftar pada program studi yang mensyaratkan mahasiswanya menguasai mata uji A.

### 2.8.2 Ujian Tulis Tahun 2018

Ujian tulis tahun 2018 Tidak jauh berbeda dengan ujian pada tahun 2017. Ujian tulis pada tahun 2018 dibedakan menjadi dua, berdasar penggunaan alat tes yaitu UTBC atau Ujian Tulis berbasis Cetak atau UTBK Ujian Tulis berbasis Komputer. Tidak ada perbedaan antara UTBC dan UTBK hanya penggunaan alat tes saja yang berbeda. Sama dengan tahun 2017, soal-soal ujian yang digunakan pada UTBC dan UTBK merupakan soal-soal yang telah dikembangkan sedemikian rupa sehingga memenuhi persyaratan tingkat kesulitan, validitas dan daya pembeda yang memadai, dengan tujuan yaitu soal-soal yang diselesaikan nanti dapat digunakan sebagai tolak ukur untuk mengukur kemampuan penalaran tingkat tinggi (*higher order thinking skills*) yang meliputi beberapa aspek antara lain potensi akademik, penguasaan bidang studi dasar bidang saintek dan sosial (Kemenristekdikti, 2018). Sedangkan untuk materi yang diujikan pada ujian tulis sama dengan ujian tulis pada tahun 2017 meliputi TKPA, TKD Saintek dan TKD Soshum. Hanya saja sistem penilaiannya berubah, jika pada tahun 2017 masih menggunakan sistem penilaian klasik +4 -1, maka pada tahun 2018 menggunakan sistem *item response theory* (IRT).

IRT adalah sistem *scoring* dengan memperhatikan bobot atau karakteristik pada tiap butir mata uji. Tetapi, meskipun menggunakan sistem IRT, penerimaan calon mahasiswa masih didasarkan pada nilai komposit atau nilai total dari semua mata uji, tanpa memperhatikan nilai permata uji. Karakteristik dari mata uji ini antara lain adalah tingkat kesulitan relatif terhadap soal yang lain (3 kategori : mudah, sedang, sulit) yang mana soal dengan kategori sulit akan memiliki bobot nilai yang lebih tinggi dibandingkan dengan soal dalam kategori mudah.

### 2.9 OpenBUGS

*OpenBugs* merupakan *software* untuk menganalisa model statistika kompleks dengan menggunakan metode MCMC. *BUGS* merupakan singkatan dari *Bayesian inference Using Gibbs Sampling* yang berarti analisis inferensi Bayesian dengan menggunakan *gibbs sampling* (salah satu algoritma dari MCMC). Sejak tahun 1980, proyek *BUGS* dikembangkan oleh Andrew Thomas

bersamaan dengan kecerdasan buatan (AI) dengan tujuan awal yaitu mencari cara yang tepat untuk menyelesaikan ketidakpastian pada struktur grafis, memahami bahwa metode simulasi dapat digunakan untuk inferensi dan memahami bahwa pemrograman yang berorientasi pada objek dapat dieksploitasi untuk algoritma simulasi.

OpenBugs adalah *software open source* yang dapat dijalankan di *windows* ataupun *linux*. OpenBUGS tidak jauh berbeda dengan *WinBUGS*, *OpenBUGS* lebih unggul untuk mengupdate algoritma *Gibbs Sampling* pada tiap *node*, selain itu *OpenBUGS* juga dapat mendefinisikan algoritma tanpa batas, sedangkan *WinBUGS* hanya dapat mendefinisikan satu algoritma di tiap-tiap kelas yang dihasilkan (Lunn, *et.al.*,2013). Sama seperti *WinBUGS*, pemodelan dilakukan dengan membuat struktur gambar yang disebut *Directed Acyclic Graphs* (DAGs) yang mana tiap *node* pada struktur gambar dibangun dari teori matematika, sehingga hal ini mampu memecah kerumitan model menjadi suatu runtutan perhitungan yang relatif lebih mudah.

*(Halaman ini sengaja dikosongkan)*

## BAB 3 METODOLOGI PENELITIAN

### 3.1 Sumber Data

Sumber data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Program Kerja Evaluasi dan Pengembangan SNMPTN dan SBMPTN 2018 (Pokja Evalbang 2018). Data ini berisi tentang data calon mahasiswa baru yang mendaftar di ITS melalui jalur ujian tulis pada bidang saintek dengan batasan hanya 2 departemen yang akan diteliti yaitu Fakultas Sains (FSAINS) dan Fakultas Matematika, Komputasi dan Sains Data (FMKSD). Unit-unit observasi yang digunakan dalam penelitian ini adalah data nilai ujian kelompok TKD dan TKPA dari calon mahasiswa baru yang memilih ITS sebagai PTN pilihan. Data terdiri dari 3,062 pengamatan pada enam program studi berbeda. Rincian pengamatan di tiap program studi akan ditunjukkan pada Tabel 3.1

**Tabel 3.1** Rincian Jumlah Pengamatan di tiap Program Studi

Departemen	Program Studi	Jumlah Pendaftar	Pendaftar Diterima
<b>Fakultas Sains</b>	Biologi	450	40
	Kimia	456	46
	Fisika	308	48
	Total	1,214	136
<b>Fakultas Matematika, Komputasi dan Sains Data</b>	Matematika	477	51
	Statistika	524	44
	Sains Aktuaria	847	27
	Total	1,848	122
<b>Total Keseluruhan</b>		3,062	258

### 3.2 Variabel Penelitian

Variabel yang digunakan pada penelitian ini meliputi variabel dependen dan independen. Data terdiri dari satu variabel dependen dengan distribusi *Bernoulli* dan 10 variabel independen yang terbagi ke dalam dua kelompok, yaitu

kelompok nilai ujian TKPA yang terdiri 6 mata ujian dan TKD yang terdiri dari 4 mata ujian pada Tabel 3.2.

**Tabel 3.2** Rincian Jumlah Pengamatan di tiap Program Studi

Variabel	Deskripsi	Skala Data	Keterangan
$y$	Status Penerimaan	Kategorikal	1 = diterima 0 = tidak diterima
$x_1$	Nilai TPA Figural	Rasio	
$x_2$	Nilai TPA Numerikal	Rasio	
$x_3$	Nilai TPA Verbal	Rasio	
$x_4$	Nilai Bahasa Inggris	Rasio	
$x_5$	Nilai Bahasa Indonesia	Rasio	
$x_6$	Nilai Matematika Dasar	Rasio	
$x_7$	Nilai TKD Biologi	Rasio	<i>Mixture</i>
$x_8$	Nilai TKD Fisika	Rasio	<i>Mixture</i>
$x_9$	Nilai TKD Kimia	Rasio	<i>Mixture</i>
$x_{10}$	Nilai TKD Matematika IPA	Rasio	<i>Mixture</i>

### 3.2.1 Variabel Respon

Variabel respon yang digunakan dalam penelitian ini merupakan data *binary* yang menunjukkan status calon penerimaan mahasiswa baru dari hasil seleksi melalui ujian tulis bidang saintek tahun 2018, dengan keterangan 0 = tidak diterima dan 1 = diterima.

### 3.2.2 Variabel Prediktor

Variabel yang digunakan pada penelitian ini terdiri dari dua kelompok nilai ujian, TKPA dan TKD yang mana masing-masing kelompok memiliki *beberapa* nilai penyusun sebagai berikut :

Kelompok nilai ujian TKPA yaitu :

1. Nilai Ujian Tulis TPA Figural ( $x_1$ )

Ujian ini digunakan untuk mengukur kegesitan atau kecepatan mental calon mahasiswa baru yang berhubungan dengan gambar. Tujuan di lakukannya tes ini adalah untuk mengungkap pencapaian seseorang secara intelektual.

2. Nilai Ujian Tulis TPA Numerikal ( $x_2$ )

Ujian ini digunakan untuk mengukur kegesitan atau kecepatan mental calon mahasiswa baru dalam hal berpikir terstruktur, logis matematis. Selain itu tes ini juga mengukur tingkat ketelitian seseorang dalam menganalisa data.

3. Nilai Ujian Tulis TPA Verbal ( $x_3$ )

Ujian ini digunakan untuk mengukur kecepatan mental calon mahasiswa baru di bidang kata dan bahasa. Seseorang yang memiliki tingkat pemahaman bahasa yang baik menunjukkan bahwa orang tersebut memiliki wawasan dan pengetahuan yang baik pula.

4. Nilai Ujian Tulis Bahasa Inggris ( $x_4$ )

Merupakan mata ujian yang wajib dikerjakan oleh seluruh peserta ujian. Karena mata ujian ini digunakan untuk mengukur kemampuan berbahasa inggris dari para calon mahasiswa baru.

5. Nilai Ujian Tulis Bahasa Indonesia ( $x_5$ )

Merupakan mata ujian yang wajib dikerjakan oleh seluruh peserta ujian. Karena mata ujian ini digunakan untuk mengukur kemampuan berbahasa indonesia dari para calon mahasiswa baru.

6. Nilai ujian tulis Matematika Dasar ( $x_6$ )

Merupakan tes yang digunakan untuk mengukur kemampuan calon mahasiswa baru dalam menyelesaikan soal matematika dasar dengan cepat dan tepat. Mata ujian ini diujikan untuk seluruh peserta ujian baik bidang saintek maupun soshum karena soal matematika dasar lebih mengukur pemahaman mahasiswa terhadap konsep dasar matematika.

Kelompok nilai mata ujian TKD

1. Nilai Ujian Tulis Biologi ( $x_7$ )

Merupakan nilai yang digunakan untuk mengukur kemampuan Biologi.

2. Nilai Ujian Tulis Fisika ( $x_8$ )

Merupakan nilai yang digunakan untuk mengukur kemampuan Fisika.

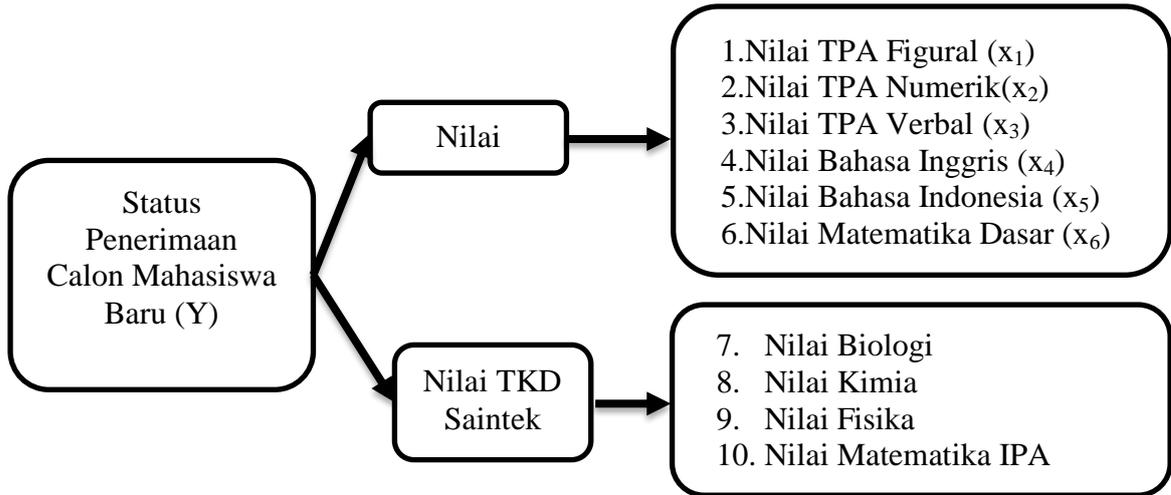
3. Nilai Ujian Tulis Kimia ( $x_9$ )

Merupakan nilai yang digunakan untuk mengukur kemampuan Kimia

4. Nilai Ujian Tulis Matematika IPA ( $x_{10}$ )

Nilai yang digunakan untuk mengukur kemampuan matematika IPA.

Untuk memahami hubungan antar variabel akan dibuat kerangka konseptual dari variable penelitian seperti pada Gambar 3.1.



Gambar 3.1 Kerangka Konseptual antar Variabel

3.3 Struktur Data Penelitian

Struktur data dari status penerimaan calon mahasiswa melalui jalur ujian tulis bidang saintek tahun 2018 ditunjukkan pada Tabel 3.3.

Tabel 3.3 Struktur Data Penelitian

Mixture	Individu	Variabel Respon (y)	Variabel Independen			
			$x_1$	$x_2$	...	$x_6$
Mixture 1	$P_{11}$	$y$	$x_{111}$	$x_{211}$		$x_{611}$
	$P_{21}$	$y_{21}$	$x_{121}$	$x_{221}$		$x_{621}$
	.	.	.	.	.	.
	.	.	.	.	.	.
Mixture 2	$P_{q1}$	$y_{q1}$	$x_{1q1}$	$x_{2q1}$		$x_{6q1}$
	$P_{12}$	$y_{12}$	$x_{112}$	$x_{212}$		$x_{612}$
	$P_{22}$	$y_{22}$	$x_{122}$	$x_{222}$		$x_{622}$
	.	.	.	.	.	.
	$P_{r2}$	$y_{r2}$	$x_{1r2}$	$x_{2r2}$		$x_{6r2}$

keterangan :

- q = banyak calon mahasiswa yang mendaftar pada *mixture* 1
- r = banyak calon mahasiswa yang mendaftar pada *mixture* 2
- $Y_{r2}$  = Nilai data Variabel y pada pendaftar ke – r untuk *mixture* 2
- $x_{1r2}$  = Nilai data Variabel  $x_1$  pada pendaftar ke-r untuk *mixture* 2

### 3.4 Tahapan Analisis Data

Secara umum tahapan analisis data dalam penelitian ini adalah sebagai berikut.

1. Melakukan *preprocessing*.

Proses *preprocessing* data dilakukan dengan cara mem-*filter* data calon mahasiswa baru seluruh Indonesia dan hanya menyisahkan calon mahasiswa yang memilih ITS (baik pilihan pertama, kedua dan ketiga) sebagai PTN pilihan yang dituju, serta mendeskripsikan karakteristik dari data.

2. Identifikasi distribusi *Bernouli Mixture*.

Identifikasi dilakukan dengan cara membentuk kriteria penerimaan (AC) yang didapat dengan cara membandingkan nilai ujian TKD yang berkoresponden dengan prodi pilihan (P) dengan status penerimaan (y). Perbandingan nilai P didapat dengan cara membandingkan nilai TKD yang berkoresponden dengan prodi pilihan dengan tiga nilai TKD lainnya. Nilai P akan bernilai satu (1) jika nilai TKD yang berkoresponden dengan prodi pilihan lebih tinggi dari ketiga nilai TKD yang lain dan sebaliknya.

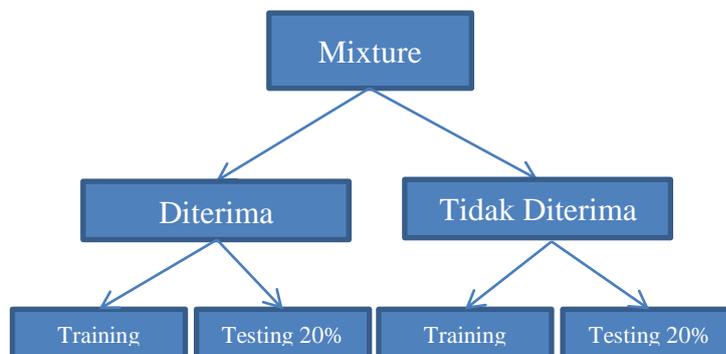
Kriteria penerimaan (AC) akan bernilai benar, jika memenuhi salah satu dari dua kondisi. Kondisi pertama yaitu calon mahasiswa baru berstatus tidak diterima, tetapi nilai TKD yang berkoresponden lebih rendah dan kondisi yang kedua yaitu calon mahasiswa baru berstatus diterima, tetapi nilai TKD yang berkoresponden dengan prodi pilihan rendah. Sebaliknya, kriteria penerimaan (AC) akan bernilai salah jika calon mahasiswa baru berstatus tidak diterima, tetapi nilai TKD yg berkoresponden lebih lebih unggul dibanding mata uji TKD lainnya. Kondisi kedua yaitu calon mahasiswa baru yang berstatus diterima, tetapi nilai TKD yg berkoresponden lebih rendah dari ketiga TKD yang lain. Untuk mempermudah dalam melakukan identifikasi *mixture*, kriteria *mixture* ditunjukkan pada Tabel 3.4.

**Tabel 3.4** Identifikasi Mixture Sesuai dengan Kriteria Penerimaan

Status	Perbandingan Nilai TKD yang berkoresponden dengan Prodi Pilihan	Kriteria Penerimaan	Interpretasi
(y)	(P)	(AC)	
0	0	Kriteria penerimaan benar (1)	calon mahasiswa tidak diterima (y=0) karena nilai TKD yang berkorespondensi dengan prodi pilihan rendah (AC=1)
0	1	kriteria penerimaan salah (0)	calon mahasiswa tidak diterima (y=0) tetapi nilai TKD yang berkorespondensi dengan prodi pilihan tinggi (AC=0)
1	0	kriteria penerimaan salah (0)	calon mahasiswa diterima (y=1) karena nilai TKD yang berkorespondensi dengan prodi pilihan rendah (AC=0)
1	1	Kriteria penerimaan benar (1)	calon mahasiswa diterima (y=1) dengan nilai TKD yang berkorespondensi dengan prodi pilihan rendah (AC=1)

3. Pembagian data *training* dan *testing*.

Pembagian data training dan testing pada penelitian ini harus dilakukan dengan cara bertahap. Pembagian data *training* dan *testing* yang tidak tepat akan mempengaruhi ketepatan akurasi untuk prediksi. Pembagian data *training* dan *testing* untuk penelitian ini perlu mempertimbangkan pembagian kategori *mixture*-nya. Pembagian data *training* dan *testing* tanpa mempertimbangkan faktor *mixture* pada data ini, akurasi yang diberikan akan semakin kecil. Alur pembagian data *training* dan *testing* pada penelitian ini ditunjukkan pada Gambar 3.2.

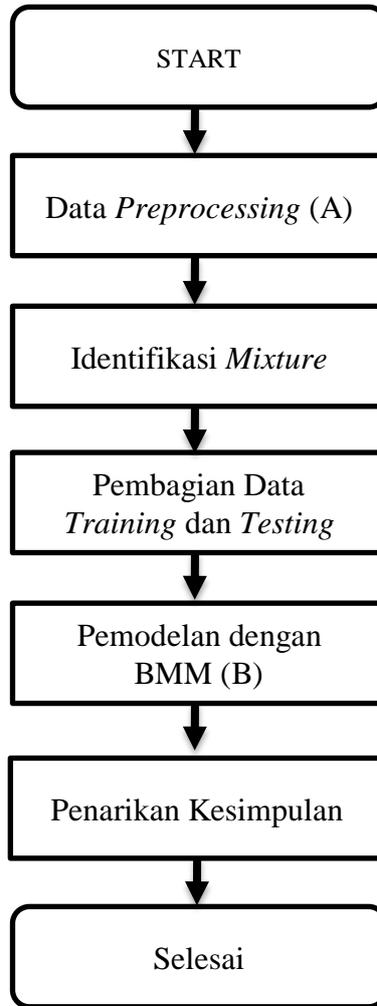


**Gambar 3.2** Alur Pembagian Data *Training* dan *Testing*

4. Melakukan estimasi parameter dengan metode BMM.

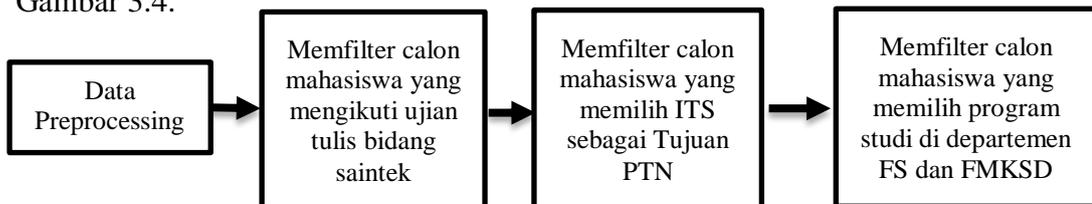
### 3.5 Diagram Alir

Gambaran diagram alir berdasarkan langkah analisis yang telah dibuat pada penelitian ini ditunjukkan pada Gambar 3.3



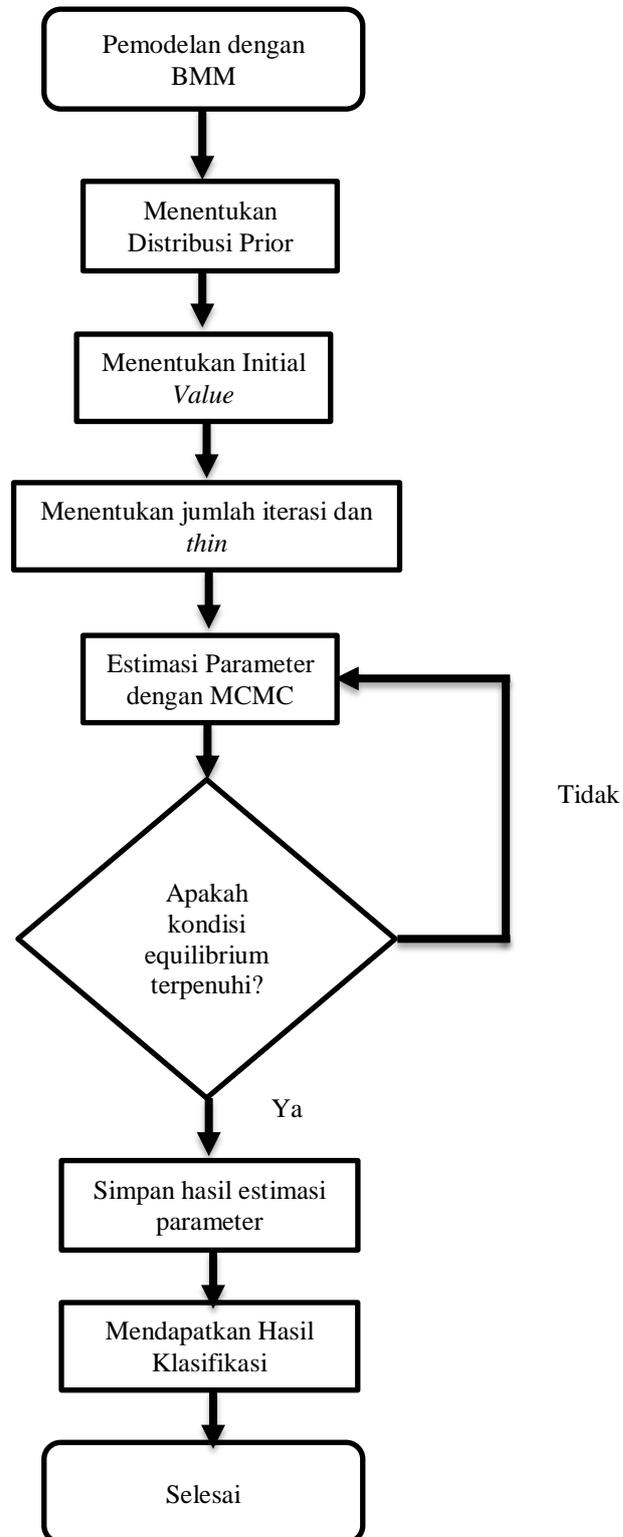
**Gambar 3.3** Diagram Alir Penelitian

Simbol (A) merupakan tahapan proses *preprocessing* data yang ditunjukkan pada Gambar 3.4.



**Gambar 3.4** Tahapan Preprocessing Data

Sedangkan simbol (B) menunjukkan pemodelan dengan BMM yang ditunjukkan pada Gambar 3.5.



**Gambar 3.5** Tahapan dalam Pemodelan dengan BMM

## BAB 4 ANALISIS DAN PEMBAHASAN

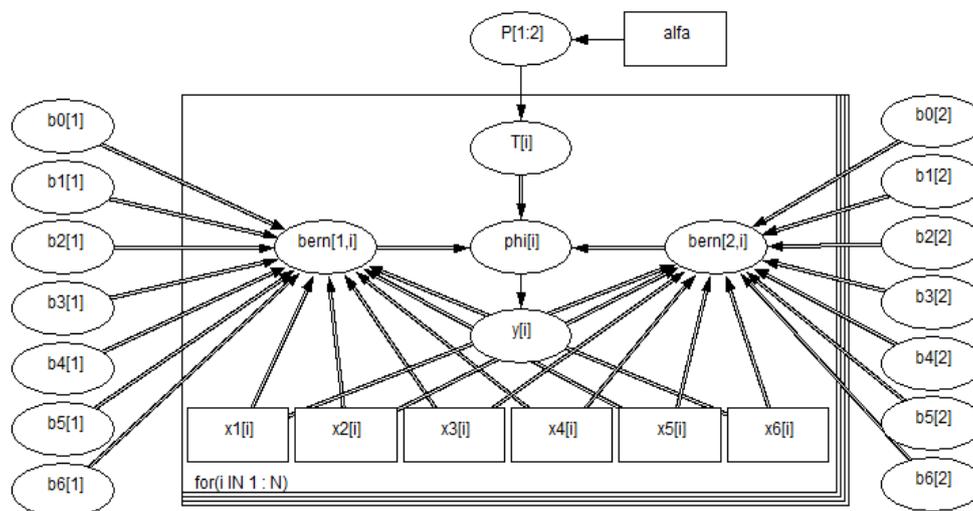
Langkah pertama yang dilakukan adalah melakukan *preprocessing* data. Proses *preprocessing* memisahkan lebih dari 800,000 dari hasil penerimaan di Indonesia, kemudian memfilter 20,089 calon mahasiswa yang memilih ITS sebagai PTN pilihan. Selanjutnya, memfilter 3,062 calon mahasiswa baru yang memilih FMKSD dan FSAINS sebagai prodi pilihan.

### 4.1 Kajian Algoritma Bernoulli Mixture Model dan Pemeriksaan Konvergensi

Pada penelitian ini akan dijelaskan pembuatan algoritma BMM untuk proses estimasi parameter, proses MCMC dua tahap dan pemeriksaan konvergensi.

#### 4.1.1 Algoritma Bernoulli Mixture Model pada OpenBUGS

Pemodelan bayesian mixture model akan dilakukan dengan bantuan *software openbugs*. Langkah pertama untuk melakukan pemodelan adalah dengan membuat *doodle*. *Doodle* terdiri dari kumpulan nodes yang dihubungkan dengan tanda panah. Tiap bentuk atau tipe *nodes* memiliki makna tersendiri, *nodes* berbentuk kotak merupakan sebuah konstan sedangkan *nodes* berbentuk oval merupakan parameter stokastik yang memiliki nilai dan distribusi tertentu. Doodle untuk Bernoulli Mixture Model ditunjukkan oleh Gambar 4.1.



**Gambar 4.1** Struktur DAG Bernoulli Mixture Model

Dari Gambar 4.1 diketahui bahwa nilai nilai  $\phi[i]$  merupakan nilai *logical* dengan *link identity* dari persamaan (2.9), yang merupakan *link identity* dari *exponential family* untuk model *Bernoulli mixture* dengan fungsi *logit*. Untuk  $y[i]$  merupakan nilai respon yang memiliki tipe stokastik dengan kepadatan atau *densitas* distribusi Bernoulli yang diperjelas dengan *nodes* `bern[1,i]` dan `bern[2,i]`. Pada BMM, parameter yang di estimasi akan berjumlah 2x dari total parameter, hal ini dikarenakan tiap parameter akan diestimasi pada masing-masing komponen *mixture 1* dan *mixture 2*. Setelah pembuatan *nodes* selesai maka selanjutnya *nodes* akan dikonversi ke dalam bentuk *syntax* yang mana *syntax* ini akan digunakan untuk mengestimasi parameter ditunjukkan pada Tabel 4.1.

**Tabel 4.1** *Syntax* dari *Nodes Bernoulli Mixture Model*

```

model;
{
  for( i in 1 : N ) {
    y[i] ~ dbern(phi[i])
  }
  for( i in 1 : N ) {
    phi[i] <- bern[T[i] , i]
  }
  for( i in 1 : N ) {
    T[i] ~ dcat(P[1:2])
  }
  P[1:2] ~ ddirch(alfa[])
  for( i in 1 : N ) {
    logit(bern[2 , i]) <- b0[2] + b1[2] * x1[i] + b2[2] * x2[i] + b3[2] * x3[i] + b4[2] *
x4[i] + b5[2] * x5[i] + b6[2] * x6[i]
  }
  for( i in 1 : N ) {
    logit(bern[1 , i]) <- b0[1] + b1[2] * x1[i] + b2[1] * x2[i] + b3[1] * x3[i] + b4[1] *
x4[i] + b5[1] * x5[i] + b6[1] * x6[i]
  }
}

```

*Prior* dan *initial* untuk proses estimasi tahap awal akan diambil dari hasil estimasi parameter dengan metode GLMs pada tiap *mixture* yang selanjutnya hasil *prior*-nya akan diintegrasikan kedalam algoritma BMM pada Tabel 4.1. Hasil integrasi *prior* dan *initial* pada algoritma BMM ditunjukkan pada Lampiran 1. Proses estimasi akan dilakukan dengan proses MCMC dua tahap yaitu mengambil hasil estimasi MCMC untuk dijadikan prior pada estimasi berikutnya hingga didapatkan hasil yang konvergen (Efendiev, *et.al.*, 2013). Penggunaan MCMC

dua tahap mampu mempercepat proses konvergensi rantai markov (Stuart, *et.al.*, 2019). Setelah mendapatkan hasil pada tahap dua dan rantai markov sudah konvergen maka selanjutnya hasil estimasi dapat digunakan untuk memodelkan status penerimaan calon mahasiswa. sebagai info tambahan, proses estimasi parameter hingga mencapai kondisi konvergen dilakukan minimal sebanyak tiga tahap. Pembahasan tentang pemeriksaan konvergensi dapat dilihat pada subbab 4.1.2

#### **4.1.2 Pemeriksaan Konvergensi**

Setelah pembuatan algoritma dan proses penentuan *prior* serta *initial* telah berakhir, maka selanjutnya proses estimasi akan dilakukan. Proses estimasi pada tiap tahap berbeda-beda berdasar pada *prior*, *initial*, *thin* dan iterasi yang digunakan. Semakin banyak iterasi dan *thin* yang digunakan, maka semakin besar jumlah iterasi yang dilakukan dan semakin banyak waktu yang digunakan. Setelah proses estimasi parameter selesai tahapan selanjutnya adalah pemeriksaan konvergensi. Pemeriksaan konvergensi akan dilakukan dengan menggunakan dua metode yaitu inferensia dan melalui *diagnostic plot*. Secara umum ada tiga uji yang digunakan dalam pemeriksaan konvergensi yaitu uji *Geweke*, *Gelman-Rubin* dan *Raftery-Lewis*.

Pemeriksaan konvergensi pada penelitian ini akan dibahas secara bertahap mengingat untuk mendapatkan keadaan konvergen peneliti memerlukan minimal tiga tahap proses MCMC. Tahap pertama dilakukan dengan melakukan iterasi sebanyak 50,000 dan *thin* sebanyak 80. Tahap pertama dilakukan pengambilan sampel rantai markov dengan iterasi sebanyak 50,000 dan *thin* sebanyak 80. Tahap kedua dilakukan pengambilan sampel rantai markov dengan iterasi sebanyak 140,000 dan *thin* sebanyak 150. Tahap ketiga dilakukan pengambilan sampel rantai markov dengan iterasi sebanyak 180,000 dan *thin* sebanyak 180.

Proses pemeriksaan konvergensi ini akan menunjukkan bahwa kondisi *prior* dan *initial* yang berbeda akan mempengaruhi hasil yang berbeda pula. nilai *thin* dan iterasi yang berbeda juga memberikan hasil yang berbeda. Proses estimasi MCMC 2 tahap juga semakin mempercepat proses konvergensi,

sedangkan jika tanpa MCMC 2 tahap, penambahan *thin* pada penelitian ini harus ditambah hingga menjadi 600 dengan iterasi 100,000, yang mana proses ini membutuhkan waktu empat hari untuk proses estimasi parameter. Sebagai ilustrasi berikut akan dijelaskan secara bertahap proses estimasi parameter hingga mencapai kondisi konvergen untuk prodi Fisika.

#### 1. Estimasi Parameter Tahap Pertama

Proses estimasi parameter tahap pertama adalah proses estimasi paling awal yang dilakukan dalam penelitian ini. *Prior* yang digunakan pada penelitian ini merupakan *pseudo prior* berdasar pada perhitungan secara *frequentist* dengan menggunakan regresi logistik. Penentuan *prior* ini dilakukan dengan cara mengestimasi parameter di tiap *mixture*. Setelah mendapatkan nilai *mean* dan *varians* pada tiap *mixture*, selanjutnya hasil estimasi parameter akan digunakan sebagai *prior* pada proses MCMC tahap pertama. *Pseudo prior* yang digunakan dalam penelitian ini ditunjukkan pada Tabel 4.2.

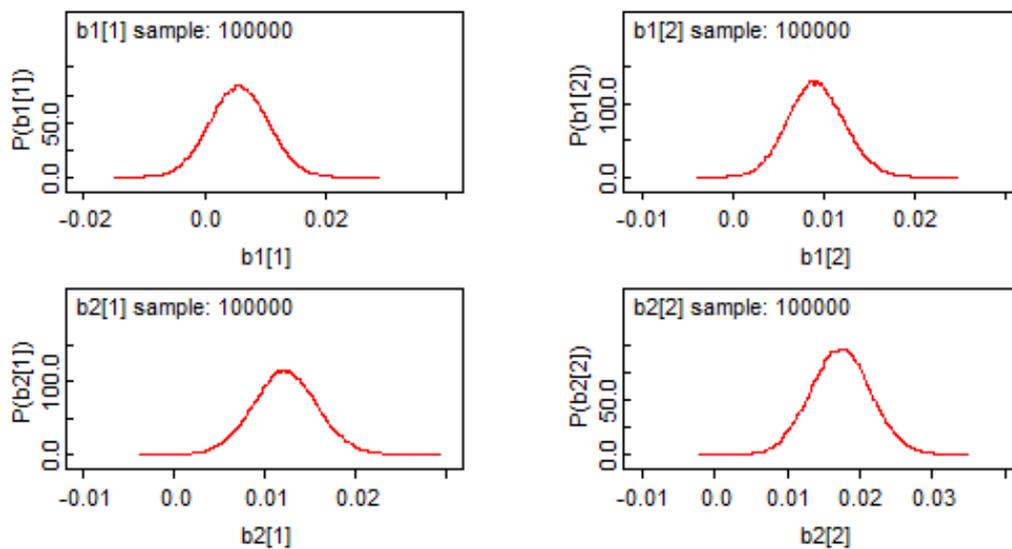
**Tabel 4.2** *Pseudo Prior* untuk Proses MCMC Tahap Pertama Prodi Fisika

Mixture 1			Mixture 2		
Parameter	Mean	Varians	Parameter	Mean	Varians
<b>b0[1]</b>	-29.5807	0.018724	<b>b0[2]</b>	-28.6633	0.021004
<b>b1[1]</b>	0.00569	43168.63	<b>b1[2]</b>	0.011797	25989.41
<b>b2[1]</b>	0.012775	35016.05	<b>b2[2]</b>	0.016418	24337.95
<b>b3[1]</b>	0.009417	41530.57	<b>b3[2]</b>	0.002474	54437.19
<b>b4[1]</b>	0.004624	140379.2	<b>b4[2]</b>	0.006958	122169.9
<b>b5[1]</b>	0.006546	78068.64	<b>b5[2]</b>	0.00647	91660.63
<b>b6[1]</b>	0.007677	109282.2	<b>b6[2]</b>	0.001588	76057.91

Tabel 4.2 merupakan *prior* yang digunakan untuk estimasi parameter tahap pertama. setelah mendapatkan *prior* maka tahapan selanjutnya adalah menentukan *initial* untuk proses MCMC. Pada tahap pertama *initial* akan disamadengankan nol. Langkah selanjutnya adalah melakukan estimasi parameter dengan MCMC, tahap pertama *thin* yang digunakan adalah sebanyak 80 dan iterasi sebanyak 50,000, artinya proses pengambilan sampel untuk rantai markov akan dilakukan sebanyak 4,000,000 iterasi dengan mengambil satu sampel pada

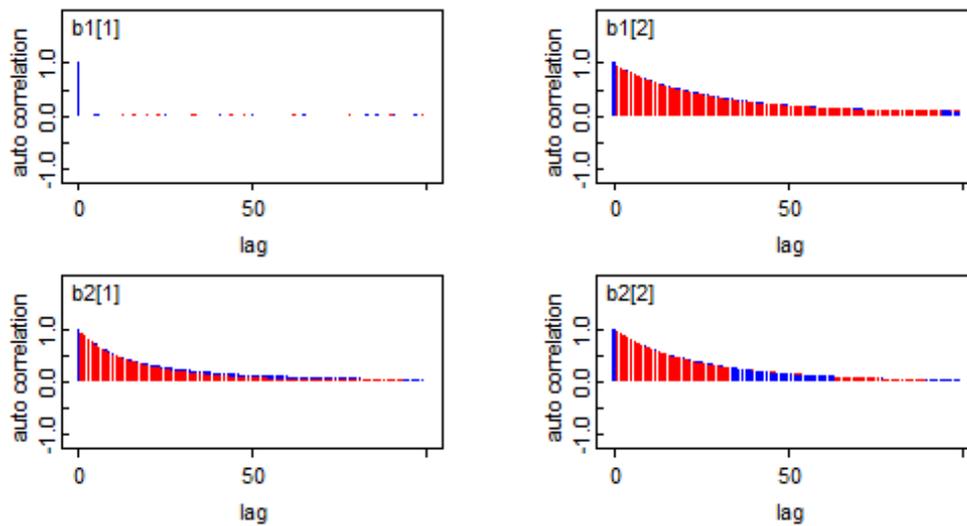
tiap-tiap 80 iterasi dengan demikian, setiap rantai markov yang terbentuk akan terdiri dari 50,000 sampel.

Proses estimasi tahap pertama memakan waktu selama 2867 detik atau selama 47 menit. Pada penelitian ini jumlah rantai markov yang digunakan adalah dua. Hal ini dikarenakan untuk pemeriksaan konvergensi dengan metode *Gelman-rubin* akan menguji dua rantai markov, selain itu penggunaan 2 rantai markov akan mempermudah dalam visualisasi *diagnostic plot*. Langkah selanjutnya yaitu memeriksa konvergensi dengan *diagnostic plot*. Jika pada proses ini *diagnostic plot* telah menunjukkan hasil yang konvergen maka pembahasan akan dilanjutkan dengan pemeriksaan secara inferensia. Visualisasi dengan *density plot* ditunjukkan pada Gambar 4.2



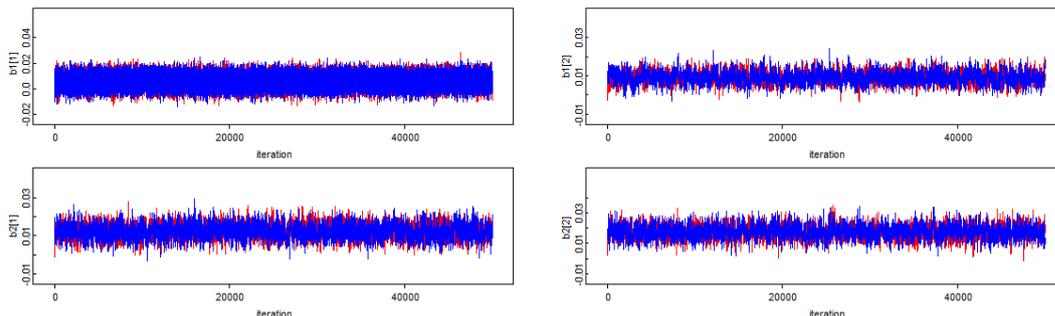
**Gambar 4.2** *Density Plot* Tahap Pertama Prodi Fisika

Hasil visualisasi dengan menggunakan *density plot* menunjukkan bahwa hasil estimasi parameter telah memiliki satu titik puncak dan sehingga dapat disimpulkan bahwa parameter yang telah konvergen. Hasil visualisasi dengan *autocorrelation plot* akan ditunjukkan pada Gambar 4.3. Hasil visualisasi dengan menggunakan *autocorrelation plot* menunjukkan bahwa proses estimasi parameter belum konvergen. Hal ini dikarenakan meskipun nilai *lag* 0 mendekati satu, tetapi nilai *lag* setelah *lag* 0 dan selanjutnya tidak mendekati nol, nilai ini juga menunjukkan bahwa autokorelasi pada tiap parameter masih tinggi dan hasil estimasi parameter belum bisa dikatakan mendekati distribusi target.



**Gambar 4.3** *Autocorrelation Plot* Tahap Pertama Prodi Fisika

Langkah selanjutnya akan dilakukan visualisasi dengan *serial plot*. Hasil visualisasi untuk *serial plot* akan ditunjukkan pada Gambar 4.4. Hasil visualisasi dengan menggunakan *serial plot* juga menunjukkan bahwa proses estimasi parameter tahap pertama juga belum konvergen.



**Gambar 4.4** *Serial Plot* Tahap Pertama Prodi Fisika

Hasil pada visualisasi *diagnostic plot* menunjukkan bahwa ketiga *plot* belum menunjukkan hasil yang konvergen. Dengan demikian, dapat dikatakan bahwa hasil estimasi pada tahap pertama belum konvergen, selanjutnya proses pemeriksaan konvergensi akan dilakukan secara inferensia. Proses ini dilakukan dalam rangka meneliti perbandingan antar metode pemeriksaan konvergensi.

Pemeriksaan konvergensi secara inferensia, akan dilakukan dengan menggunakan tiga metode. Yang mana tiap-tiap metode menguji rantai markov dengan cara yang berbeda-beda dan memiliki sensitivitas yang berbeda-beda pada tiap perubahan *thin* ataupun iterasi. Untuk dapat melihat perbedaan ini pemeriksaan konvergensi secara inferensia pada tahap pertama, kedua dan ketiga

akan dibandingkan diakhir pembahasan subbab ini. Pemeriksaan konvergensi dengan menggunakan metode *Geweke* akan ditunjukkan pada Tabel 4.3.

**Tabel 4.3** Hasil Uji *Geweke* Tahap Pertama Prodi Fisika

<i>Geweke : Zscores</i>							
	<b>Par.</b>	<i>Chain 1</i>	<i>Chain 2</i>		<b>Par.</b>	<i>Chain 1</i>	<i>Chain 2</i>
<b>Mixture</b> <b>1</b>	<b>P[1]</b>	-1.4707	0.76483	<b>Mixture</b> <b>2</b>	<b>P[2]</b>	1.4707	-0.76483
	<b>b0[1]</b>	1.8205	-0.00474		<b>b0[2]</b>	0.1868	-0.09223
	<b>b1[1]</b>	-0.021	<b>-2.42367</b>		<b>b1[2]</b>	-1.1385	0.69461
	<b>b2[1]</b>	-1.9586	0.37704		<b>b2[2]</b>	1.5466	-0.07595
	<b>b3[1]</b>	-1.4787	-0.62732		<b>b3[2]</b>	<b>-2.026</b>	-0.41027
	<b>b4[1]</b>	-1.7053	-1.5892		<b>b4[2]</b>	0.0539	0.27534
	<b>b5[1]</b>	<b>-2.3641</b>	-0.99192		<b>b5[2]</b>	0.3124	-0.0876
	<b>b6[1]</b>	<b>-2.4476</b>	-0.16809		<b>b6[2]</b>	-1.0452	-1.08963

Hasil pemeriksaan dengan menggunakan metode *Geweke* menunjukkan bahwa hasil uji Z dari rantai markov melebihi batas yang ditetapkan yaitu  $Z = |2|$ , sehingga dapat dikatakan proses rantai marov belum konvergen. Selanjutnya, pemeriksaan konvergensi akan dilakukan dengan menggunakan metode *Gelman-Rubin* dengan mengukur nilai *potential scale reduction factor* (PSRF).

**Tabel 4.4** Hasil Uji *Raftery-Lewis* Tahap Pertama Prodi Fisika

<b>Parameter</b>	<i>Dependence Factor</i> <i>(I)</i>	
	<i>Chain 1</i>	<i>Chain 2</i>
<b>P[1]</b>	1.01	1.01
<b>b0[1]</b>	<b>23.2</b>	<b>28.6</b>
<b>b1[1]</b>	0.994	0.999
<b>b2[1]</b>	<b>10.4</b>	<b>12.2</b>
<b>b3[1]</b>	<b>8.2</b>	<b>6.78</b>
<b>b4[1]</b>	2.6	2.51
<b>b5[1]</b>	4.02	4.38
<b>b6[1]</b>	4.63	3.9
<b>P[2]</b>	0.983	1
<b>b0[2]</b>	<b>18.6</b>	<b>20.7</b>
<b>b1[2]</b>	<b>18.3</b>	<b>16.9</b>
<b>b2[2]</b>	<b>12.1</b>	<b>15.7</b>
<b>b3[2]</b>	5	<b>6.37</b>
<b>b4[2]</b>	2.77	3.71
<b>b5[2]</b>	4.32	5.25
<b>b6[2]</b>	<b>5.65</b>	4.53

Hasil pemeriksaan dengan menggunakan metode *gelman-rubin* menunjukkan bahwa nilai PSRF  $\hat{R} = 1.01$  sehingga dapat dikatakan bahwa proses estimasi sudah konvergen, selanjutnya rantai markov akan di uji dengan metode *Raftery-Lewis* ditunjukkan pada Tabel 4.4. Hasil pemeriksaan dengan metode *Raftery-Lewis* menunjukkan bahwa masih banyak nilai  $I \geq 5$ . Nilai ini menunjukkan bahwa pemeriksaan dengan metode *Raftery-Lewis* menunjukkan hasil yang belum konvergen. Dari ketiga metode yang digunakan untuk pemeriksaan konvergensi, hanya satu metode yang menunjukkan bahwa hasil rantai markov sudah konvergen yaitu metode *Gelman-Rubin*. Ini menunjukkan bahwa hasil pemeriksaan masih belum bisa dikatakan konvergen sehingga proses estimasi akan dilanjutkan pada tahap kedua dan hasil estimasi parameter pada tahap pertama akan digunakan sebagai *prior* untuk estimasi parameter tahap kedua. Proses estimasi dengan menggunakan *prior* dari hasil estimasi parameter sebelumnya hingga didapatkan hasil yang konvergen disebut dengan proses MCMC dua tahap (Efendiev, *et.al.*, 2013).

#### 1. Estimasi Parameter Tahap Kedua

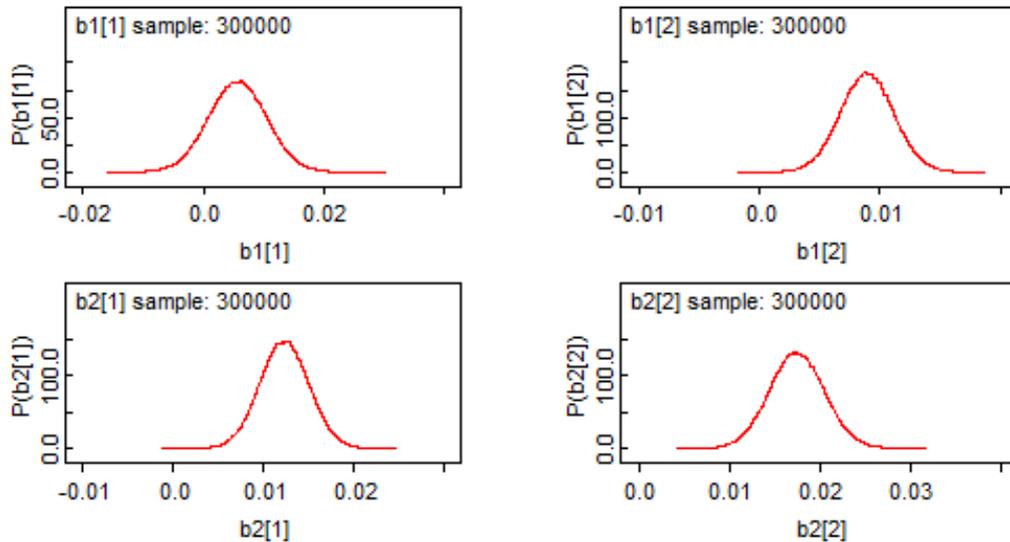
Proses estimasi tahap kedua merupakan rangkaian dari proses estimasi parameter dengan MCMC hingga rantai dinyatakan konvergen. *Prior* yang digunakan pada tahap kedua merupakan hasil estimasi parameter yang belum konvergen pada tahap pertama. *Prior* yang digunakan untuk tahap kedua ditunjukkan pada Tabel 4.5.

**Tabel 4.5** *Prior* dan *Initial* untuk Proses MCMC Tahap Kedua Prodi Fisika

<i>Mixture 1</i>			<i>Mixture 2</i>			<i>Initial</i>		
<b>b0[1]</b>	-31.84	0.062065	<b>b0[2]</b>	-28.04	0.073046	<b>b0</b>	-31.84	-28.04
<b>b1[1]</b>	0.005695	43132.78	<b>b1[2]</b>	0.009223	102596.9	<b>b1</b>	0.005695	0.009223
<b>b2[1]</b>	0.01224	79349.34	<b>b2[2]</b>	0.01738	57701.51	<b>b2</b>	0.01224	0.01738
<b>b3[1]</b>	0.009448	88524.4	<b>b3[2]</b>	0.002704	113596.5	<b>b3</b>	0.009448	0.002704
<b>b4[1]</b>	0.004727	283838.6	<b>b4[2]</b>	0.007051	273542.1	<b>b4</b>	0.004727	0.007051
<b>b5[1]</b>	0.006784	173900.8	<b>b5[2]</b>	0.00661	197005.2	<b>b5</b>	0.006784	0.00661
<b>b6[1]</b>	0.007945	224613.1	<b>b6[2]</b>	0.00163	168103.4	<b>b6</b>	0.007945	0.00163

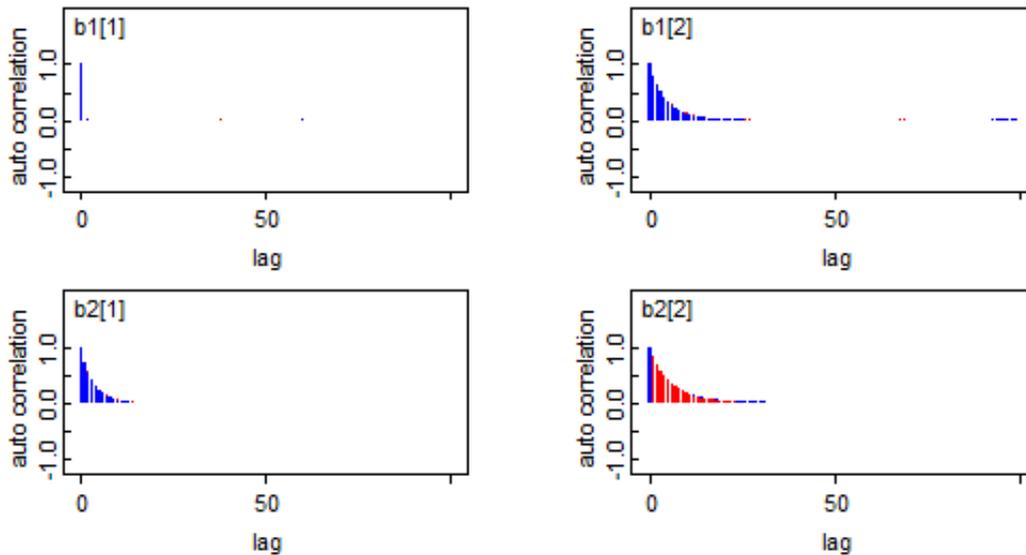
Tabel 4.5 merupakan *prior* dan *initial* yang digunakan untuk estimasi parameter tahap kedua, setelah mendapatkan *prior* maka tahapan selanjutnya adalah melakukan estimasi parameter dengan MCMC, untuk tahap kedua *thin* yang digunakan adalah sebanyak 140 dan iterasi sebanyak 150,000, artinya proses pengambilan sampel untuk rantai markov akan dilakukan sebanyak 21,000,000 iterasi dengan mengambil satu sampel pada tiap-tiap 140 iterasi dengan demikian, setiap rantai markov yang terbentuk akan terdiri dari 150,000 sampel.

Proses estimasi tahap kedua memakan waktu yang lebih lama mengingat jumlah iterasi yang dilakukan sebanyak 21,000,000. Proses estimasi tahap kedua dilakukan selama 14852 detik atau selama 4 jam, setelah proses estimasi selesai, maka langkah selanjutnya yaitu memeriksa konvergensi dengan *diagnostic plot*. Visualisasi dengan *density plot* ditunjukkan pada Gambar 4.5. Hasil visualisasi dengan menggunakan *density plot* menunjukkan bahwa proses estimasi parameter telah konvergen, hal ini ditunjukkan dengan distribusi parameter yang telah berdistribusi normal.



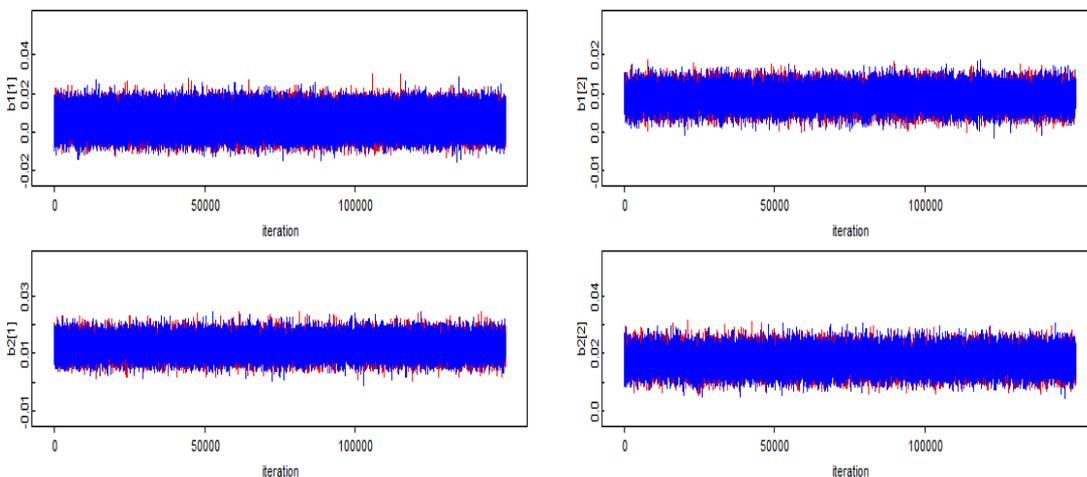
**Gambar 4.5** *Density plot* Tahap Kedua Prodi Fisika

Hasil visualisasi dengan *autocorrelation plot* akan ditunjukkan pada Gambar 4.6. Hasil visualisasi dengan menggunakan *autocorrelation plot* menunjukkan bahwa proses estimasi sudah konvergen konvergen. Hal ini ditunjukkan dari nilai *lag* 0 mendekati satu, dan nilai *lag* setelah *lag* 0 dan selanjutnya mendekati nol.



**Gambar 4.6** *Autocorrelation Plot* Tahap Kedua Prodi Fisika

Hasil visualisasi untuk *serial plot* akan ditunjukkan pada Gambar 4.7. Hasil visualisasi dengan menggunakan *serial plot* juga menunjukkan bahwa proses estimasi parameter juga belum konvergen.



**Gambar 4.7** *Serial Plot* Tahap Kedua Prodi Fisika

Hasil pada visualisasi *diagnostic plot* menunjukkan bahwa *density plot*, *serial plot* dan *autocorrelation plot* telah menunjukkan hasil yang konvergen. Dengan demikian, secara visual proses estimasi sudah konvergen. Selanjutnya pemeriksaan konvergensi akan dilanjutkan secara inferensia. Pemeriksaan konvergensi dengan menggunakan metode *Geweke* akan ditunjukkan pada Tabel 4.6.

**Tabel 4.6** Hasil Uji *Geweke* Tahap Kedua Prodi Fisika

<i>Geweke : ZScores</i>							
	Parameter	Chain 1	Chain 2		Parameter	Chain 1	Chain 2
<b>Mixture 1</b>	<b>P[1]</b>	-1.2151	0.23032	<b>Mixture 2</b>	<b>P[2]</b>	1.2151	-0.23038
	<b>b0[1]</b>	<b>-2.8632</b>	1.23287		<b>b0[2]</b>	-1.6134	1.3266
	<b>b1[1]</b>	0.088	1.14896		<b>b1[2]</b>	0.7934	-0.00844
	<b>b2[1]</b>	<b>2.6644</b>	-1.57928		<b>b2[2]</b>	1.0257	-0.21512
	<b>b3[1]</b>	0.0643	-0.99251		<b>b3[2]</b>	0.0766	-1.35581
	<b>b4[1]</b>	<b>2.4153</b>	-0.11291		<b>b4[2]</b>	0.6654	-1.21014
	<b>b5[1]</b>	1.3797	0.02011		<b>b5[2]</b>	<b>3.032</b>	-1.2435
	<b>b6[1]</b>	<b>2.136</b>	0.35562		<b>b6[2]</b>	-1.0624	-1.44711

Hasil pemeriksaan dengan menggunakan metode *Geweke* menunjukkan bahwa hasil uji Z dari rantai markov melebihi batas yang ditetapkan yaitu  $Z = |2|$ , sehingga dapat dikatakan proses rantai marov belum konvergen. Selanjutnya, pemeriksaan konvergensi akan dilakukan dengan menggunakan metode *Gelman-Rubin* dengan mengukur nilai *potential scale reduction factor* (PSRF).

**Tabel 4.7** Hasil Uji *Raftery-Lewis* Tahap Kedua Prodi Fisika

	Parameter	<i>Dependence Factor (I)</i>	
		Chain 1	Chain 2
<b>Mixture 1</b>	P[1]	1	0.994
	b0[1]	<b>7.95</b>	<b>7.46</b>
	b1[1]	0.991	0.999
	b2[1]	<b>5.33</b>	<b>5.36</b>
	b3[1]	3.71	3.9
	b4[1]	2.13	1.19
	b5[1]	2.2	2.24
	b6[1]	2.16	2.23
<b>Mixture 1</b>	P[2]	1	1.01
	b0[2]	<b>7.62</b>	<b>6.96</b>
	b1[2]	<b>7.58</b>	<b>6.03</b>
	b2[2]	<b>7.26</b>	<b>6.82</b>
	b3[2]	4.43	3.34
	b4[2]	2.1	2.09
	b5[2]	2.14	2.18
	b6[2]	2.31	3.21

Hasil pemeriksaan dengan menggunakan metode *gelman-rubin* menunjukkan bahwa nilai PSRF  $\hat{R} = 1$  sehingga dapat dikatakan bahwa rantai

markov sudah konvergen, selanjutnya rantai markov akan di uji dengan metode *Raftery-Lewis*, ditunjukkan pada Tabel 4.7. Hasil pemeriksaan dengan metode *Raftery-Lewis* menunjukkan bahwa masih banyak nilai  $I \geq 5$ . Nilai ini menunjukkan bahwa pemeriksaan dengan metode *Raftery-Lewis* menunjukkan hasil yang belum konvergen. Dari ketiga metode yang digunakan untuk pemeriksaan konvergensi, hanya satu metode yang menunjukkan bahwa hasil rantai markov sudah konvergen. Ini menunjukkan bahwa hasil pemeriksaan masih belum bisa dikatakan konvergen sehingga harus dilanjutkan pada tahap ketiga.

## 2. Estimasi Parameter Tahap Ketiga

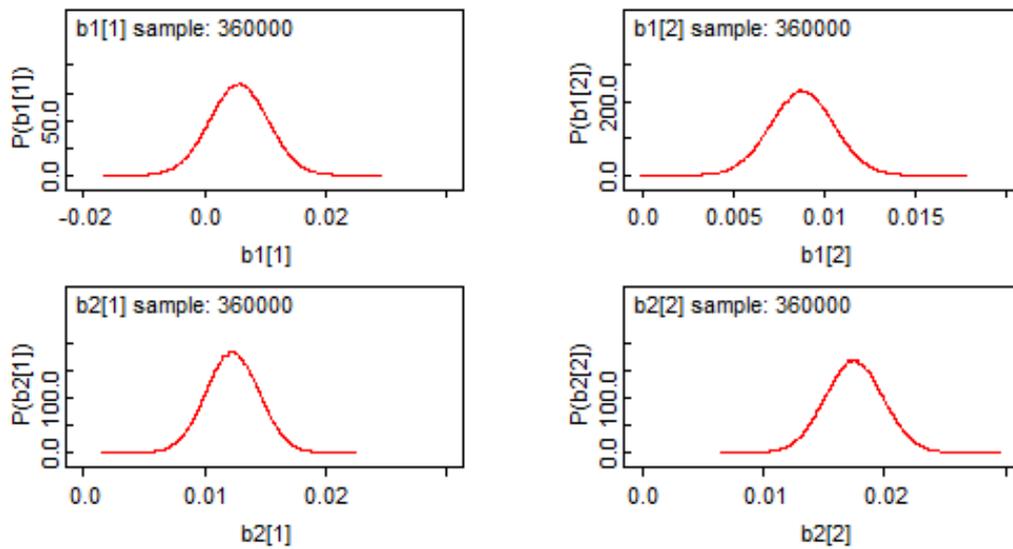
Proses estimasi tahap kedua merupakan rangkaian dari proses estimasi parameter dengan MCMC hingga rantai dinyatakan konvergen. *Prior* yang digunakan pada tahap kedua merupakan hasil estimasi parameter yang belum konvergen pada tahap pertama. *Prior* dan *initial* yang digunakan untuk tahap ketiga ditunjukkan pada Tabel 4.8.

**Tabel 4.8** *Prior* dan *Initial* Tahap Ketiga Prodi Fisika

<i>Mixture 1</i>			<i>Mixture 2</i>			<i>Initial</i>		
<b>b0[1]</b>	-31.84	0.062065	<b>b0[2]</b>	-28.04	0.073046	<b>b0</b>	-31.84	-28.04
<b>b1[1]</b>	0.005695	43132.78	<b>b1[2]</b>	0.009223	102596.9	<b>b1</b>	0.005695	0.009223
<b>b2[1]</b>	0.01224	79349.34	<b>b2[2]</b>	0.01738	57701.51	<b>b2</b>	0.01224	0.01738
<b>b3[1]</b>	0.009448	88524.4	<b>b3[2]</b>	0.002704	113596.5	<b>b3</b>	0.009448	0.002704
<b>b4[1]</b>	0.004727	283838.6	<b>b4[2]</b>	0.007051	273542.1	<b>b4</b>	0.004727	0.007051
<b>b5[1]</b>	0.006784	173900.8	<b>b5[2]</b>	0.00661	197005.2	<b>b5</b>	0.006784	0.00661
<b>b6[1]</b>	0.007945	224613.1	<b>b6[2]</b>	0.00163	168103.4	<b>b6</b>	0.007945	0.00163

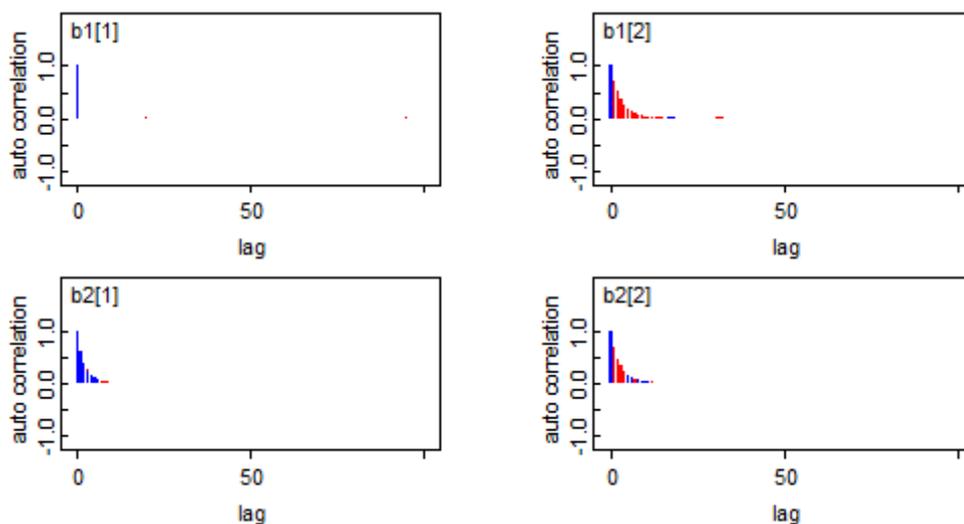
Tabel 4.8 merupakan *prior* dan *initial* yang digunakan untuk estimasi parameter tahap ketiga, selanjutnya adalah melakukan estimasi parameter dengan MCMC, pada tahap ketiga *thin* yang digunakan adalah sebanyak 180 dan iterasi sebanyak 180,000, artinya proses pengambilan sampel untuk rantai markov akan dilakukan sebanyak 32,400,000 iterasi dengan mengambil satu sampel pada tiap-tiap 180 iterasi dengan demikian, setiap rantai markov yang terbentuk akan terdiri dari 180,000 sampel.

Proses estimasi tahap kedua memakan waktu yang lebih lama yaitu selama 49,696 detik atau selama 14 jam, setelah proses estimasi selesai, maka langkah selanjutnya yaitu memeriksa konvergensi dengan *diagnostic plot*. Visualisasi dengan *density plot* ditunjukkan pada Gambar 4.8.



**Gambar 4.8** *Density plot* Tahap Ketiga Prodi Fisika

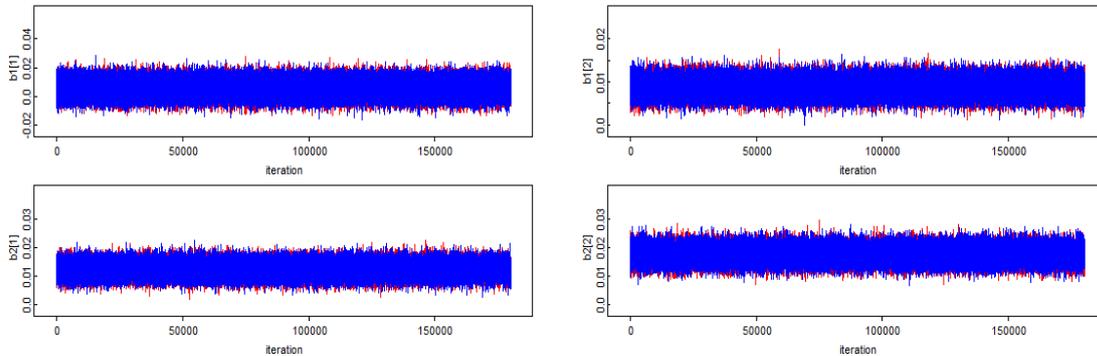
Hasil visualisasi dengan menggunakan *density plot* menunjukkan bahwa estimasi parameter telah berdistribusi normal dan hanya memiliki satu titik puncak sehingga dapat disimpulkan bahwa hasil estimasi parameter sudah konvergen. Hasil visualisasi *autocorrelation plot* akan ditunjukkan pada Gambar 4.9.



**Gambar 4.9** *Autocorrelation Plot* Tahap Ketiga Prodi Fisika

Hasil visualisasi dengan menggunakan *autocorrelation plot* menunjukkan bahwa proses estimasi sudah konvergen. Hasil visualisasi untuk

*serial plot* akan ditunjukkan pada Gambar 4.10. Hasil visualisasi dengan menggunakan *serial plot* juga menunjukkan bahwa proses estimasi parameter telah konvergen.



**Gambar 4.10** *Serial Plot* Tahap Ketiga Prodi Fisika

Hasil pada visualisasi *diagnostic plot* menunjukkan bahwa ketiga *plot* telah menunjukkan hasil yang konvergen, selanjutnya proses pemeriksaan konvergensi akan dilakukan dengan metode inferensia. Pemeriksaan konvergensi dengan menggunakan metode *Geweke* akan ditunjukkan pada Tabel 4.9.

**Tabel 4.9** Hasil Uji *Geweke* Tahap Ketiga Prodi Fisika

		<i>Geweke : Zscores</i>	
		Chain 1	Chain 2
<b>Mixture 1</b>	P[1]	-1.1097	0.258
	b0[1]	0.2038	0.245
	b1[1]	-0.2152	-0.404
	b2[1]	-0.0904	0.517
	b3[1]	-1.2521	-0.253
	b4[1]	1.164	1.599
	b5[1]	1.3467	0.937
	b6[1]	-0.2514	-0.332
<b>Mixture 2</b>	P[2]	1.1097	-0.258
	b0[2]	0.4276	0.891
	b1[2]	0.0575	-1.314
	b2[2]	0.3015	1.33
	b3[2]	-0.3884	-1.483
	b4[2]	-0.1601	0.787
	b5[2]	-0.5636	-0.867
	b6[2]	-1.5166	-1.233

Hasil pemeriksaan dengan menggunakan metode *Geweke* menunjukkan bahwa hasil uji  $Z$  dari rantai markov berada dalam batas ditetapkan yaitu  $Z = |2|$ , sehingga dapat dikatakan proses estimasi parameter sudah konvergen.

**Tabel 4.10** Hasil Uji *Raftery-Lewis* Tahap Ketiga Prodi Fisika

<i>Raftery Lewis : Dependence Factor (I)</i>			
	Chain 1	Chain 2	
<b>Mix 1</b>	P[1]	1.01	1
	b0[1]	3.49	3.5
	b1[1]	0.999	0.995
	b2[1]	2.53	3.33
	b3[1]	2.3	2.35
	b4[1]	1.07	1.08
	b5[1]	1.1	1.1
	b6[1]	1.13	1.13
<b>Mix 2</b>	P[2]	1	1.01
	b0[2]	3.36	3.49
	b1[2]	3.7	3.7
	b2[2]	3.74	3.56
	b3[2]	2.28	2.17
	b4[2]	1.06	1.07
	b5[2]	1.12	1.12
	b6[2]	2.08	1.18

Langkah berikutnya yaitu memeriksa rantai markov dengan metode *Raftery-Lewis* yang ditunjukkan pada Tabel 4.10. Hasil pemeriksaan dengan metode *Raftery-Lewis* menunjukkan bahwa seluruh nilai  $I \leq 5$ . Nilai ini menunjukkan bahwa hasil estimasi parameter telah konvergen dan tidak ada autokorelasi di tiap parameter. Selanjutnya, pemeriksaan konvergensi akan dilakukan dengan menggunakan metode *Gelman-Rubin* dengan mengukur nilai *potential scale reduction factor* (PSRF). Hasil pemeriksaan dengan menggunakan metode *Gelman-Rubin* menunjukkan bahwa nilai PSRF  $\hat{R} = 1$  sehingga dapat dikatakan bahwa proses estimasi parameter sudah konvergen. Dengan demikian dapat disimpulkan bahwa baik dengan *diagnostic plot* maupun ketiga metode inferensia estimasi parameter telah konvergen, sehingga hasil estimasi parameter yang dihasilkan dianggap mampu menggambarkan pemodelan BMM.

Hasil estimasi dengan *thin* dan iterasi yang berbeda perlu untuk dilakukan, mengingat setiap perubahan *thin*, iterasi dan *prior* akan mempengaruhi

hasil akhir pada analisis Bayesian. Setiap penambahan *thin* dan iterasi akan mendekatkan hasil estimasi parameter terhadap target distribusi, sehingga harapan untuk mendapatkan hasil parameter yang mendekati sebenarnya akan tercapai. Perbandingan hasil konvergensi pada tiap tahap akan dirangkum pada Tabel 4.11.

**Tabel 4.11** Perbandingan Hasil Konvergensi

Metode yang Digunakan	Tahap 1 – Thin 80 Iterasi 50,000	Tahap 2 – Thin 140 Iterasi 150,000	Tahap 3 – Thin 180 Iterasi 180,000
<b>Diagnostic Plot</b>			
1. <i>Density plot</i>	Belum Konvergen	Konvergen	Konvergen
2. <i>Autocorrelation Plot</i>	Belum Konvergen	Konvergen	Konvergen
3. <i>Serial Plot</i>	Belum Konvergen	Konvergen	Konvergen
<b>Metode Inferensia</b>			
1. <i>Geweke</i>	Belum Konvergen	Belum Konvergen	Konvergen
2. <i>Gelman-Rubin</i>	Konvergen	Konvergen	Konvergen
3. <i>Raftery-Lewis</i>	Belum Konvergen	Belum Konvergen	Konvergen

Hasil rangkuman perbandingan konvergensi pada Tabel 4.11 menunjukkan bahwa, semakin besar nilai *thin* dan iterasi maka semakin cepat sebuah nilai mendekati target distribusinya. Tahap kedua dan ketiga *diagnostic plot* menunjukkan bahwa secara visual proses estimasi menunjukkan kecenderungan hasil yang konvergen, tetapi jika diperiksa secara inferensia proses estimasi belum menunjukkan hasil yang konvergen hingga tahap ketiga. Penggunaan metode inferensia sebaiknya dilakukan dengan membandingkan lebih dari satu metode hal ini dikarenakan, satu metode belum bisa menunjukkan hasil yang akurat.

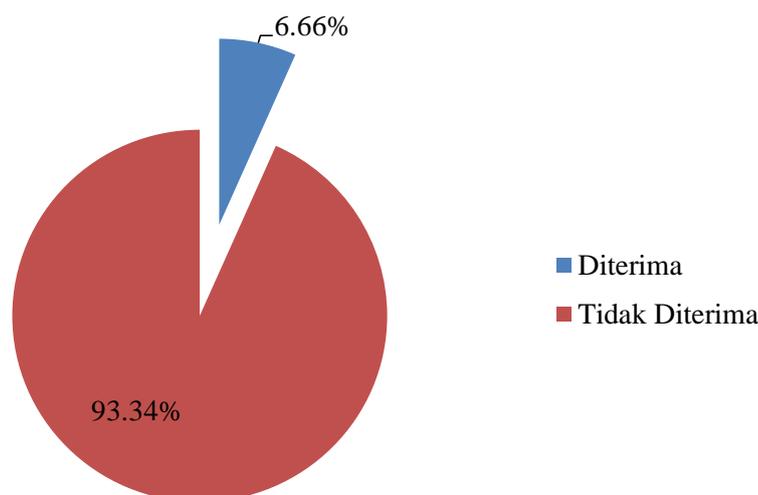
Hasil pemeriksaan dengan berbagai metode ini menunjukkan bahwa metode *Gelman-Rubin* memberikan hasil yang signifikan pada tiap tahap, artinya metode *Gelman-Rubin* tidak sensitif terhadap perubahan *thin* dan iterasi. Sedangkan metode *Geweke* dan *Raftery-Lewis* menunjukkan bahwa semakin bertambah nilai *thin* dan iterasi, maka nilai *dependence factor (i)* dan *Zscores* akan semakin kecil atau bisa diartikan memberikan hasil yang konvergen. Pada subbab selanjutnya akan ditunjukkan hasil pemodelan pada tiap prodi, dengan menunjukkan hasil estimasi pada tahap ketiga saja.

## 4.2 Karakteristik data dan Pemodelan dengan Bernoulli Mixture Model

Pada subbab ini, akan dijelaskan tentang karakteristik serta tahapan dalam pemodelan dengan *Bernoulli Mixture Model*.

### 4.2.1 Karakteristik Calon Mahasiswa Baru di FSAINS dan FMKSD

Setelah menyelesaikan tahap *pre-processing*, tahapan kedua adalah melakukan analisis deskriptif terhadap variabel status penerimaan mahasiswa baru dan karakteristik variabel independen yang akan digunakan untuk analisa dalam penelitian ini. Secara keseluruhan jumlah calon mahasiswa yang mendaftar di ITS melalui jalur ujian tulis pada tahun 2018 ada sebanyak 20,089 calon mahasiswa dengan total calon mahasiswa yang diterima adalah sebanyak 1,338 atau hanya sekitar 6.66% mahasiswa yang diterima di seluruh fakultas yang ada di ITS. proporsi penerimaan mahasiswa di ITS disajikan pada Gambar 4.11.



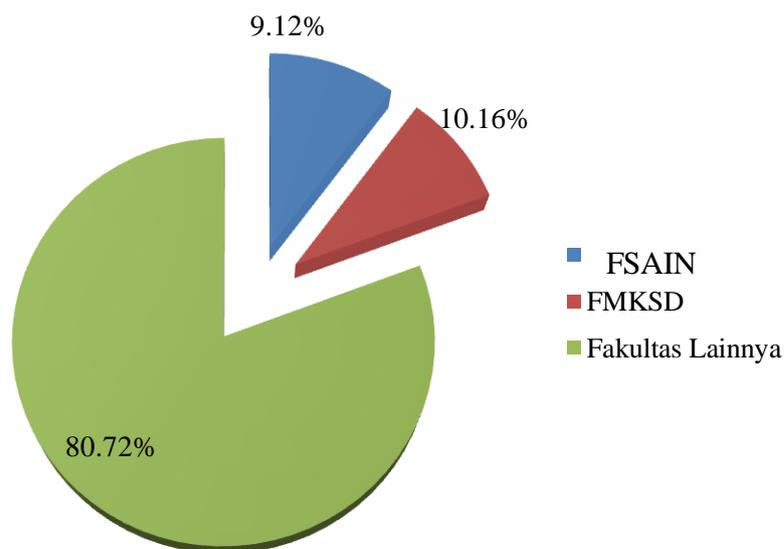
**Gambar 4.11** Proporsi Penerimaan Calon Mahasiswa di ITS Secara Keseluruhan

Dari total calon mahasiswa yang mendaftar, proporsi calon mahasiswa yang memilih FSAINS dan FMKSD adalah 15.24% atau sebanyak 3,062 calon mahasiswa. Proporsi penerimaan di FMKSD dan FSAINS adalah sebesar 8.4% dari pendaftar yaitu sebanyak 258 calon mahasiswa baru. Fakultas Sains terdiri dari tiga prodi yaitu Fisika, Kimia dan Biologi. Pada tahun 2018 peminat FSAINS tercatat ada sebanyak 1,214 pendaftar atau sebanyak 6.04% dari total pendaftar di ITS, dengan rincian 450 calon mahasiswa baru mendaftar pada prodi Biologi, 456

calon mahasiswa baru mendaftar pada prodi Kimia dan 308 calon mahasiswa baru mendaftar pada prodi Fisika. Untuk Prodi FMKSD, memiliki tiga prodi yaitu matematika, statistika dan sains aktuaria. Pada tahun 2018 tercatat bahwa peminat atau calon mahasiswa baru yang mendaftar ada sebanyak 1.848 calon mahasiswa dari yang mendaftar di FMKSD atau 9.2% dari seluruh calon mahasiswa yang mendaftar di ITS, dengan rincian 477 calon mahasiswa mendaftar pada prodi Matematika, 524 calon mahasiswa mendaftar pada prodi Statistika dan 847 calon mahasiswa mendaftar pada prodi Aktuaria.

#### 4.2.2 Proporsi dan Keketatan Ujian Tulis di FSAINS dan FMKSD

Proporsi penerimaan calon mahasiswa di FSAINS dan FMKSD terhadap total calon mahasiswa yang diterima di ITS ditunjukkan oleh *pie chart* pada Gambar 4.12.



**Gambar 4.12** Proporsi Penerimaan Calon Mahasiswa Di FSAINS Dan FMKSD

Dapat dikatakan bahwa dari 1,338 mahasiswa yang diterima di ITS, 136 mahasiswa diterima di FSAINS dan 122 mahasiswa lainnya diterima di FMKSD. Berdasar data yang diperoleh, setiap program studi memiliki jumlah pendaftar dan daya tampung atau kuota penerimaan yang berbeda-beda. Jumlah pendaftar dan daya tampung tiap program studi ditunjukkan pada Tabel 4.12.

Dari Tabel 4.12 dapat disimpulkan bahwa setiap program studi memiliki karakteristik penerimaan yang berbeda dikarenakan jumlah pendaftar dan kuota

yang penerimaan yang berbeda-beda. dapat juga diartikan bahwa untuk pendaftar yang memiliki nilai yang sama, akan memiliki probabilitas penerimaan yang berbeda pada tiap-tiap prodi. Hasil pada Tabel 4.12 juga menunjukkan bahwa data status penerimaan calon mahasiswa baru merupakan variabel kategorik yang *imbalance*, dengan proporsi penerimaan tidak lebih dari 10% pada tiap prodi. Perbandingan antara jumlah pendaftar dan pendaftar diterima inilah yang kemudian disebut keketatan.

**Tabel 4.12** Rincian Jumlah Pendaftar dan Daya Tampung Tiap Program Studi

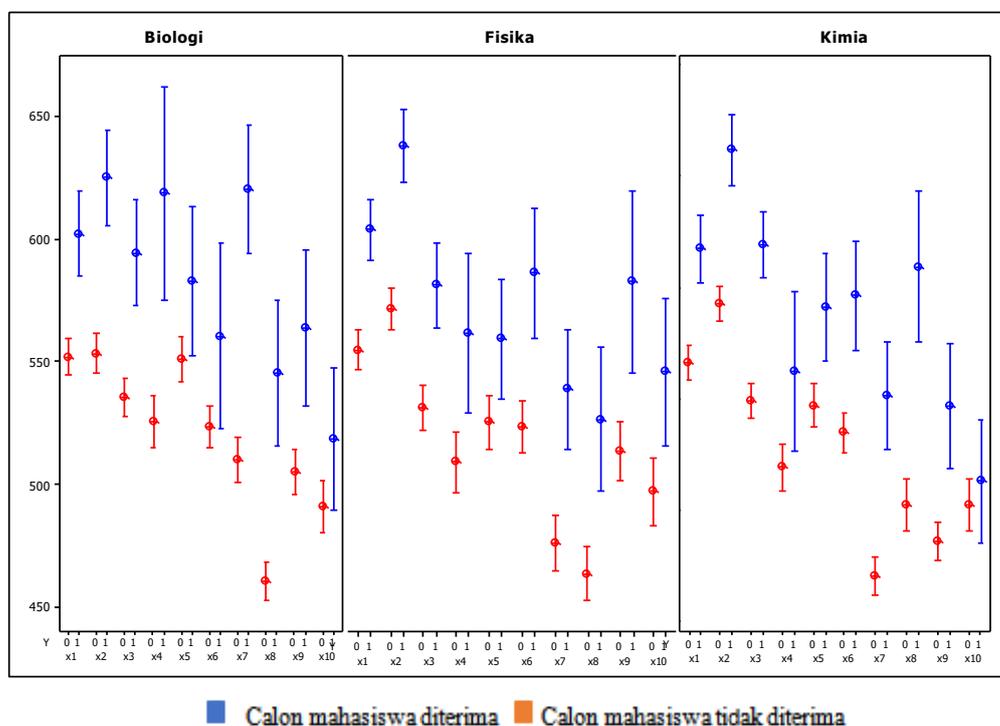
Departemen	Program Studi	Jumlah Pendaftar	Pendaftar Diterima	Keketatan
FSAINS	Biologi	450	40	1 : 11
	Kimia	456	46	1 : 10
	Fisika	308	48	1 : 6
FMKSD	Matematika	477	51	1 : 9
	Statistika	524	44	1 : 12
	Aktuaria	847	27	1 : 31

Hasil pada Tabel 4.12 menunjukkan bahwa setiap prodi memiliki keketatan yang berbeda-beda. Prodi yang memiliki keketatan tertinggi di departemen FSAINS keketatan adalah program studi Biologi, yaitu 11. Nilai keketatan 11 dapat memberikan informasi bahwa untuk dapat masuk di prodi biologi setiap satu orang calon mahasiswa harus bersaing dengan 11 calon mahasiswa lainnya. Untuk Departemen FMKSD, prodi yang memiliki keketatan tertinggi adalah prodi sains aktuaria, yaitu 31. Nilai keketatan 31 dapat memberikan informasi bahwa untuk dapat masuk di prodi Aktuaria setiap satu orang calon mahasiswa harus bersaing dengan 31 calon mahasiswa lainnya.

#### 4.2.3 Karakteristik Calon Mahasiswa Baru dengan *Boxplot* dan *Interval Plot*

Karakteristik calon mahasiswa baru akan dianalisa berdasar nilai ujian yang mereka dapat. Hal ini dilakukan agar peneliti bisa memberikan gambaran yang lebih jelas tentang latar belakang pada penelitian ini. Visualisasi *Interval*

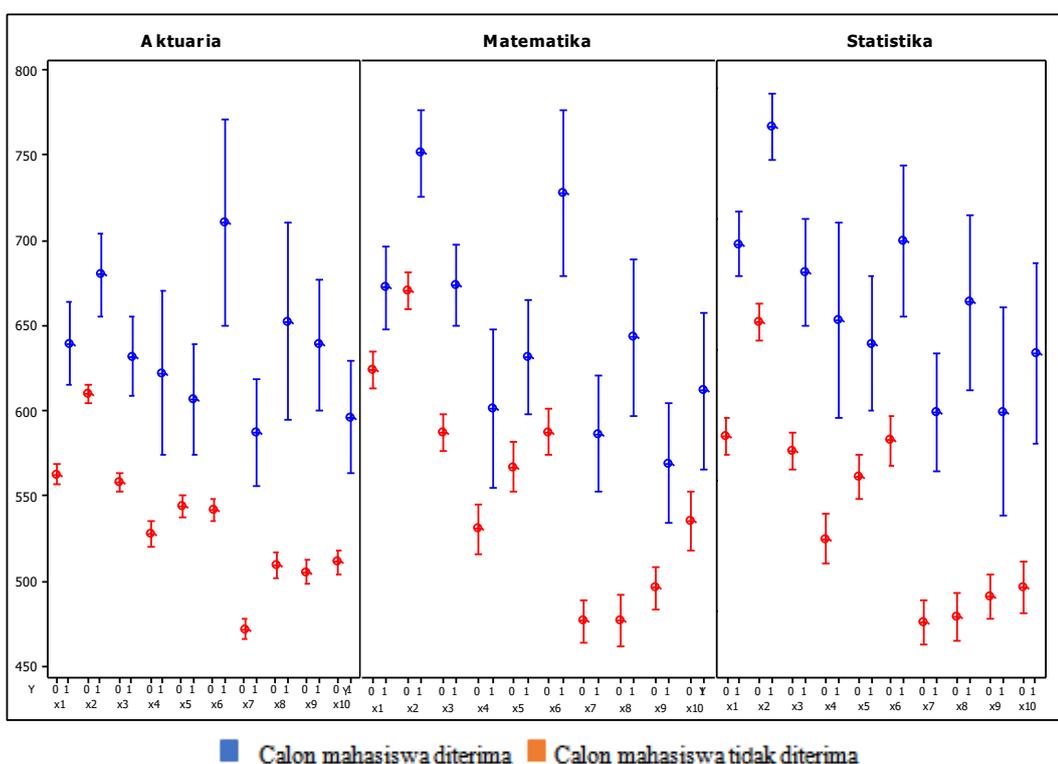
*plot* pada analisis ini digunakan untuk membandingkan rata-rata antara kelompok nilai calon mahasiswa yang diterima dan tidak diterima. Visualisasi dengan *interval plot* pada FSAINS ditampilkan pada Gambar 4.13.



**Gambar 4.13** *Interval Plot* pada Tiap Prodi di FSAINS

Berdasar Gambar 4.13 dapat disimpulkan bahwa calon mahasiswa yang diterima memiliki nilai rata-rata yang lebih tinggi dibandingkan dengan calon mahasiswa yang tidak diterima. Jika dilihat secara individual pada tiap grafik dapat diketahui bahwa untuk calon mahasiswa yang mendaftar dan diterima pada prodi Biologi ( $x_7;1$ ), Fisika ( $x_8;1$ ) dan Kimia ( $x_9;1$ ) menunjukkan bahwa masing-masing prodi telah menunjukkan rata-rata nilai TKD yang secara rata-rata mata uji yang berkorespondensi dengan prodi pilihan memiliki nilai yang lebih tinggi dibanding mahasiswa yang tidak diterima. Selain itu, nilai TPA numerikal untuk calon mahasiswa baru yang diterima ( $x_2;1$ ) pada seluruh prodi di FSAINS memiliki nilai rata-rata yang tinggi, artinya calon mahasiswa baru memiliki kemampuan numerikal yang tinggi. Pada penelitian ini nilai rata-rata hanya digunakan dalam visualisasi data saja dan tidak digunakan dalam proses inferens, mengingat variabilitas yang tinggi pada tiap nilai mata uji.

Gambar 4.13 menunjukkan bahwa calon mahasiswa baru yang mendaftar di FSAINS cenderung memiliki nilai matematika IPA ( $x_{10;1}$ ) yang rendah, dibandingkan nilai mata uji lainnya, hal ini tidak menjadi masalah karena identitas untuk masuk FSAINS bukanlah nilai matematika IPA. Sama halnya dengan *interval plot* pada FSAINS, *interval plot* di FMKSD juga memberikan gambaran bahwa calon mahasiswa yang diterima, memiliki nilai TKD rata-rata yang lebih tinggi dibandingkan dengan yang tidak diterima. Visualisasi dengan *interval plot* pada FMKSD ditampilkan pada Gambar 4.14.

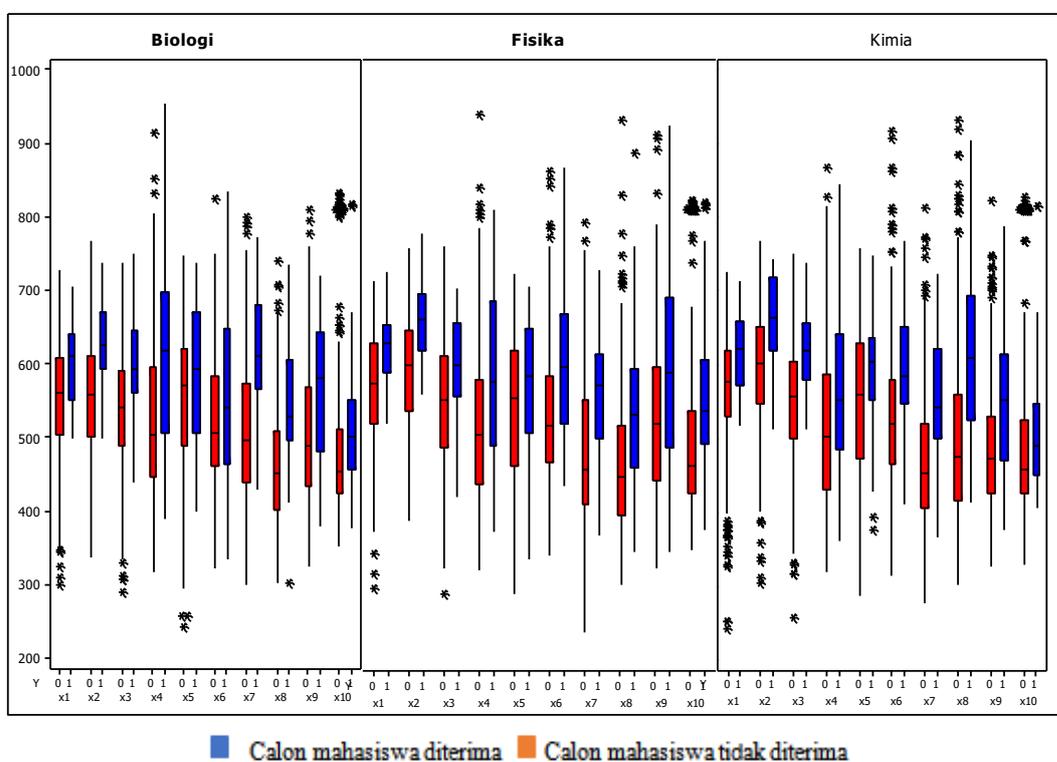


**Gambar 4.14** *Interval Plot* pada Tiap Prodi di FMKSD

Untuk prodi Aktuaria, Matematika dan Statistika, calon mahasiswa baru yang diterima cenderung memiliki nilai matematika dasar dan nilai TPA numerikal pada masing-masing prodi yang tinggi, meskipun demikian, nilai matematika IPA di ketiga prodi nilai rata-rata untuk mata uji matematika IPA tidak lebih dari 650, ini dapat diartikan bahwa calon mahasiswa yang memiliki TPA numerikal dan matematika dasar yang tinggi, belum tentu memiliki nilai matematika dasar yang tinggi juga mengingat ketiga mata uji ini sangat berbeda. Nilai matematika untuk calon mahasiswa baru yang memilih prodi matematika diibaratkan sebuah

identitas, jika calon mahasiswa baru tidak jelas atau tidak punya identitas maka harusnya, calon mahasiswa baru tidak dapat masuk, kenali dan digali potensinya di prodi matematika. Gambar 4.14 menunjukkan bahwa calon mahasiswa baru yang diterima memiliki nilai biologi yang rendah dibanding mata uji lain. Hal ini tidaklah sebuah masalah yang besar karena tidak ada hubungan antara nilai biologi dengan nilai matematika IPA.

Untuk FSAINS dan FMKSD dapat diperoleh kesimpulan bahwa ada kecenderungan nilai rata-rata pada mata ujian TKPA ( $x_1- x_6$ ), yaitu nilai ujian TKPA lebih tinggi dibanding nilai TKD ( $x_7- x_{10}$ ). Adanya kecenderungan ini dikarenakan pada tahun 2018, system penilaian masih berdasarkan pada nilai total. System yang berfokus pada nilai total menyebabkan banyaknya “manipulasi” untuk mendapatkan nilai yang tinggi. Manipulasi ini disarankan oleh lembaga bimbingan belajar dengan cara mengerjakan soal yang lebih mudah dan berfokus pada materi uji yang dikuasai saja. Langkah berikutnya data akan divisualisasikan ke dalam *boxplot*. Visualisasi ini dilakukan dengan tujuan mengukur variabilitas dan median dari data.



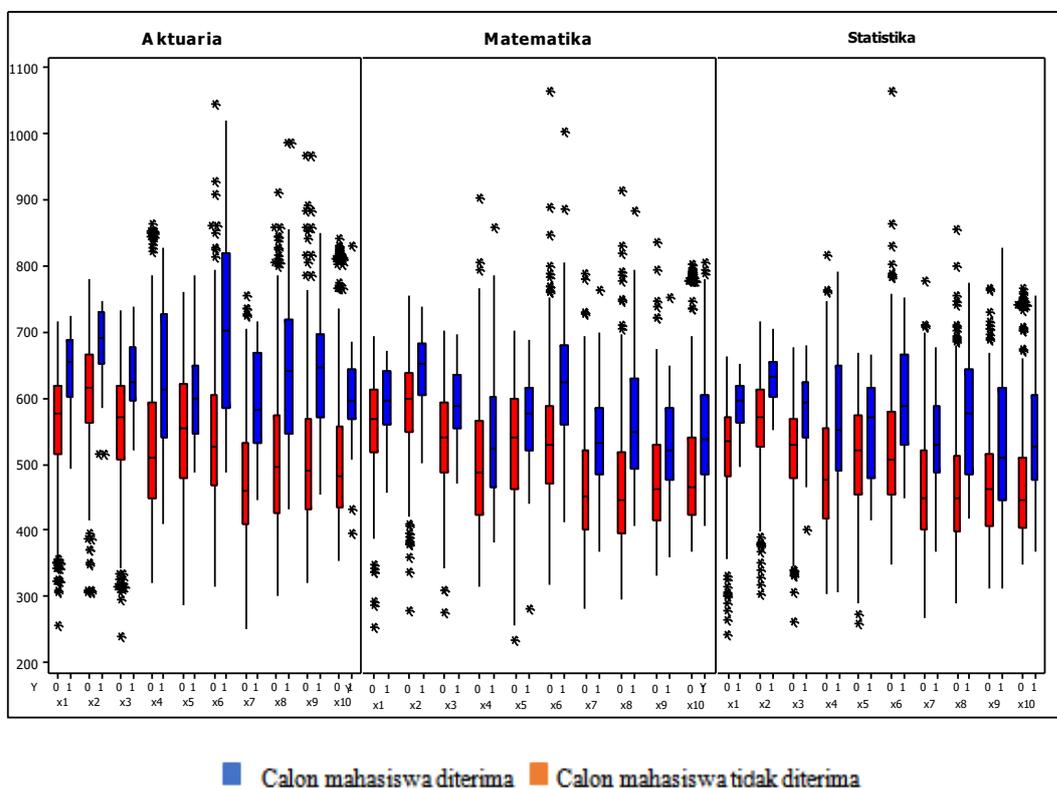
**Gambar 4.15** *Boxplot* pada Tiap Prodi di FSAINS

Visualisasi data dengan *boxplot* pada FSAINS ditunjukkan pada Gambar 4.15. Berdasar pada Gambar 4.15, secara umum dapat dilihat bahwa variabilitas dari nilai calon mahasiswa sangat tinggi. Variabilitas tertinggi untuk program studi biologi adalah mata ujian Bahasa Inggris ( $x_4$ ) dengan nilai interkuartil sebesar 192.662. sedangkan untuk prodi Fisika, variabilitas tertinggi ditunjukkan oleh mata ujian kimia dengan nilai interkuartil sebesar 202.372. Untuk prodi Kimia, variabilitas tertinggi ditunjukkan oleh mata ujian kimia ( $x_9$ ) dengan nilai interkuartil sebesar 167.303. Selain nilai variabilitas, *boxplot* pada Gambar 4.15 menunjukkan bahwa untuk semua prodi memiliki range nilai yang besar, ditunjukkan dengan panjangnya *boxplot* diseluruh mata uji.

Untuk prodi Biologi, *boxplot* dari nilai mata uji Biologi cenderung tidak saling tumpang tindih dan jika dibandingkan dengan nilai TKD lainnya dapat dilihat bahwa nilai biologi cenderung lebih tinggi dibandingkan nilai TKD lain. Data ini mengindikasikan bahwa secara tidak langsung prodi biologi menerapkan suatu keadaan dimana calon mahasiswa baru yang diterima memiliki nilai biologi lebih tinggi daripada TKD lainnya. Sedangkan untuk prodi Fisika, *boxplot* menginformasikan bahwa calon mahasiswa baru yang masuk prodi Fisika memiliki nilai kimia yang tinggi dan untuk prodi Kimia, calon mahasiswa baru cenderung memiliki nilai fisika yang tinggi. Adanya ketidaksesuaian ini menjadi pertanyaan, bagaimana calon mahasiswa yang kompetensi Kimia tinggi tetapi masuk prodi Fisika, dan sebaliknya? Oleh sebab itu, evaluasi akan dilakukan dengan menggunakan BMM. Yang menarik untuk dibahas adalah, hampir diseluruh nilai mata uji pada status penerimaan  $y=0$  memiliki *otliers* bagian atas yang tinggi, utamanya pada mata uji yang menjadi *mixture* di penelitian ini. Kondisi ini menunjukkan bahwa masih banyak calon mahasiswa yang punya nilai tinggi tetapi tidak diterima. Penyebab tidak diterima ini belum diketahui secara pasti bisa saja memang kuota penerimaan yang penuh ataupun faktor lain.

Gambar 4.16 adalah visualisasi *boxplot* pada tiap Prodi di FMKSD yang menunjukkan bahwa nilai memiliki variabilitas yang tinggi hampir di seluruh mata uji. Variabilitas tertinggi untuk program studi Aktuaria adalah mata ujian Matematika Dasar ( $x_6$ ) dengan nilai interkuartil sebesar 233.351. sedangkan untuk prodi Matematika, variabilitas tertinggi ditunjukkan oleh mata ujian Fisika ( $x_8$ )

dengan nilai interkuartil sebesar 145.433. Untuk prodi Statistika, variabilitas tertinggi ditunjukkan oleh mata ujian kimia ( $x_9$ ) dengan nilai interkuartil sebesar 193.971.



**Gambar 4.16** Box Plot pada Tiap Prodi di FMKSD

Dapat dilihat secara umum pada seluruh prodi bahwa calon mahasiswa baru yang diterima di FMKSD memiliki nilai TPA numerikal dan matematika dasar yang lebih tinggi dibanding mata uji lainnya. Tetapi nilai matematika yang dimiliki calon mahasiswa baru cenderung rendah, hal ini karena mata uji matematika IPA berbeda dengan TPA numerikal dan matematika dasar. Variabilitas nilai penerimaan yang tinggi khususnya pada prodi Aktuaria, menunjukkan bahwa nilai matematika dasar calon mahasiswa baru yang diterima sangat variatif dengan nilai minimum adalah 487 dan nilai tertinggi adalah 1021. Meskipun nilai sangat variatif, tetapi jika di bandingkan antara prodi Aktuaria, Matematika dan Statistika calon mahasiswa baru yang mendaftar di prodi Aktuaria memiliki nilai yang lebih tinggi dibanding prodi Statistika ataupun Matematika. Pada fakultas FMKSD terlihat jelas bahwa ada kecenderungan nilai TKPA yang lebih tinggi dibanding nilai TKD, ini bisa saja mengindikasikan jika

fokus calon mahasiswa baru yang memilih FMKSD bukan menyelesaikan nilai TKD.

Gambar 4.16 menunjukkan kecenderungan nilai TKPA yang lebih tinggi di banding nilai TKD. Hal ini menandakan bahwa ada indikasi mencapai nilai total tinggi dengan fokus pada nilai TKPA. Secara umum, visualisasi pada Gambar 4.15 ataupun Gambar 4.16 memberikan informasi yaitu masih banyak terdapat *upper outlier* untuk kelompok calon mahasiswa yang tidak diterima pada semua uji, khususnya pada bagian nilai TKD yang merupakan fokus pada penelitian ini. Adanya *upper outlier* dapat diartikan bahwa, sebenarnya masih banyak calon mahasiswa yang potensial untuk diterima (karena nilai TKD matematika IPA tinggi) tetapi menjadi tidak diterima karena tidak adanya standar ketentuan lain selain nilai total. Sehingga penelitian ini akan mengevaluasi status penerimaan mahasiswa baru yang mempertimbangkan nilai TKD dan prodi yang berkoresponden, dengan tujuan system penilaian yang akan datang tidak hanya berfokus pada nilai tetapi juga mempertimbangkan kompetensi siswa dan prodi pilihan. Tabel 4.13 menunjukkan daftar nilai mata ujian TKD dan prodi yang berkorespondensi.

**Tabel 4.13** Nilai TKD dan Prodi yang Berkoresponden

<b>Nilai TKD</b>	<b>Program Studi yang koresponden</b>
<b>Biologi</b>	Biologi
<b>Kimia</b>	Kimia
<b>Fisika</b>	Fisika
<b>Matematika IPA</b>	Matematika, Statistika dan Sains Aktuarial

#### 4.2.4 Pemodelan Bernoulli Mixture

Pemodelan BMM akan dilakukan dengan menganalisa data dengan menggunakan *software* OpenBUGS. Proses terpenting pada pemodelan BMM adalah proses identifikasi *mixture* dan penentuan *prior*. Berikut akan ditunjukkan tahapan pada pemodelan BMM, sedangkan penentuan *prior* dilakukan secara bertahap sesuai dengan penjabaran pada subbab 4.1.2.

### A. Identifikasi *Mixture*

Langkah-langkah identifikasi *mixture* yang ditunjukkan pada subbab 3.4 poin dua merupakan proses untuk membagi data sesuai dengan kriteria penerimaan. Tahap identifikasi *mixture* merupakan tahap yang paling penting pada penelitian ini. Hasil dari identifikasi *mixture* ditunjukkan pada Tabel 4.14.

**Tabel 4.14** Deskripsi Data Penerimaan Calon Mahasiswa Baru Berdasar Kriteria P

Status Prodi	Diterima (Y=1)		Tidak Diterima (Y=0)	
	Nilai Tinggi (P=1)	Nilai Rendah (P=0)	Nilai Tinggi (P=1)	Nilai Rendah (P=0)
<b>Biologi</b>	20 siswa	20 siswa	132 siswa	278 siswa
<b>Fisika</b>	21 siswa	27 siswa	110 siswa	150 siswa
<b>Kimia</b>	23 siswa	23 siswa	124 siswa	286 siswa
<b>Aktuarial</b>	3 siswa	24 siswa	220 siswa	600 siswa
<b>Matematika</b>	11 siswa	40 siswa	125 siswa	301 siswa
<b>Statistika</b>	9 siswa	35 siswa	98 siswa	382 siswa

Dari Tabel 4.14 diketahui calon mahasiswa yang masuk pada program studi biologi tetapi masih memiliki nilai biologi bukan yang tertinggi ada sebanyak 50% atau sebanyak 20 calon mahasiswa. Dari tabel terlihat bahwa masih ada 132 calon mahasiswa yang memiliki nilai biologi yang lebih baik, tetapi tidak diterima pada program studi biologi. Untuk prodi kimia, calon mahasiswa baru yang berstatus diterima tetapi memiliki nilai rendah, tetapi diterima ada 23 calon mahasiswa. Dari tabel terlihat bahwa masih ada 124 calon mahasiswa yang memiliki nilai kimia yang lebih baik, tetapi tidak diterima pada program studi kimia. Untuk prodi Fisika, diketahui bahwa dari total 48 calon mahasiswa yang diterima, 27 calon tidak memiliki nilai fisika tertinggi, padahal masih ada 110 calon mahasiswa yang memiliki nilai fisika yang lebih baik.

Berbeda dengan FSAINS, FMKSD persentase calon mahasiswa yang diterima dengan nilai matematika tidak tertinggi hampir 50% bahkan lebih. Untuk prodi matematika, 51 mahasiswa yang diterima, hanya 11 orang yang memiliki nilai matematika tertinggi. Untuk prodi statistika dari 44 mahasiswa yang diterima, hanya 9 orang yang saja yang dianggap memenuhi kriteria, sedangkan 35 siswa lainnya masih berada pada kategori diterima tetapi memiliki nilai lebih rendah dibanding nilai TKD lainnya. Untuk prodi Aktuarial, dari 27 mahasiswa

yang diterima, hanya 3 orang yang saja yang dianggap memenuhi kriteria, sedangkan 24 siswa lainnya masih berada pada kategori diterima tetapi memiliki nilai lebih rendah dibanding nilai TKD lainnya. Berdasarkan hasil analisis inilah, akan dilakukan penelitian untuk memodelkan status penerimaan calon mahasiswa baru dengan menggunakan kriteria penerimaan dengan menggunakan metode *Bernoulli mixture* model. Setelah proses identifikasi *mixture* selesai pada semua prodi, maka tahapan selanjutnya adalah melakukan pembagian data *training* dan *testing*.

### B. Pembagian Data *Training* dan *Testing*

Pembagian data *training* dan *testing* pada penelitian ini harus dilakukan dengan cara bertahap. Pembagian data *training* dan *testing* dilakukan dengan tujuan mendapatkan hasil ketepatan klasifikasi dari hasil prediksi dari model. Pembagian data yang tidak tepat akan mempengaruhi ketepatan akurasi untuk prediksi. Pembagian data *training* dan *testing* untuk penelitian ini perlu mempertimbangkan pembagian kategori *mixture*-nya.

**Tabel 4.15** Pembagian Data *Training* dan *Testing* pada FMKSD

	Training			Total	Testing		Total
	Tidak diterima	Diterima			Tidak diterima	Diterima	
Aktuaria	Mix 1	480	2	482	120	1	121
	Mix 2	176	19	195	44	5	49
	Total	656	21	677	164	6	170
Matematika	Mix 1	240	9	249	61	2	63
	Mix 2	100	32	132	25	8	33
	Total	340	41	381	86	10	96
Statistika	Mix 1	305	7	312	77	2	79
	Mix 2	78	28	106	20	7	27
	Total	383	35	418	97	9	106

Pembagian data *training* dan *testing* tanpa mempertimbangkan faktor *mixture* pada data ini, akurasi yang diberikan akan semakin kecil. Hasil pembagian data *training* dan *testing* pada FMKSD akan ditunjukkan pada Tabel

4.15. Hasil pembagian data training dan *testing* menunjukkan bahwa variabel respon yang digunakan pada penelitian ini merupakan variabel respon dengan proporsi penerimaan yang *imbalance*. Hal ini menyebabkan perbandingan pengambilan data untuk status diterima dan tidak diterima sangat kecil. Bahkan untuk prodi Aktuaria, data *testing* yang digunakan pada *mixture* 1 dengan status diterima hanya satu orang. Selanjutnya, hasil pembagian data *training* dan *testing* untuk Fakultas Sains ditunjukkan pada Tabel 4.16.

**Tabel 4.16** Pembagian Data *Training* dan *Testing* pada FSAINS

	Training			Testing			
	Tidak diterima	diterima	Total	Tidak diterima	diterima	Total	
<b>Fisika</b>	Mix 1	119	17	136	31	4	35
	Mix 2	88	21	109	22	6	28
	Total	207	38	245	53	10	63
<b>Kimia</b>	Mix 1	229	18	247	57	5	62
	Mix 2	99	18	117	25	5	30
	Total	328	36	364	82	10	92
<b>Biologi</b>	Mix 1	222	16	238	56	4	60
	Mix 2	105	16	121	27	4	31
	Total	327	32	359	83	8	91

#### 4.2.5 Hasil Pemodelan dengan Metode BMM Pada Fakultas Sains (Fisika, Biologi, Kimia)

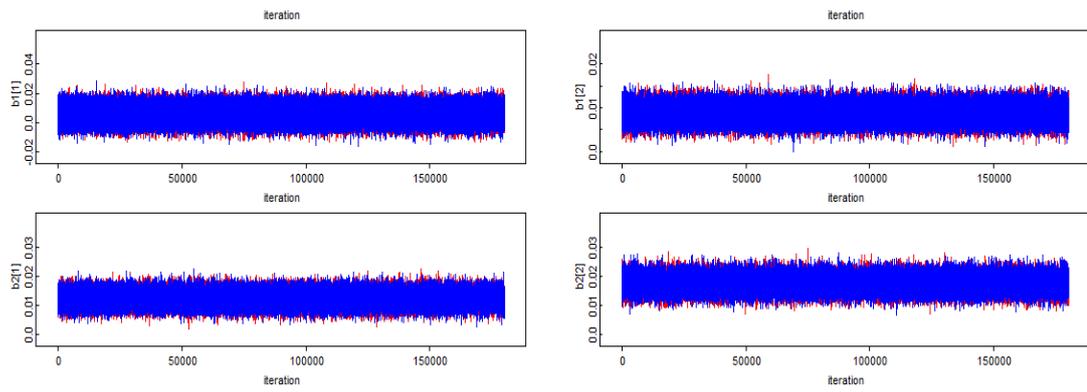
Tahap berikutnya akan dilakukan estimasi pada parameter Bernoulli Mixture pada tiap-tiap prodi di Fakultas Sains. Hasil yang akan ditunjukkan pada estimasi ini meliputi hasil estimasi parameter, uji konvergensi, *confusion matrix* dan akurasi. Parameter model *Bernoulli Mixture* akan diestimasi secara komputasional berdasar struktur DAG pada Gambar 4.1 yang diterjemahkan ke dalam koding yang ditunjukkan pada Lampiran 3. Detail proses dan hasil estimasi parameter akan dijelaskan pada masing-masing prodi.

##### A. Prodi Fisika

Estimasi parameter dengan melibatkan enam variabel prediktor menghasilkan 12 parameter, yang mana keseluruhan parameter memiliki

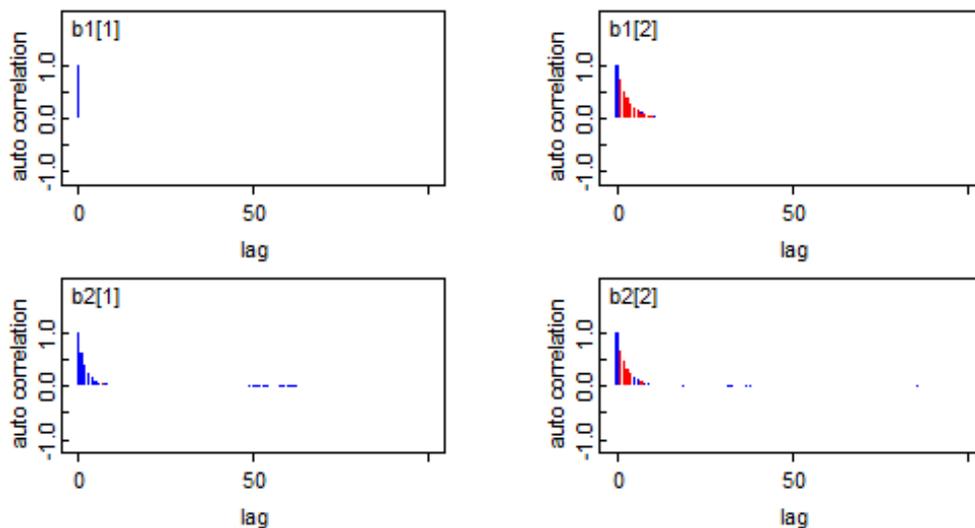
karakteristik yang menunjukkan sifat konvergen pada MCMC yaitu *irreducible*, *aperiodic* yang akan ditunjukkan MCMC *diagnostic plot* sebagai berikut.

1. Serial *plot* dari rantai MCMC yang dibangun, terlihat bahwa sampel hasil proses MCMC tidak menunjukkan kecenderungan pada pola tertentu, *stasioner* dan bersifat random. Hasil visualisasi Gambar 4.17 menunjukkan bahwa *serial plot* menunjukkan hasil yang random dan tidak berpola tertentu.



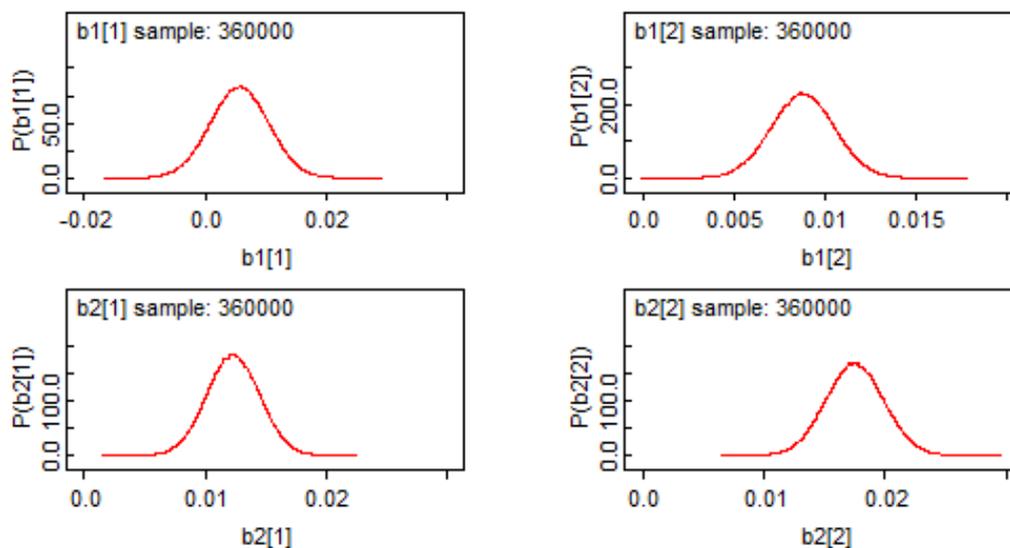
**Gambar 4.17** *Serial Plot* dari Hasil Estimasi Parameter BMM Prodi Fisika

2. Plot *autocorrelation* yang dihasilkan proses estimasi juga membuktikan bahwa sampel yang dihasilkan proses MCMC telah bersifat random, ditunjukkan dengan autokorelasi yang bernilai satu hanya pada *lag 0* dan bernilai mendekati nol pada *lag* selanjutnya. *Autocorrelation* plot ditunjukkan pada Gambar 4.18.



**Gambar 4.18** *Autocorrelation Plot* Prodi Fisika

3. *Density plots* yang dihasilkan pada proses estimasi dengan *OpenBUGS* juga menunjukkan hasil yang random dan secara stabil berada pada satu zona tanpa adanya kecenderungan pola tertentu. Selain itu *density plot* kedua rantai yang memiliki *initial value* berbeda juga saling bersinggungan. Hasil *density plots* ditunjukkan pada Gambar 4.19.



**Gambar 4.19** *Density plot* dari Hasil Estimasi Parameter BMM Prodi Fisika

Untuk memperkuat hasil visualisasi *Diagnostic plot*, metode inferensia akan digunakan sebagai bukti penguat untuk menunjukkan bahwa parameter yang dihasilkan telah memenuhi syarat konvergen. Metode inferensia yang digunakan yaitu dengan menggunakan uji *Raftery-Lewis*. Uji *Raftery-Lewis* menunjukkan dependensi pada variabel dalam penelitian. Hasil uji akan dikatakan telah memenuhi syarat konvergen jika nilai *dependence factor* ( $I$ ) pada masing-masing variabel di tiap-tiap *chain* tidak lebih dari 5 ( $I \leq 5$ ). Hasil uji yang menunjukkan nilai  $I > 5$  mengindikasikan tingginya autokorelasi pada tiap parameter (Ntzoufras, 2009). Hasil Uji *Raftery-Lewis* ditunjukkan pada Tabel 4.17. Hasil pada Tabel 4.17 menunjukkan bahwa nilai *dependence factor* ( $I$ ) tidak lebih besar dari lima ( $I \leq 5$ ), sehingga dapat dikatakan bahwa rantai telah konvergen. Dari hasil *diagnostic plot* yaitu *serial*, *autocorrelation* dan *density plot* serta pemeriksaan secara inferensia, dapat ditarik kesimpulan bahwa proses estimasi parameter telah mencapai kondisi yang konvergen.

**Tabel 4.17** Pemeriksaan Konvergensi dengan *Raftery-Lewis* Prodi Fisika

<b>Raftery Lewis : Dependence Factor (I)</b>			
	Chain 1	Chain 2	
<b>Mix 1</b>	P[1]	1.01	1
	b0[1]	3.49	3.5
	b1[1]	0.999	0.995
	b2[1]	2.53	3.33
	b3[1]	2.3	2.35
	b4[1]	1.07	1.08
	b5[1]	1.1	1.1
<b>Mix 2</b>	b6[1]	1.13	1.13
	P[2]	1	1.01
	b0[2]	3.36	3.49
	b1[2]	3.7	3.7
	b2[2]	3.74	3.56
	b3[2]	2.28	2.17
	b4[2]	1.06	1.07
b5[2]	1.12	1.12	
b6[2]	2.08	1.18	

Dengan demikian, proses estimasi parameter yang dihasilkan dapat menggambarkan model BMM untuk prodi Fisika. Hasil parameter pada kondisi konvergen untuk prodi Fisika ditunjukkan pada Tabel 4.18.

**Tabel 4.18** Hasil Estimasi Parameter Untuk Prodi Fisika

	<b>Par.</b>	<b>Estimasi</b>	<b>Standar Deviasi</b>	<b>2.50%</b>	<b>97.50%</b>	<b>MC Error</b>	<b>Sig.</b>
<b>Mix 1</b>	P[1]	0.5547	0.0316	0.4925	0.6160	5.43E-05	ya
	b0[1]	-31.96	1.9670	-35.83	-28.10	7.45E-03	ya
	TPA Figural	0.0057	0.0048	-0.0037	0.0151	8.02E-06	tidak
	TPA Numerikal	0.0123	0.0022	0.0081	0.0166	7.73E-06	ya
	TPA Verbal	0.0096	0.0022	0.0054	0.0139	6.38E-06	ya
	Bahasa Inggris	0.0048	0.0013	0.0023	0.0073	2.43E-06	ya
	Bahasa Indonesia	0.0070	0.0015	0.0040	0.0100	3.48E-06	ya
	Matematika Dasar	0.0081	0.0014	0.0053	0.0109	3.08E-06	ya
<b>Mix 2</b>	P[2]	0.4453	0.0316	0.3840	0.5075	5.43E-05	ya
	b0[2]	-27.85	1.8490	-31.48	-24.23	7.27E-03	ya
	TPA Figural	0.0088	0.0017	0.0054	0.0122	7.09E-06	ya
	TPA Numerikal	0.0175	0.0024	0.0129	0.0222	9.21E-06	ya
	TPA Verbal	0.0028	0.0020	-0.0011	0.0066	5.77E-06	tidak
	Bahasa Inggris	0.0071	0.0012	0.0046	0.0095	2.55E-06	ya
	Bahasa Indonesia	0.0066	0.0015	0.0038	0.0095	3.24E-06	ya
	Matematika Dasar	0.0016	0.0016	-0.0016	0.0047	3.81E-06	tidak

Kolom estimasi pada Tabel 4.18 menunjukkan hasil estimasi untuk tiap parameter. Kolom standar deviasi akan dibandingkan dengan nilai pada kolom MC Error. Nilai MC Error merupakan sebuah nilai yang digunakan untuk mengukur variabilitas pada tiap hasil estimasi, jika nilai MC Error lebih kecil dibandingkan nilai standar deviasi, maka dapat dikatakan hasil estimasi di estimasi dengan presisi yang tinggi. Estimasi parameter dengan presisi tinggi menunjukkan bahwa hasil estimasi mendekati nilai sebenarnya (Ntzoufras, 2009). Pada penelitian ini, menunjukkan bahwa nilai MC Error sangat kecil dan lebih kecil dari nilai standar deviasi, sehingga dapat dikatakan bahwa hasil estimasi mendekati nilai sebenarnya.

Uji signifikansi parameter BMM menggunakan *credible interval*. Nilai *credible interval* merupakan sebuah nilai yang dibentuk dari *Highest Posterior Density* (HPD) dengan ketentuan jika *credible interval* memuat nilai nol (0) maka hasil estimasi parameter tidak signifikan dan sebaliknya, jika *credible interval* tidak memuat nilai nol (0) maka hasil estimasi parameter dikatakan signifikan. Hasil uji signifikansi parameter menunjukkan bahwa pada *mixture 1* ada satu variabel yang tidak signifikan yaitu nilai TPA Figural. Kemudian untuk *mixture 2* ada dua variabel yang tidak signifikan yaitu nilai TPA verbal dan matematika dasar. Hal ini dapat menunjukkan bahwa selain memiliki kompetensi yang tinggi di bidang fisika, seorang calon mahasiswa yang ingin masuk prodi Fisika harus memiliki kompetensi yang tinggi pada empat mata uji lain meliputi nilai mata uji TPA numerik, TPA verbal, bahasa indonesia, bahasa inggris dan matematika dasar. Model yang terbentuk akan memenuhi persamaan berikut.

$$\begin{aligned}
 &P_{mix}(\pi, x, \beta) \\
 &= 0.555 \frac{\exp(-31.960 + 0.006 x_1 + 0.012 x_2 + 0.010 x_3 + 0.005 x_4 + 0.007 x_5 + 0.008 x_6)}{1 + \exp(-31.960 + 0.006 x_1 + 0.012 x_2 + 0.010 x_3 + 0.005 x_4 + 0.007 x_5 + 0.008 x_6)} \\
 &+ 0.445 \frac{\exp(-27.850 + 0.009 x_1 + 0.018 x_2 + 0.003 x_3 + 0.007 x_4 + 0.007 x_5 + 0.002 x_6)}{1 + \exp(-27.850 + 0.009 x_1 + 0.018 x_2 + 0.003 x_3 + 0.007 x_4 + 0.007 x_5 + 0.002 x_6)}
 \end{aligned}$$

Hasil model yang terbentuk menunjukkan bahwa model akan cenderung mengelompokkan calon mahasiswa baru ke dalam kelompok *mixture 1* dengan probabilitas 0.555 kali lebih besar dibanding *mixture 2*. Artinya kemampuan model untuk mendeteksi status penerimaan calon mahasiswa baru pada prodi

Fisika dengan kriteria benar akan lebih tinggi 0.555 kali dibanding mendeteksi status penerimaan calon mahasiswa baru dengan kriteria salah.

Nilai positif pada tiap-tiap variabel yang signifikan berarti bahwa semakin tinggi nilai TPA numerik, TPA verbal, bahasa indonesia, bahasa inggris dan matematika dasar maka kecenderungan untuk masuk kategori *mixture* 1 akan semakin besar dan sebaliknya. Dari hasil model yang terbentuk akan digunakan untuk memprediksi status penerimaan calon mahasiswa. Hasil prediksi didapat dengan mengintegrasikan nilai-nilai pada data *testing* dengan model. Hasil ketepatan klasifikasi dari model ditunjukkan pada Tabel 4.19.

**Tabel 4.19** Hasil Klasifikasi Pada Tiap *Mixture* Prodi Fisika

Aktual	Prediksi	
	Mixture 1	Mixture 2
Mixture 1	32	3
Mixture 2	14	14

Hasil ketepatan klasifikasi pada tiap *mixture* didapat dari membandingkan *mixture* aktual dengan *mixture* hasil prediksi model. Ketepatan klasifikasi menunjukkan bahwa model BMM memberikan akurasi sebesar 73.02%. Model BMM memprediksi bahwa 32 calon mahasiswa baru tepat diklasifikasikan untuk masuk pada *mixture* 1 dan 14 calon mahasiswa baru tepat diprediksi masuk pada kelompok *mixture* 2. Rincian status penerimaan pada tiap *mixture* ditunjukkan pada Table 4.20 dan Tabel 4.21.

**Tabel 4.20** Rincian Prediksi Status Penerimaan Pada *Mixture* 1 Prodi Fisika

Status	Nilai TKD Fisika	
	Tinggi	Rendah
Diterima	1	0
Tidak Diterima	3	31

Tabel 4.20 menunjukkan rincian status penerimaan pada *mixture* 1 yang memberikan gambaran bahwa hasil prediksi pada *mixture* 1 adalah mahasiswa yang tidak diterima dengan kondisi benar (0,0), artinya model lebih banyak memprediksi calon mahasiswa baru yang benar dikategorikan dengan kriteria

yang benar. Model salah memprediksi 3 calon mahasiswa baru yang tidak diterima tetapi nilai tinggi.

**Tabel 4.21** Rincian Prediksi Status Penerimaan Pada *Mixture 2* Prodi Fisika

Status	Nilai TKD Fisika	
	Tinggi	Rendah
Diterima	11	3
Tidak Diterima	11	3

Tabel 4.21 menunjukkan rincian status penerimaan pada *mixture 2*. Pada kelompok *mixture 2* merupakan kelompok status penerimaan dengan kriteria salah, yang mana model memprediksi bahwa ada 24 calon mahasiswa yang memang salah di klasifikasikan dan memprediksi ada 11 calon mahasiswa baru yang akan diterima dengan nilai tinggi dan 3 calon mahasiswa baru yang akan tidak diterima dengan nilai rendah.

#### B. Prodi Kimia

Menggunakan analogi yang sama diberikan pada proses estimasi pada prodi Kimia, hasil *diagnostic plot* dan pemeriksaan *Raftery-Lewis* menunjukkan bahwa proses estimasi parameter telah menunjukkan hasil yang konvergen.

**Tabel 4.22** Pemeriksaan Konvergensi dengan *Raftery-Lewis* Prodi Kimia

Raftery Lewis : Dependence Factor (I)			
		Chain 1	Chain 2
Mixture 1	P[1]	1	1
	b0[1]	2.51	3.46
	b1[1]	0.995	1
	b2[1]	2.56	3.35
	b3[1]	3.26	3.3
	b4[1]	1.07	1.06
	b5[1]	2.16	2.12
	b6[1]	1.2	1.2
Mixture 2	P[2]	0.999	1.01
	b0[2]	3.63	3.66
	b1[2]	3.93	3.71
	b2[2]	4.58	3.83
	b3[2]	3.49	3.48
	b4[2]	1.05	1.05
	b5[2]	2.22	2.11
	b6[2]	1.14	1.13

Hasil *diagnostic plot* akan ditunjukkan pada Lampiran 4. Nilai I pada Tabel 4.22 menunjukkan bahwa proses estimasi parameter sudah konvergen dan tidak terjadi autokorelasi. Dengan demikian, dari hasil pemeriksaan uji konvergensi dapat ditarik kesimpulan bahwa hasil estimasi parameter telah memenuhi kondisi konvergen dan dapat menggambarkan memodelkan status penerimaan mahasiswa baru untuk prodi Kimia.

Hasil estimasi parameter untuk prodi Kimia pada kondisi konvergen ditunjukkan pada Tabel 4.23. Hasil estimasi pada prodi Kimia menunjukkan bahwa nilai MC Error sangat kecil dan lebih kecil dari nilai standar deviasi, sehingga dapat dikatakan bahwa hasil estimasi mendekati nilai sebenarnya.

**Tabel 4.23** Hasil Estimasi Parameter Untuk Prodi Kimia

	<b>Par.</b>	<b>Estimasi</b>	<b>Standar Deviasi</b>	<b>2.50%</b>	<b>97.50%</b>	<b>MC Error</b>	<b>Sig.</b>
<b>Mix 1</b>	P[1]	0.6776	0.0245	0.6287	0.7245	4.10E-05	Ya
	b0[1]	-27.63	1.7250	-31.02	-24.26	0.006373	Ya
	TPA Figural	0.0076	0.0050	-0.0022	0.0174	8.65E-06	Tidak
	TPA Numerikal	0.0073	0.0020	0.0034	0.0112	7.06E-06	Ya
	TPA Verbal	0.0137	0.0021	0.0096	0.0179	7.08E-06	Ya
	Bahasa Inggris	0.0011	0.0011	-0.0012	0.0033	2.21E-06	Tidak
	Bahasa Indonesia	0.0054	0.0016	0.0022	0.0085	4.06E-06	Ya
	Matematika Dasar	0.0085	0.0014	0.0058	0.0112	3.23E-06	Ya
<b>Mix 2</b>	P[2]	0.3224	0.0245	0.2755	0.3713	4.10E-05	Ya
	b0[2]	-25.74	1.9360	-29.54	-21.95	0.008163	Ya
	TPA Figural	0.0064	0.0017	0.0032	0.0097	6.75E-06	Ya
	TPA Numerikal	0.0134	0.0025	0.0085	0.0183	1.05E-05	Ya
	TPA Verbal	0.0091	0.0023	0.0045	0.0136	8.48E-06	Ya
	Bahasa Inggris	0.0039	0.0012	0.0016	0.0062	2.16E-06	Ya
	Bahasa Indonesia	0.0042	0.0017	0.0008	0.0076	4.46E-06	Ya
	Matematika Dasar	0.0029	0.0015	-0.0001	0.0058	3.28E-06	Tidak

Hasil uji signifikansi menunjukkan bahwa untuk *mixture 1* terdapat empat variabel yang berpengaruh untuk memodelkan status penerimaan dengan kriteria benar. Keempat nilai tersebut yaitu nilai mata uji TPA numerik, TPA verbal, bahasa indonesia dan matematika dasar. Model yang terbentuk akan memenuhi persamaan berikut.

$$\begin{aligned}
& P_{kim}(\pi, x, \beta) \\
&= 0.6776 \frac{\exp(-27.63 + 0.0076x_1 + 0.0073x_2 + 0.0137x_3 + 0.0011x_4 + 0.0054x_5 + 0.0085x_6)}{1 + \exp(-27.63 + 0.0076x_1 + 0.0073x_2 + 0.0137x_3 + 0.0011x_4 + 0.0054x_5 + 0.0085x_6)} \\
&+ 0.3224 \frac{\exp(-25.74 + 0.0064x_1 + 0.0134x_2 + 0.0091x_3 + 0.0039x_4 + 0.0042x_5 + 0.0029x_6)}{1 + \exp(-25.74 + 0.0064x_1 + 0.0134x_2 + 0.0091x_3 + 0.0039x_4 + 0.0042x_5 + 0.0029x_6)}
\end{aligned}$$

Hasil pemodelan menunjukkan bahwa model cenderung mengelompokkan calon mahasiswa baru pada *mixture* 1 dengan probabilitas 0.6776 lebih besar dibanding pada *mixture* 2. Artinya untuk model lebih cenderung mendeteksi status penerimaan calon mahasiswa baru dengan kriteria benar pada prodi Kimia dengan probabilitas lebih tinggi dibandingkan mendeteksi status penerimaan calon mahasiswa baru dengan kriteria salah. Variabel nilai mata uji yang mempengaruhi status penerimaan dengan kriteria benar pada prodi Kimia adalah variabel nilai TPA numerik, TPA verbal, bahasa indonesia dan matematika dasar. Selanjutnya hasil pemodelan akan digunakan untuk memprediksi status penerimaan calon mahasiswa baru. Hasil klasifikasi ditunjukkan pada Tabel 4.24.

**Tabel 4.24** Hasil Klasifikasi Pada Tiap *Mixture* Prodi Kimia

	Prediksi	
Aktual	Mixture 1	Mixture 2
Mixture 1	46	16
Mixture 2	8	22

Hasil pada Tabel 4.24 menghasilkan prediksi dari model BMM yang terbentuk, dengan memberikan ketepatan klasifikasi sebesar 73.91%. Dari hasil, diketahui bahwa model memprediksi secara tepat 46 calon mahasiswa baru untuk diklasifikasikan pada *mixture* 1 dan 22 calon mahasiswa baru tepat diklasifikasikan pada *mixture* 2. Rincian status penerimaan pada *mixture* 1 dan *mixture* 2 ditunjukkan pada Tabel 4.25 dan Tabel 4.26.

**Tabel 4.25** Rincian Prediksi Status Penerimaan Pada *Mixture* 1 Prodi Kimia

Status	Nilai TKD Kimia	
	Tinggi	Rendah
Diterima	4	1
Tidak Diterima	15	42

Tabel 4.25 memberikan informasi bahwa 62 dari 92 calon mahasiswa baru dikategorikan ke dalam *mixture* 1. Ini menunjukkan bahwa model cenderung mengkategorikan calon mahasiswa baru ke dalam kategori *mixture* 1 dengan memprediksi 42 calon mahasiswa baru tidak diterima karena nilai rendah dan 4 calon mahasiswa baru diterima dengan nilai tinggi. Selain itu model *mixture* 1 juga memprediksi 16 calon mahasiswa baru dengan kriteria salah. sedangkan untuk tabel 4.26, menunjukkan prediksi pada kelompok *mixture* 2. Hasil prediksi menunjukkan bahwa model memprediksi 22 calon mahasiswa baru pada kelompok salah ke dalam status tidak diterima dengan nilai rendah dan 8 calon mahasiswa baru yang salah di kategorikan.

**Tabel 4.26** Rincian Prediksi Status Penerimaan Pada *Mixture* 2 Prodi Kimia

Status	Nilai TKD Kimia	
	Tinggi	Rendah
Diterima	0	5
Tidak Diterima	3	22

### C. Prodi Biologi

Menggunakan analogi yang sama sesuai pada prodi Fisika dan Kimia, proses estimasi parameter pada prodi Biologi juga diestimasi dengan MCMC algoritma *Gibbs sampling*. Iterasi pada proses MCMC dilakukan sebanyak 180.000 kali dan *thin* 180. Hasil pemeriksaan konvergensi diketahui bahwa, hasil pemeriksaan dengan *diagnostic plot* yang ditunjukkan pada Lampiran 5 telah menunjukkan hasil yang konvergen. Sedangkan hasil pemeriksaan konvergensi dengan metode *Raftery-Lewis* telah menunjukkan hasil yang konvergen, dengan nilai  $I \leq 5$ , pada masing-masing rantai markov. Nilai  $I \leq 5$  menunjukkan bahwa sudah tidak ada autokorelasi di tiap-tiap parameter. Hasil pemeriksaan dengan metode *Raftery-Lewis* akan ditunjukkan pada Tabel 4.27. Dari hasil pemeriksaan *diagnostic plot*, dengan metode *Raftery-Lewis* dapat ditarik kesimpulan bahwa proses estimasi parameter telah mencapai kondisi yang konvergen. Dengan demikian, proses estimasi parameter yang dihasilkan dapat digunakan untuk pemodelan status penerimaan calon mahasiswa baru.

**Tabel 4.27** Pemeriksaan Konvergensi dengan *Raftery-Lewis* Prodi Biologi

<b>Raftery Lewis : Dependence Factor (I)</b>			
	Chain 1	Chain 2	
<b>Mixture 1</b>	P[1]	1	1
	b0[1]	1.01	0.999
	b1[1]	2.27	2.28
	b2[1]	2.3	2.27
	b3[1]	1	0.995
	b4[1]	3.49	3.52
	b5[1]	2.28	2.27
	b6[1]	2.25	2.26
<b>Mixture 2</b>	P[2]	2.23	2.26
	b0[2]	2.34	2.27
	b1[2]	1.05	1.04
	b2[2]	1.11	1.1
	b3[2]	1.13	1.13
	b4[2]	1.16	1.15
	b5[2]	1.11	1.09
	b6[2]	1.14	1.11

Hasil estimasi parameter yang telah mencapai kondisi konvergen untuk prodi Biologi ditunjukkan pada Tabel 4.28

**Tabel 4.28** Hasil Estimasi Parameter untuk Prodi Biologi

	<b>Par.</b>	<b>Estimasi</b>	<b>Standar Deviasi</b>	<b>2.50%</b>	<b>97.50%</b>	<b>MC Error</b>	<b>Sig.</b>
<b>Mix 1</b>	P[1]	0.6621	0.0249	0.6125	0.7099	4.28E-05	ya
	b0[1]	-21.72	1.5140	-24.69	-18.76	0.004657	ya
	TPA Figural	0.0060	0.0044	-0.0027	0.0147	7.25E-06	tidak
	TPA Numerikal	0.0095	0.0018	0.0059	0.0131	5.24E-06	ya
	TPA Verbal	0.0100	0.0019	0.0063	0.0137	5.61E-06	ya
	Bahasa Inggris	0.0024	0.0011	0.0003	0.0045	1.95E-06	ya
	Bahasa Indonesia	0.0017	0.0014	-0.0011	0.0045	2.96E-06	tidak
	Matematika Dasar	0.0015	0.0015	-0.0014	0.0043	2.95E-06	tidak
<b>Mix 2</b>	P[2]	0.3379	0.0249	0.2901	0.3875	4.28E-05	ya
	b0[2]	-17.88	1.6240	-21.07	-14.70	0.005083	ya
	TPA Figural	0.0080	0.0016	0.0049	0.0111	5.7E-06	ya
	TPA Numerikal	0.0085	0.0019	0.0048	0.0123	5.23E-06	ya
	TPA Verbal	0.0016	0.0022	-0.0027	0.0059	6.3E-06	tidak
	Bahasa Inggris	0.0065	0.0013	0.0040	0.0090	2.69E-06	ya
	Bahasa Indonesia	-0.0003	0.0016	-0.0034	0.0027	3.51E-06	tidak
	Matematika Dasar	0.0031	0.0016	0.0000	0.0063	3.43E-06	tidak

Hasil estimasi pada prodi Biologi menunjukkan bahwa nilai MC Error mendekati nol dan lebih kecil dari nilai standar deviasi, sehingga dapat dikatakan bahwa hasil estimasi parameter mendekati nilai sebenarnya. Dari hasil uji signifikansi menunjukkan bahwa empat variabel yang berpengaruh untuk memodelkan status penerimaan dengan kriteria benar. Keempat nilai tersebut yaitu nilai mata uji TPA numerik, TPA verbal, bahasa Inggris dan matematika dasar. Model yang terbentuk akan memenuhi persamaan berikut. Hasil pemodelan menunjukkan bahwa model cenderung mengelompokkan calon mahasiswa baru pada *mixture* 1 dengan probabilitas 0.6621 lebih besar dibanding pada *mixture* 2.

$$P_{mix}(\pi, x, \beta) = 0.6621 \frac{\exp(-21.72 + 0.0060x_1 + 0.0095x_2 + 0.0100x_3 + 0.0024x_4 + 0.0017x_5 + 0.0015x_6)}{1 + \exp(-21.72 + 0.0060x_1 + 0.0095x_2 + 0.0100x_3 + 0.0024x_4 + 0.0017x_5 + 0.0015x_6)} + 0.3379 \frac{\exp(-17.88 + 0.0080x_1 + 0.0085x_2 + 0.0016x_3 + 0.0065x_4 + -0.0003x_5 + 0.0031x_6)}{1 + \exp(-17.88 + 0.0080x_1 + 0.0085x_2 + 0.0016x_3 + 0.0065x_4 + -0.0003x_5 + 0.0031x_6)}$$

Hasil pemodelan memberikan arti bahwa model lebih cenderung mendeteksi status penerimaan calon mahasiswa baru dengan kriteria benar pada prodi Biologi dengan probabilitas 0.6621 lebih tinggi dibandingkan mendeteksi status penerimaan calon mahasiswa baru dengan kriteria salah. Selanjutnya, hasil pemodelan akan digunakan untuk memprediksi pengelompokkan *mixture*. Hasil ketepatan klasifikasi dari model ditunjukkan pada Tabel 4.29.

**Tabel 4.29** Hasil Klasifikasi Pada Tiap *Mixture* Prodi Biologi

		Prediksi	
Aktual	Mixture 1	Mixture 2	
Mixture 1		46	14
Mixture 2		15	16

Hasil perbandingan antara *mixture* aktual dengan *mixture* hasil prediksi dari model BMM yang terbentuk, akan memperoleh ketepatan klasifikasi sebesar 68.13% dengan hasil klasifikasi ditunjukkan pada Tabel 4.29. Sedangkan rincian status penerimaan pada tiap *mixture* ditunjukkan pada Tabel 4.30 dan Tabel 4.31.

**Tabel 4.30** Rincian Prediksi Status Penerimaan Pada *Mixture 1* Prodi Biologi

Status	Nilai TKD Biologi	
	Tinggi	Rendah
Diterima	3	1
Tidak Diterima	13	43

Hasil Tabel 4.30 menunjukkan bahwa 46 dari 60 calon mahasiswa baru di prediksi sesuai dengan kriteria yang diinginkan, yaitu diterima dengan nilai tinggi ataupun tidak diterima dengan nilai rendah. Model juga memprediksi 14 calon mahasiswa baru yang di kategorikan karena tidak sesuai dengan kriteria.

**Tabel 4.31** Rincian Prediksi Status Penerimaan Pada *Mixture 2* Prodi Biologi

Status	Nilai TKD Biologi	
	Tinggi	Rendah
Diterima	2	2
Tidak Diterima	13	14

Untuk *mixture 2* merupakan kelompok penerimaan calon mahasiswa baru dengan kriteria salah, yang mana model memprediksi 16 calon mahasiswa baru yang di kategorikan ke dalam kriteria benar yaitu diterima dengan nilai tinggi sebanyak 2 calon dan tidak diterima dengan nilai rendah sebanyak 14 calon. Selain itu model juga memprediksi 15 calon mahasiswa baru yang salah di kategorikan kedalam kelompok salah. Secara umum untuk prodi FSAINS diketahui bahwa variabel yang berpengaruh di masing-masing prodi ditunjukkan pada Tabel 4.32.

Hasil pada Tabel 4.32 menunjukkan bahwa nilai TPA numerikal signifikan pada seluruh prodi. Selain itu mata uji bahasa indonesia dan bahasa inggris juga signifikan pada model. Nilai TPA figural tidak signifikan pada *mixture 1*, tetapi nilai TPA figural signifikan pada *mixture 2*, ini mengindikasikan bahwa TPA figural cenderung membuat calon mahasiswa baru untuk dikategorikan dengan kriteria salah.

**Tabel 4.32** Variabel yang Signifikan pada Tiap Mixture

Prodi	Variabel yang Signifikan	
	Mixture 1	Mixture 2
Fisika	TPA Numerikal	TPA Figural
	TPA Verbal	TPA Numerikal
	Bahasa Inggris	Bahasa Inggris
	Bahasa Indonesia	Bahasa Indonesia
	Matematika Dasar	
Kimia	TPA Numerikal	TPA Figural
	TPA Verbal	TPA Numerikal
	Bahasa Indonesia	TPA Verbal
	Matematika Dasar	Bahasa Inggris
		Bahasa Indonesia
Biologi	TPA Figural	TPA Figural
	TPA Numerikal	TPA Numerikal
	Bahasa Inggris	Bahasa Inggris

#### 4.2.6 Hasil Pemodelan dengan Metode BMM Pada Fakultas Matematika, Komputasi dan Sains Data (Aktuaria, Matematika, Statistika)

Berikut ini akan ditunjukkan hasil pemodelan dengan metode BMM untuk fakultas FMKSD.

##### A. Prodi Aktuaria

Dengan analogi yang sama diberikan pada proses estimasi pada prodi Fisika, hasil *diagnostic plot* yang ditunjukkan pada Lampiran 7 menunjukkan bahwa secara visual proses estimasi sudah mencapai kondisi yang konvergen. Pemeriksaan dengan metode *Raftery-Lewis* juga menunjukkan bahwa proses estimasi parameter telah konvergen ditunjukkan pada Tabel 4.33. Hasil metode *Raftery-Lewis* Pada Tabel 4.33 menunjukkan bahwa semua nilai  $I \leq 5$ , artinya tidak ada autokoreasi di tiap parameter. Dengan demikian dapat ditarik kesimpulan bahwa metode *Raftery-Lewis* membuktikan bahwa hasil estimasi parameter memenuhi syarat konvergen. Hasil estimasi pada prodi Aktuaria ditunjukkan pada Tabel 4.34. Hasil estimasi menginformasikan bahwa nilai *MC Error* mendekati nol dan lebih kecil dari nilai standar deviasi, sehingga dapat dikatakan bahwa hasil estimasi diestimasi dengan presisi yang tinggi dan memiliki nilai estimasi yang mendekati nilai sebenarnya.

**Tabel 4.33** Hasil Uji *Raftery-Lewis* Prodi Aktuaria

<b>Raftery Lewis : Dependence Factor (I)</b>			
	Chain 1	Chain 2	
<b>Mixture 1</b>	P[1]	1.01	0.999
	b0[1]	3.57	3.49
	b1[1]	1.01	0.996
	b2[1]	2.37	2.33
	b3[1]	2.13	2.08
	b4[1]	1.02	1.01
	b5[1]	2.14	2.03
	b6[1]	1.02	1.02
<b>Mixture 2</b>	P[2]	1	1.01
	b0[2]	4.76	4.81
	b1[2]	3.62	4.51
	b2[2]	3.5	3.51
	b3[2]	3.5	3.61
	b4[2]	1.08	1.11
	b5[2]	2.09	2.12
	b6[2]	2.03	1.16

Dari hasil uji signifikansi menunjukkan bahwa untuk *mixture* 1 hanya terdapat satu variabel yang berpengaruh untuk memodelkan status penerimaan dengan kriteria benar.

**Tabel 4.34** Hasil Estimasi Parameter Prodi Aktuaria

	<b>Par.</b>	<b>Estimasi</b>	<b>Standar Deviasi</b>	<b>2.50%</b>	<b>97.50%</b>	<b>MC Error</b>	<b>Sig.</b>
<b>Mix 1</b>	P[1]	0.7113	0.0174	0.6766	0.7447	2.87E-05	ya
	b0[1]	-39.26	5.6320	-50.42	-28.33	0.02209	ya
	TPA Figural	0.0462	0.0335	-0.0194	0.1120	5.75E-05	tidak
	TPA Numerikal	0.0066	0.0067	-0.0066	0.0198	2.27E-05	tidak
	TPA Verbal	0.0089	0.0062	-0.0031	0.0211	1.65E-05	tidak
	Bahasa Inggris	0.0003	0.0036	-0.0067	0.0073	6.14E-06	tidak
	Bahasa Indonesia	0.0085	0.0060	-0.0033	0.0204	1.56E-05	tidak
	Matematika Dasar	0.0120	0.0028	0.0065	0.0177	4.88E-06	ya
<b>Mix 2</b>	P[2]	0.2887	0.0174	0.2553	0.3234	2.87E-05	ya
	b0[2]	-41.25	2.4180	-45.99	-36.52	0.01169	ya
	TPA Figural	0.0173	0.0025	0.0123	0.0222	1.04E-05	ya
	TPA Numerikal	0.0087	0.0024	0.0040	0.0133	9.05E-06	ya
	TPA Verbal	0.0159	0.0025	0.0109	0.0209	9.15E-06	ya
	Bahasa Inggris	0.0045	0.0014	0.0018	0.0072	2.84E-06	ya
	Bahasa Indonesia	0.0060	0.0021	0.0019	0.0100	5.65E-06	ya
	Matematika Dasar	0.0108	0.0013	0.0083	0.0135	2.98E-06	ya

Variabel nilai tersebut yaitu nilai mata uji matematika dasar. Artinya untuk bisa diterima di prodi Aktuaria, seorang calon mahasiswa baru wajib memiliki nilai matematika dasar dan matematika IPA yang baik. Persamaan model yang terbentuk akan memenuhi persamaan berikut.

$$P_{mix}(\pi, x, \beta) = 0.7113 \frac{\exp(-39.26 + 0.0462x_1 + 0.0066x_2 + 0.0089x_3 + 0.0003x_4 + 0.0085x_5 + 0.0120x_6)}{1 + \exp(-39.26 + 0.0462x_1 + 0.0066x_2 + 0.0089x_3 + 0.0003x_4 + 0.0085x_5 + 0.0120x_6)} + 0.2887 \frac{\exp(-41.25 + 0.0173x_1 + 0.0087x_2 + 0.0159x_3 + 0.0045x_4 + 0.0060x_5 + 0.0108x_6)}{1 + \exp(-41.25 + 0.0173x_1 + 0.0087x_2 + 0.0159x_3 + 0.0045x_4 + 0.0060x_5 + 0.0108x_6)}$$

Hasil pemodelan menunjukkan bahwa model cenderung mengelompokkan data pada *mixture* 1 dengan probabilitas 0.7113 lebih besar dibandingkan *mixture* 2. Dengan variabel yang mempengaruhi penerimaan dengan kriteria benar adalah variabel nilai mata uji matematika dasar. Model dari hasil estimasi akan digunakan untuk memprediksi *mixture* yang akan dibentuk. Hasil ketepatan klasifikasi dari model ditunjukkan pada Tabel 4.35.

**Tabel 4.35** Hasil Klasifikasi Pada Tiap *Mixture* Prodi Aktuaria

Aktual	Prediksi	
	Mixture 1	Mixture 2
Mixture 1	116	5
Mixture 2	44	5

Hasil perbandingan antara *mixture* aktual dengan *mixture* hasil prediksi dari model BMM yang terbentuk, akan diperoleh ketepatan sebesar 71.18%. Hasil Tabel 4.35 menunjukkan bahwa model cenderung dikategorikan pada *mixture* 1 dibanding *mixture* 2. Hal ini menunjukkan bahwa hasil pemodelan cenderung dikelompokkan pada *mixture* 1. Rincian prediksi status penerimaan pada tiap *mixture* ditunjukkan pada Tabel 4.36 dan Tabel 4.37.

**Tabel 4.36** Rincian Prediksi Status Penerimaan Pada *Mixture* 1 Prodi Aktuaria

Status	Nilai TKD Matematika IPA	
	Tinggi	Rendah
	Diterima	0
Tidak Diterima	1	116

Hasil Tabel 4.36 menunjukkan bahwa *mixture* 1 memprediksi 116 calon mahasiswa baru ke dalam kategori tidak diterima dengan kriteria benar. Sedangkan 5 calon mahasiswa baru di kategorikan ke dalam kriteria salah.

**Tabel 4.37** Rincian Prediksi Status Penerimaan Pada *Mixture* 2 Prodi Aktuaria

Status	Nilai TKD Matematika IPA	
	Tinggi	Rendah
Diterima	44	5
Tidak Diterima	0	0

Hasil pada Tabel 4.37 merupakan kelompok *mixture* 2 yaitu kelompok status penerimaan dengan kriteria salah. Hasil prediksi pada *mixture* 2 menunjukkan bahwa model memprediksi 44 calon mahasiswa yang diterima, memiliki nilai tinggi.

#### B. Prodi Matematika

Menggunakan analogi yang sama sesuai urutan pada prodi Fisika, hasil pemeriksaan konvergensi dengan *diagnostic plot* dan uji *Raftery-Lewis* menunjukkan bahwa proses estimasi parameter telah memenuhi syarat konvergen.

**Tabel 4.38** Hasil Uji *Raftery-Lewis* Prodi Matematika

Raftery Lewis : Dependence Factor (I)			
		Chain 1	Chain 2
Mixture 1	P[1]	1.01	1.01
	b0[1]	2.25	2.3
	b1[1]	0.999	1
	b2[1]	2.32	2.25
	b3[1]	1.12	1.13
	b4[1]	1.05	1.06
	b5[1]	2.19	2.2
	b6[1]	1.06	1.06
Mixture 2	P[2]	1	1
	b0[2]	4.51	4.83
	b1[2]	3.52	3.46
	b2[2]	3.45	3.48
	b3[2]	3.53	3.51
	b4[2]	1.15	1.13
	b5[2]	2.09	2.1
	b6[2]	2.15	2.08

Hasil *diagnostic plot* yang ditunjukkan pada Lampiran 8, membuktikan bahwa sampel yang dihasilkan proses MCMC telah memenuhi syarat konvergen. Selain itu, hasil dari uji *Raftery–Lewis* yang ditunjukkan pada Tabel 4.38 telah menunjukkan hasil yang konvergen. Faktor dependen ( $I$ ) menunjukkan nilai  $I \leq 5$  yang mana telah memenuhi syarat konvergen dan sudah tidak terjadi autokorelasi pada tiap parameternya. Dengan demikian, dapat ditarik kesimpulan bahwa proses estimasi parameter telah mencapai kondisi yang konvergen dan hasil estimasi parameter dianggap dapat menggambarkan model yang terbentuk. Hasil dari proses estimasi parameter yang konvergen pada prodi Matematika ditunjukkan pada Tabel 4.39.

**Tabel 4.39** Hasil Estimasi Parameter untuk Prodi Matematika

	Parameter	Estimasi	Standar Deviasi	2.50%	97.50%	Signifikansi
<b>Mix 1</b>	P[1]	0.6527	0.0243	0.6044	0.6996	ya
	b0[1]	-16.32	2.0100	-20.28	-12.41	ya
	b1[1]	0.0101	0.0079	-0.0053	0.0256	tidak
	b2[1]	0.0076	0.0026	0.0026	0.0127	ya
	b3[1]	-0.0056	0.0023	-0.0101	-0.0011	ya
	b4[1]	-0.0035	0.0019	-0.0072	0.0003	tidak
	b5[1]	0.0110	0.0025	0.0062	0.0159	ya
	b6[1]	0.0082	0.0014	0.0056	0.0109	ya
<b>Mix 2</b>	P[2]	0.3473	0.0243	0.3004	0.3956	ya
	b0[2]	-33.00	1.8560	-36.64	-29.37	ya
	b1[2]	0.0024	0.0016	-0.0007	0.0054	tidak
	b2[2]	0.0096	0.0020	0.0057	0.0135	ya
	b3[2]	0.0203	0.0022	0.0160	0.0246	ya
	b4[2]	0.0049	0.0013	0.0024	0.0075	ya
	b5[2]	0.0052	0.0014	0.0025	0.0080	ya
	b6[2]	0.0106	0.0013	0.0081	0.0131	ya

Hasil estimasi menginformasikan bahwa nilai *MC Error* mendekati nol dan lebih kecil dari nilai standar deviasi, sehingga dapat dikatakan bahwa hasil estimasi diestimasi dengan presisi yang tinggi dan memiliki nilai estimasi yang mendekati nilai sebenarnya. Hasil uji signifikansi menunjukkan bahwa untuk *mixture 1* terdapat empat variabel yang signifikan untuk memodelkan status penerimaan dengan kriteria benar. Variabel nilai tersebut yaitu nilai mata uji TPA

numerik, TPA verbal, bahasa inggris dan matematika dasar. Artinya untuk bisa diterima di prodi Matematika, seorang calon mahasiswa baru wajib memiliki nilai kompetensi matematika IPA yang baik. Selain nilai matematika IPA, calon mahasiswa juga wajib memiliki nilai TPA numerik, TPA verbal, bahasa inggris dan matematika dasar yang baik. Persamaan model yang terbentuk akan memenuhi persamaan berikut.

$$P_{mix}(\pi, x, \beta) = 0.6527 \frac{\exp(-16.32 + 0.0101x_1 + 0.0076x_2 + -0.0056x_3 + -0.0035x_4 + 0.0110x_5 + 0.0082x_6)}{1 + \exp(-16.32 + 0.0101x_1 + 0.0076x_2 + -0.0056x_3 + -0.0035x_4 + 0.0110x_5 + 0.0082x_6)} + 0.3473 \frac{\exp(-33.00 + 0.0024x_1 + 0.0096x_2 + 0.0203x_3 + 0.0049x_4 + 0.0052x_5 + 0.0106x_6)}{1 + \exp(-33.00 + 0.0024x_1 + 0.0096x_2 + 0.0203x_3 + 0.0049x_4 + 0.0052x_5 + 0.0106x_6)}$$

Hasil pemodelan menunjukkan bahwa model cenderung mengelompokkan data pada *mixture* 1 dengan probabilitas 0.6527 lebih besar dibandingkan *mixture* 2. Artinya, hasil pengelompokkan *mixture* yang dihasilkan akan cenderung berada pada kelompok *mixture* 1. Selanjutnya hasil pemodelan akan digunakan untuk memprediksi *mixture*. Hasil prediksi kemudian akan dibandingkan dengan data aktualnya sehingga akan didapat ketepatan klasifikasi yang ditunjukkan pada Tabel 4.40.

**Tabel 4.40** Hasil Klasifikasi Pada Tiap *Mixture* Prodi Matematika

	Prediksi	
Aktual	Mixture 1	Mixture 2
Mixture 1	55	8
Mixture 2	6	27

Hasil prediksi menunjukkan bahwa model cenderung memprediksi calon mahasiswa baru untuk masuk kategori *mixture* 1. Artinya kemungkinan untuk diterima dengan kriteria benar dan tidak diterima dengan kriteria benar lebih tinggi dibandingkan dengan penerimaan dengan kriteria salah. Ketepatan klasifikasi yang diberikan untuk prodi matematika adalah sebesar 63.54%. Rincian jumlah calon mahasiswa baru pada tiap *mixture* ditunjukkan pada Tabel 4.41 dan Tabel 4.42.

**Tabel 4.41** Rincian Prediksi Status Penerimaan Pada *Mixture 1* Prodi Matematika

Status	Nilai TKD Matematika IPA	
	Tinggi	Rendah
Diterima	0	6
Tidak Diterima	2	55

Tabel 4.41 merupakan hasil prediksi untuk status penerimaan pada *mixture 1*, hasil menunjukkan bahwa model mengkategorikan secara benar 55 calon mahasiswa baru untu tidak diterima secara rendah. Sedangkan 8 calon mahasiswa lainnya salah dikategorikan.

**Tabel 4.42** Rincian Prediksi Status Penerimaan Pada *Mixture 2* Prodi Matematika

Status	Nilai TKD Matematika IPA	
	Tinggi	Rendah
Diterima	6	2
Tidak Diterima	25	0

Tabel 4.42 merupakan hasil prediksi untuk penerimaan pada *mixture 2*, hasil menunjukkan bahwa, model menunjukkan 27 calon mahasiswa baru dikategorikan ke dalam kategori yang salah sedangkan 6 calon mahasiswa baru lainnya diterima dengan kriteria benar.

### C. Prodi Statistika

Menggunakan analogi yang sama diberikan pada proses estimasi pada prodi Fisika, hasil *diagnostic plot* yang ditunjukkan pada Lampiran 9 telah memenuhi syarat konvergen. Hasil uji *Raftery-Lewis* yang ditunjukkan pada Tabel 4.43 menunjukkan sudah tidak terjadi autokorelasi di tiap parameter yang ditunjukkan dengan nilai  $I \leq 5$ , sehingga dapat ditarik kesimpulan bahwa metode *Raftery-Lewis* menunjukkan hasil estimasi parameter memenuhi syarat konvergen. Dari hasil *diagnostic plot* dan metode *Raftery-Lewis* yang telah dilakukan, dapat ditarik kesimpulan bahwa proses estimasi sudah memenuhi syarat konvergen. Dengan demikian, hasil estimasi parameter pada proses yang konvergen ini sudah dapat digunakan untuk memodelkan status penerimaan calon mahasiswa baru pada prodi Statistika.

**Tabel 4.43** Pemeriksaan Konvergensi dengan *Raftery-Lewis* Prodi Statistika

<b>Raftery Lewis : Dependence Factor (I)</b>			
	Chain 1	Chain 2	
<b>Mixture 1</b>	P[1]	1.01	1.01
	b0[1]	1	1.02
	b1[1]	3.22	3.36
	b2[1]	3.37	3.28
	b3[1]	1	0.996
	b4[1]	3.48	3.68
	b5[1]	2.29	2.3
	b6[1]	3.26	3.22
<b>Mixture 2</b>	P[2]	2.21	2.27
	b0[2]	2.11	2.15
	b1[2]	1.07	1.07
	b2[2]	1.04	1.03
	b3[2]	2.17	2.17
	b4[2]	1.14	1.12
	b5[2]	1.02	1.02
	b6[2]	1.05	1.04

Hasil estimasi parameter dengan kondisi konvergen ditunjukkan pada Tabel 4.44.

**Tabel 4.44** Hasil Estimasi Parameter untuk Prodi Statistika

	<b>Par.</b>	<b>Estimasi</b>	<b>Standar Deviasi</b>	<b>2.50%</b>	<b>97.50%</b>	<b>MC Error</b>	<b>Sig.</b>
<b>Mix 1</b>	P[1]	0.7452	0.0213	0.7024	0.7857	3.61E-05	Ya
	b0[1]	-48.75	2.9190	-54.50	-43.05	0.01017	Ya
	TPA Figural	0.0333	0.0137	0.0065	0.0601	2.27E-05	Ya
	TPA Numerikal	0.0038	0.0032	-0.0024	0.0100	8.79E-06	tidak
	TPA Verbal	0.0199	0.0037	0.0128	0.0271	1.08E-05	Ya
	Bahasa Inggris	0.0114	0.0020	0.0074	0.0154	3.96E-06	Ya
	Bahasa Indonesia	0.0143	0.0031	0.0083	0.0204	8.13E-06	Ya
	Matematika Dasar	0.0005	0.0017	-0.0029	0.0038	3.03E-06	tidak
<b>Mix 2</b>	P[2]	0.2548	0.0213	0.2143	0.2976	3.61E-05	Ya
	b0[2]	-48.36	1.8930	-52.08	-44.66	0.006655	Ya
	TPA Figural	0.0231	0.0021	0.0189	0.0273	8.24E-06	Ya
	TPA Numerikal	0.0151	0.0022	0.0108	0.0194	7.36E-06	Ya
	TPA Verbal	0.0134	0.0019	0.0097	0.0170	4.68E-06	Ya
	Bahasa Inggris	0.0089	0.0012	0.0065	0.0113	2.32E-06	Ya
	Bahasa Indonesia	0.0098	0.0016	0.0067	0.0130	3.51E-06	Ya
	Matematika Dasar	0.0069	0.0011	0.0046	0.0091	2.22E-06	Ya

Hasil estimasi pada prodi Statistika menunjukkan bahwa nilai MC Error mendekati nol dan bernilai lebih kecil dari standar deviasi, sehingga dapat dikatakan bahwa hasil estimasi mendekati nilai sebenarnya. Dari hasil uji signifikansi menunjukkan bahwa untuk *mixture* 1 terdapat empat variabel yang berpengaruh untuk memodelkan status penerimaan dengan kriteria benar. Keempat nilai tersebut yaitu nilai mata uji TPA figural, TPA numerik, TPA verbal, bahasa indonesia. Pada prodi Statistika, nilai matematika dasar tidak menjadi variabel yang berpengaruh pada model. Sehingga dapat diartikan bahwa selain harus mendapat kompetensi yang baik untuk mata uji matematika IPA, seorang calon mahasiswa baru juga harus memiliki kompetensi yang baik paa mata uji TPA figural, TPA numerik, TPA verbal, bahasa indonesia. Model BMM yang terbentuk akan memenuhi persamaan berikut.

$$P_{mix}(\pi, x, \beta) = 0.7452 \frac{\exp(-48.75 + 0.0333 x_1 + 0.0038 x_2 + 0.0199 x_3 + 0.0114 x_4 + 0.0143 x_5 + 0.0005 x_6)}{1 + \exp(-48.75 + 0.0333 x_1 + 0.0038 x_2 + 0.0199 x_3 + 0.0114 x_4 + 0.0143 x_5 + 0.0005 x_6)} + 0.2548 \frac{\exp(-48.36 + 0.0231 x_1 + 0.0151 x_2 + 0.0134 x_3 + 0.0089 x_4 + 0.0098 x_5 + 0.0069 x_6)}{1 + \exp(-48.36 + 0.0231 x_1 + 0.0151 x_2 + 0.0134 x_3 + 0.0089 x_4 + 0.0098 x_5 + 0.0069 x_6)}$$

Hasil prediksi menunjukkan bahwa model cenderung memprediksi calon mahasiswa baru untuk masuk kategori *mixture* 1 dengan probabilitas 0.7452. Artinya kemungkinan untuk diterima dengan kriteria benar dan tidak diterima dengan kriteria benar lebih tinggi dibandingkan dengan penerimaan dengan kriteria salah. Selanjutnya hasil pemodelan akan digunakan untuk memprediksi *mixture* dengan hasil ketepatan klasifikasi dapat dilihat pada Tabel 4.45.

**Tabel 4.45** Hasil Klasifikasi Pada Tiap *Mixture* Prodi Statistika

<b>Prediksi</b>		
<b>Aktual</b>	<b>Mixture 1</b>	<b>Mixture 2</b>
<b>Mixture 1</b>	30	49
<b>Mixture 2</b>	4	23

Hasil perbandingan antara *mixture* aktual dengan *mixture* hasil prediksi dari model BMM yang terbentuk, akan memperoleh ketepatan klasifikasi sebesar 50% dengan hasil klasifikasi ditunjukkan pada Tabel 4.45. Hasil ketepatan

klasifikasi yang rendah ini, karena banyak terjadi misklasifikasi pada data. Model memprediksi 49 calon mahasiswa baru untuk masuk pada *mixture* 2 dari aktual *mixture* 1. Rincian prediksi status penerimaan tiap *mixture* untuk prodi Statistika ditunjukkan pada Tabel 4.46 dan Tabel 4.47.

**Tabel 4.46** Rincian Prediksi Status Penerimaan Pada *Mixture* 1 Prodi Statistika

Status	Nilai TKD Matematika IPA	
	Tinggi	Rendah
Diterima	2	49
Tidak Diterima	0	28

Tabel 4.46 menunjukkan bahwa model memprediksi 49 calon mahasiswa baru untuk diterima tetapi nilai rendah. Sedangkan untuk Tabel 4.47 menunjukkan bahwa ada 23 orang yang berada pada kategori *mixture* 2 yaitu 18 orang diprediksi tidak diterima tetapi memiliki nilai tinggi dan 5 orang diterima tetapi memiliki nilai rendah.

**Tabel 4.47** Rincian Prediksi Status Penerimaan Pada *Mixture* 2 Prodi Statistika

Status	Nilai TKD Matematika IPA	
	Tinggi	Rendah
Diterima	2	5
Tidak Diterima	18	2

Secara umum dapat diperoleh kesimpulan untuk FMKSD adalah hasil pemodelan menunjukkan bahwa meskipun prodi Statistika, Matematika dan Aktuaria berada pada satu rumpun yang sama tetapi model yang dihasilkan pada tiap prodi tidak menunjukkan signifikansi parameter yang sama. Variabel yang signifikan pada masing-masing prodi ditunjukkan pada Tabel 4.48. Untuk prodi Aktuaria dan Matematika, variabel nilai matematika dasar berpengaruh pada penerimaan dengan kondisi benar, sedangkan untuk prodi Statistika, nilai matematika dasar tidak signifikan. Untuk prodi Statistika, variabel yang mempengaruhi adalah nilai TPA figural, TPA verbal, bahasa Inggris dan bahasa Indonesia. Adanya perbedaan ini dipengaruhi oleh beberapa faktor yaitu jumlah

pendaftar, keketatan dan kualitas dan kompetensi individual dari pendaftar. Selain itu, untuk prodi FMKSD, model cenderung memberikan ketepatan klasifikasi yang rendah. Ketepatan klasifikasi yang rendah dapat dipengaruhi oleh beberapa faktor, yang pertama yaitu data *training* memiliki data yang *imbalance* dan kurang dari 10. Yang kedua *mixture* yang digunakan dalam penelitian, penentuan *mixture* baik untuk FMKSD berdasar pada pertimbangan peneliti sehingga penentuan *mixture* pada penelitian selanjutnya perlu untuk di lakukan secara statistik.

**Tabel 4.48** Variabel yang Signifikan pada Tiap Prodi di FMKSD

Prodi	Variabel yang Signifikan	
	Mixture 1	Mixture 2
Aktuaria	Matematika Dasar	TPA Figural
		TPA Numerikal
		TPA Verbal
		Bahasa Inggris
		Bahasa Indonesia
Matematika		Matematika Dasar
	TPA Numerikal	TPA Numerikal
	TPA Verbal	TPA Verbal
	Bahasa Indonesia	Bahasa Inggris
	Matematika Dasar	Bahasa Indonesia
Statistika		Matematika Dasar
	TPA Figural	TPA Figural
	TPA Verbal	TPA Numerikal
	Bahasa Inggris	TPA Verbal
	Bahasa Indonesia	Bahasa Inggris
	Bahasa Indonesia	
	Matematika Dasar	

*(Halaman ini sengaja dikosongkan)*

## BAB 5 KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Dari hasil diskusi pada bab analisis dan pembahasan, akan diperoleh kesimpulan sebagai berikut:

1. Dari hasil pembentukan algoritma dapat diketahui bahwa perbedaan mendasar metode *Bernoulli Mixture Model* dengan regresi dengan respon biner lainnya seperti ; regresi logistik *polythomus*, regresi logistik biner adalah pada pembagian kriteria. Jika regresi *polythomus* membagi kedalam empat kriteria, BMM hanya membagi kedalam dua kriteria, yang mana pada tiap kriterianya dapat diuraikan kembali sesuai dengan *mixture* yang ditentukan. Penggunaan *prior*, *thin* dan iterasi yang berbeda mempengaruhi hasil estimasi parameter. *Prior* yang digunakan secara umum adalah nilai *means* dan *varians*. Pada penelitian ini diketahui bahwa semakin besar *thin* dan iterasi yang digunakan, maka proses estimasi parameter akan cepat konvergen. Selain itu dari hasil perbandingan antara pemeriksaan konvergensi melalui *diagnostic plot* dan secara inferensia, dapat diketahui bahwa ada kesulitan tersendiri dalam memeriksa konvergensi secara visual. Kesulitan ini dikarenakan, penentuan secara visual berdasar pada kemampuan dan pengalaman peneliti dalam memeriksa konvergensi. Sementara itu, pemeriksaan inferensia dirasa lebih mudah karena lebih terukur dengan adanya batasan-batasan kriteria yang jelas.
2. Karakteristik dari data status penerimaan calon mahasiswa baru ditunjukkan pada poin berikut.
  - A. Pada tiap prodi masih banyak variabel yang memiliki *outlier* yang berada diatas nilai median, artinya masih banyak calon mahasiswa baru yang memiliki kompetensi tinggi tetapi tidak diterima.
  - B. Variabel respon dari data merupakan variabel respon yang *imbalance* dengan perbandingan calon mahasiswa baru yang diterima hanya 6.67%.
  - C. Variabel yang signifikan pada penelitian berbeda-beda pada tiap prodi. Hal ini dikarenakan berbagai macam faktor, yang pertama yaitu kompetensi

individual dari pendaftar, jumlah pendaftar, kuota penerimaan dan keketatan di tiap prodi.

- D. Untuk Fakultas Sains, prodi yang memberikan hasil ketepatan klasifikasi tertinggi adalah prodi Kimia yaitu sebesar 73.91%, prodi Fisika 73.02% dan prodi Biologi 68.13%. Sedangkan untuk Fakultas Matematika, Komputasi dan Sains Data, prodi yang memberikan hasil ketepatan klasifikasi tertinggi adalah prodi Aktuaria yaitu sebesar 71.18%. kemudian prodi Matematika sebesar 63.54% dan prodi Statistika yaitu sebesar 50%. Model BMM untuk prodi Statistika ditunjukkan pada persamaan berikut. Ketepatan klasifikasi menunjukkan bahwa semakin besar persentase ketepatan klasifikasi maka semakin besar kemampuan model untuk memprediksi penerimaan calon mahasiswa baru dengan menggunakan *mixture* yang ditentukan. Sebaliknya, semakin rendah persentase ketepatan klasifikasi maka dapat diartikan bahwa *mixture* yang digunakan tidak sesuai dengan keadaan di lapangan dan perlu untuk di evaluasi.

## 5.2 Saran

Berdasarkan hasil dan pembahasan pemodelan status penerimaan calon mahasiswa baru, saran yang dapat diberikan adalah untuk meningkatkan akurasi klasifikasi yaitu menangani kasus *imbalance* pada variabel respon di penelitian ini. Peneliti telah mencoba untuk menangani kasus *imbalance* dengan menggunakan SMOTE, tetapi dikarenakan penambahan jumlah sampel yang banyak menyebabkan proses estimasi berjalan sangat lama yaitu  $\pm 23$  jam, hasil ketepatan klasifikasi yang diberikan menunjukkan bahwa hasil tidak memberikan hasil ketepatan klasifikasi yang signifikan. Hal ini bisa diakibatkan oleh beberapa faktor, yang pertama yaitu penentuan *prior* yang belum tepat atau jumlah *thin* dan iterasi perlu ditambah. Untuk penelitian selanjutnya, peneliti dapat menangani kasus *imbalance* ataupun mengubah *prior* serta melakukan penambahan *thin* dan iterasi agar mendapat hasil ketepatan klasifikasi yang lebih tinggi. Yang kedua yaitu mengevaluasi penentuan *mixture*, pada penelitian ini khususnya untuk prodi pada FMKSD, ada kecenderungan bahwa nilai matematika IPA tidak berpengaruh signifikan untuk dijadikan *mixture*. Dengan demikian, untuk penelitian

selanjutnya, selain didasarkan pada pertimbangan peneliti, penentuan *mixture* juga perlu didasarkan pada cara-cara statistik seperti *feature selection* dan *feature extraction*.

*(Halaman ini sengaja dikosongkan)*

## DAFTAR PUSTAKA

- Bain, L., & Engelhardt, M. (1992). *Introduction to Probability and Mathematical Statistics, Second Edition*. California: Cengage Learning, Inc.
- Bishop, C. (2006). *Pattern Recognition and Machine Learning, Information Science and Statistics*. Springer.
- Boldstad, W. (2007). *Introduction to Bayesian Statistics. 2nd Edition*. Wiley: New Jersey.
- Box, G., & Tiao, G. (1973). *Bayesian Inference in Statistical Analysis*. John Willey And Sons, Inc: New York.
- Brooks, S., & Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7, 434-455.
- Carlin, B., & Chib, S. (1995). Bayesian Model Choice Via Markov Chain Monte Carlo Methods. *Journal of The Royal Statistical Society, Series B (Methodological)*, 473-484.
- Casella, G., & George, I. (1992). Explaining The Gibbs Sampler. *Journal of The American Statistical Association. Vol.46 No.3*, Hal.167-174.
- Celeux, G., & Govaert, G. (1991). Clustering criteria for discrete data and latent class models. *Journal of Classification*, 157-176.
- Cowles, M. K., & Carlin, B. P. (1996). Markov Chain Monte Carlo Convergence Diagnostics:. *Journal of the American Statistical Association*, 883-904.
- Dapiah. (2018). *Pemodelan Rata-rata Nilai Ujian di SMAN Jawa Barat dengan menggunakan Pendekatan Bayesian Mixture Normal Hirarki*. Surabaya: Pascasarjana Thesis.
- Efendiev, Y., Datta-Gupta, A., & Ginting, V. (2013). An efficient two-stage Markov chain Monte Carlo method for dynamic data integration. 49.
- Gelman, A., Charlin, J., & Stern, H. (1995). *Bayesian Data Analysis*. London: Chapman and Hall.
- Geman, S., & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Explore*.
- Gilks, W. (1996). Full Conditional Distribution. In W. Gilks, *Markov Chain Monte Carlo in Practice* (pp. 75-88). UK: Springer.

- Gonzales, J., Juan, J., Dupont, P., Vidal, E., & Casacuberta, F. (2001). A Bernoulli Mixture Model For Word Categorization. *Proceedings of the IX Spanish Symposium on Pattern Recognition and Image Analysis*. Spain.
- Govaert, G., & Nadif, M. (2008). Block Clustering With Bernoulli Mixture Models : Comparison of Different Approaches. *Computational Statistics and Data Analysis*, 3233-3245.
- Grantham, N. (2015). Clustering Binary Data with Bernoulli Mixture Models. *Semantic Scholars*.
- Grim, J., Somol, P., & Pudil, P. (2000). Multivariate structural Bernoulli mixtures for Recognition. *15th International Conference on*, vol. 2, 585-589.
- Han, J., & Kamber, M. (2006). *Data Mining : Concept and Techniques Second Edition*. Waltham, USA: Morgan Kaufmann Publishers.
- Hollmen, J., & Tikka, J. (2007). Compact and Understandable Descriptions of Mixtures of Bernoulli Distributions. (221460769).
- Iriawan, N. (2001). Penaksiran Model Mixture Normal Univariabel : Suatu Pendekatan Bayesian dengan MCMC. *Prosiding Seminar Nasional dan Konferda VII Matematika Wilayah DIY dan Jawa Tengah, Yogyakarta*, 105-110.
- Iriawan, N. (2012). *Pemodelan dan Analisis Data*. Surabaya: ITS press.
- Iriawan, N., SSU, B., Suryaningtyas, W., Susanto, I., & A. Pravitasari, A. (2018). Bayesian Bernoulli Mixture Regression Model For Bidikmisi Scholarship Classification. *Journal of Science and Information*, 67-76.
- Juan, A., & Vidal, E. (2002). On the use of Bernoulli mixture models for text classification. *Pattern Recognition*.
- Juan, A., Vidal, E., Casacuberta, F., & Casacuberta, F. (2001). A Bernoulli Mixture Model for Word Categorization. *Proceedings of the IX Spanish Symposium on*. Spain.
- Koop, G. (2003). *Bayesian Econometrics*. Chicester. England: John Wiley & Sons.
- L., A., J.L, G., & Mallet, A. (1993). An autologistic model for the genetic analysis of familial binary data. *American journal of human genetics* , 894.

- Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2013). *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Boca Raton: Taylor & Francis Group, LLC.
- Marchand, E., & Strawderman, W. (2012). On Bayesian credible sets in restricted parameter space problems and lower bounds for frequentist coverage. *arxiv*.
- McLachlan, G., & Peel, S. N. (2003). On Clustering by Mixture Models. *Exploratory Data Analysis in Empirical Research*, 141-148.
- McCullagh, P., & Nelder, J. (1989). *Generalized Linear Models, Second Edition*. Boca Raton: Chapman and Hall/CRC.
- McLachlan, G., & Basford, K. (1988). *Mixture Model : Inference and Application To Clustering*. Marcel and Decker Inc.
- McLachlan, G., & Peel, D. (2000). *Finite Mixture Models*. New York: John Wiley and Sons, Inc.
- McLachlan, G., Peel, D., & Prado, P. (1998). Clustering Via Normal Mixture Models. *Research Gate*.
- Miftahurrohmah, B., Iriawan, N., & Fithriasari, K. (2017). On The Value at Risk Using Bayesian Mixture Laplace Autoregressive Approach for Modelling the Islamic Stock Autoregressive Approach for Modelling the Islamic Stock Risk Investment. *Journal of Physics: Conference Series*.
- Nadif, M., & Govaert, G. (1998). Clustering For Binary Data And Mixture Models - Choice Of The Model. *Applied Stochastic Models and Data Analysis*, 269-278.
- Ntzoufras, I. (2009). *Bayesian Modeling Using WinBugs*. John Wiley and Sons, Inc.: New York.
- Pratola, M. T. (2017, November 19). *matthewpratola.com*. Retrieved from Matthew T. Pratola : Assistant Professor, Dept. Of Statistics, The Ohio State University: <http://www.matthewpratola.com/wp-content/uploads/2017/11/stat8810-slides15.pdf>
- Ristekdikti. (2016, April 8). *Kabar : Ristekdikti*. Retrieved from ristekdikti.go.id: <https://ristekdikti.go.id/kabar/kompetensi-ciri-pendidikan-tinggi-berkualitas/>

- Saeed, M., Javed, K., & Babri, H. A. (2013). Machine learning using Bernoulli mixture models : Clustering, Rule, Extraction and Dimensionality Reduction. *Neurocomputing*119, 366-374.
- Salim, A. (2016). *Pengoptimalan Naïve Bayes Dan Regresi Logistik Menggunakan Algoritma Genetika Untuk Data Klasifikasi (Studi Kasus : Pembuangan Limbah Domestik di Surabaya Timur)*. Surabaya.
- Satriya, A. M. (2016). *Small Area Estimation Terhadap Pengeluaran Per Kapita Di Kabupaten Bangkalan Dengan Metode Hierarchical Bayes*. Tesis, Institut Teknologi Sepuluh Nopember: Surabaya.
- Souza, D. F., & Moura, F. A. (2016). Multivariate Beta Regression with Application in Small Area Estimation. *Journal of Official Statistics*, v.32, Hal. 747-768.
- Spiegelhalter, D., Best, N., Carlin, B., & van der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of The Royal Statistical Society. B* 64, 583-639.
- Spiegelhalter, D., Best, N., Charlin, B., & Vanderlinde, A. (2002). Bayesian Measures of Model Complexity and Fit (with Discussion). *Journal Of The Royal Statistiactal Society*, 583-639.
- Stephens, M. (1997). *Bayesian Methods for Mixture of Normal Distribution*. England: University of Oxford.
- Stuart, G. K., Minkoff, S. E., & Pereira, F. (2019). A Two-Stage Markov Chain Monte Carlo Method for Seismic Inversion and Uncertainty Quantification. 84(6).
- Suryaningtyas , W., Iriawan, N., Ulama, B., Fithriasari, K., Susanto, I., & Pravitasari, A. (2018). On The Bernoulli Mixture Model for Bidikmisi Scholarship. *Journal of Physics: Conference Series*.
- Susanto, I., Iriawan, N., Kuswanto , H., Suhartono, Fithriasari, K., Ulama, B., et al. (2018). On The Markov Chain Monte Carlo Convergence Diagnostic of Bayesian Finite Mixture Model Of Income Distribution. *Journal of Physics: Conference Series*.
- Thomas, A., & Thorne, G. (2009, February 1). *The Center For Development and Training*. Retrieved from CDL: <https://www.cdl.org/articles/higher-order-thinking-its-hot/>

Uspensky, J. V. (1937). *Introduction to Mathematical Probability*. New York: Mc.GrawHill.

Wang X, & Kaban A. (2005). Finding Uninformative Features in Binary and Automated Learning. *Intelligent Data Engineering*, 40-47.

Ye, M., Zhang, P., & Nie, L. (2018). Clustering sparse binary data with hierarchical Bayesian Bernoulli mixture model. *Computational Statistics and Data Analysis*, 32-49.

*(Halaman ini sengaja dikosongkan)*

## LAMPIRAN

### Lampiran 1 *R-Script* Untuk Metode GLM

```
#Set Lokasi Data
setwd("F:/2511 THESIS/2511 - SuperFIX/0412-Data RUN/CSV ")

#baca data
datatrain<-read.csv("data train smote nested akt.csv",header=T)

View(datatrain)

#Pembagian Data
datatrain.Mix1 = subset(datatrain, datatrain$Mix==1)
datatrain.Mix2 = subset(datatrain, datatrain$Mix==2)

#Metode GLM Mixture 1
reglog1 <- glm(Status..Y.~TPA.Fig+TPA.Num +TPA.Ver+B.Ing
+B.Ind+Mat.Das,
              family = binomial(link='logit'), data=datatrain.Mix1)
summary(reglog1)

#Metode GLM Mixture 2
reglog2 <- glm(Status..Y.~TPA.Fig+TPA.Num +TPA.Ver+B.Ing
+B.Ind+Mat.Das,
              family = binomial(link='logit'), data=datatrain.Mix2)
summary(reglog2)
```

## Lampiran 2 R-Script Untuk Pembagian Data Training dan Testing

```
setwd("F:/2511 THESIS/2511 - SuperFIX/2612-HASIL FIX ALL/Data RUN")
data=read.csv("Biologi.csv",header=T)
dim(data)
names(data)
datamix_1<-subset(data, data$Mix==1); #View(datamix_1)
dim(datamix_1)
datamix_2<-subset(data, data$Mix==2); #View(datamix_2)
dim(datamix_2)

#Stratified Mix 1
n_mix1 = nrow(datamix_1)
n.train_mix1 = floor(n_mix1*0.8)
n.test_mix1 = n_mix1-n.train_mix1
target_mix1 = as.factor(datamix_1$Status..Y.)

#Perhitungan proporsi
ktgmix1_Status..Y. <- table(datamix_1$Status..Y.); ktmix1_Status..Y.
prop.ktmix1_Status..Y. <- ktmix1_Status..Y./length(datamix_1$Status..Y.);
prop.ktmix1_Status..Y.

#pengambilan data pada status tidak diterima + pembagian training testing
n.ktmix1_Status..Y.1 = ktmix1_Status..Y.[1]
prop.ktmix1_Status..Y.1 = prop.ktmix1_Status..Y.[1]
data.ktmix1_Status..Y.1 = subset(datamix_1, datamix_1$Status..Y.==0)
n.train.ktmix1_Status..Y.1 = round(n.train_mix1*prop.ktmix1_Status..Y.1)
n.test.ktmix1_Status..Y.1 = n.ktmix1_Status..Y.1-n.train.ktmix1_Status..Y.1

#pengambilan data pada status diterima + pembagian training testing
n.ktmix1_Status..Y.2 = ktmix1_Status..Y.[2]
prop.ktmix1_Status..Y.2 = prop.ktmix1_Status..Y.[2]
data.ktmix1_Status..Y.2 = subset(datamix_1, datamix_1$Status..Y.==1)
n.train.ktmix1_Status..Y.2 = round(n.train_mix1*prop.ktmix1_Status..Y.2)
n.test.ktmix1_Status..Y.2 = n.ktmix1_Status..Y.2-n.train.ktmix1_Status..Y.2

#penggabungan data
{partisi2 = rbind(cbind(n.ktmix1_Status..Y.1, n.ktmix1_Status..Y.2),
                cbind(n.train.ktmix1_Status..Y.1, n.train.ktmix1_Status..Y.2),
                cbind(n.test.ktmix1_Status..Y.1, n.test.ktmix1_Status..Y.2))
  colnames(partisi2)=c("Mixture 1","Mixture 2")
  rownames(partisi2)=c("Jumlah Data","Jumlah Data Training","Jumlah Data Testing")
  cbind(partisi2,rowSums(partisi2))}

index.ktmix1_Status..Y.1 = which(datamix_1$Status..Y.==0, arr.ind=TRUE)
train.index.ktmix1_Status..Y.1 = sample(index.ktmix1_Status..Y.1,
size=n.train.ktmix1_Status..Y.1)
index.ktmix1_Status..Y.2 = which(datamix_1$Status..Y.==1, arr.ind=TRUE)
train.index.ktmix1_Status..Y.2 = sample(index.ktmix1_Status..Y.2,
size=n.train.ktmix1_Status..Y.2)

train.index_mix1 = c(train.index.ktmix1_Status..Y.1, train.index.ktmix1_Status..Y.2)
data.train_mix1 = datamix_1[train.index_mix1,]
data.test_mix1 = datamix_1[-train.index_mix1,]
target.train_mix1 = target_mix1[train.index_mix1]
target.test_mix1 = target_mix1[-train.index_mix1]
```

```

#Stratified Mix 2
n_mix2 = nrow(datamix_2)
n.train_mix2 = floor(n_mix2*0.8)
n.test_mix2 = n_mix2-n.train_mix2
target_mix2 = as.factor(datamix_2$Status..Y.)

#Perhitungan proporsi
ktgmix2_Status..Y. <- table(datamix_2$Status..Y.); ktgmix2_Status..Y.
prop.ktgmix2_Status..Y. <- ktgmix2_Status..Y./length(datamix_2$Status..Y.);
prop.ktgmix2_Status..Y.

#pengambilan data pada status tidak diterima + pembagian training testing
n.ktgmix2_Status..Y.1 = ktgmix2_Status..Y.[1]
prop.ktgmix2_Status..Y.1 = prop.ktgmix2_Status..Y.[1]
data.ktgmix2_Status..Y.1 = subset(datamix_2, datamix_2$Status..Y.==0)
n.train.ktgmix2_Status..Y.1 = round(n.train_mix2*prop.ktgmix2_Status..Y.1)
n.test.ktgmix2_Status..Y.1 = n.ktgmix2_Status..Y.1-n.train.ktgmix2_Status..Y.1

#pengambilan data pada status diterima + pembagian training testing
n.ktgmix2_Status..Y.2 = ktgmix2_Status..Y.[2]
prop.ktgmix2_Status..Y.2 = prop.ktgmix2_Status..Y.[2]
data.ktgmix2_Status..Y.2 = subset(datamix_2, datamix_2$Status..Y.==1)
n.train.ktgmix2_Status..Y.2 = round(n.train_mix2*prop.ktgmix2_Status..Y.2)
n.test.ktgmix2_Status..Y.2 = n.ktgmix2_Status..Y.2-n.train.ktgmix2_Status..Y.2

#penggabungan data
{partisi2 = rbind(cbind(n.ktgmix2_Status..Y.1, n.ktgmix2_Status..Y.2),
                cbind(n.train.ktgmix2_Status..Y.1, n.train.ktgmix2_Status..Y.2),
                cbind(n.test.ktgmix2_Status..Y.1, n.test.ktgmix2_Status..Y.2))
  colnames(partisi2)=c("Mixture 1", "Mixture 2")
  rownames(partisi2)=c("Jumlah Data", "Jumlah Data Training", "Jumlah Data Testing")
  cbind(partisi2,rowSums(partisi2))}

index.ktgmix2_Status..Y.1 = which(datamix_2$Status..Y.==0, arr.ind=TRUE)
train.index.ktgmix2_Status..Y.1 = sample(index.ktgmix2_Status..Y.1,
size=n.train.ktgmix2_Status..Y.1)
index.ktgmix2_Status..Y.2 = which(datamix_2$Status..Y.==1, arr.ind=TRUE)
train.index.ktgmix2_Status..Y.2 = sample(index.ktgmix2_Status..Y.2,
size=n.train.ktgmix2_Status..Y.2)

train.index_mix2 = c(train.index.ktgmix2_Status..Y.1, train.index.ktgmix2_Status..Y.2)
data.train_mix2 = datamix_2[train.index_mix2,]
data.test_mix2 = datamix_2[-train.index_mix2,]
target.train_mix2 = target_mix2[train.index_mix2]
target.test_mix2 = target_mix2[-train.index_mix2]

#Export data
data.train<-rbind(data.train_mix1,data.train_mix2); dim(data.train)
data.test<-rbind(data.test_mix1,data.test_mix2); dim(data.test)

write.csv(data.train, "data train nested kim.csv")
write.csv(data.test, "data test nested kim.csv")

```

### Lampiran 3 Script Untuk Proses Estimasi Parameter dengan OpenBUGS

```

model;
{
  for( i in 1 : N ) {
    y[i] ~ dbern(phi[i])
  }
  for( i in 1 : N ) {
    phi[i] <- bern[T[i] , i]
  }
  for( i in 1 : N ) {
    T[i] ~ dcat(P[1:2])
  }
  P[1:2] ~ ddirch(alfa[])
  for( i in 1 : N ) {
    logit(bern[2 , i]) <- b0[2] + b1[2] * x1[i] + b2[2] * x2[i] + b3[2] * x3[i] + b4[2] *
x4[i] + b5[2] * x5[i] + b6[2] * x6[i]
  }
  for( i in 1 : N ) {
    logit(bern[1 , i]) <- b0[1] + b1[2] * x1[i] + b2[1] * x2[i] + b3[1] * x3[i] + b4[1] *
x4[i] + b5[1] * x5[i] + b6[1] * x6[i]
  }

  b0[1]~dnorm(-48.85,0.016923346)
  b1[1]~dnorm(0.03319,5320.164861)
  b2[1]~dnorm(0.003499,26971.68332)
  b3[1]~dnorm(0.02028,17688.04483)
  b4[1]~dnorm(0.01173,76818.69804)
  b5[1]~dnorm(0.01472,27713.07621)
  b6[1]~dnorm(0.0004588,126285.0132)

  b0[2]~dnorm(-47.62,0.038102308)
  b1[2]~dnorm(0.0223,48112.82033)
  b2[2]~dnorm(0.0151,54845.91041)
  b3[2]~dnorm(0.01322,86251.32342)
  b4[2]~dnorm(0.008735,225467.1454)
  b5[2]~dnorm(0.009784,124508.8747)
  b6[2]~dnorm(0.00674,262446.251)

}

inits
list(
  b0=c(-48.85,-47.62),
  b1=c(0.03319,0.0223),
  b2=c(0.003499,0.0151),
  b3=c(0.02028,0.01322),
  b4=c(0.01173,0.008735),
  b5=c(0.01472,0.009784),
  b6=c(0.0004588,0.00674))

```

Bagian ini merupakan Prior untuk analisis Bayesian, prior dapat diubah-ubah berdasar pada ketentuan dari peneliti.

Bagian ini adalah inisialisasi pada algoritma *Gibbs Sampling*.

**DATA**

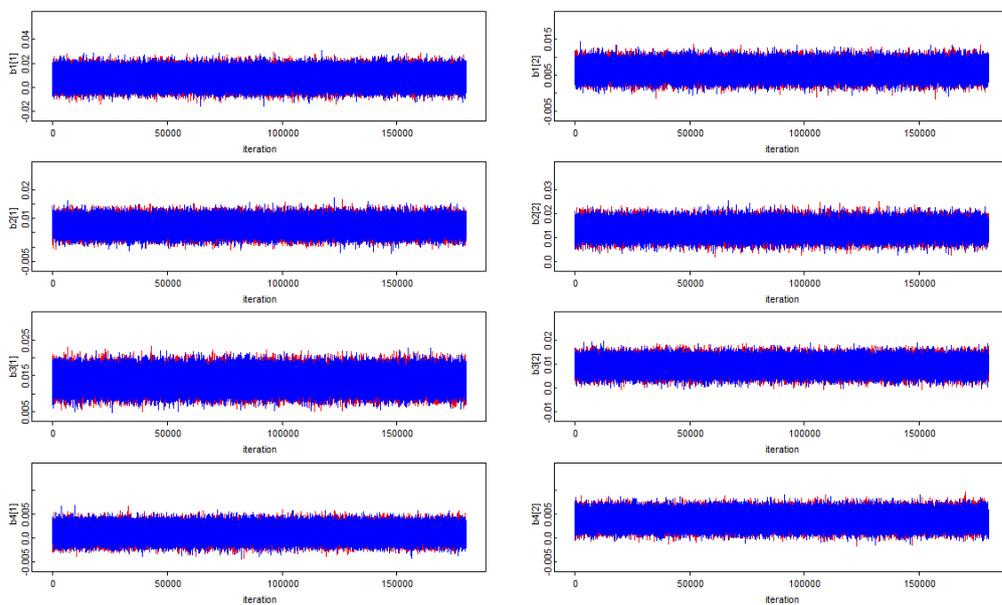
```
list(  
  alfa=c(1,1),  
  N=418,
```

```
  x1 = c ( 456.876, 564.789,.....,445.765, 678.987),  
  x2 = c ( 456.876, 564.789,.....,445.765, 678.987),  
  x3 = c ( 456.876, 564.789,.....,445.765, 678.987),  
  x4 = c ( 456.876, 564.789,.....,445.765, 678.987),  
  x5 = c ( 456.876, 564.789,.....,445.765, 678.987),  
  x6 = c ( 456.876, 564.789,.....,445.765, 678.987))
```

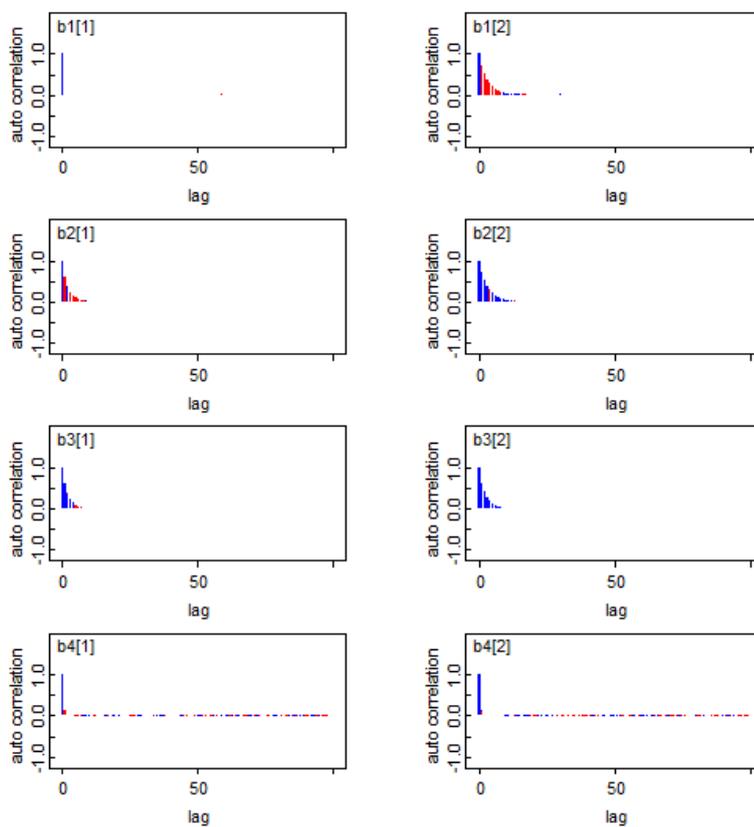
Bagian ini merupakan list dari data. Jumlah N merupakan banyak data yang dianalisis.

## Lampiran 4 Bukti Proses Estimasi Parameter Telah Konvergen Untuk Prodi Kimia

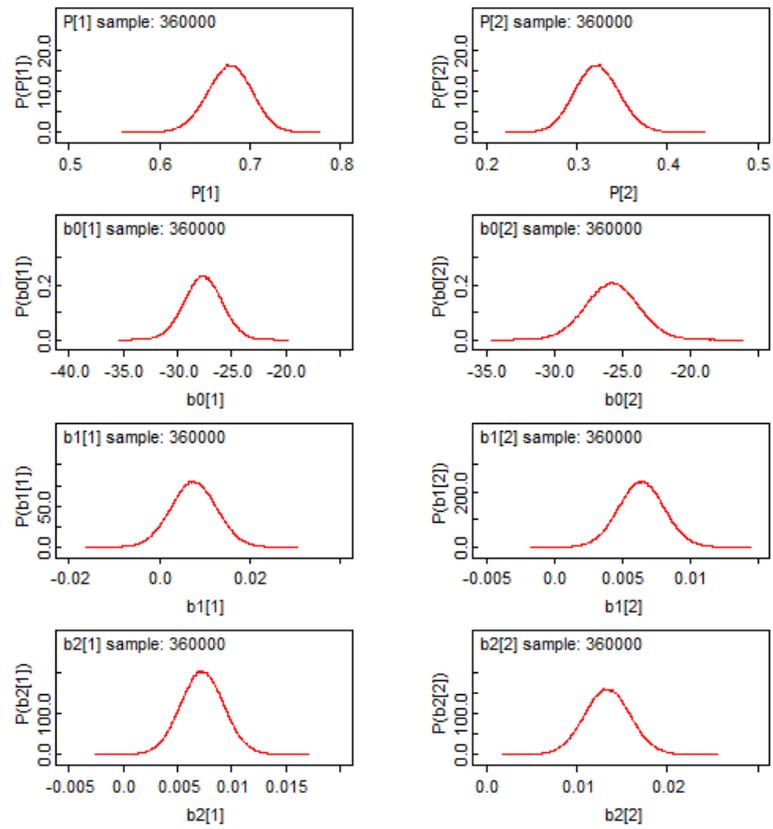
### A. Serial Plot



### B. Autocorrelation Plot

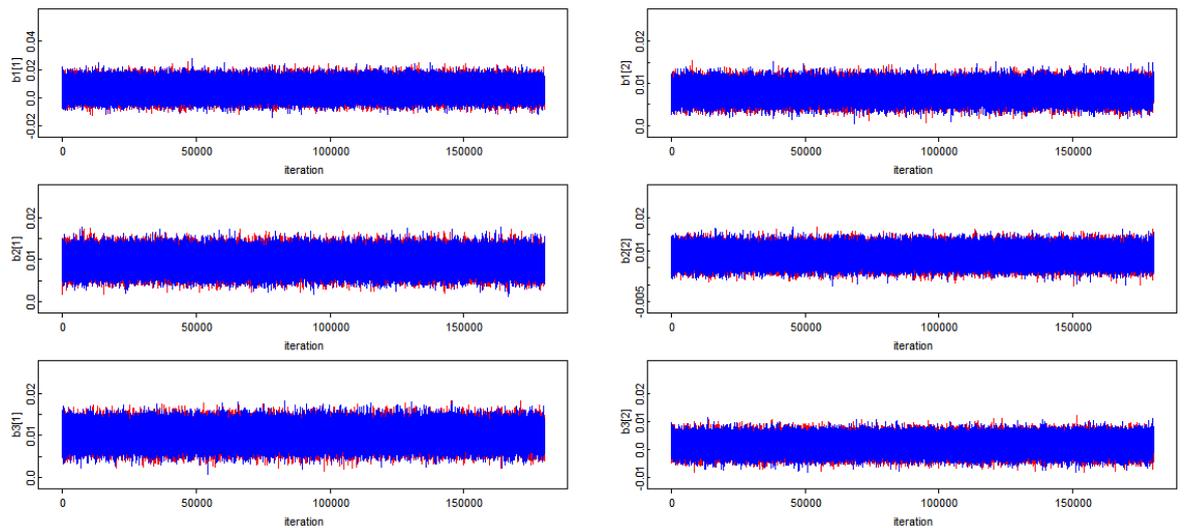


### C. Density plot

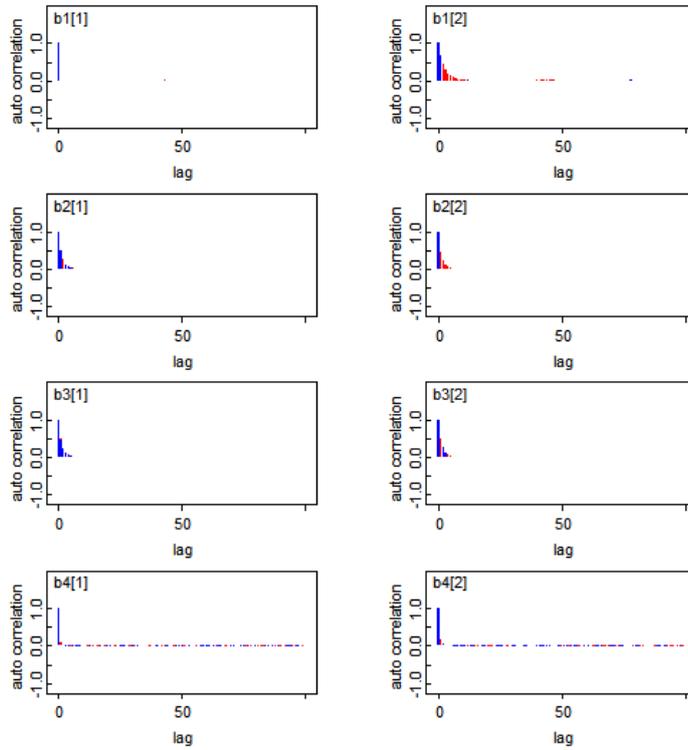


## Lampiran 5 Bukti Proses Estimasi Parameter Telah Konvergen Untuk Prodi Biologi

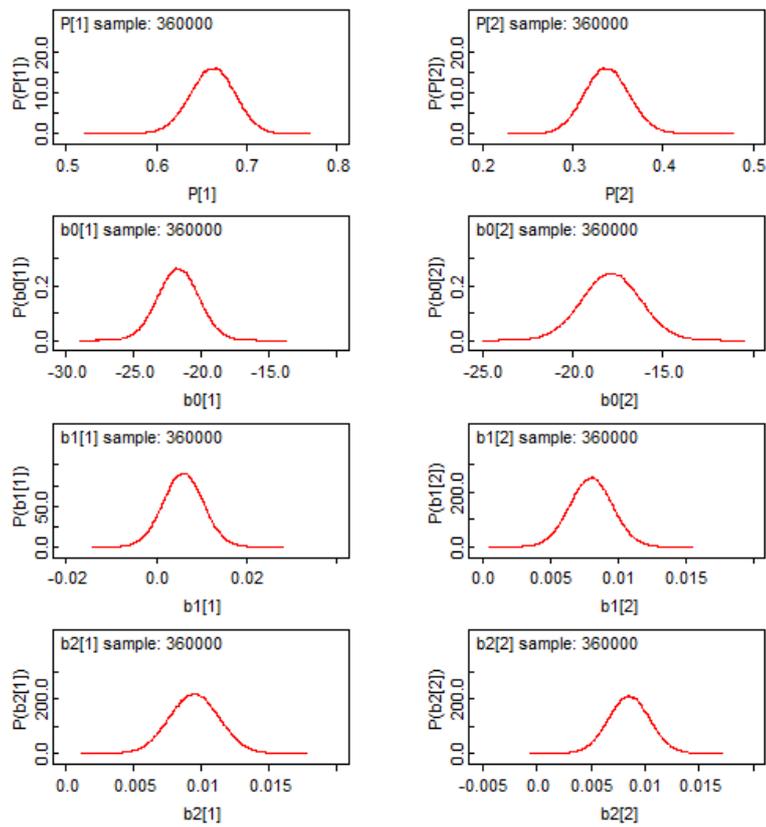
### A. Serial Plot



## B. Autocorrelation Plot

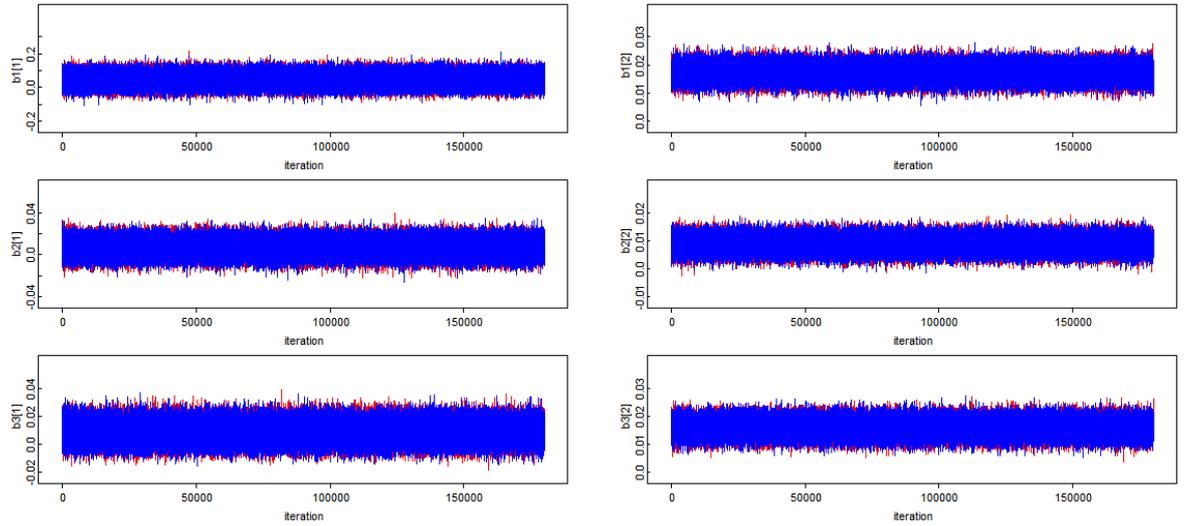


## C. Density plot

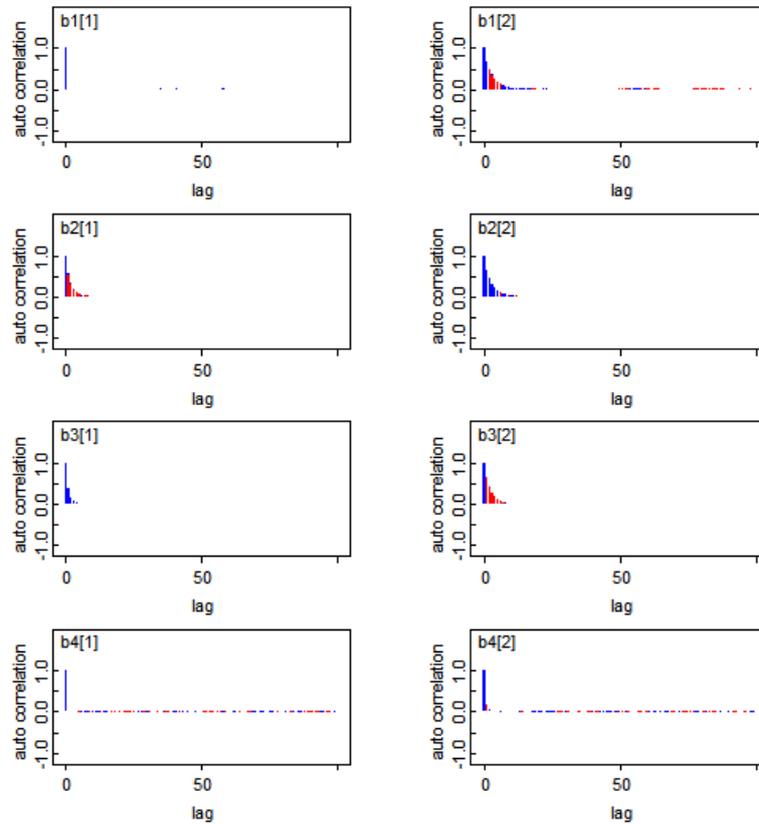


## Lampiran 6 Bukti Proses Estimasi Parameter Telah Konvergen Untuk Prodi Aktuaria

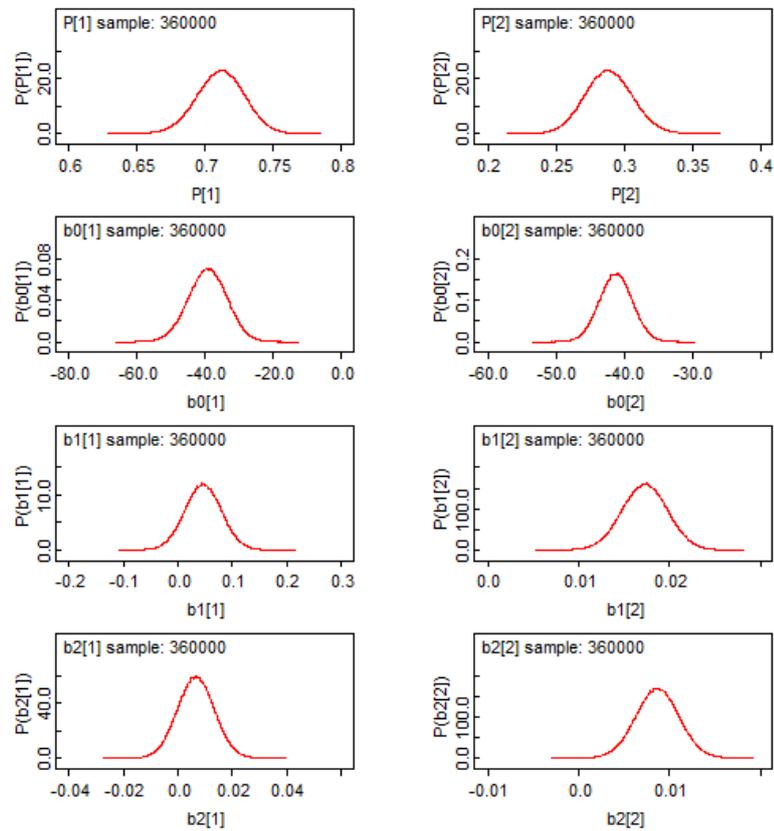
### A. Serial Plot



### B. Autocorrelation Plot

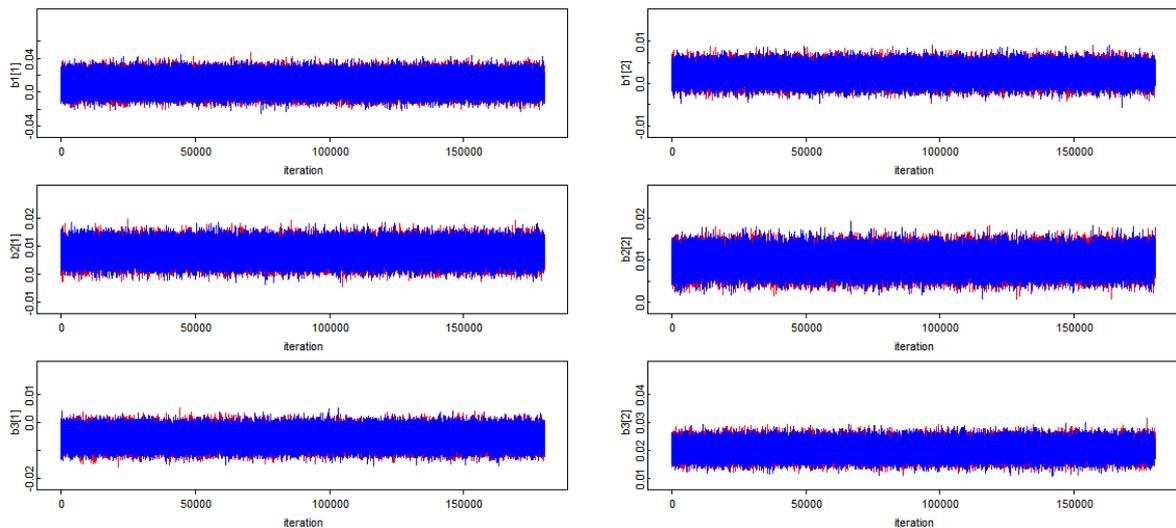


### C. Density plot

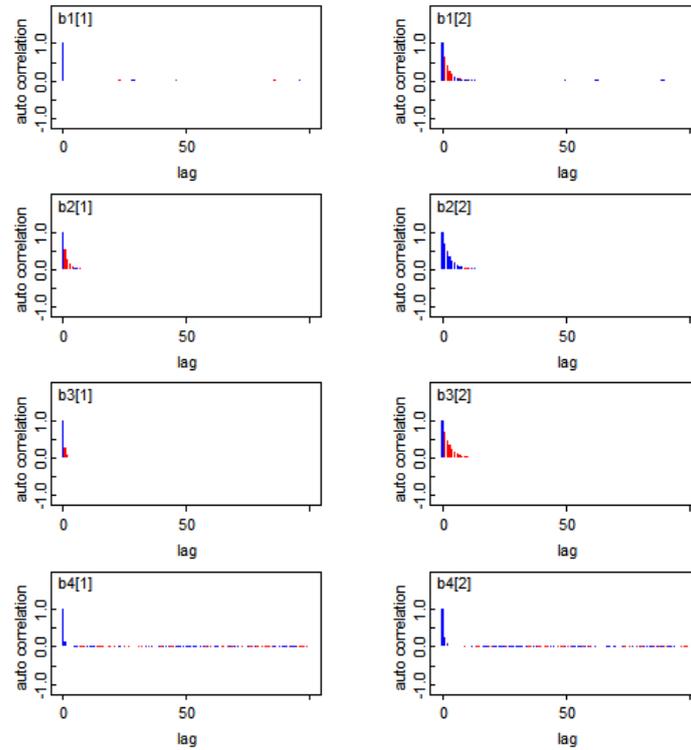


## Lampiran 7 Bukti Proses Estimasi Parameter Telah Konvergen Untuk Prodi Matematika

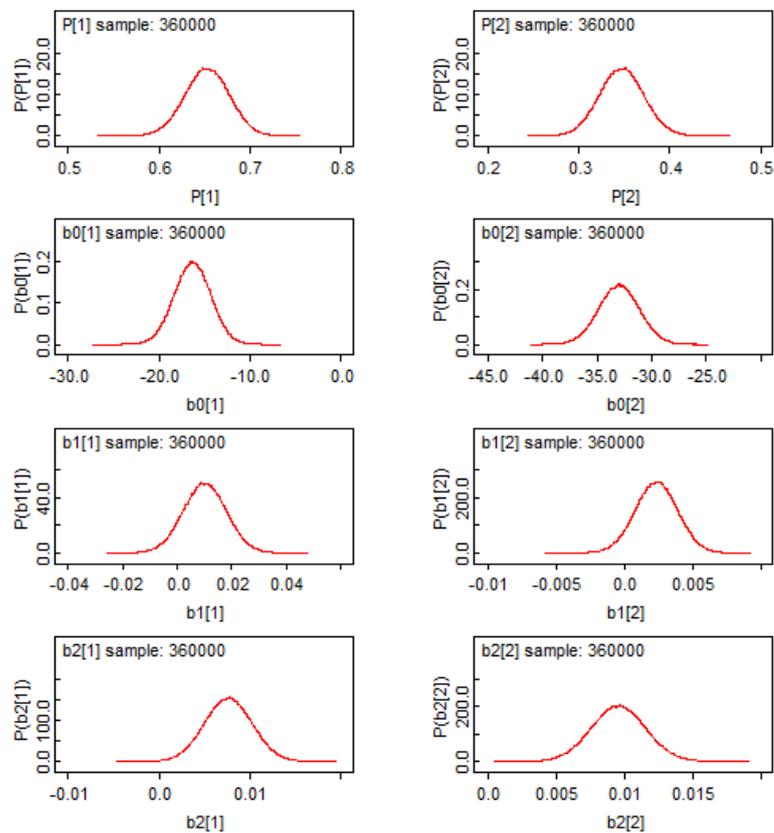
### A. Serial Plot



### B. Autocorrelation Plot

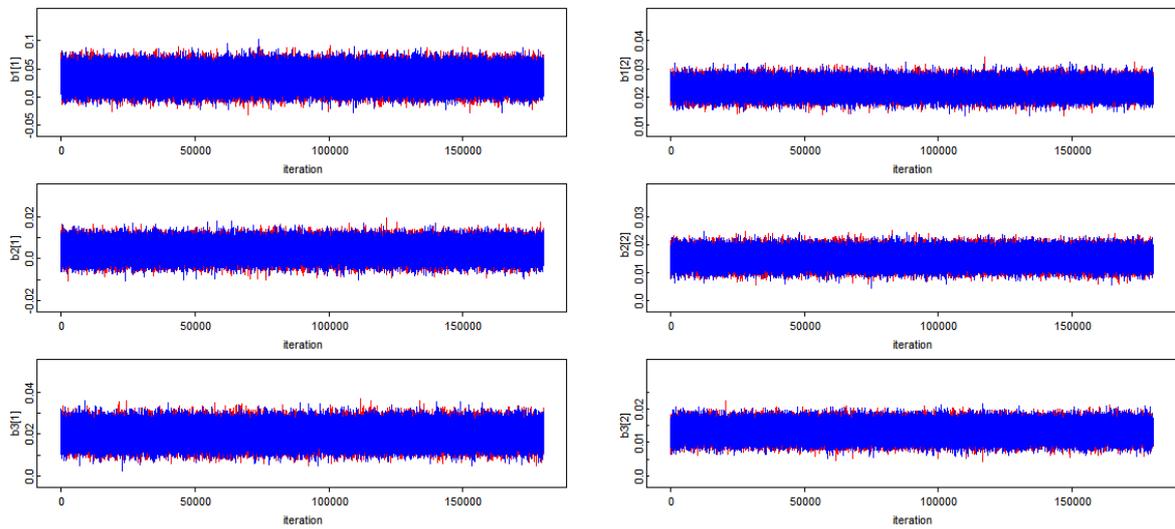


### C. Density plot

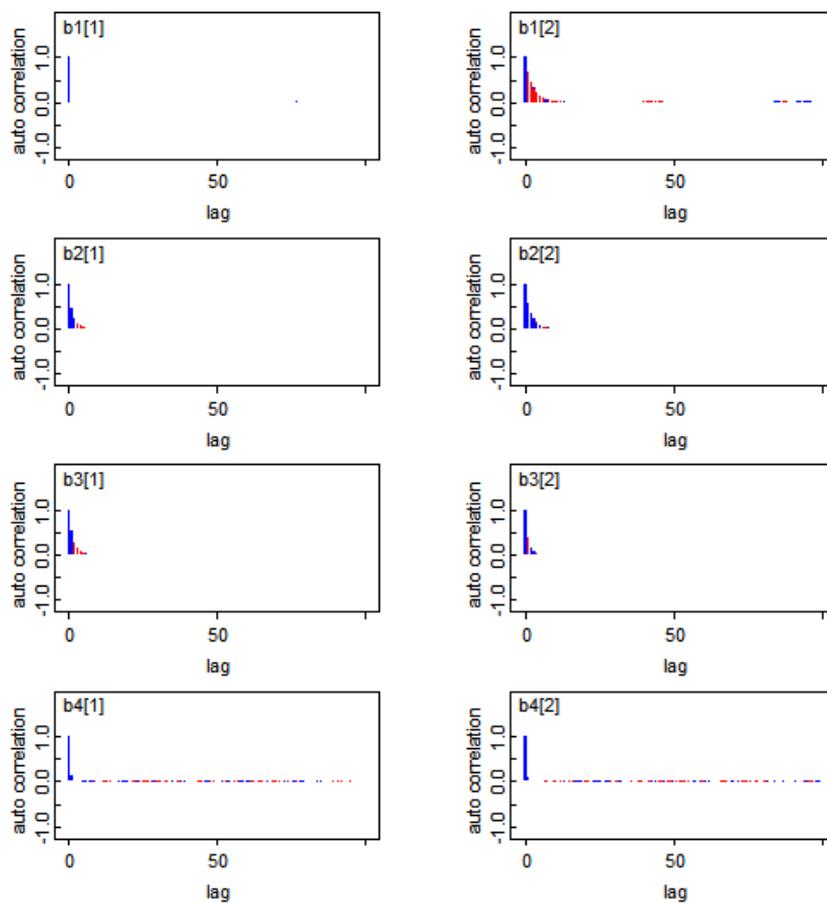


## Lampiran 8 Bukti Proses Estimasi Parameter Telah Konvergen Untuk Prodi Statistika

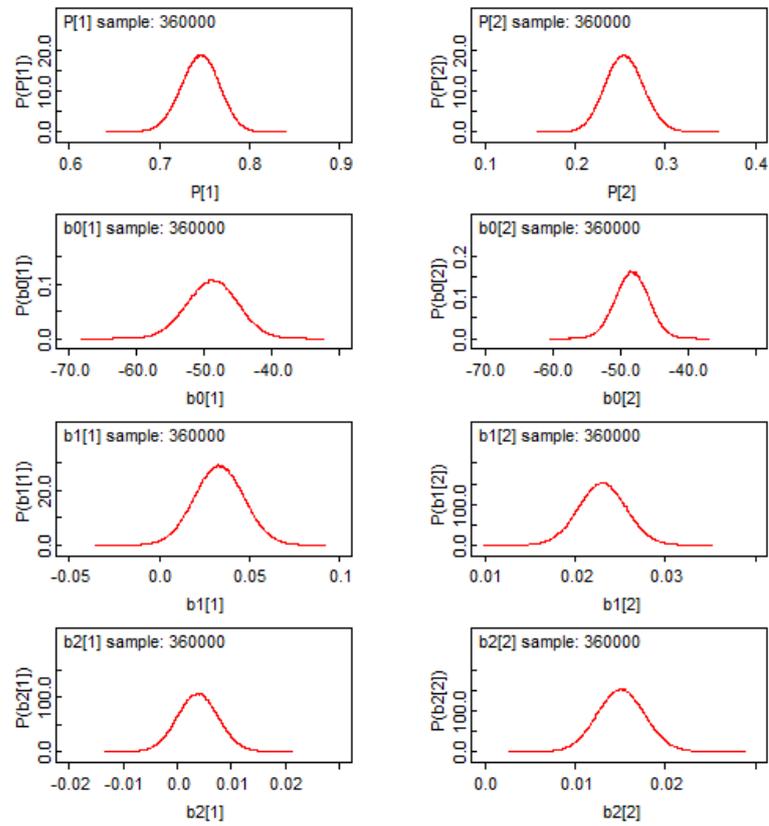
### A. Serial Plot



### B. Autocorrelation Plot



C. Density plot



## Biodata Penulis



Penulis merupakan mahasiswi S2 Statistika angkatan 2017. Drama perkuliahan yang dialami penulis membuat penulis memperdalam ilmu Statistika selama lima semester. Penulis merupakan alumni S1 Statistika ITS 2012 yang kemudian melanjutkan studi S2 Statistika ITS setelah vakum dari dunia pendidikan selama satu tahun. Selama masa perkuliahan berlangsung, penulis juga mengambil beberapa *job freelance* untuk mengisi kekosongan dalam hidup. Bagi penulis, menjadi mahasiswa saja tidak cukup untuk menghadapi kerasnya dunia kerja, sehingga penulis mencoba berbagai jenis pekerjaan yang dirasa sesuai dengan kebutuhan penulis. Beberapa pekerjaan yang pernah diambil penulis antara lain yaitu *Sales Executive Bangkok Tour and Travel* untuk wilayah Indonesia, surveyor beberapa project PDPM-ITS, *Quality Control* untuk Program Gapura Digital by Google Indonesia, Koordinator untuk program *Google Supercharges* dan *Training Coordinator* untuk Program *Google Digital Agency Training*. Dari seluruh pekerjaan yang pernah diambil, penulis ingin berbagi sedikit pengalaman dengan pembaca bahwa ilmu statistika sangat dibutuhkan dalam segala macam jenis pekerjaan, karena sejatinya ilmu statistika tidak hanya sekedar rumus, tapi pola pikir. Penulis menerapkan pola pikir ini untuk menyelesaikan tesis yang menurut penulis sangat berat dan Alhamdulillah, dengan bantuan segala pihak penulis bisa menyelesaikan dengan cukup baik. Penulis tidak menutup mata jika dalam penulisan tesis ini masih banyak yang belum sempurna. Jika pembaca berkenan untuk memberikan kritik, saran, cacian dan makian silahkan sampaikan via *email* di : [shielanoveliadp@gmail.com](mailto:shielanoveliadp@gmail.com). Akhir kata tidak ada yang mustahil ketika kita mau berubah, tidak ada yang sulit ketika kita mau berusaha. Semua sudah ada jalannya, tinggal bagaimana cara kita untuk menemukan jalannya. Tetap Semangat!