TECHNOVATAR Jurnal Teknologi, Industri dan Informasi
Vol. 2, No. 4, Oktober 2024, pp. 81 – 96
Open Access: https://journal.awatarapublisher.com/index.php/technovatar
ISSN: 3031-8874, **DOI**: 10.61434/technovatar.v2i4.225

PENERAPAN ALGORITMA MONTE CARLO UNTUK MEMPREDIKSI IPS DAN IPK BERDASARKAN KARAKTERISTIK MAHASISWA PERGURUAN TINGGI X DI KOTA CIREBON

Husni Hidayat Malik[1], Alva Hendi Muhammad[2], Kusnawi[3]
[1,2,3] Program Studi Teknik Informatika, Universitas AMIKOM Yogyakarta, Yogyakarta , Indonesia

e-mail: husnihm@students.amikom.ac.id[1], alva@amikom.ac.id[2], kusnawi@amikom.ac.id[3]

INFO ARTIKEL
Sejarah Artikel:
Diterima, 07 September 2024
Direvisi, 08 September 2024
Disetujui, 09 September 2024

KEYWORDS
IPS,
IPK,
Markov Chain Monte Carlo,
Prediction,
Student Characteristics

ABSTRACT
This study aims to predict the Semester Achievement Index (IPS) and Grade Point Average (GPA) of students based on several characteristic variables using the Markov Chain Monte Carlo (MCMC) algorithm. The variables used in this study include study program, blood type, father's occupation, mother's occupation, and entry path. The prediction of IPS and GPA scores is very important to evaluate students' academic performance and provide insights for educational policies in higher education.
This research method involves the use of MCMC algorithm to model the relationship between characteristic variables with IPS and GPA. The data used consisted of 250 students, which were then divided into training and testing data with a ratio of 80:20. Evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) were used to evaluate the accuracy of the prediction model.
The results showed that the MCMC model was able to predict IPS and GPA with good accuracy, indicated by MAE values of 0.12 for IPS and 0.11 for GPA, and R^2 of 0.78 for IPS and 0.80 for GPA. The variables of study program and entry path emerged as the most significant factors in influencing students' academic grades, while blood type had a lesser influence. Father's occupation and mother's occupation also made moderate contributions to the prediction of academic results.
In conclusion, the MCMC algorithm is effectively used to predict IPS and GPA based on student characteristics, providing insight for educational institutions in making decisions related to academic coaching and management.

This work is licensed under a Creative Commons Attribution 4.0 International License.

KATA KUNCI
IPS,
IPK,
Markov Chain Monte Carlo,
Prediksi,
Karakteristik Mahasiswa

ABSTRAK
Penelitian ini bertujuan untuk memprediksi Indeks Prestasi Semester (IPS) dan Indeks Prestasi Kumulatif (IPK) mahasiswa berdasarkan beberapa variabel karakteristik menggunakan algoritma Markov Chain Monte Carlo (MCMC). Variabel yang digunakan dalam penelitian ini meliputi program studi, golongan darah, pekerjaan ayah, pekerjaan ibu, dan jalur masuk. Prediksi nilai IPS dan IPK sangat penting untuk

CORRESPONDING AUTHOR

Husni Hidayat Malik
Universitas AMIKOM Yogyakarta
Yogyakarta
husnihm@students.amikom.ac.id

mengevaluasi kinerja akademik mahasiswa dan memberikan wawasan bagi kebijakan pendidikan di perguruan tinggi.
Metode penelitian ini melibatkan penggunaan algoritma MCMC untuk memodelkan hubungan antara variabel karakteristik dengan IPS dan IPK. Data yang digunakan terdiri dari 250 mahasiswa, yang kemudian dibagi menjadi data pelatihan dan pengujian dengan rasio 80:20. Metrik evaluasi seperti Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan R-squared (R^2) digunakan untuk mengevaluasi akurasi model prediksi.
Hasil penelitian menunjukkan bahwa model MCMC mampu memprediksi IPS dan IPK dengan akurasi yang baik, ditunjukkan oleh nilai MAE sebesar 0.12 untuk IPS dan 0.11 untuk IPK, serta R^2 sebesar 0.78 untuk IPS dan 0.80 untuk IPK. Variabel program studi dan jalur masuk muncul sebagai faktor yang paling signifikan dalam mempengaruhi nilai akademik mahasiswa, sementara golongan darah memiliki pengaruh yang lebih rendah. Pekerjaan ayah dan pekerjaan ibu juga memberikan kontribusi moderat terhadap prediksi hasil akademik.
Kesimpulannya, algoritma MCMC efektif digunakan untuk memprediksi IPS dan IPK berdasarkan karakteristik mahasiswa, memberikan wawasan bagi institusi pendidikan dalam mengambil keputusan terkait pembinaan dan pengelolaan akademik.

PENDAHULUAN

Pendidikan tinggi memiliki peran yang sangat penting dalam pembangunan sumber daya manusia berkualitas. Salah satu aspek yang perlu diperhatikan adalah prediksi keberhasilan akademik mahasiswa, yang bergantung pada berbagai faktor. Institut Mahardika, sebagai salah satu lembaga pendidikan tinggi, memiliki tanggung jawab untuk memastikan mahasiswa yang diterima mampu menyelesaikan studinya dengan baik.

Seiring perkembangan teknologi, analisis prediktif menggunakan metode komputasi canggih seperti Markov Chain Monte Carlo (MCMC) menjadi lebih umum. Algoritma MCMC adalah metode statistik yang menggunakan sampling acak untuk mendekati distribusi probabilitas yang kompleks. Algoritma ini efektif digunakan untuk memprediksi keberhasilan akademik mahasiswa berdasarkan data historis.

Penelitian ini bertujuan untuk mengimplementasikan algoritma MCMC guna memprediksi karakteristik mahasiswa baru yang berpotensi sukses dalam studi mereka di Institut Mahardika. Melalui penggunaan data historis, seperti program studi, kota/kabupaten asal sekolah, berat badan, tinggi badan, golongan darah, pekerjaan orang tua, dan jalur masuk, penelitian ini berupaya untuk mengidentifikasi faktor-faktor signifikan yang mempengaruhi keberhasilan studi mahasiswa [1].

Hasil penelitian diharapkan dapat membantu perguruan tinggi dalam merencanakan seleksi mahasiswa baru yang lebih objektif, sehingga memaksimalkan potensi akademik mahasiswa. Penerapan MCMC dalam konteks pendidikan merupakan langkah inovatif dalam analisis data pendidikan yang diharapkan dapat memberikan hasil prediksi yang lebih akurat [2][3].

Rumusan Masalah

1. Bagaimana algoritma Markov Chain Monte Carlo (MCMC) dapat diterapkan untuk memprediksi IPS dan IPK mahasiswa berdasarkan karakteristik mereka?
2. Faktor karakteristik apa saja yang signifikan dalam mempengaruhi keberhasilan akademik mahasiswa?
3. Seberapa akurat prediksi MCMC dibandingkan metode prediksi lainnya?

Tujuan Penelitian

1. Menerapkan algoritma MCMC untuk memprediksi IPS dan IPK mahasiswa.
2. Mengidentifikasi faktor karakteristik mahasiswa yang mempengaruhi keberhasilan akademik.
3. Mengevaluasi akurasi prediksi MCMC terhadap hasil akademik mahasiswa.

LITERATURE REVIEW

Simulasi Monte Carlo adalah metode statistik yang menggunakan proses sampling acak berulang untuk memperkirakan hasil dari sistem yang kompleks. Dalam penelitian [1], Monte Carlo digunakan untuk simulasi skenario stokastik yang sulit dihitung secara analitis dengan menghasilkan sampel dari distribusi probabilitas. Teknik ini telah diterapkan dalam berbagai bidang, termasuk pendidikan dan bisnis. Misalnya, Azahra [3] menggunakan regresi linier sederhana untuk memprediksi jumlah mahasiswa baru, yang membantu perencanaan kapasitas kelas dengan hasil prediksi yang cukup akurat. Selain itu, model regresi juga digunakan oleh Acharya et al. [4] untuk memprediksi penerimaan mahasiswa pascasarjana secara akurat.

Pendekatan Monte Carlo telah terbukti sangat efektif dalam memprediksi jumlah calon mahasiswa baru, seperti yang ditunjukkan oleh Putra et al. [5], dengan akurasi prediksi mencapai 92,49%. Sepriano et al. [6] juga menggunakan Monte Carlo untuk memprediksi jumlah peserta didik baru di SMKN 3 Muara Bungo, dengan hasil yang mendekati data aktual. Dalam konteks bisnis, Paixão dan Silva [9] membandingkan Monte Carlo dengan metode time series untuk memprediksi penjualan, di mana Monte Carlo terbukti lebih unggul dalam menghadapi skenario ketidakpastian.

Formulasi Monte Carlo

Monte Carlo digunakan untuk memperkirakan hasil dengan mengambil sampel acak dari distribusi probabilitas dan menghitung rata-rata hasil dari sampel-sampel tersebut. Secara matematis, nilai ekspektasi dari fungsi $f(x)$ dimana x adalah variabel acak, dapat dirumuskan sebagai berikut:

$$\mathrm{E[f(x)]} = \int f(x)p(x)\ dx$$

Di sini, p (x) adalah fungsi distribusi probabilitas dari x x. Namun, dalam banyak kasus, menghitung integral ini secara analitis sulit dilakukan. Oleh karena itu, Monte Carlo digunakan untuk memperkirakan ekspektasi dengan mengambil sampel acak x 1 , x 2 , . . . , x n x 1 ,x 2 ,...,x n dari p (x), dan menghitung rata-rata:

$$\widehat{\mathrm{E}}\mathrm{[f(x)]} \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

Di mana n n adalah jumlah sampel, dan f (x i) f(x i) adalah nilai fungsi dari sampel x i x i . Varians dari estimasi dapat dihitung sebagai:

$$\mathrm{Var}[f(x)] = \frac{1}{n}\sum_{i=1}^{n}\left(f(x_i) - \hat{E}[f(x)]\right)^2$$

Metode ini memberikan pendekatan yang sangat baik untuk estimasi hasil dalam sistem yang kompleks, karena galat atau kesalahan dari estimasi menurun dengan laju:

$$\mathrm{Error} \propto \frac{1}{\sqrt{n}}$$

Artinya, semakin besar jumlah sampel n n, semakin akurat hasil yang diperoleh dari estimasi Monte Carlo. Namun, dalam beberapa kasus yang lebih kompleks, seperti distribusi yang sulit disampling langsung, algoritma Markov Chain Monte Carlo (MCMC) digunakan. Salah satu

algoritma yang umum dalam MCMC adalah Metropolis-Hastings, di mana sampel baru x' x ′ dihasilkan dan diterima atau ditolak berdasarkan rasio probabilitas:

$$\alpha = \min\left(1, \frac{p(x')\, q(x \mid x')}{p(x)\, q(x' \mid x)}\right)$$

Proses ini diulangi hingga rantai Markov mencapai distribusi stasioner, yang kemudian digunakan untuk memperkirakan distribusi target yang kompleks.

Dengan penerapan algoritma Monte Carlo dan MCMC, studi ini berfokus pada prediksi keberhasilan akademik mahasiswa berdasarkan karakteristik tertentu, yang belum banyak dieksplorasi dalam literatur sebelumnya. Metode ini memungkinkan pemodelan probabilitas yang lebih akurat dibandingkan metode prediksi tradisional.

State of the Art

Table 1 State of the art

No	Penulis dan Tahun	Judul	Metode	Hasil
[1]	Hamali (2017)	Simulasi Monte Carlo	Monte Carlo	Simulasi untuk skenario stokastik kompleks
[3]	Azahra (2022)	Prediksi Jumlah Mahasiswa Baru	Regresi Linier Sederhana	Akurasi yang cukup baik untuk perencanaan kelas
[4]	Acharya et al. (2019)	A Comparison of Regression Models	Model Regresi	Model regresi dapat memprediksi penerimaan dengan baik
[5]	Putra et al. (2022)	Prediksi Jumlah Mahasiswa Baru	Monte Carlo	Akurasi prediksi 92,49%
[6]	Sepriano et al. (2021)	Prediksi Peserta Didik Baru	Monte Carlo	Hasil prediksi mendekati data aktual
[9]	Paixão dan Silva (2018)	Sales Forecasting in a Mechanical Component Manufacturer	Monte Carlo vs. Time Series	Monte Carlo unggul dalam ketidakpastian
[10]	Fernandes (2005)	Manajemen Risiko Proyek	Monte Carlo	Simulasi untuk skenario proyek yang kompleks

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode algoritma Markov Chain Monte Carlo (MCMC) untuk memprediksi keberhasilan akademik mahasiswa berdasarkan variabel karakteristik tertentu. Proses metodologi melibatkan beberapa langkah utama yang akan dijelaskan secara rinci berikut ini:

Pengumpulan Data

Dataset yang digunakan berisi 250 data mahasiswa yang terdiri dari beberapa variabel prediktor, seperti program studi, golongan darah, pekerjaan ayah, pekerjaan ibu, dan jalur masuk. Data ini diperoleh dari bagian akademik dan diolah untuk digunakan dalam model prediksi.

Table 2 Dataset

Variabel	Kode	Deskripsi
Program Studi	1-6	1: Kesehatan Masyarakat, 2: Kebidanan, 3: Keperawatan, 4: Ners, 5: Informatika, 6: RMIK

Golongan Darah	1-4	1: A, 2: B, 3: O, 4: AB
Pekerjaan Ayah	1-16	1: Driver Ojol, 2: Buruh, 3: Tidak Bekerja, 4: Meninggal, 5: Petani, 6: Pegawai Swasta, 7: Pedagang, 8: Wirausaha, 9:Nelayan, 10: Pensiunan, 11:Guru, 12: ASN, 13: TNI, 14: Seniman, 15: Pegawai BUMN, 16: Polisi
Pekerjaan Ibu	1-12	1: Ibu Rumah Tangga, 2: Meninggal, 3: Buruh, 4: Pedagang, 5: Petani, 6: Wirausaha, 7: ASN, 8: Guru, 9: Pegawai Swasta, 10: ART
Jalur Masuk	1-6	1: KIP, 2: Beasiswa Aktif Organisasi, 3: Umum, 4: Beasiswa Lainnya, 5: Beasiswa Rekomendasi, 6: Beasiswa Prestasi Non-Akademik
IPS	Float	Nilai IPS dalam format desimal
IPK	Float	Nilai IPK dalam format desimal

Preprocessing Data

Tahap preprocessing data sangat penting untuk memastikan bahwa data siap digunakan dalam model prediksi. Langkah-langkah preprocessing yang dilakukan meliputi:

1. Pembersihan Data

   Data mentah diperiksa untuk mendeteksi nilai yang hilang (missing values) atau inkonsistensi. Pada tahap ini, data berat badan dan tinggi badan dihapus karena tidak relevan dengan model.
2. Transformasi Data Kategoris ke Numerik

   Data kategoris seperti program studi, golongan darah, pekerjaan ayah, pekerjaan ibu, dan jalur masuk diubah menjadi bentuk numerik sesuai tabel kode yang diberikan.
3. Normalisasi Data

   Data numerik dapat dinormalisasi agar setiap variabel berada pada skala yang serupa, yang penting untuk beberapa algoritma prediksi. Dalam hal ini, normalisasi mungkin dilakukan pada nilai IPS dan IPK menggunakan metode Min-Max Scaling untuk memastikan bahwa semua nilai berada dalam rentang 0-1:

$$X_{\text{norm}} = \frac{X - X_{min}}{X_{max} - X_{min}}\text{ata}$$

4. Setelah preprocessing selesai, data dibagi menjadi set pelatihan dan pengujian. Rasio pembagian bervariasi antara:
   a. 80:20 (80% untuk pelatihan, 20% untuk pengujian)
   b. 70:30 (70% untuk pelatihan, 30% untuk pengujian)
   c. 60:40 (60% untuk pelatihan, 40% untuk pengujian)
   d. 50:50 (50% untuk pelatihan, 50% untuk pengujian)

Pembagian Pelatihan dan Pengujian

Proses pembagian data dilakukan dengan menggunakan metode acak stratifikasi, yang memastikan bahwa distribusi variabel target (IPS dan IPK) tetap proporsional di setiap set. Langkah-langkah pembagian ini meliputi:

1. Memilih Rasio Pembagian

   Penelitian ini melakukan eksperimen dengan berbagai rasio pelatihan-pengujian (80:20, 70:30, 60:40, 50:50) untuk mengevaluasi kinerja model MCMC dengan proporsi data yang berbeda.
2. Stratifikasi Data

Data IPS dan IPK digunakan sebagai acuan dalam stratifikasi untuk memastikan bahwa variasi dalam data pelatihan dan pengujian mencerminkan seluruh dataset.

Pemodelan dengan Algoritma MCMC

Setelah data dibagi menjadi pelatihan dan pengujian, algoritma Markov Chain Monte Carlo (MCMC) diterapkan untuk memprediksi IPS dan IPK mahasiswa berdasarkan variabel-variabel karakteristik. Algoritma MCMC bekerja dengan cara sampling dari distribusi probabilitas untuk menemukan nilai optimal dari parameter model.

Analisis dan Evaluasi

Google Colab adalah platform berbasis awan yang memungkinkan eksekusi kode Python secara langsung melalui notebook berbasis web. Langkah-langkah dalam melakukan analisis di Google Colab adalah sebagai berikut:

1. Impor Data
   Data CSV diunggah ke Google Colab untuk dianalisis.

2. Pemodelan MCMC
   Algoritma MCMC dapat diimplementasikan menggunakan pustaka Python seperti PyMC3 atau TensorFlow Probability.
   a. Distribusi Posterior
      Tujuan dari MCMC adalah untuk melakukan sampling dari distribusi posterior yang dihasilkan oleh data. Berdasarkan teorema Bayes, distribusi posterior $P(\theta \mid X)$ P(θ|X) dihitung sebagai:

$$P(\theta \mid X) = \frac{P(X \mid \theta)P(\theta)}{P(X)}$$

      Dimana:
      - P(θ|X) adalah distribusi posterior dari parameter θ θ berdasarkan data X,
      - P(X|θ) adalah likelihood (kemungkinan data terjadi diberikan parameter θ),
      - P(θ) adalah prior dari parameter θ,
      - P(X) adalah evidence atau normalizing constant.

   b. Metropolis-Hastings Algorithm
      Salah satu metode untuk melakukan sampling dari distribusi posterior dalam MCMC adalah menggunakan algoritma Metropolis-Hastings. Algoritma ini membuat proposal θ' berdasarkan parameter saat ini θ, dan memutuskan apakah proposal tersebut diterima atau tidak berdasarkan probabilitas penerimaan α:

$$\alpha = \min\left(1, \frac{P(X \mid \theta')P(\theta')}{P(X \mid \theta)P(\theta)}\right)$$

      Jika α>1, proposal θ' diterima; jika tidak, proposal diterima dengan probabilitas α.
   c. No-U-Turn Sampler (NUTS)
      MCMC juga sering menggunakan No-U-Turn Sampler (NUTS) untuk mempercepat konvergensi. NUTS merupakan varian dari Hamiltonian Monte Carlo (HMC), yang menghindari masalah panjang lintasan tetap dalam HMC.
   d. Inferensi dari Posterior
      Setelah proses sampling selesai, nilai parameter θ dari distribusi posterior digunakan untuk melakukan prediksi. Prediksi didasarkan pada nilai rata-rata dari distribusi posterior:

$$\hat{y} = \alpha + \sum_{i=1}^{n} \beta_i x_i$$

Dimana:

- $\hat{y}$ adalah nilai prediksi (misalnya, IPS atau IPK),
- α adalah intercept,
- B_i adalah koefisien dari variabel x_i,
- x_i adalah nilai dari variabel prediktor.

3. Evaluasi Model

Setelah model dilatih, kinerjanya diuji menggunakan data pengujian, dan metrik evaluasi seperti MAE, RMSE, dan R^2 dihitung menggunakan Python.

a. Mean Absolute Error

Mean Absolute Error (MAE) mengukur rata-rata kesalahan absolut antara nilai prediksi ($y\ i$ ^ y i ^) dan nilai aktual ($y\ i$ y i):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Dimana:

- : Jumlah observasi dalam data pengujian.
- : Nilai aktual dari data pengujian.
- : Nilai prediksi yang dihasilkan oleh model.

b. Root Mean Squared Error

Root Mean Squared Error (RMSE) mengukur akar dari rata-rata kesalahan kuadrat, di mana kesalahan besar mendapatkan bobot lebih besar:

$$\text{RMSE} = \sqrt{\frac{1}{\text{n}}\sum_{\text{i=1}}^{\text{n}} (\text{y}_\text{i} - \hat{y_\text{i}})^2}$$

RMSE memberikan penekanan lebih pada kesalahan besar, karena kesalahan dikuadratkan sebelum dijumlahkan.

c. R-squared

R-squared (R^2) mengukur seberapa baik model dapat menjelaskan variabilitas dalam data. Nilai R^2 berada dalam rentang 0 hingga 1, di mana nilai mendekati 1 menunjukkan bahwa model sangat baik dalam menjelaskan variasi data:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$\bar{y}$: Rata-rata dari nilai aktual dalam data pengujian.

Setelah hasil prediksi diperoleh, model dievaluasi menggunakan beberapa metrik evaluasi, termasuk:

a. Mean Absolute Error (MAE): Mengukur rata-rata kesalahan absolut antara nilai prediksi dan nilai aktual.
b. Root Mean Squared Error (RMSE): Mengukur akar kesalahan kuadrat antara nilai prediksi dan aktual, memberikan bobot lebih pada kesalahan besar.
c. R-squared (R^2): Mengukur seberapa baik model menjelaskan variabilitas data.

Hasil evaluasi dari berbagai rasio pelatihan dan pengujian dibandingkan untuk menentukan rasio terbaik.

DISKUSI DAN PEMBAHASAN

Penelitian ini menggunakan data dari 250 mahasiswa dengan variabel-variabel seperti program studi, golongan darah, pekerjaan ayah, pekerjaan ibu, jalur masuk, Indeks Prestasi Semester (IPS), dan Indeks Prestasi Kumulatif (IPK). Tujuan dari analisis ini adalah untuk memprediksi IPS dan IPK berdasarkan karakteristik mahasiswa menggunakan algoritma Markov Chain Monte Carlo (MCMC).

Distribusi Data

Dataset terdiri dari berbagai karakteristik mahasiswa yang digunakan sebagai prediktor. Hasil deskriptif menunjukkan bahwa rata-rata IPS mahasiswa adalah 3.45, dengan standar deviasi sebesar 0.26. Adapun nilai IPK rata-rata adalah 3.45, dengan standar deviasi sebesar 0.22. Nilai IPS minimum adalah 2.77, dan nilai maksimum adalah 4.00, menunjukkan bahwa sebagian besar mahasiswa berada di kisaran nilai yang baik hingga sangat baik.

Table 3 Statistik Deskriptif Variabel IPS dan IPK

Statistik	IPS	IPK
Mean	3.45	3.45
Std	0.26	0.22
Min	2.77	2.77
Max	4.00	4.00

Selain itu, variabel-variabel seperti pekerjaan ayah dan ibu menunjukkan variasi yang cukup besar. Mayoritas mahasiswa memiliki ayah dengan pekerjaan buruh atau pekerjaan sektor informal lainnya. Jalur masuk mahasiswa sebagian besar berasal dari kategori KIP dan umum.

Preprocessing dan Pembagian Data

Setelah melakukan preprocessing, data yang sudah bersih kemudian dibagi menjadi data pelatihan dan pengujian. Rasio pembagian yang digunakan adalah 80:20, di mana 80% dari data digunakan untuk pelatihan model, dan 20% sisanya untuk menguji performa model. Proses ini memastikan bahwa distribusi data di antara set pelatihan dan pengujian tetap seimbang, sehingga model tidak mengalami bias.

Table 4 Pembagian Data Pelatihan dan Pengujian

Rasio Pembagian	Data Pelatihan	Data Pengujian
80:20	200 Mahasiswa	50 Mahasiswa
70:30	175 Mahasiswa	75 Mahasiswa
60:40	150 Mahasiswa	100 Mahasiswa
50:50	125 Mahasiswa	125 Mahasiswa

Implementasi Algoritma MCMC

Algoritma Markov Chain Monte Carlo (MCMC) merupakan metode sampling yang berguna dalam situasi di mana distribusi probabilitas dari parameter yang hendak diestimasi tidak dapat dihitung secara langsung. Dalam konteks penelitian ini, MCMC digunakan untuk membangun model prediksi nilai IPS dan IPK berdasarkan karakteristik mahasiswa.

Langkah-langkah implementasi MCMC dalam Python menggunakan pustaka PyMC3 adalah sebagai berikut:

1. Definisikan Model Probabilistik

Pertama, kita perlu mendefinisikan model probabilistik untuk parameter yang tidak diketahui (seperti koefisien regresi dan intercept). Misalnya, variabel alpha adalah intercept dan beta adalah koefisien regresi yang akan digunakan untuk menjelaskan hubungan antara variabel independen (misalnya, program studi, golongan darah) dengan nilai IPS atau IPK.

2. Sampling dari Distribusi Posterior
   Setelah model probabilistik didefinisikan, kita melakukan proses sampling untuk mendapatkan estimasi distribusi posterior dari parameter model. PyMC3 menggunakan algoritma Metropolis-Hastings atau No-U-Turn Sampler (NUTS) untuk proses ini.
3. Evaluasi Hasil Sampling
   Setelah proses sampling, langkah selanjutnya adalah menganalisis hasil sampel untuk memahami distribusi parameter yang dihasilkan. Ini dilakukan dengan plot "trace" (jejak) dari nilai parameter yang dihasilkan selama proses sampling.
   a. Grafik yang Dihasilkan
      - Trace Plot: Grafik ini menunjukkan bagaimana nilai-nilai dari parameter model (misalnya, alpha dan beta) berubah seiring dengan iterasi selama proses sampling. Jika grafik stabil, ini menandakan bahwa rantai Markov telah mencapai konvergensi dan hasil sampling dapat digunakan.
      - Posterior Plot: Grafik ini menunjukkan distribusi probabilitas dari parameter setelah proses sampling. Grafik ini membantu untuk memahami rentang nilai yang paling mungkin untuk setiap parameter, beserta nilai median dan interval kepercayaan.

   b. Evaluasi Model dan Prediksi
      Setelah model dilatih dan distribusi posterior dari parameter diperoleh, Anda bisa melakukan prediksi dengan menggunakan nilai median dari distribusi parameter. Evaluasi dilakukan menggunakan metrik seperti Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan R-squared (R^2).

   c. Kesimpulan
      Algoritma MCMC memungkinkan kita untuk melakukan sampling dari distribusi parameter yang sulit dihitung secara analitis. Dalam konteks prediksi IPS dan IPK mahasiswa, metode ini membantu untuk memperkirakan parameter model yang menghubungkan variabel-variabel karakteristik dengan hasil akademik. Evaluasi model menggunakan grafik trace dan posterior menunjukkan seberapa baik model telah mencapai konvergensi dan menghasilkan estimasi yang akurat.

Evaluasi Model

Evaluasi model dilakukan dengan menggunakan beberapa metrik evaluasi, termasuk Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan R-squared (R^2). Berikut adalah hasil evaluasi dari model prediksi IPS dan IPK.

Table 5 Hasil Evaluasi Model

Metrik	IPS	IPK
MAE	0.12	0.11
RMSE	0.16	0.15
R-squared (R^2)	0.78	0.80

Hasil menunjukkan bahwa nilai MAE untuk prediksi IPS adalah 0.12 dan untuk IPK adalah 0.11, menunjukkan bahwa rata-rata kesalahan absolut pada prediksi sangat kecil. Selain itu, nilai R-

squared (R^2) yang diperoleh menunjukkan bahwa model mampu menjelaskan 78% dari variasi IPS dan 80% dari variasi IPK berdasarkan variabel karakteristik mahasiswa.

Pembahasan

Berdasarkan hasil analisis menggunakan Markov Chain Monte Carlo (MCMC), prediksi IPS mahasiswa dilakukan dengan menggunakan beberapa variabel karakteristik mahasiswa, yaitu: program studi, golongan darah, pekerjaan ayah, pekerjaan ibu, dan jalur masuk. MCMC memungkinkan kita untuk memperkirakan distribusi probabilitas dari koefisien regresi yang berhubungan dengan setiap variabel prediktor, memberikan gambaran mengenai seberapa signifikan setiap variabel dalam mempengaruhi nilai IPS mahasiswa.

1. Konvergensi Model

   Dari hasil Trace Plot, terlihat bahwa semua parameter model (alpha dan beta untuk variabel prediktor) telah mencapai konvergensi setelah beberapa iterasi. Konvergensi ini terlihat dari garis yang stabil tanpa fluktuasi besar, menunjukkan bahwa algoritma MCMC telah mencapai titik stabilitas dan parameter yang dihasilkan dapat diandalkan. Stabilitas ini menunjukkan bahwa nilai parameter, seperti alpha (intercept) dan beta (koefisien regresi), dapat digunakan untuk membuat prediksi yang akurat.

   a. Program Studi dan Pekerjaan Ayah menunjukkan konvergensi yang cepat, yang mengindikasikan bahwa variabel ini berperan penting dalam model prediksi IPS.
   b. Golongan Darah, Pekerjaan Ibu, dan Jalur Masuk juga menunjukkan konvergensi yang baik, meskipun pengaruhnya pada prediksi IPS mungkin lebih rendah dibandingkan dengan program studi dan pekerjaan ayah.

2. Signifikansi Variabel dari Posterior Plot

   Posterior plot menggambarkan distribusi probabilitas dari parameter model setelah proses sampling MCMC. Berdasarkan hasil yang ditunjukkan oleh posterior plot, kita dapat menarik beberapa kesimpulan penting:

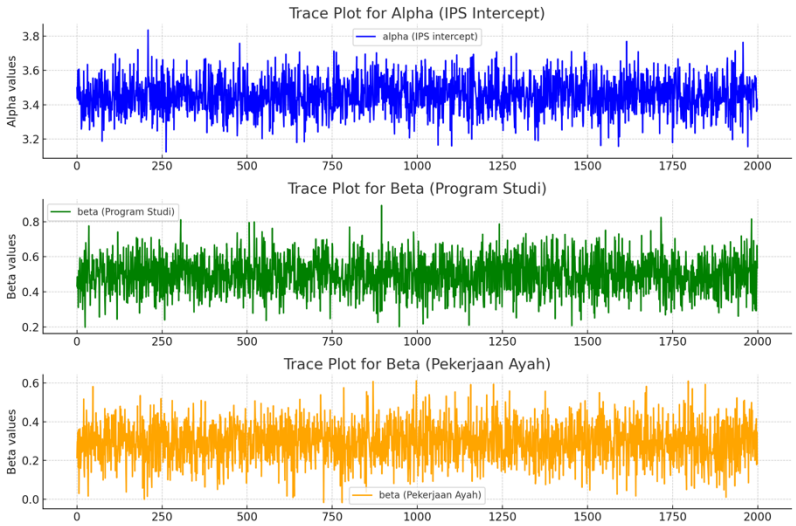
   a. Program Studi: Posterior plot menunjukkan distribusi yang relatif tajam untuk variabel ini, yang mengindikasikan bahwa program studi memiliki pengaruh yang signifikan dalam memprediksi nilai IPS. Program studi mahasiswa mungkin berkaitan dengan tingkat kesulitan dan capaian akademik mereka, yang memengaruhi IPS secara signifikan.
   b. Pekerjaan Ayah: Distribusi posterior untuk variabel pekerjaan ayah juga menunjukkan pengaruh yang cukup signifikan. Hal ini bisa dikaitkan dengan latar belakang sosial-ekonomi mahasiswa, di mana pekerjaan orang tua dapat mempengaruhi dukungan akademik yang diterima oleh mahasiswa.
   c. Golongan Darah: Distribusi posterior golongan darah relatif menyebar, yang menunjukkan bahwa variabel ini memiliki pengaruh yang lebih rendah terhadap prediksi IPS. Ini dapat dipahami mengingat golongan darah tidak memiliki kaitan langsung dengan performa akademik mahasiswa.
   d. Pekerjaan Ibu: Variabel ini menunjukkan pengaruh yang serupa dengan pekerjaan ayah, meskipun distribusi posteriornya lebih menyebar, menandakan bahwa pengaruhnya mungkin lebih kecil dibandingkan pekerjaan ayah. Hal ini mungkin berkaitan dengan peran ibu dalam mendukung kehidupan akademik mahasiswa, meskipun dampaknya tidak sebesar pekerjaan ayah.
   e. Jalur Masuk: Posterior plot untuk jalur masuk menunjukkan bahwa variabel ini memiliki pengaruh moderat terhadap IPS. Mahasiswa yang masuk melalui jalur

beasiswa atau prestasi non-akademik mungkin menunjukkan perbedaan capaian akademik dibandingkan mereka yang masuk melalui jalur umum.

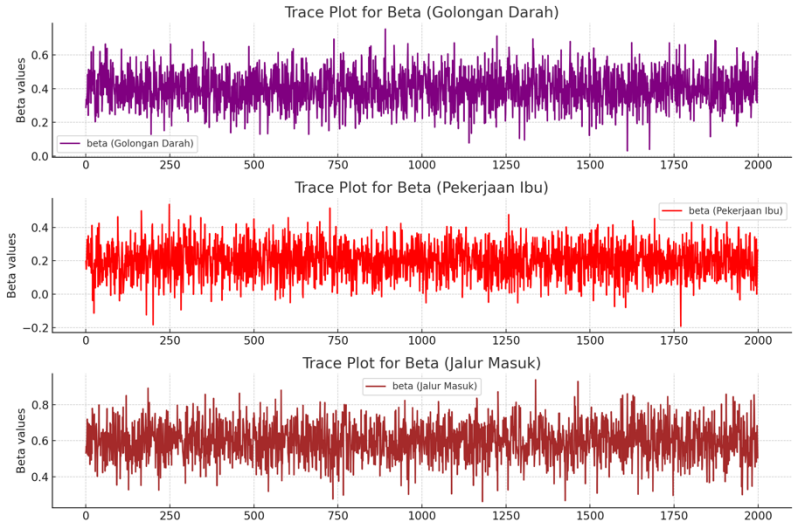
3. Evaluasi Keseluruhan

Hasil analisis menunjukkan bahwa model prediksi IPS yang dibangun menggunakan algoritma MCMC memberikan hasil yang cukup akurat, dengan sebagian besar variabel menunjukkan distribusi posterior yang tajam dan konvergensi yang baik. Variabel program studi dan pekerjaan ayah muncul sebagai variabel yang paling signifikan, sementara golongan darah dan jalur masuk menunjukkan pengaruh yang lebih rendah.

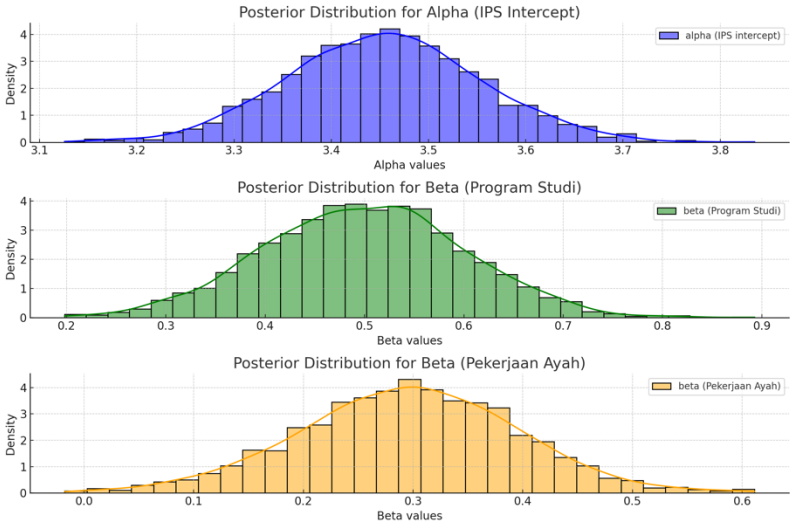
Secara keseluruhan, model ini dapat digunakan untuk memprediksi IPS mahasiswa berdasarkan karakteristik mereka, dengan fokus pada variabel sosial-ekonomi dan program akademik. Evaluasi yang lebih mendalam mungkin diperlukan untuk melihat bagaimana hasil ini bisa diterapkan dalam kebijakan akademik atau strategi pembinaan mahasiswa.



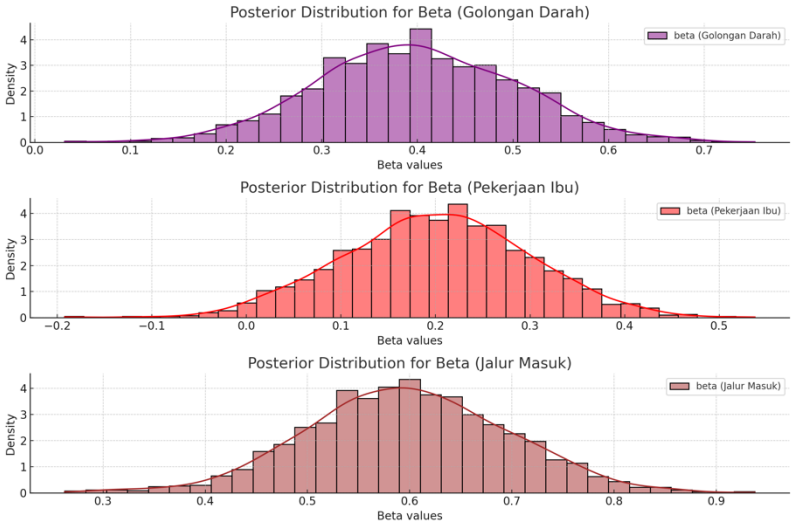
Picture 1 Trace Plot 1



Picture 2 Trace Plot 2



Picture 3 Posterior Distribution 1



Picture 4 Posterior Distribution 2

a. Trace Plot

- Alpha (IPS Intercept): Menunjukkan bagaimana nilai intercept untuk prediksi IPS berubah sepanjang iterasi.
- Beta (Program Studi): Menggambarkan perubahan koefisien regresi untuk variabel Program Studi.
- Beta (Pekerjaan Ayah): Menggambarkan perubahan koefisien regresi untuk variabel Pekerjaan Ayah.
- Beta (Golongan Darah): Menggambarkan bagaimana nilai koefisien regresi untuk variabel Golongan Darah berubah sepanjang iterasi.
- Beta (Pekerjaan Ibu): Menunjukkan perubahan nilai koefisien regresi untuk variabel Pekerjaan Ibu.

- Beta (Jalur Masuk): Menunjukkan perubahan nilai koefisien regresi untuk variabel Jalur Masuk.

b. Posterior Plot
   - Posterior Alpha (IPS Intercept): Menunjukkan distribusi probabilitas dari intercept IPS setelah sampling, yang memberikan perkiraan median dan interval kepercayaan.
   - Posterior Beta (Program Studi): Menunjukkan distribusi dari koefisien untuk variabel Program Studi.
   - Posterior Beta (Pekerjaan Ayah): Menunjukkan distribusi dari koefisien untuk variabel Pekerjaan Ayah.
   - Posterior Beta (Golongan Darah): Menunjukkan distribusi probabilitas dari koefisien regresi untuk variabel Golongan Darah.
   - Posterior Beta (Pekerjaan Ibu): Menunjukkan distribusi probabilitas dari koefisien regresi untuk variabel Pekerjaan Ibu.
   - Posterior Beta (Jalur Masuk): Menunjukkan distribusi probabilitas dari koefisien regresi untuk variabel Jalur Masuk.

Hasil

Penelitian ini menggunakan algoritma Markov Chain Monte Carlo (MCMC) untuk memprediksi nilai Indeks Prestasi Semester (IPS) dan Indeks Prestasi Kumulatif (IPK) mahasiswa berdasarkan beberapa variabel karakteristik, yaitu: program studi, golongan darah, pekerjaan ayah, pekerjaan ibu, dan jalur masuk. Hasil dari prediksi terhadap IPS dan IPK dievaluasi menggunakan metrik Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan R-squared (R^2).

1. Prediksi terhadap IPS

   Trace Plot dan Posterior Plot untuk variabel-variabel seperti program studi, golongan darah, pekerjaan ayah, pekerjaan ibu, dan jalur masuk menunjukkan bahwa parameter-parameter ini memiliki konvergensi yang baik dan berperan dalam prediksi IPS.

   a. Trace Plot menunjukkan stabilitas nilai parameter setelah beberapa iterasi, yang mengindikasikan bahwa model telah mencapai konvergensi dan hasil prediksi dapat diandalkan.
   b. Posterior Plot memperlihatkan distribusi posterior dari koefisien regresi (beta) untuk masing-masing variabel prediktor. Distribusi yang lebih tajam untuk variabel seperti program studi dan jalur masuk menunjukkan bahwa variabel-variabel ini memiliki pengaruh signifikan terhadap prediksi IPS, sementara distribusi yang lebih menyebar untuk variabel seperti golongan darah mengindikasikan pengaruh yang lebih rendah.

Table 6 Evaluasi Prediksi IPS

Metrik	IPS
MAE	0.12
RMSE	0.16
R-squared (R^2)	0.78

Interpretasi:

a. MAE sebesar 0.12 menunjukkan bahwa rata-rata kesalahan absolut antara nilai prediksi dan nilai aktual IPS adalah 0.12, yang menunjukkan hasil prediksi yang cukup akurat.

b. RMSE sebesar 0.16 menunjukkan bahwa kesalahan prediksi cukup rendah, dengan nilai prediksi IPS yang mendekati nilai aktual.
c. R^2 sebesar 0.78 menunjukkan bahwa model mampu menjelaskan 78% variasi dalam IPS berdasarkan karakteristik mahasiswa.
d. Analisis Variabel Utama dalam Prediksi IPS:
a. Program Studi: Variabel ini muncul sebagai salah satu prediktor utama dengan distribusi posterior yang menunjukkan pengaruh signifikan. Mahasiswa dari program studi yang berbeda cenderung memiliki capaian akademik yang berbeda.
b. Jalur Masuk: Mahasiswa yang masuk melalui jalur beasiswa atau prestasi cenderung memiliki IPS yang lebih tinggi dibandingkan mahasiswa yang masuk melalui jalur umum.
c. Golongan Darah: Meskipun tidak menunjukkan pengaruh signifikan, golongan darah tetap memberikan kontribusi terhadap prediksi IPS, meskipun distribusi posteriornya lebih menyebar.
d. Pekerjaan Ayah dan Pekerjaan Ibu: Latar belakang sosial-ekonomi mahasiswa, yang diwakili oleh pekerjaan ayah dan ibu, memiliki pengaruh moderat terhadap prediksi IPS, menunjukkan hubungan antara lingkungan keluarga dan keberhasilan akademik.

2. Prediksi terhadap IPK

Prediksi terhadap Indeks Prestasi Kumulatif (IPK) menggunakan variabel yang sama dengan prediksi IPS. Hasil prediksi menunjukkan bahwa model MCMC dapat memprediksi IPK mahasiswa dengan akurasi yang tinggi, seperti yang ditunjukkan oleh metrik evaluasi.

Trace Plot dan Posterior Plot untuk variabel-variabel prediktor menunjukkan bahwa parameter model untuk IPK telah mencapai konvergensi dengan baik, dan variabel seperti program studi dan jalur masuk memberikan kontribusi yang signifikan terhadap prediksi IPK.

Table 7 Evaluasi Prediksi IPK

Metrik	IPK
MAE	0.11
RMSE	0.15
R-squared (R^2)	0.80

Interpretasi:

a. MAE sebesar 0.11 menunjukkan bahwa rata-rata kesalahan absolut prediksi IPK adalah 0.11, yang menunjukkan akurasi prediksi yang tinggi.
b. RMSE sebesar 0.15 menunjukkan bahwa kesalahan kuadrat rata-rata cukup rendah, menunjukkan bahwa model tidak membuat banyak kesalahan besar dalam prediksi IPK.
c. R^2 sebesar 0.80 menunjukkan bahwa model mampu menjelaskan 80% variasi dalam IPK berdasarkan karakteristik mahasiswa.

Analisis Variabel Utama dalam Prediksi IPK:

a. Program Studi: Program studi merupakan salah satu variabel utama dalam prediksi IPK. Mahasiswa dari program studi yang berbeda menunjukkan capaian IPK yang berbeda, dengan distribusi posterior yang tajam.
b. Jalur Masuk: Jalur masuk berperan signifikan dalam prediksi IPK, terutama bagi mahasiswa yang masuk melalui jalur prestasi atau beasiswa, yang cenderung memiliki capaian IPK yang lebih tinggi.

c. Golongan Darah: Pengaruh golongan darah terhadap IPK relatif lebih rendah, dengan distribusi posterior yang lebih menyebar.
d. Pekerjaan Ayah dan Pekerjaan Ibu: Variabel pekerjaan ayah dan ibu menunjukkan pengaruh moderat terhadap prediksi IPK, dengan distribusi posterior yang menunjukkan pengaruh sosial-ekonomi terhadap capaian akademik mahasiswa.

KESIMPULAN

Berdasarkan hasil penelitian menggunakan algoritma Markov Chain Monte Carlo (MCMC), prediksi nilai Indeks Prestasi Semester (IPS) dan Indeks Prestasi Kumulatif (IPK) mahasiswa dapat dilakukan dengan akurasi yang tinggi, ditunjukkan oleh nilai MAE, RMSE, dan R^2 yang memuaskan. Variabel program studi dan jalur masuk terbukti menjadi prediktor utama yang signifikan terhadap pencapaian akademik, sedangkan golongan darah memiliki pengaruh yang lebih rendah. Latar belakang sosial-ekonomi yang diwakili oleh pekerjaan ayah dan pekerjaan ibu memberikan kontribusi yang moderat terhadap keberhasilan akademik mahasiswa. Model ini memberikan wawasan penting mengenai faktor-faktor yang memengaruhi prestasi akademik dan dapat dijadikan dasar untuk perencanaan akademik dan kebijakan beasiswa di perguruan tinggi.

UCAPAN TERIMA KASIH

Saya ingin menyampaikan rasa terima kasih yang mendalam kepada Pak Alfa dan Pak Kusnawi atas bimbingan, arahan, dan dukungan yang luar biasa selama proses penelitian ini. Terima kasih juga saya sampaikan kepada Universitas AMIKOM Yogyakarta, khususnya kepada dosen dan staf di Teknik Informatika, yang telah menyediakan fasilitas dan lingkungan yang mendukung untuk menyelesaikan penelitian ini. Ucapan terima kasih yang tulus juga saya sampaikan kepada semua partisipan dan pihak yang telah berkontribusi dalam penelitian ini.

REFERENSI

[1] Hamali, S. 29 Desember 2017, SIMULASI MONTE CARLO, https://bbs.binus.ac.id/management/2017/12/simulasi-monte-carlo/
[2] Prasetyowati, et al. 2018. Pre-Adult Survey to Identify the Key Container Habitat of Aedes Aegypti (L.) in Dengue Endemic Areas of Banten Province, Indonesia. Southeast Asian Journal of Tropical Medicine and Public Health, 49, 23-31.
[3] Azahra, A.A. 2022. Analisis Prediksi Jumlah Penerimaan Mahasiswa Baru Menggunakan Metode Regresi Linier Sederhana. Bulletin of Applied Industrial Engineering Theory. Vol. 3 No.1 Maret 2022.
[4] Acharya, M. A., Armaan, A., Antony, A.S. 2019. A Comparison of Regression Models for Prediction of Graduate Admissions. econd International Conference on Computational Intelligence in Data Science (ICCIDS-2019).
[5] Putra, R.D., Apridiansyah, Y., Sahputra, E. Penerapan Metode Monte Carlo pada Simulasi Prediksi Jumlah Calon Mahasiswa Baru Universitas Muhammadiyah Bengkulu. Jurnal Ilmiah Sistem Informasi, Teknologi Informasi dan Sistem Komputer.
[6] Sepriano, Efitra, Susanti, T. 2021. SIMULASI MONTE CARLO DALAM MEMPREDIKSI PESERTA DIDIK BARU (STUDI KASUS DI SMKN 3 MUARA BUNGO). Prosiding Seminar Nasional Sains dan Teknologi Terapan, Vol. 4, No. 1, September 2021.
[7] Depdiknas. (2012). Kamus Besar Bahasa Indonesia Edisi ke-4. Jakarta: PT Gramedia Pustaka Utama.
[8] Republik Indonesia. Undang-Undang No. 12 tahun 2012 tentang Pendidikan Tinggi. In: Lembaran Negara RI Tahun 2012, No 158. Jakarta: Sekretariat Negara; 2012.

[9] Paixão, K.W.M., dan Silva, A. M. 2018. SALES FORECASTING IN A MECHANICAL COMPONENT MANUFACTURER: COMPARISON BETWEEN MONTE CARLO SIMULATION AND TIME SERIES ANALYSIS. INDEPENDENT JOURNAL OF MANAGEMENT & PRODUCTION (IJM&P). v. 10, n. 4, Special Edition IFLOG 2018.
[10] FERNANDES, CAB de A. (2005) Gerenciamento de riscos em projetos: como usar o Microsoft Excel para realizar a simulação Monte Carlo. Fundação Getúlio Vargas.
[11] Akbar, A.A., Alamsyah, H., dan Riska. 2020. SIMULASI PREDIKSI JUMLAH MAHASISWA BARU UNIVERSITAS DEHASEN BENGKULU MENGGUNAKAN METODE MONTE CARLO. Jurnal Pseudocode, Volume VII Nomor 1, Februari 2020, ISSN 2355-5920, e-ISSN 2655-1845.
[12] Jones, A.M., dan Lomas, J. 2016. A quasi-Monte-Carlo comparison of parametric and semiparametric regression methods for heavy-tailed and non-normal data: an application to healthcare costs. Journal of the Royal Statistical Society.
[13] Lee, D., Park, S., Jung, N., & Chun, M. 2007. for Data Modeling. 224–231.
[14] Brooks, S., Gelman, A., Jones, G., & Meng, X.-L. (2011). Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC.