

Implementasi Seleksi Fitur Klasifikasi Waktu Kelulusan Mahasiswa Menggunakan Correlation Matrix With Heatmap

Amiruddin
Program Studi Teknik Informatika
Universitas Ichsan Gorontalo
Gorontalo, Indonesia
amier.76@gmail.com

Rezqiwati Ishak
Program Studi Teknik Informatika
Universitas Ichsan Gorontalo
Gorontalo, Indonesia
rezqi.uig@gmail.com

Diterima : Mei 2022
Disetujui : Juni 2022
Dipublikasi : Juli 2022

Abstrak—Klasifikasi waktu kelulusan Mahasiswa biasanya menggunakan lebih dari 10 atribut input, namun tidak semua atribut tersebut memiliki relevansi dengan atribut output, sehingga sebelum dilakukan klasifikasi perlu dilakukan pemilihan atribut yang relevan melalui seleksi fitur terlebih dahulu. Seleksi fitur merupakan bagian dalam tahapan *data preparation* pada siklus hidup *data science*, seleksi Fitur merupakan salah satu konsep inti dalam pembelajaran mesin yang sangat mempengaruhi kinerja model data mining. Dimana pada tahapan ini dilakukan seleksi fitur yang berkontribusi paling besar pada variabel prediksi atau output, memilih fitur yang tidak relevan dalam dataset dapat mengurangi keakuratan model dan membuat model belajar berdasarkan fitur yang tidak relevan. Dengan demikian melakukan seleksi fitur sebelum memodelkan data akan didapatkan manfaat seperti : mengurangi *Overfitting*, meningkatkan akurasi dan mengurangi waktu pelatihan. Teknik yang digunakan dalam seleksi fitur pada penelitian ini adalah *Correlation Matrix with Heatmap* karena teknik ini mudah digunakan dan juga memberikan hasil yang baik. Setelah dilakukan implementasi seleksi fitur pada klasifikasi waktu kelulusan Mahasiswa dengan menggunakan 13 variabel/atribut awal didapatkan hasil akhir sebanyak 9 atribut terseleksi yang memiliki relevansi atau kontribusi terhadap variabel output (waktu kelulusan) yaitu : jenis kelamin, kelas, umur, SKS1, IPS1, SKS2, SKS3, SKS4 dan SKS5, sedangkan atribut input yang kurang relevan terdapat 4 atribut yaitu : IPS2, IPS3, IPS4 dan IPS5, sehingga untuk melakukan klasifikasi waktu kelulusan Mahasiswa cukup menggunakan 9 atribut yang terseleksi agar bisa mendapatkan hasil akurasi yang maksimal, proses klasifikasi akan dilakukan pada penelitian selanjutnya.

Kata Kunci— *Seleksi Fitur; Correlation Matrix; Heatmap; Python*

Abstract — *Classification of graduation time Students usually use more than 10 input attributes, but not all of these attributes have relevance to output attributes, so before classification is carried out, it is necessary to select relevant attributes through feature selection first. Feature selection is part of the data preparation stages in the data science lifecycle, Feature selection is one of the core concepts in machine learning that greatly affects the performance of data mining models. Where at this stage a selection of features that contribute the most to the prediction or*

output variables is carried out, choosing irrelevant features in the dataset can reduce the accuracy of the model and make the model learn based on irrelevant features. Thus, selecting features before modeling data will get benefits such as: reducing Overfitting, increasing accuracy and reducing training time. The technique used in the feature selection in this study is correlation matrix with heatmap because this technique is easy to use and also gives good results. After the implementation of feature selection in the classification of student graduation time using 13 variables / initial attributes, the final results of 9 selected attributes were obtained that had relevance or contribution to the output variables (graduation time) namely: gender, class, age, SKS1, IPS1, SKS2, SKS3, SKS4 and SKS5, while the less relevant input attributes were 4 attributes, namely: IPS2, IPS3, IPS4 and IPS5, so that to classify the graduation time Students only need to use the 9 attributes selected in order to get maximum accuracy results, the classification process will be carried out in the next study.

Keywords— *Feature selection; Correlation Matrix; Heatmap; Python*

I. PENDAHULUAN

Tahapan *data preparation* merupakan tahapan kedua dalam *Knowledge Discovery and Data Mining* (KDD) atau tahapan ketiga dalam *Cross Industry Standard Process Data Mining* (CRISP-DM). KDD dan CRISP-DM merupakan metodologi *data science* yang banyak digunakan dalam pengembangan data mining. Tahapan *data preparation* atau istilah lain *data pre-processing/data manipulation/ data cleansing* merupakan tahapan yang paling banyak memakan waktu sekitar 60%-80% [1] dan langkah paling penting dalam seluruh siklus hidup dalam perancangan model data mining. Adapun jenis kegiatan pada tahap *data preparation* ini diantaranya memilih data yang relevan, pembersihan data, seleksi fitur, ekstraksi fitur dan lain-lain, dengan melakukan tahapan ini, maka akan diperoleh beberapa manfaat diantaranya kompilasi data menjadi efisien dan efektif atau menghindari duplikasi, identifikasi dan memperbaiki *error*, mudah perubahan secara global dan menghasilkan informasi yang akurat untuk pengambilan keputusan [1].

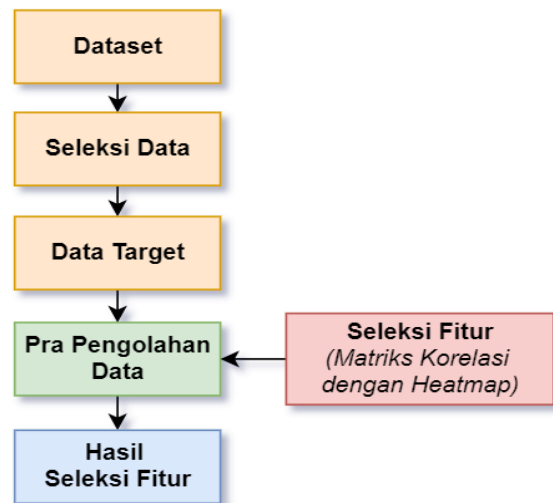
Diantara *data preparation* yang disebutkan di atas, pada kasus ini difokuskan pada seleksi fitur, karena seleksi/pemilihan fitur merupakan salah satu konsep inti dalam pembelajaran mesin yang sangat mempengaruhi kinerja model yang sedang dibangun. Seleksi fitur merupakan proses yang dilakukan secara otomatis atau manual memilih fitur-fitur yang berkontribusi paling besar pada variabel prediksi atau output, fitur yang tidak relevan atau hanya sebagian relevan dapat berdampak negatif terhadap kinerja model[2]. Dengan menggunakan seleksi fitur, maka resiko *overfitting* dapat dihindari, akurasi menjadi lebih baik karena menghilangkan data redundansi dan fitur-fitur yang tidak signifikan atau relevan, sehingga dapat menghemat waktu yang dibutuhkan dalam proses klasifikasi karena berkurangnya dimensi data. Hal ini sudah dibuktikan oleh [2] dalam penelitiannya dimana sebelum dilakukan seleksi fitur didapatkan akurasi hanya 65%, namun setelah dilakukan seleksi fitur akurasinya meningkat menjadi 81%, begitu juga penelitian yang dilakukan oleh[3] dengan melakukan seleksi fitur dapat meningkatkan hasil akurasi menjadi 90.81%. selain seleksi fitur dapat juga dilakukan ekstraksi fitur seperti yang sudah dilakukan oleh[4], namun pada penelitian ini hanya membahas seleksi fitur.

Seleksi fitur jika dilakukan secara manual maka akan memakan waktu yang lama dan hasilnya sering kali tidak akurat atau malah ada atribut penting yang dihilangkan, secara teori untuk data dengan n atribut akan terdapat sekitar 2^n kemungkinan subhimpunan atribut, misalnya memiliki data dengan 10 atribut, maka harus memilih satu dari 1,024 kemungkinan subhimpunan atribut[5][6], untuk itu diperlukan Teknik seleksi fitur secara otomatis diantaranya (a). *Univariate Selection*, (b). *Feature Importance*, (c). *Correlation Matrix with Heatmap*[2][7], dan lainnya. Untuk kasus ini akan pilih teknik *Correlation Matrix with Heatmap* karena teknik ini mudah digunakan dan juga memberikan hasil yang baik[2].

Teknik *Correlation Matrix with Heatmap* akan diimplementasikan pada seleksi fitur klasifikasi waktu kelulusan Mahasiswa. Beberapa penelitian yang sudah mengangkat kasus ini namun belum menggunakan teknik seleksi fitur diantaranya penelitian yang dilakukan oleh [8] dengan atribut yang digunakan yaitu : jenis kelamin, asal sekolah, jalur masuk, nilai ujian nasional, gaji orang tua, IP semester 1-4, IPK semester 1-4 dan keterangan lulus, penelitian berikutnya yang dilakukan oleh[9] dengan atribut yang digunakan yaitu : jenis kelamin, usia, jurusan, kelas, pekerjaan, IPS 1, IPS 2, IPS 3, IPS 4 dan keterangan karena peneliti kesulitan dalam mendapatkan dataset dari kedua penelitian tersebut di atas maka akan digunakan dataset yang dikumpulkan sendiri dari Universitas Ichsan Gorontalo dengan menggunakan atribut kurang lebih sama dengan atribut yang digunakan peneliti sebelumnya diantaranya : jenis kelamin, kelas, umur, SKS semester 1, IPS semester 1, SKS semester 2, IPS semester 2, SKS Semester 3, IPS semester 3, SKS semester 4, IPS semester 4, SKS semester 5, IPS semester 5 dan status kelulusan[10].

II. METODE

Tahapan penelitian yang digunakan diuraikan pada gambar 1 berikut :



Gambar 1. Tahapan Penelitian

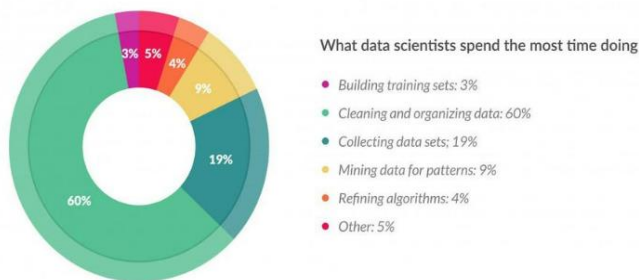
1. Dataset : tahap pertama yang dilakukan adalah mengumpulkan dataset yang terdiri dari data Mahasiswa, aktivitas kuliah dan data alumni pada Universitas Ichsan Gorontalo.
2. Seleksi Data : pada tahap ini dilakukan seleksi data dari dataset yang telah dikumpulkan dengan memilih atribut awal yang akan digunakan sebanyak 13 atribut yaitu : jenis kelamin, kelas, umur, SKS semester 1, IPS semester 1, SKS semester 2, IPS semester 2, SKS Semester 3, IPS semester 3, SKS semester 4, IPS semester 4, SKS semester 5, IPS semester 5 dan status kelulusan.
3. Data Target : pada tahap ini dilakukan penggabungan data berdasarkan atribut yang sudah dipilih dari setiap dataset yang sudah dikumpulkan yaitu dari data Mahasiswa, aktivitas kuliah dan data alumni.
4. Pra Pengolahan Data : pada tahap ini dilakukan pra pengolahan data berdasarkan data target yang telah didapatkan dari proses sebelumnya dengan beberapa tahapan yaitu :
 - a. Pengecekan data kosong untuk setiap atribut
 - b. Pengecekan data yang tidak sesuai misalnya data SKS yang lebih besar dari 24 SKS dan IPS yang lebih dari nilai 4
 - c. Mengkonversi nilai atribut yang berisi non numerik ke numerik agar dapat dilakukan seleksi fitur dengan *Correlation Matrix with Heatmap* karena syaratnya semua atribut harus berjenis numerik.
 - d. Melakukan seleksi fitur dengan Teknik *Correlation Matrix with Heatmap*.
5. Hasil Seleksi Fitur : pada tahap ini akan didapatkan matriks korelasi dengan peta *Heatmap* yang menunjukkan hubungan korelasi untuk setiap atribut, sehingga didapatkan atribut yang berpengaruh atau relevan terhadap atribut output (status kelulusan). Atribut yang relevan akan digunakan untuk proses selanjutnya yaitu proses klasifikasi.

A. Data Preparation

Data Preparation merupakan tahapan ke 2 atau ke 3 dalam siklus hidup dalam pengembangan *Data Science*. Pada tahapan ini dilakukan beberapa kegiatan diantaranya memilih data yang relevan, mengintegrasikan data dengan

menggabungkan kumpulan data, membersihkannya, melengkapi nilai yang hilang dengan menghapusnya atau menguraikannya, memeriksa data outlier, mengubah format data ke dalam struktur yang diinginkan, seleksi fitur dan lainnya. *Data Preparation* adalah tahapan yang paling memakan waktu serta langkah paling penting dalam seluruh siklus hidup dalam data mining[11].

Sebuah survey yang dilakukan oleh CrowdFlower menyatakan bahwa sekitar 80% waktu yang digunakan untuk melakukan tahapan *Data Preparation*[12][1]. Hasil survey dapat dilihat pada gambar 2 berikut ini :



Gambar 2. Hasil Survey Riset *Data Scientists*[12]

B. Seleksi Fitur

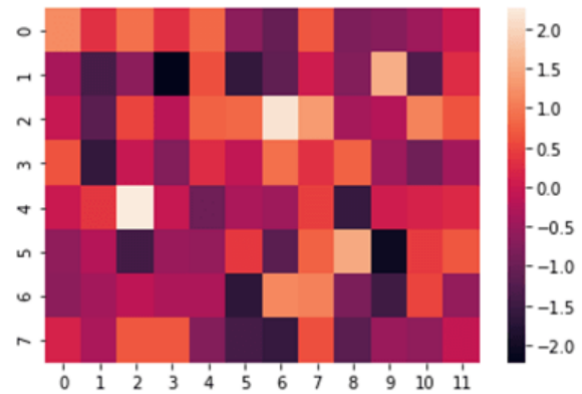
Seleksi Fitur adalah salah satu konsep inti dalam pembelajaran mesin yang sangat mempengaruhi kinerja model data mining. Fitur data yang digunakan untuk melatih model pembelajaran mesin yang dibangun memiliki pengaruh besar pada kinerja yang akan dicapai[7].

Seleksi Fitur adalah proses dimana seleksi dilakukan secara otomatis atau manual memilih fitur-fitur yang berkontribusi paling besar pada variabel prediksi atau output. Memilih fitur yang tidak relevan dalam dataset dapat mengurangi keakuratan model dan membuat model belajar berdasarkan fitur yang tidak relevan. Manfaat melakukan seleksi fitur sebelum memodelkan data yaitu : mengurangi *Overfitting*, meningkatkan akurasi dan mengurangi waktu pelatihan[2].

C. Correlation Matrix with Heatmap

Heatmap adalah metode visualisasi dua dimensi yang menunjukkan variasi dalam besarnya fenomena tertentu dalam hal warna yang diwakili warna berbeda. *Heatmap* (peta panas) adalah teknik visualisasi data yang menunjukkan bentuk dan arah nilai panas yang berbeda pada tingkat suhu yang berbeda untuk satu set titik data[13]. Korelasi bernilai positif artinya terjadi peningkatan satu nilai fitur meningkatkan nilai variabel target atau bernilai negatif artinya peningkatan satu nilai fitur menurunkan nilai variabel target[2].

Heatmap memudahkan untuk mengidentifikasi fitur mana yang paling terkait dengan variabel target, berikut contoh *Correlation Matrix with Heatmap* menggunakan *library seaborn* pada gambar 3 berikut :



Gambar 3. Contoh Visualisasi *Heatmap*[13]

III. HASIL DAN PEMBAHASAN

Proses *seleksi Fitur* dilakukan sesuai dengan tahapan yang sudah diuraikan di atas, mulai tahapan awal sampai dengan tahapan seleksi fitur menggunakan *tools Colaboratory Notebook* dengan bahasa pemrograman *Python* menggunakan beberapa *library* seperti *Numpy*, *Pandas*, *Matplotlib* dan *Seaborn*[14][15]. Berikut tahapan yang dilakukan dalam seleksi fitur pada klasifikasi waktu kelulusan Mahasiswa.

A. Pengumpulan Dataset

Pada tahapan ini dilakukan seleksi dataset yang akan digunakan yaitu diambil dari data Mahasiswa, data aktivitas kuliah per semester dan data alumni pada laman <https://siakun.unisan.ac.id/> data diambil secara acak dari 11 Program Studi dengan jumlah dataset sebanyak 1.095 record[10].

B. Seleksi Data

Pada tahap ini dilakukan seleksi data dari dataset yang sudah dikumpulkan yaitu memilih atribut yang akan digunakan dari tiga file dataset yang digunakan, setelah dilakukan serangkaian pengolahan data, maka didapatkan atribut sebanyak 14 atribut, berikut contoh sampel hasil seleksi data seperti pada tabel 1 berikut ini :

TABEL 1. SAMPLE HASIL SELEKSI DATASET

JK	Kelas	Umur	SKS1	IPS1	SKS2	IPS2	SKS3
P	N	18	22	3,64	19	3,84	22
P	N	21	22	3,5	19	3,89	22
P	N	18	22	3,36	19	3,84	22
P	R	21	22	3,64	19	3,74	22
L	N	23	22	4	19	3,84	22
P	R	19	22	3,45	19	2,11	18
P	R	19	22	3,64	19	3,74	22
P	R	18	22	3,36	19	3,53	22
P	R	18	22	3,73	19	3,74	22
L	R	17	22	3,59	19	2,84	20

IPS3	SKS4	IPS4	SKS5	IPS5	Status
------	------	------	------	------	--------

3,77	24	3,25	24	3,75	Tepat Waktu
3,36	24	3,38	24	3,25	Tepat Waktu
3,36	24	3,38	24	3,13	Tepat Waktu
3,41	24	3,63	24	3,5	Tepat Waktu
4	24	3,63	24	3,88	Tepat Waktu
2,77	18	2,35	18	2,75	Tidak Tepat
3	24	3,13	24	3,13	Tepat Waktu
3,45	24	3,13	24	3,25	Tepat Waktu
3,59	24	3,75	24	3,13	Tepat Waktu
3,55	24	2,88	20	3,4	Tepat Waktu

Atribut	Nilai Atribut	Nilai Konversi
	Tidak Tepat	2

Setelah dilakukan konversi nilai, maka didapatkan dataset seperti pada tabel 4 berikut :

TABEL 4. SAMPLE DATASET AKHIR

JK	Kelas	Umur	SKS1	IPS1	SKS2	IPS2	SKS3
2	2	18	22	3,64	19	3,84	22
2	2	21	22	3,5	19	3,89	22
2	2	18	22	3,36	19	3,84	22
2	1	21	22	3,64	19	3,74	22
2	2	23	22	4	19	3,84	22
1	1	19	22	3,45	19	2,11	22
2	1	19	22	3,64	19	3,74	22
2	1	18	22	3,36	19	3,53	22
2	1	18	22	3,73	19	3,74	22
1	1	17	22	3,59	19	2,84	20

C. Penentuan Data Target

Pada tahap ini dilakukan penentuan data target yaitu menentukan atribut input dan atribut output yang diuraikan pada tabel 2 berikut ini :

TABEL 2. ATRIBUT INPUT DAN OUTPUT

No	Nama Atribut	Keterangan	Jenis Atribut
1	JK	Jenis Kelamin	Input
2	Kelas	Kelas	Input
3	Umur	Umur	Input
4	SKS1	SKS Semester 1	Input
5	IPS1	Indeks Prestasi Semester 1	Input
6	SKS2	SKS Semester 2	Input
7	IPS2	Indeks Prestasi Semester 2	Input
8	SKS3	SKS Semester 3	Input
9	IPS3	Indeks Prestasi Semester 3	Input
10	SKS4	SKS Semester 4	Input
11	IPS4	Indeks Prestasi Semester 4	Input
12	SKS5	SKS Semester 5	Input
13	IPS5	Indeks Prestasi Semester 5	Input
14	Status	Status Kelulusan	Output

IPS3	SKS4	IPS4	SKS5	IPS5	Status
3,77	24	3,25	24	3,75	1
3,36	24	3,38	24	3,25	1
3,36	24	3,38	24	3,13	1
3,41	24	3,63	24	3,5	1
4	24	3,63	24	3,88	1
2,77	18	2,35	18	2,75	2
3	24	3,13	24	3,13	1
3,45	24	3,13	24	3,25	1
3,59	24	3,75	24	3,13	1
3,55	24	2,88	20	3,4	1

3. Melakukan seleksi fitur menggunakan Teknik *Correlation Matrix with Heatmap* dengan tahapan sebagai berikut :

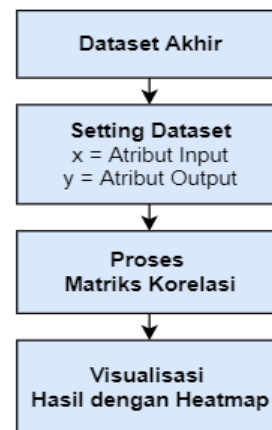
D. Data Preparation

Pada tahap ini dilakukan serangkaian proses untuk mendapatkan dataset yang bisa digunakan dalam seleksi fitur dengan tahapan sebagai berikut :

- Melakukan pembersihan data, pengecekan data kosong, data yang tidak sesuai dan lainnya
- Melakukan konversi nilai atribut untuk atribut non numerik ke numerik agar bisa diproses pada seleksi fitur seperti pada tabel 3 di bawah ini :

TABEL 3. KONVERSI NILAI ATRIBUT

Atribut	Nilai Atribut	Nilai Konversi
JK	L	1
	P	2
Kelas	R	1
	N	2
Status	Tepat Waktu	1



Gambar 4. Tahapan Proses Matriks Korelasi

- Berdasarkan dataset akhir yang telah diolah seperti pada tabel 4 di atas, selanjutnya dilakukan setting dataset yaitu atribut input dimasukkan ke dalam

variabel x dan atribut output dimasukkan ke dalam variabel y

- 2) Proses perbandingan matriks korelasi antara setiap atribut input dengan atribut output
- 3) Visualisasi hasil proses matriks korelasi dengan *Heatmap*.

Berikut ini perintah untuk melakukan seleksi fitur dengan menggunakan bahasa pemrograman Python :

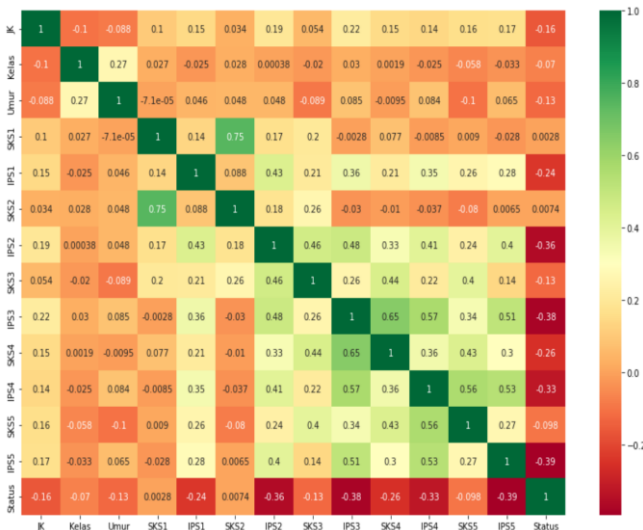
```
#setting dataset
X = data[['JK','Kelas','Umur','SKS1','IPS1','SKS2','IPS2',
'SKS3','IPS3','SKS4','IPS4','SKS5','IPS5']].values #variabel input
y = data[['Status']].values #variabel target/output

#proses perbandingan matriks korelasi untuk setiap atribut
corrmat = data.corr()
top_corr_features = corrmat.index
plt.figure(figsize=(15,10))

#visualisasi dengan heatmap
g=sns.heatmap(data[top_corr_features].corr(),annot=True,cmap="RdYlGn")
```

E. Hasil Seleksi Fitur

Pada tahap ini dilakukan analisis terhadap atribut yang berkaitan atau relevan dengan atribut output dengan melihat visualisasi *Heatmap* pada gambar 5 berikut ini :



Gambar 5. Visualisasi Hasil *Correlation Matrix with Heatmap*

Pada gambar 5 di atas dapat dianalisa bahwa atribut yang memiliki warna merah muda adalah atribut yang relavan dengan atribut aoutput dan sebaliknya atribut yang berwarna merah tua adalah atribut yang kurang relevan dengan atribut aoutput. Hasil analisis diuraikan pada tabel 5 berikut :

TABEL 5. HASIL ANALISIS ATRIBUT

No	Nama Atribut Input	Ket
1	JK	Relevan
2	Kelas	Relevan
3	Umur	Relevan
4	SKS1	Relevan
5	IPS1	Relevan
6	SKS2	Relevan

No	Nama Atribut Input	Ket
7	IPS2	Kurang Relevan
8	SKS3	Relevan
9	IPS3	Kurang Relevan
10	SKS4	Relevan
11	IPS4	Kurang Relevan
12	SKS5	Relevan
13	IPS5	Kurang Relevan

Berdasarkan tabel 5 di atas, dapat dilihat bahwa dari 13 atribut input hanya 9 atribut yang ada relevansi dengan atribut output, sisanya 4 atribut kurang relevan, sehingga untuk melakukan klasifikasi waktu kelulusan Mahasiswa pada Universitas Ichsan Gorontalo cukup menggunakan 9 atribut input yaitu : JK (jenis kelamin), kelas, umur, SKS1, IPS1, SKS2, SKS3, SKS4 dan SKS5.

Implementasi seleksi fitur menggunakan *Correlation Matrix with Heatmap* sangat ditentukan dari jumlah dataset yang digunakan dan nilai setiap atribut input, sehingga atribut yang terseleksi pada penelitian ini bisa saja berbeda dengan penelitian lain atau di kampus lain.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah diuraikan di atas, maka hasil seleksi fitur menggunakan teknik *Correlation Matrix with Heatmap* pada klasifikasi waktu kelulusan Mahasiswa dari 13 atribut awal yang digunakan terseleksi menjadi 9 atribut yang memiliki relevansi atau berkontribusi terhadap hasil output yaitu : jenis kelamin, kelas, umur, SKS1, IPS1, SKS2, SKS3, SKS4 dan SKS5, sedangkan atribut input yang kurang relevan terdapat 4 atribut yaitu : IPS2, IPS3, IPS4 dan IPS5, sehingga untuk melakukan klasifikasi waktu kelulusan Mahasiswa cukup menggunakan 9 atribut yang terseleksi agar bisa mendapatkan hasil akurasi yang maksimal, khususnya di Universitas Ichsan Gorontalo. Proses klasifikasi akan dilakukan pada penelitian selanjutnya.

REFERENSI

- [1] K. Walch, "Data Preparation for Machine Learning Still Requires Humans," 2019. <https://www.techtarget.com/searchenterpriseai/feature/Data-preparation-for-machine-learning-still-requires-humans> (accessed May 21, 2022).
- [2] R. Shaikh, "Feature Selection Techniques in Machine Learning with Python," 2018. <https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e> (accessed May 21, 2022).
- [3] Jupriyadi, "Implementasi Seleksi Fitur Menggunakan Algoritma FVBRM Untuk Klasifikasi Serangan Pada Intrusion Detection System (IDS)," *Semmastek*, pp. 1–6, 2018.
- [4] Z. Y. Lamasigi, "DCT Untuk Ekstraksi Fitur Berbasis GLCM Pada Identifikasi Batik Menggunakan K-NN," *JJEE*, vol. 3, pp. 1–6, 2021.
- [5] Suyanto, *Data Mining untuk Klasifikasi dan Klusterisasi Data*. Bandung: Informatika, 2019.
- [6] Suyanto, *Machine Learning Tingkat Dasar dan Lanjut*. Bandung: Informatika, 2018.
- [7] P. Pedamkar, "Machine Learning Feature Selection," 2019. <https://www.educba.com/machine-learning-feature-selection/> (accessed May 20, 2022).
- [8] E. P. Rohmawan, "Prediksi Kelulusan Mahasiswa Tepat Waktu

- Menggunakan Metode Decision Tree Dan Artificial Neural Network,” *MATRIK*, vol. 20, no. 1, pp. 21–30, 2018.
- [9] A. Budiyantra and Irwansyah, “Prediksi Mahasiswa Lulus Tepat Waktu Menggunakan Algoritma Decision Tree (C4.5) Pada STMIK Widuri Jakarta,” *Infotech*, vol. 5, no. 2, pp. 7–13, 2019.
- [10] BAAK UNISAN, “Data Aktivitas Mahasiswa,” Gorontalo, 2022. [Online]. Available: <https://siakun.unisan.ac.id/>.
- [11] P. Pedamkar, “Data Science Lifecycle,” *EDUCBA. All Rights Reserved*. <https://www.educba.com/data-science-lifecycle/> (accessed May 15, 2022).
- [12] CrowdFlower, “Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says,” 2016. <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/?sh=5df53b386f63> (accessed May 15, 2022).
- [13] Educba, “Seaborn heatmap.” <https://www.educba.com/seaborn-heatmap> (accessed May 20, 2022).
- [14] T. Wahyono, *Fundamental of Python for Machine Learning*. Yogyakarta: Gava Media, 2018.
- [15] Y. Heryadi and T. Wahyono, *Machine Learning Konsep dan Implementasi*. Yogyakarta: Gava Media, 2020.