

Perbandingan Metode Klasifikasi Data Mining Untuk Mengukur Tingkat Kepuasan Mahasiswa Terhadap Sistem Informasi Penilaian Nonakademik UNESA (SIPENA)

Ananda Rizky Abidin¹, I Kadek Dwi Nuryana²

^{1,2} Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya

¹ananda.19035@mhs.unesa.ac.id

²dwinuryana@unesa.ac.id

Abstrak— Sistem Informasi Penilaian NonAkademik UNESA (SIPENA) adalah suatu sistem yang didesain untuk sarana atau fasilitas yang mendukung dalam proses penilaian nonakademik mahasiswa. Dalam upaya mengoptimalkan kepuasan mahasiswa terhadap SIPENA, dibutuhkan sebuah model dengan menggunakan metode dari *machine learning* dengan membandingkan algoritma terbaik untuk klasifikasi data kepuasan mahasiswa pengguna SIPENA. Pada penelitian ini, peneliti membuat dan menyebarkan kuesioner kepuasan pengguna kepada mahasiswa Universitas Negeri Surabaya berdasar dengan variabel dan indikator yang dibutuhkan yaitu, kualitas sistem, kualitas informasi, kegunaan yang dirasakan dan kepuasan pengguna itu sendiri. Hasil dari kuesioner tersebut diolah dengan perangkat lunak SPSS untuk uji validitas dan reliabilitas. Setelah didapatkan hasil yang valid dan reliabel peneliti melanjutkan mengolah dataset kepuasan pengguna SIPENA pada *tools* Jupyter Notebook dengan *library* PyCaret untuk dilakukan klasifikasi dan perbandingan. Hasil penelitian ini menunjukkan bahwa dari 16 model algoritma klasifikasi pada *library* PyCaret yang dibandingkan pada dataset kepuasan pengguna SIPENA, model dari algoritma *Extra Trees Classifier* adalah yang terbaik dengan nilai akurasi 0.9743, yang kedua adalah algoritma *Logistic Regression* dengan nilai akurasi 0.9714, dan yang ketiga adalah algoritma *Random Forest Classifier* dengan nilai akurasi 0.9657.

Kata Kunci— Perbandingan, Klasifikasi, Data Mining, PyCaret, Kepuasan, Akurasi

I. PENDAHULUAN

Perkembangan teknologi informasi dapat memberikan dampak yang luar biasa dalam mengubah kinerja sebuah sistem menjadi lebih efektif dan efisien. Suatu sistem harus diketahui apakah komponen dari sistem informasi tersebut sudah berjalan dengan baik atau membutuhkan evaluasi untuk memperbaiki agar sistem informasi tersebut dapat lebih berkembang menjadi lebih baik [1]. Sistem informasi juga memiliki peran yang penting pada bidang pendidikan dalam mempermudah segala kegiatan baik itu akademik dan nonakademik.

Berdasarkan *user guide* sipena untuk mahasiswa Sistem Informasi Penilaian Non Akademik UNESA adalah suatu sistem yang didesain sebagai sarana atau fasilitas yang mendukung dalam proses penilaian non akademik mahasiswa. Sehingga Universitas Negeri Surabaya dapat menyediakan layanan informasi yang lebih baik dan efektif untuk proses penilaian nonakademik sebagai prasyarat yudisium dan Surat

Keterangan Pendamping Ijasah (SKPI). SIPENA menyediakan beberapa layanan, seperti tambah KRP (kartu rencana prestasi), cetak KRP dan KHP (Kartu Hasil Prestasi), TKM (Transkrip Kegiatan Mahasiswa) serta dilengkapi dengan menu panduan yang berisi dokumen pedoman pelaksanaan penilaian nonakademik Universitas Negeri Surabaya.

Evaluasi sistem dibutuhkan dalam meningkatkan kinerja sistem agar lebih baik. Tingkat kualitas sistem dapat mempengaruhi kepuasan pengguna terhadap sistem informasi yang digunakan. Evaluasi merupakan sebuah proses yang menghasilkan informasi mengenai kemajuan suatu kegiatan yang telah diraih untuk mengetahui ada atau tidaknya selisih antara bentuk manfaat yang telah diselesaikan dengan harapan yang diinginkan [1].

Kepuasan pengguna menggambarkan keselarasan antara harapan seseorang dan hasil yang diperoleh dengan adanya suatu sistem dimana tempat orang tersebut berpartisipasi dalam pengembangan sistem informasi. Kepuasan pengguna sistem informasi merupakan tolak ukur tingkat keberhasilan sebuah sistem informasi, kepuasan tersebut merupakan sebuah penilaian yang menyangkut apakah kinerja suatu sistem informasi terbilang baik atau buruk, dan apakah sistem informasi yang digunakan cocok atau tidak dengan tujuan penggunaannya [2]. Kepuasan pengguna dapat dipengaruhi oleh kualitas sistem dan kualitas informasi yang dihasilkan. Beberapa faktor yang berkaitan dengan kepuasan pengguna sistem informasi yaitu kualitas sistem informasi, kualitas informasi, dan *perceived usefulness* [3].

Dalam upaya mengoptimalkan kepuasan mahasiswa terhadap SIPENA, dibutuhkan sebuah model dengan menggunakan metode dari *machine learning* dengan membandingkan algoritma terbaik untuk klasifikasi data kepuasan mahasiswa pengguna sistem penilaian nonakademik unesa. Teknik data mining, atau yang lebih umum dikenal sebagai teknik *Machine Learning*, adalah metode yang digunakan untuk meramalkan situasi berdasarkan pemodelan terhadap data [4]. Pada penelitian ini menggunakan salah satu jenis *supervised learning* yakni metode klasifikasi.

Pada penelitian yang dilakukan [5] menghasilkan Algoritma *Decision Tree C4.5* memiliki performa terbaik dengan akurasi 88% dan presisi 98% dibandingkan algoritma *Naïve Bayes* dan *Support Vector Machine* ketika dibuat model dari data latih dan diuji menggunakan data uji dari dataset kuesioner kepuasan

mahasiswa terhadap sarana dan prasana pembelajaran. Kesimpulan dari penelitian fatt[6] menghasil nilai akurasi algoritma *C45* sebesar sebesar 96,50%, sedangkan untuk nilai akurasi menggunakan Algoritma *Support Vector Machine* sebesar 89,66%, Akurasi menggunakan Algoritma *Naïve Bayes* sebesar 89,88% dan nilai akurasi dengan metode Optimasi *Naïve Bayes* dengan pemilihan fitur *Particle Swarm Optimization* sebesar 95,85%.

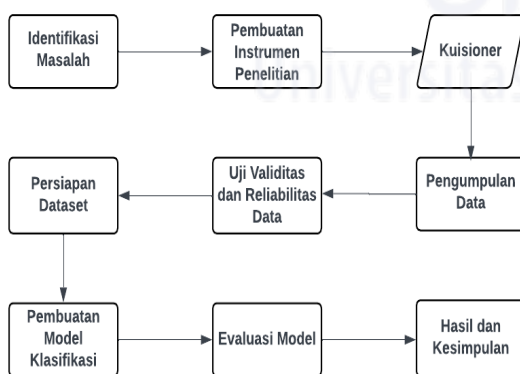
Dalam proses untuk mendapatkan model terbaik, penelitian ini menggunakan *library* PyCaret. *Library* PyCaret adalah kumpulan *library open source* yang digunakan untuk membangun model machine learning dengan mudah. *Library* PyCaret berjalan di bahasa pemrograman Python dan menyediakan *tools* untuk *preprocessing* data, membangun model, membandingkan model, mengevaluasi, menampilkan visualisasi hasil training sampai menyimpan model dalam format tertentu [7]. PyCaret memungkinkan untuk menguji banyak model secara mudah dan cepat. Dengan menggunakan metode tersebut dapat memprediksi apa yang dibutuhkan berdasarkan kepuasan mahasiswa yang menggunakan SIPENA pada Universitas Negeri Surabaya sehingga akan meningkat kualitas layanan dan meningkatnya kepuasan yang dirasakan mahasiswa.

Berdasarkan latar belakang tersebut, penelitian ini akan membuat sebuah perbandingan hasil dari metode klasifikasi Data Mining dengan menggunakan PyCaret untuk komparasi atau perbandingan algoritma terbaik dan melakukan pemilihan model terbaik dalam mengukur tingkat kepuasan mahasiswa terhadap sistem informasi penilaian nonakademik unesa (SIPENA).

II. METODE PENELITIAN

A. Jenis dan Rancangan Penelitian

Penelitian ini adalah penelitian kuantitatif dengan berfokus untuk pengukuran data statistika hasil kuesioner untuk menjawab rumusan masalah yang ada. Dengan tahapan penelitian pada (Gbr. 1).



Gbr. 1 Tahapan Penelitian

B. Populasi dan Sampel

Populasi pada penelitian ini adalah mahasiswa UNESA yang menggunakan SIPENA. Dengan fokus dari penelitian ini yaitu SIPENA dari sisi pengguna mahasiswa. Sehingga populasi dari penelitian ini adalah seluruh mahasiswa UNESA sebagai pengguna SIPENA. Penentuan sampel dalam penelitian ini dilakukan dengan menggunakan metode purposive sampling yaitu pemilihan sampel yang dipilih sesuai dengan dasar kriteria yang dinilai dari peneliti [8]. Kriteria yang digunakan dalam penelitian ini adalah mahasiswa UNESA yang berkaitan dengan SIPENA sebagai sistem yang digunakan. Sehingga berdasarkan penelitian terkait [6], untuk jumlah sampel yang dibutuhkan dalam penelitian ini kurang lebih adalah sebanyak 500 sampel.

C. Variabel dan Definisi Oprasional

TABEL I
 VARIABEL DAN DEVINISI OPRASIONAL

| Variabel | Indikator | Definisi Oprasional |
|-----------------------------|---------------------------|--|
| Kualitas Sistem | Kecepatan Akses | Waktu yang diperlukan sistem dalam merespon perintah <i>user</i> . |
| | Kehandalan Sistem | Sistem sudah beroperasi sesuai dengan fungsi yang diharapkan |
| | Kemudahan untuk Digunakan | Kemudahan pengguna dalam memahami sistem. |
| | Kemudahan untuk Diakses | Kemudahan pengguna dalam mengakses sistem |
| | Keamanan Sistem | Keamanan data yang ada pada sistem. |
| Kualitas Informasi | Akurasi | Tingkat ketelitian akurasi dalam proses komputasi sistem. |
| | Ketepatan Waktu | Menyajikan suatu informasi yang relevan dan tepat waktu |
| | Kelengkapan | Seberapa lengkap informasi atau fungsi yang ada pada sistem |
| | Relevansi | Kesesuaian informasi yang dihasilkan terhadap kebutuhan pengguna. |
| | Konsistensi | Keseragaman atau konsistensi dalam desain dan teknik dokumentasi yang ada pada sistem. |
| <i>Perceived Usefulness</i> | Berguna | Seberapa berguna sistem yang berjalan bagi pengguna. |

| | | |
|-------------------|--------------------------|---|
| | Bermanfaat | Seberapa bermanfaat sistem yang berjalan bagi pengguna. |
| | Efektifitas | Seberapa efektif sistem yang berjalan bagi pengguna. |
| | Produktifitas | Seberapa produktif sistem yang berjalan bagi pengguna. |
| Kepuasan Pengguna | <i>Perceived Value</i> | Kesesuaian antara nilai keuntungan yang didapatkan dan kualitas pelayanan yang diberikan. |
| | <i>User Expectation</i> | Perasaan yang timbul pada pengguna setelah harapan terhadap layanan terpenuhi atau tidak. |
| | <i>Perceived Quality</i> | Keandalan kualitas layanan dalam memberikan pelayanan yang disesuaikan dengan kebutuhan pengguna. |

D. Instrumen Penelitian

Instrumen penelitian adalah sebuah alat untuk mengumpulkan data atau mengukur objek dari suatu penelitian. Instrumen yang valid dan konsisten dalam memberikan hasil penelitian akan mendapatkan data yang sesuai dengan keadaan sebenarnya. Instrumen penelitian yang digunakan yaitu berupa kuesioner terkait dengan penelitian yang dilakukan.

Kuesioner pada penelitian ini menggunakan lima poin skala likert. Dengan menggunakan lima poin skala likert, dapat membuat indikator lebih rinci dalam mengukur variabelnya. Skala likert yang digunakan penelitian ini ditunjukkan pada Tabel II.

TABEL II
SKALA LIKERT

| No. | Simbol | Keterangan | Nilai |
|-----|--------|---------------------|-------|
| 1. | STS | Sangat Tidak Setuju | 1 |
| 2. | TS | Tidak Setuju | 2 |
| 3. | CS | Cukup Setuju | 3 |
| 4. | S | Setuju | 4 |
| 5. | SS | Sangat Setuju | 5 |

E. Teknik Pengumpulan Data

Penulis melakukan pengumpulan data melalui penyebaran kuesioner yang dibuat menggunakan Google Form ke media sosial seperti WhatsApp, dan Instagram mahasiswa Universitas Negeri Surabaya.

F. Teknik Analisis Data

Pada analisis data untuk mengolah data hasil dari kuesioner responden dalam uji validitas dan reliabilitas menggunakan

software SPSS dan menggunakan library PyCaret untuk mengolah data yang sudah valid dan reliabel untuk mendapatkan hasil komparasi atau perbandingan metode klasifikasi *Data Mining* terbaik untuk mengukur tingkat kepuasan mahasiswa pengguna SIPENA.

III. HASIL DAN PEMBAHASAN

A. Kuesioner Penelitian

Pada penelitian ini kuesioner yang telah dibuat menggunakan Google Form dengan pertanyaan yang diadopsi dari penelitian-penelitian terdahulu yang terkait dengan variabel dan indikator yang disesuaikan dengan kebutuhan dari kualitas sistem, kualitas informasi, *perceived usefulness* dan kepuasan pengguna.

1) Variabel Kualitas Sistem

Kualitas sistem dari SIPENA pada penelitian ini diukur dari sisi kecepatan akses (*response time*), kehandalan sistem (*system reliability*), kemudahan untuk digunakan (*ease to use*), kemudahan untuk diakses (*system flexibility*), dan keamanan sistem (*system security*). Untuk pertanyaan variabel kualitas sistem ada pada Tabel II.

TABEL III
KUESIONER VARIABEL KUALITAS SISTEM

| No. | Pertanyaan | Sumber |
|-----|---|---------------------------------|
| 1 | SIPENA mudah diakses oleh pengguna. | (Mangun Buana & Wirawati, 2018) |
| 2 | SIPENA dapat mengoperasikan sejumlah perintah dalam waktu yang relatif singkat, tanpa mengalami hambatan. | |
| 3 | SIPENA dapat dengan cepat merespon suatu perintah pembatalan maupun permintaan. | |
| 4 | Menu dan navigasi yang disediakan SIPENA dapat dijalankan dengan mudah dan interaktif. | |
| 5 | SIPENA memiliki batasan hak akses yang berbeda antara mahasiswa dan dosen | |
| 6 | Data yang ada pada SIPENA terjamin kerahasiaannya | |

2) Variabel Kualitas Informasi

Untuk menilai kualitas informasi SIPENA pada penelitian ini menggunakan lima dimensi yaitu akurasi (*accuracy*), ketepatan waktu (*timeliness*), kelengkapan (*completeness*),

relevansi (*relevance*), dan konsistensi (*consistency*). Untuk pertanyaan variabel kualitas informasi ada pada Tabel IV.

TABEL IV
KUESIONER VARIABEL KUALITAS INFORMASI

| No. | Pertanyaan | Sumber |
|-----|---|----------------------------------|
| 1 | Informasi yang disajikan SIPENA memiliki ketepatan yang tinggi. | (Mangun Buana & Wirawati, 2018)) |
| 2 | Informasi yang disajikan SIPENA sesuai dengan kebutuhan. | |
| 3 | Informasi yang disajikan SIPENA mudah untuk dipelajari dan dipahami. | |
| 4 | Menu-menu yang tersedia pada SIPENA, secara instan dapat menampilkan informasi yang sesuai dengan yang diinginkan | |
| 5 | Output yang dihasilkan SIPENA mempunyai tampilan yang sama (konsisten). *Contoh: KRP, KHP, TKM | |

3) Variabel Perceived Usefulness

Dalam penelitian ini, variabel *perceived usefulness* diukur dari indikator *useful*, *beneficial*, *effectiveness*, dan *productivity* untuk menambah produktivitas, efektif, bermanfaat, mempercepat pekerjaan, dan mengembangkan kinerja pekerjaan. Untuk pertanyaan variabel *perceived usefulness* ada pada Tabel V.

TABEL V
KUESIONER VARIABEL PERCEIVED USEFULNESS

| No. | Pertanyaan | Sumber |
|-----|--|---------------------------------|
| 1 | SIPENA berguna dalam proses mempermudah pekerjaan mahasiswa | (Mangun Buana & Wirawati, 2018) |
| 2 | SIPENA memberikan keuntungan bagi mahasiswa. | |
| 3 | Dengan adanya SIPENA, pekerjaan menjadi lebih efektif dalam proses penilaian | |
| 4 | SIPENA dapat digunakan pada seluruh proses penilaian nonakademik mahasiswa | |

4) Variabel Kepuasan Pengguna

Dalam penelitian ini, variabel kepuasan pengguna diukur dari indikator *perceived value*, *user expectation*, dan *perceived quality*. Untuk pertanyaan variabel kepuasan pengguna ada pada Tabel VI.

TABEL VI
KUESIONER VARIABEL KEPUASAN PENGGUNA

| No. | Pertanyaan | Sumber |
|-----|------------|--------|
|-----|------------|--------|

| | | |
|---|--|---------------------------------|
| 1 | Layanan yang tersedia pada SIPENA, dapat memberikan manfaat yang sesuai dalam proses penilaian | (Mangun Buana & Wirawati, 2018) |
| 2 | Layanan yang diberikan SIPENA sesuai dengan yang diinginkan | |
| 3 | Informasi yang dihasilkan SIPENA dapat diandalkan/dipercaya | |

B. Uji Validitas dan Reliabilitas

Uji validitas dan reliabilitas dilakukan pada 30 sampel data yang dipilih secara acak berdasarkan pada penelitian yang dilakukan oleh [5]. Berikut adalah hasil dari uji validitas pada Tabel VII.

TABEL VII
UJI VALIDITAS

| Kode Indikator | Corrected Item-Total Correlation | Keterangan Valid (Ya atau Tidak) |
|----------------|----------------------------------|----------------------------------|
| KS1 | 0.707 | Ya |
| KS2 | 0.601 | Ya |
| KS3 | 0.559 | Ya |
| KS4 | 0.536 | Ya |
| KS5 | 0.513 | Ya |
| KS6 | 0.623 | Ya |
| KI1 | 0.831 | Ya |
| KI2 | 0.840 | Ya |
| KI3 | 0.729 | Ya |
| KI4 | 0.829 | Ya |
| KI5 | 0.700 | Ya |
| PU1 | 0.723 | Ya |
| PU2 | 0.742 | Ya |
| PU3 | 0.767 | Ya |
| PU4 | 0.734 | Ya |
| KP1 | 0.714 | Ya |
| KP2 | 0.851 | Ya |
| KP3 | 0.848 | Ya |

Hipotesis untuk uji validitas kuesioner ini adalah sebagai berikut:

- H_0 : Jika nilai korelasi (r hitung) $>$ r tabel, maka data kuesioner penelitian valid
 H_1 : Jika nilai korelasi (r hitung) $<$ r tabel, maka data kuesioner penelitian tidak valid

Menurut hasil yang ada pada Tabel VII, maka dapat disimpulkan bahwa hipotesis uji validitas kuesioner menerima H_0 sebab nilai korelasi (r hitung) $>$ r tabel (0,361). Dengan demikian dapat diartikan jika seluruh data instrumen penelitian dinyatakan valid sebab mempunyai nilai korelasi $>$ r tabel (0,361).

TABEL VIII
UJI RELIABILITAS

| Cronbach's Alpha | N of Items |
|------------------|------------|
| 0.952 | 18 |

Hipotesis untuk uji reliabilitas kuesioner ini adalah sebagai berikut:

- H_0 : Jika nilai Alpha $>$ 0,70, maka data kuesioner penelitian reliabel atau konsisten.
 H_1 : Jika nilai Alpha $<$ 0,70, maka data kuesioner penelitian tidak reliabel atau konsisten.

Dapat disimpulkan berdasarkan Tabel VIII, bahwa karena nilai Alpha $>$ 0,70. Dengan demikian dapat diartikan bahwa hipotesis uji reliabilitas kuesioner menerima H_0 seluruh data instrumen penelitian dinyatakan reliabel atau konsisten karena memiliki nilai Alpha 0.952.

C. Persiapan Dataset

Sebelum meolah data hasil kuesioner kepuasan pengguna SIPENA, dilakukan persiapan dataset untuk dapat di klasifikasikan dengan mudah nantinya. Proses dalam persiapan dataset ini dilakukan dengan menghapus data responden yang tidak memenuhi kriteria, serta menghapus kolom data yang tidak diperlukan dalam proses pengolahan data nantinya seperti kolom nomor induk mahasiswa, nomor telepon / email, dan kolom kriteria pengguna.

Setelah data sudah menjadi bersih dan siap untuk diolah pertama yang dilakukan adalah memberi label pada data kepuasan pengguna SIPENA. Label diberikan berdasarkan hasil rata-rata dari nilai skala likert kepuasan pengguna SIPENA [9]. Ketentuan label seperti pada Tabel IX

TABEL IX
NILAI RATA-RATA KEPUASAN DAN LABEL DATA

| Nilai Rata-Rata Kepuasan | Label Data |
|-------------------------------|-------------------|
| $1 \leq \text{rata-rata} < 2$ | Sangat Tidak Puas |
| $2 \leq \text{rata-rata} < 3$ | Tidak Puas |
| $3 \leq \text{rata-rata} < 4$ | Cukup Puas |
| $4 \leq \text{rata-rata} < 5$ | Puas |
| $\text{rata-rata} = 5$ | Sangat Puas |

TABEL X
DESKRIPSI KOLOM DATASET

| No | Kolom | Deskripsi |
|----|-------|-----------|
|----|-------|-----------|

| | | |
|----|---------------|--|
| 1 | nama | Nama responden / mahasiswa |
| 2 | jenis_kelamin | Jenis kelamin responden / mahasiswa |
| 3 | fakultas | Fakultas dari responden / mahasiswa |
| 4 | KS1 | Idikator pertanyaan kualitas sistem pada kuesioner nomor 1 |
| 5 | KS2 | Idikator pertanyaan kualitas sistem pada kuesioner nomor 2 |
| 6 | KS3 | Idikator pertanyaan kualitas sistem pada kuesioner nomor 3 |
| 7 | KS4 | Idikator pertanyaan kualitas sistem pada kuesioner nomor 4 |
| 8 | KS5 | Idikator pertanyaan kualitas sistem pada kuesioner nomor 5 |
| 9 | KS6 | Idikator pertanyaan kualitas sistem pada kuesioner nomor 6 |
| 10 | KI1 | Idikator pertanyaan kualitas informasi pada kuesioner nomor 1 |
| 11 | KI2 | Idikator pertanyaan kualitas informasi pada kuesioner nomor 2 |
| 12 | KI3 | Idikator pertanyaan kualitas informasi pada kuesioner nomor 3 |
| 13 | KI4 | Idikator pertanyaan kualitas informasi pada kuesioner nomor 4 |
| 14 | KI5 | Idikator pertanyaan kualitas informasi pada kuesioner nomor 5 |
| 15 | PU1 | Idikator pertanyaan <i>perceived usefulness</i> pada kuesioner nomor 1 |
| 16 | PU2 | Idikator pertanyaan <i>perceived usefulness</i> pada kuesioner nomor 2 |
| 17 | PU3 | Idikator pertanyaan <i>perceived usefulness</i> pada kuesioner nomor 3 |
| 18 | PU4 | Idikator pertanyaan <i>perceived usefulness</i> pada kuesioner nomor 4 |
| 19 | KP1 | Idikator pertanyaan kepuasan pengguna pada kuesioner nomor 1 |
| 20 | KP2 | Idikator pertanyaan kepuasan pengguna pada kuesioner nomor 1 |
| 21 | KP3 | Idikator pertanyaan kepuasan pengguna pada kuesioner nomor 1 |
| 22 | rata_rata | Jumlah rata-rata tingkat kepuasan responden / mahasiswa |

| | | |
|----|--------------|--|
| 23 | Satisfaction | Tingkat kepuasan responden / mahasiswa berdasarkan hasil rata-rata |
|----|--------------|--|

D. Pembuatan Model Klasifikasi

Setelah dilakukan proses persiapan data / data *preprocessing* menghasilkan dataset dengan jumlah 499 baris dan 23 kolom. Dimana kolom nama, jenis_kelamin, fakultas, dan rata_rata tidak akan dimasukkan ke dalam proses *modeling* sedangkan kolom satisfaction akan digunakan sebagai target.

Proses pembuatan model disini merupakan tahapan untuk menentukan model atau algoritma terbaik dalam proses klasifikasi. Proses *modeling* yang dilakukan menggunakan *tools* Jupyter Notebook dengan *library* PyCaret. Dalam proses untuk mendapatkan model terbaik untuk klasifikasi dilakukan tahapan antara lain melakukan setup dataset pada PyCaret, membandingkan semua metode klasifikasi untuk mendapatkan model terbaik.

1) Setup Dataset Pada PyCaret

Fungsi *setup* pada PyCaret harus dipanggil sebelum menjalankan proses atau fungsi lainnya. Dalam fungsi *setup* terdapat 2 parameter wajib yaitu data dan target yang mana keduanya wajib di definisikan, untuk parameter lainnya bersifat opsional. Pada langkah yang pertama melakukan *load* data dengan memanggil dataset kepuasan pengguna SIPENA yang sudah diolah sebelumnya dengan nama filenya adalah dataset.csv untuk di proses kedalam proses selanjutnya.

```
import pandas as pd
from pycaret.classification import*

data = pd.read_csv("dataset.csv")
```

Gbr. 2 Load Dataset

Selanjutnya akan dilakukan *setup* data untuk proses *modeling* dengan menggunakan parameter data dan target. Untuk mengabaikan kolom nama, jenis_kelamin, fakultas, dan rata_rata menggunakan parameter *ignore_features* agar kolom tersebut tidak masuk kedalam proses *modeling* dan berfokus pada klasifikasi hasil dari kepuasan pengguna SIPENA di setiap indikatornya.

```
exp = setup(data, target = 'satisfaction', session_id = 8610,
            ignore_features = ['nama', 'rata_rata', 'jenis_kelamin', 'fakultas'],
            )
```

| | Description | Value |
|----|-----------------------------|---|
| 0 | Session id | 8610 |
| 1 | Target | satisfaction |
| 2 | Target type | Multiclass |
| 3 | Target mapping | CUKUP PUAS: 0, PUAS: 1, SANGAT PUAS: 2, TIDAK PUAS: 3 |
| 4 | Original data shape | (499, 23) |
| 5 | Transformed data shape | (499, 19) |
| 6 | Transformed train set shape | (349, 19) |
| 7 | Transformed test set shape | (150, 19) |
| 8 | Ignore features | 4 |
| 9 | Numeric features | 18 |
| 10 | Preprocess | True |
| 11 | Imputation type | simple |
| 12 | Numeric imputation | mean |
| 13 | Categorical imputation | mode |
| 14 | Fold Generator | StratifiedKFold |
| 15 | Fold Number | 10 |
| 16 | CPU Jobs | -1 |
| 17 | Use GPU | False |
| 18 | Log Experiment | False |
| 19 | Experiment Name | clf-default-name |
| 20 | USI | 713b |

Gbr. 3 Setup data untuk memulai proses *modeling*

Hasil dari proses *setup* dapat dilihat pada (Gbr. 3), dalam tahapan ini juga dilakukan pembagian data *training* dan data *testing* serta parameter *fold* yang digunakan disini adalah *default* dari PyCaret yaitu 10. Berikut adalah hasil pembagian data *training* dan data *testing* dari keseluruhan jumlah pada dataset.

TABEL XI
 PEMBAGIAN DATA TRAINING DAN DATA TESTING

| Keterangan | Data Training | Data Testing | Total |
|------------|---------------|--------------|-------|
| Presentase | 70% | 30% | 100% |
| Jumlah | 349 | 150 | 499 |

2) Membandingkan Metode Klasifikasi

Pada *library* PyCaret terdapat sekitar 16 algoritma klasifikasi yang tersedia untuk proses perbandingan metode. Algoritma klasifikasi tersebut antara lain *Logistic Regression*, *K Neighbors Classifier*, *Decision Tree Classifier*, *Quadratic Discriminant Analysis*, *Linear Discriminant Analysis*, *Light Gradient Boosting Machine*, *Gradient Boosting Classifier*, *Naive Bayes*, *Extra Trees Classifier*, *CatBoost Classifier*, *Extreme Gradient Boosting*, *Random Forest Classifier*, *AdaBoost Classifier*, *Ridge Classifier*, *SVM - Linear Kernel*, dan *Dummy Classifier*. Dari beberapa metode tersebut, akan dilakukan perbandingan atau komparasi dengan menggunakan fungsi *compare_models()*. Fungsi ini berguna untuk melatih dan mengevaluasi performa model algoritma klasifikasi dengan

menggunakan *cross validation*. Keluaran dari fungsi ini adalah rata-rata skor dari *cross validation*.

```
best = compare_models()
```

| Model | Accuracy | AUC | Recall | Prec. | F1 | Kappa | MCC | TT (Sec) |
|----------|----------|--------|--------|--------|--------|--------|--------|----------|
| et | 0.9743 | 0.1992 | 0.9743 | 0.9758 | 0.9730 | 0.9494 | 0.9513 | 1.6500 |
| lr | 0.9714 | 0.1987 | 0.9714 | 0.9684 | 0.9672 | 0.9434 | 0.9461 | 4.1440 |
| rf | 0.9657 | 0.1989 | 0.9657 | 0.9697 | 0.9658 | 0.9350 | 0.9371 | 2.7790 |
| catboost | 0.9629 | 0.1989 | 0.9629 | 0.9655 | 0.9623 | 0.9288 | 0.9311 | 1.5160 |
| nb | 0.9541 | 0.1998 | 0.9541 | 0.9600 | 0.9535 | 0.9134 | 0.9172 | 2.4860 |
| lda | 0.9513 | 0.1996 | 0.9513 | 0.9573 | 0.9503 | 0.9065 | 0.9110 | 2.5740 |
| lightgbm | 0.9486 | 0.1987 | 0.9486 | 0.9563 | 0.9490 | 0.9018 | 0.9053 | 1.5530 |
| xgboost | 0.9457 | 0.1983 | 0.9457 | 0.9511 | 0.9455 | 0.8965 | 0.8996 | 1.4620 |
| gbc | 0.9313 | 0.1980 | 0.9313 | 0.9454 | 0.9351 | 0.8716 | 0.8749 | 2.9550 |
| dt | 0.9170 | 0.1781 | 0.9170 | 0.9350 | 0.9224 | 0.8456 | 0.8493 | 2.6070 |
| knn | 0.9113 | 0.1867 | 0.9113 | 0.9050 | 0.9026 | 0.8278 | 0.8337 | 2.5900 |
| ada | 0.8685 | 0.1801 | 0.8685 | 0.8601 | 0.8572 | 0.7426 | 0.7516 | 2.7810 |
| ridge | 0.8369 | 0.0000 | 0.8369 | 0.7521 | 0.7883 | 0.6451 | 0.6785 | 3.0300 |
| svm | 0.8226 | 0.0000 | 0.8226 | 0.7689 | 0.7789 | 0.6292 | 0.6697 | 2.7420 |
| dummy | 0.6304 | 0.1000 | 0.6304 | 0.3975 | 0.4875 | 0.0000 | 0.0000 | 1.4150 |

Gbr. 4 Perbandingan Model Klasifikasi

Setelah menjalankan fungsi *compare_models()* didapatkan hasil seperti pada (Gbr. 4), dengan hasil dari proses perbandingan menghasilkan 15 model yang didapatkan bahwa algoritma *Extra Trees Classifier* berada pada urutan pertama dengan nilai *accuracy* tertinggi dibandingkan algoritma lainnya. Pada urutan kedua dan ketiga secara berurut ada algoritma *Logistic Regression* dan algoritma *Random Forest Classifier*.

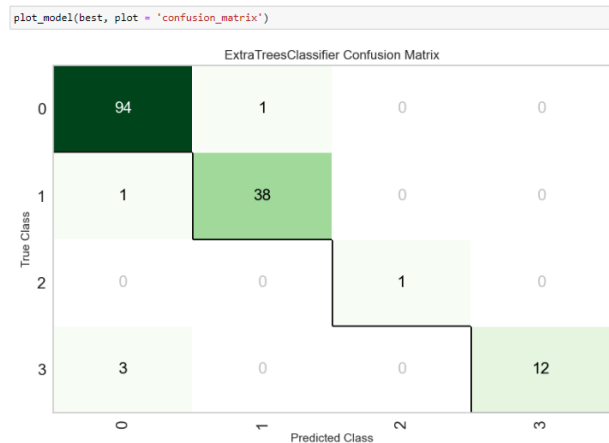
TABEL XII
HASIL PERBANDINGAN MODEL TERBAIK

| Model | Accuracy |
|---------------------------------|----------|
| <i>Extra Trees Classifier</i> | 0.9743 |
| <i>Logistic Regression</i> | 0.9714 |
| <i>Random Forest Classifier</i> | 0.9657 |

E. Evaluasi Model

Evaluation adalah tahapan dimana model terbaik yang sudah didapatkan dan dibuat sebelumnya dilakukan analisis dan evaluasi. Dalam evaluasi model akan menggunakan fungsi *plot_model()* dengan menggunakan dua parameter yaitu variabel dan nama tipe plot yang ingin ditampilkan. Pada *library* PyCaret terdapat 21 tipe plot yang tersedia, tetapi dalam penelitian ini hanya menggunakan beberapa plot saja. Secara rinci, hasil plot model dari dataset yang diolah adalah sebagai berikut :

1) Plot Confusion Matrix

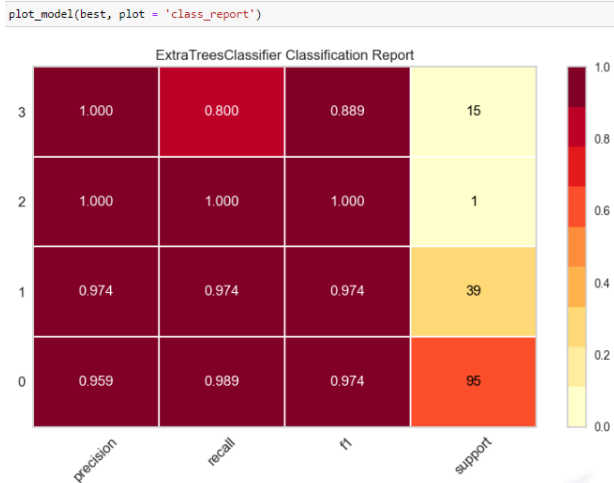


Gbr. 5 Plot Confusion matrix

Dari total data *testing* dalam data dataset kepuasan pengguna SIPENA yang berjumlah 150 terdapat 95 data yang berlabel 0 (Cukup Puas), 39 data yang berlabel 1 (Puas), 1 data yang berlabel 2 (Sangat Puas), dan 15 data yang berlabel 3 (Tidak Puas). Tetapi pada hasil plot *confusion matrix*, Terdapat 94 data yang diprediksi benar masuk ke dalam label 0 (Cukup Puas) dan 1 data yang diprediksi salah karena masuk kedalam label 1 (Puas), untuk 39 data label 1 (Puas) terdapat 1 data yang dipredikasi salah karena masuk label 0 (Cukup Puas), untuk 1 data yang berlabel 2 (Sangat Puas) dipredikasi benar, dan yang terakhir 15 data label 3 (Tidak Puas) terdapat 3 data yang dipredikasi salah dan masuk ke label 0 (Cukup Puas). Berdasarkan hasil plot model terbaik (Gbr. 5) berikut adalah nilai TP, TN, FP, FN dari masing- masing *class* :

- Class 0* (Cukup Puas) nilai *True Positive* (TP) adalah 94, nilai *False Positive* (FP) adalah $1 + 3 = 4$, nilai *False Negative* (FN) adalah 1, nilai *True Negative* (TN) adalah $38 + 1 + 12 = 51$.
- Class 1* (Puas) nilai *True Positive* (TP) adalah 38, nilai *False Positive* (FP) adalah 1, nilai *False Negative* (FN) adalah 1, nilai *True Negative* (TN) adalah $94 + 3 + 1 + 12 = 110$.
- Class 2* (Sangat Puas) nilai *True Positive* (TP) adalah 1, nilai *False Positive* (FP) adalah 0, nilai *False Negative* (FN) adalah 0, nilai *True Negative* (TN) adalah $94 + 1 + 1 + 38 + 3 + 12 = 149$.
- Class 3* (Tidak Puas) nilai *True Positive* (TP) adalah 12, nilai *False Positive* (FP) adalah 0, nilai *False Negative* (FN) adalah 3, nilai *True Negative* (TN) adalah $94 + 1 + 1 + 38 + 1 = 135$.

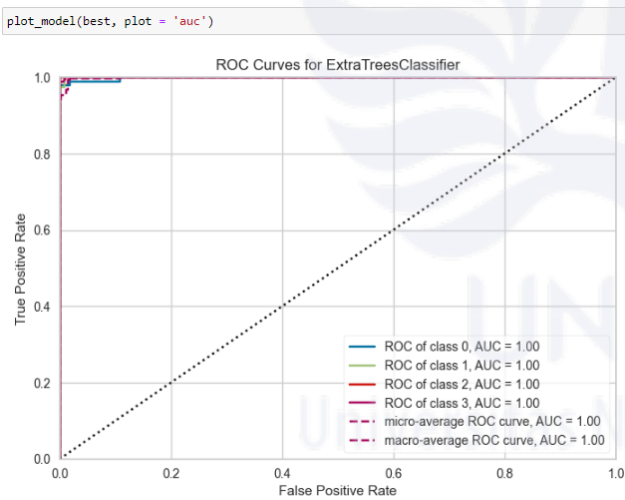
2) Plot Class Report



Gbr. 6 Class Report Kepuasan Pengguna SIPENA

Pada (Gbr. 6) merupakan hasil dari plot *class_report* dari model terbaik yang menampilkan nilai dari evaluasi metrik *precision*, *recall*, *f1 score* dan *support* dari masing-masing label/kelas. Dalam *class report* semakin baik nilainya maka akan semakin gelap warna merah yang ditampilkan serta mendekati angka 1.0 dari hasil plot dapat diketahui bahwa hampir semua mendapat nilai diatas 0,8 yang berarti ini menghasilkan skor yang baik.

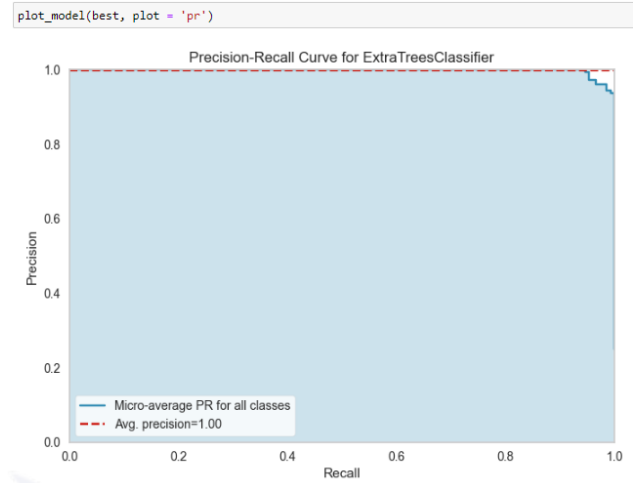
3) Plot AUC



Gbr. 7 Grafik AUC Kepuasan Pengguna SIPENA

Pada (Gbr. 7) merupakan hasil dari grafik *AUC (Area Under the Curve)* dari model terbaik. Sumbu y menampilkan tingkat positif sebenarnya atau *true positive rate* dan sumbu x akan menampilkan tingkat positif palsu atau *false positive rate*. Pada hasil plot *AUC* ini dihasilkan nilai ROC pada setiap label/kelas, nilai rata-rata micro dan macro dengan nilai hampir 1.0 semua, sehingga model dikatakan baik. Titik ideal pada grafik *AUC* adalah pada sudut kiri atas ketika *false positive rate* bernilai 0 dan *true positive rate* bernilai 1.

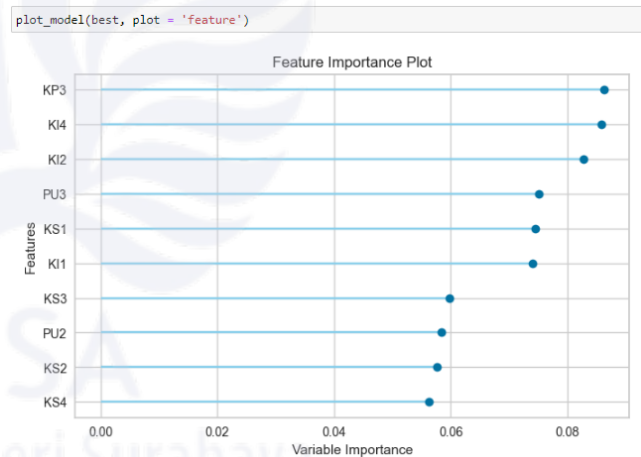
4) Plot Precision



Gbr. 8 Precision Recall Kepuasan Pengguna SIPENA

Pada (Gbr. 8) merupakan hasil dari plot *precision recall* dari model terbaik yang menampilkan hasil dari nilai *precision* pada sumbu y dan hasil dari nilai *recall* pada sumbu x. Nilai antara *precision* dan *recall* adalah antara 0 hingga 1. Pada model yang dibuat mendapatkan rata-rata nilai 1.0 sehingga dapat dikatakan sangat baik. Dimana model yang baik adalah yang mendapatkan maksimal nilai hingga mendekati 1.0.

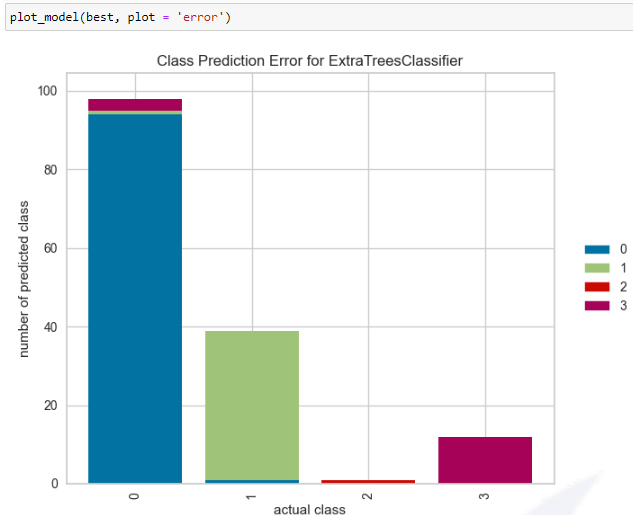
5) Plot Features Importance



Gbr. 9 Features Importance Kepuasan Pengguna SIPENA

Pada (Gbr. 9) merupakan hasil dari plot *features importance* dari model terbaik yang menampilkan fitur-fitur yang paling penting dan berpengaruh. Pada model ini didapatkan bahwa ada 3 fitur yang mendapatkan nilai diatas 0.08 yaitu, fitur KP3 (Kepuasan Pengguna), KI4 (Kualitas Informasi), dan KI2 (Kualitas Informasi). Jika dilihat berdasarkan indikator pertanyaan pada kuesioner kepuasan pengguna SIPENA fitur KP3 mewakili variabel Kepuasan Pengguna pada indikator *Perceived Quality*, untuk KI4 dan KI2 mewakili variabel Kualitas informasi pada indikator Relevansi dan Ketepatan Waktu.

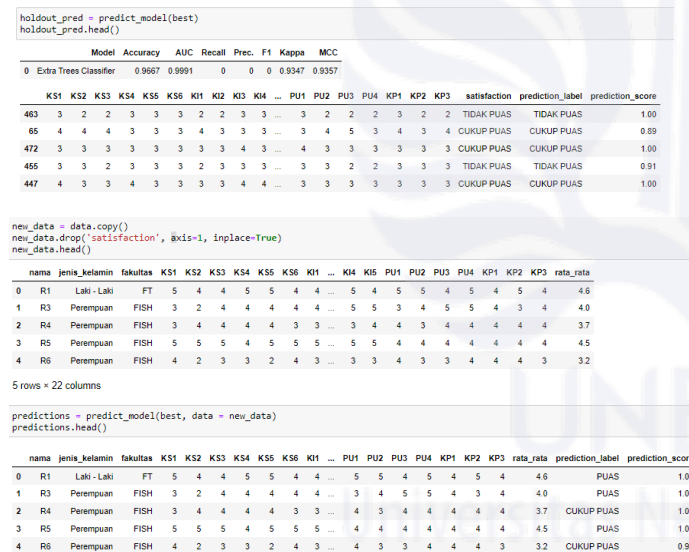
6) Plot Prediction Error



Gbr. 10 Prediction Error Kepuasan Pengguna SIPENA

Pada (Gbr. 10) merupakan hasil dari *prediction error* dari model terbaik yang menampilkan hampir sama seperti *class_report*, pada plot ini menampilkan diagram batang bertumpuk yang memuat jumlah data untuk setiap label/kelas.

7) Hasil Prediksi Label

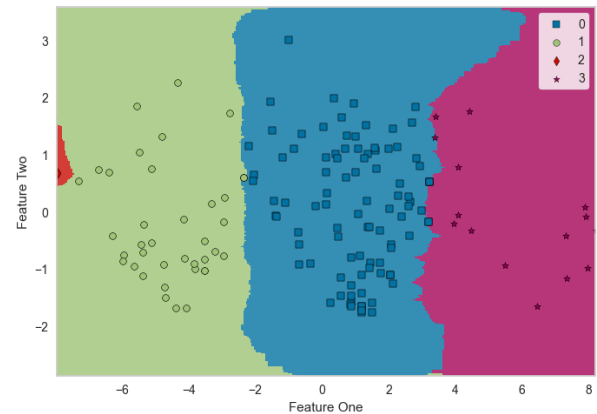


Gbr. 11 Hasil Prediksi Label

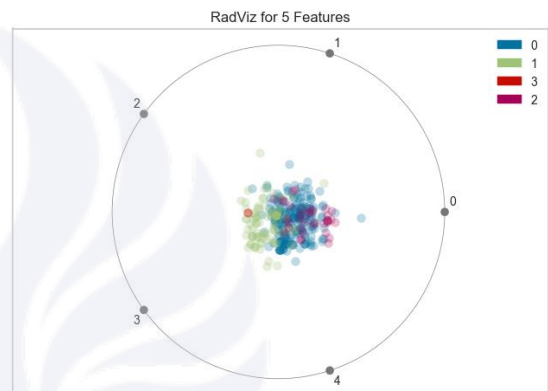
Pada (Gbr. 11) merupakan hasil dari prediksi label pada data kepuasan pengguna SIPENA yang didapatkan *prediction_label* dan *prediction_score*.

8) Visualisasi Data

Decision Boundary dan *Dimension* dari model yang didapatkan pada proses evaluasi model menampilkan visualisasi data dan persebaran data setiap class atau labelnya seperti pada (Gbr. 12) dan (Gbr. 13).



Gbr. 12 Visualisasi Data Decision Boundary



Gbr. 13 Visualisasi Data Dimension

IV. PENUTUP

A. Kesimpulan

Berdasarkan dari hasil penelitian perbandingan metode klasifikasi data mining untuk mengukur tingkat kepuasan mahasiswa terhadap sistem informasi penilaian nonakademik unesa (SIPENA) menggunakan *library* PyCaret dengan *tools* Jupyter Notebook di dapatkan kesimpulan sebagai berikut :

1. Pada proses perbandingan metode klasifikasi data mining untuk mengukur tingkat kepuasan mahasiswa terhadap sistem informasi penilaian nonakademik unesa (SIPENA) menghasilkan 3 algoritma klasifikasi dengan akurasi tertinggi yaitu yang pertama adalah algoritma *Extra Trees Classifier* dengan nilai akurasi 0.9743, yang kedua adalah algoritma *Logistic Regression* dengan nilai akurasi 0.9714, dan yang ketiga adalah algoritma *Random Forest Classifier* dengan nilai akurasi 0.9657.
2. Pada tahap elavusi model dengan menggunakan plot model *confusion_matrix* dari 150 data *testing* menghasilkan 94 data yang diprediksi benar masuk ke dalam label (Cukup Puas) dan 1 data yang diprediksi salah karena masuk kedalam label (Puas), untuk 39 data label (Puas) terdapat 1 data yang dipredikasi salah karena masuk label (Cukup Puas), untuk 1 data yang berlabel (Sangat Puas) dipredikasi benar, dan yang

terakhir 15 data label (Tidak Puas) terdapat 3 data yang diprediksi salah dan masuk ke label (Cukup Puas).

3. Pada hasil plot *features importance* menghasilkan beberapa faktor yang cukup mempengaruhi dalam proses klasifikasi untuk menentukan kepuasan pengguna SIPENA yang pertama KP3 mewakili variabel Kepuasan Pengguna pada indikator *Perceived Quality*, untuk KI4 dan KI2 mewakili variabel Kualitas informasi pada indikator Relevansi (*Relevance*) dan Ketepatan Waktu (*Timeliness*).

B. Saran

Saran yang dapat diberikan dan dijadikan bahan pertimbangan untuk perbaikan bagi penelitian dengan topik yang sama dalam perbandingan metode klasifikasi khususnya yang menggunakan *library* PyCaret adalah pada proses setup data disarankan untuk lebih mengeksplorasi parameter yang ada agar mendapatkan hasil akurasi klasifikasi yang lebih baik serta dataset sebisa mungkin dibuat sederhana agar lebih mudah dalam proses pengolahannya..

REFERENSI

- [1] N. Kinanti, A. Putri1, dan A. Dwi, "Penerapan PIECES Framework sebagai Evaluasi Tingkat Kepuasan Mahasiswa terhadap Penggunaan Sistem Informasi Akademik Terpadu (SIKADU) pada Universitas Negeri Surabaya," *JEISBI*, vol. 02, 2021, [Daring]. Tersedia pada: <https://siakadu.unesa.ac.id>
- [2] P. L. Lokapitasari Belluano, I. Indrawati, H. Harlinda, F. A. R. Tuasamu, dan D. Lantara, "ANALISIS TINGKAT KEPUASAN PENGGUNA SISTEM INFORMASI PERPUSTAKAAN MENGGUNAKAN PIECES FRAMEWORK," *ILKOM Jurnal Ilmiah*, vol. 11, no. 2, hlm. 118–128, Sep 2019, doi: 10.33096/ilkom.v11i2.398.118-128.
- [3] I. B. G. M. Mangun Buana dan N. G. P. Wirawati, "Influence Quality of Information System, Quality of Information, And Perceived Usefulness On User Accounting Information System Satisfaction," *E-Jurnal Akuntansi*, hlm. 683, Jan 2018, doi: 10.24843/eja.2018.v22.i01.p26.
- [4] M. T. Anwar dan D. R. A. Permana, "Perbandingan Performa Model Data Mining untuk Prediksi Dropout Mahasiswa," *Jurnal Teknologi dan Manajemen*, vol. 19, no. 2, hlm. 33–40, Agu 2021, doi: 10.52330/jtm.v19i2.34.
- [5] E. Mariati, A. Lestari, dan Widiatry, "MODEL KLASIFIKASI KEPUASAN MAHASISWA TEKNIK TERHADAP SARANA PEMBELAJARAN MENGGUNAKAN DATA MINING," 2020, doi: 10.47111/JTI.
- [6] Fatty Ariani dan Andi Taufik, "Perbandingan Metode Klasifikasi Data Mining untuk Prediksi Tingkat Kepuasan Pelanggan Telkomsel Prabayar," *SATIN - Sains dan Teknologi Informasi*, vol. 6, no. 2, hlm. 46–55, Des 2020, doi: 10.33372/stn.v6i2.666.
- [7] M. C. Mustaqim, "Membangun Model Machine Learning dengan Mudah Menggunakan PyCaret | by MOCH. CHAMDANI MUSTAQIM | Medium," 2020. <https://medium.com/@chamdani/membangun-model-machine-learning-dengan-mudah-menggunakan-pycaret-29f36d2c283c> (diakses 15 Juni 2023).
- [8] I. Prayanthi, E. Lompoliu, dan R. L. Klabat, "THE EFFECT OF SYSTEM QUALITY, INFORMATION QUALITY AND PERCEIVED USEFULNESS ON ACCOUNTING INFORMATION SYSTEM USER SATISFACTION," *Accounting Review* /, vol. 1, no. 2, 2020.
- [9] A. Aditya Santika, T. Hamonangan Saragih, D. Kartini, dan R. Ramadhani, "Penerapan Skala Likert Pada Klasifikasi Tingkat

Kepuasan Pelanggan Agen BRILink Menggunakan Random Forest," vol. 11, no. 3, hlm. 405–4011, 2023, doi: 10.26418/justin.v11i3.