

## APLIKASI PENGOLAHAN DATA PREDIKSI KEMISKINAN BERBASIS E-COMMERCE MENGGUNAKAN DECISION TREE DAN WRAPPER FEATURE SELECTION

### APPLICATION BASED OF E-COMMERCE POVERTY PREDICTION DATA PROCESSING DECISION TREE AND WRAPPER FEATURE SELECTION

Ade Pangestu<sup>1</sup>, Dr. Dedy Rahman Wijaya, S.T., M.T.<sup>2</sup>, Elis Hernawati, S.T., M.Kom.<sup>3</sup>

<sup>1,2,3</sup>Program Studi D3 Sistem Informasi, Fakultas Ilmu Terapan Universitas Telkom

pangestuade@student.telkomuniversity.ac.id<sup>1</sup>, [dedyrw@tass.telkomuniversity.ac.id](mailto:dedyrw@tass.telkomuniversity.ac.id)<sup>2</sup>,

[elishernawati@tass.telkomuniversity.ac.id](mailto:elishernawati@tass.telkomuniversity.ac.id)<sup>3</sup>

#### Abstrak

Kemiskinan adalah suatu kondisi seseorang yang tidak dapat memenuhi kebutuhannya, baik material maupun non-material. Badan Pusat Statistik menggunakan metode dengan konsep pendekatan kebutuhan dasar atau dapat disebut kemampuan untuk memenuhi kebutuhan dasar untuk mengukur tingkat kemiskinan di suatu daerah. Konsep ini menggunakan pendekatan kebutuhan dasar yang dibutuhkan oleh manusia seperti kebutuhan makanan dan kebutuhan bukan makanan untuk mengukur tingkat kemiskinan di suatu daerah melalui survei sosial ekonomi nasional (SUSENAS). Adapun hambatan yang dilakukan pada saat SUSENAS diperkirakan memakan waktu yang tidak singkat karena dalam kegiatan tersebut diperlukannya tahap wawancara terhadap setiap kepala rumah tangga dengan menggunakan kuesioner oleh karena itu dapat dipastikan membutuhkan biaya yang cukup besar serta para kepala rumah tangga tidak mudah untuk ditemui bahkan menghindari karena para kepala rumah tangga takut jika terjadi penipuan. Metode lain untuk melengkapi hasil survei dan sensus BPS adalah dengan menggunakan *machine learning decision tree* dan *wrapper feature selection* dibantu dengan data e-commerce yang diperoleh dari salah satu perusahaan e-commerce di Indonesia untuk mendapatkan data prediksi kemiskinan. Hasil dari eksperimen yang dilakukan dapat disimpulkan bahwa relevan dengan hasil BPS apabila terdapat banyak fitur, jika dengan 10 - 40 dan 80-90 fitur maka hasilnya tidak relevan.

Kata Kunci: Kemiskinan, BPS, *machine learning*, *decision tree*, *wrapper*.

#### Abstract

Poverty is a problem of someone who cannot meet their needs, both material and non-material. The Central Statistics Agency uses methods with basic needs or can be called needs for basic needs to measure the level of welfare in an area. This concept uses the basic needs needed by humans as food and non-food needs to measure poverty levels in the region through a national socioeconomic survey (SUSENAS). What is done when SUSENAS is estimated not to spend a short amount of time because in these activities interviews are needed for each household using a questionnaire therefore it can be ascertained that it requires a significant amount of money with the heads of households not easy to meet. trouble. Another method for completing the results of BPS surveys and censuses is to use machine learning decision trees and wrapper feature selection assisted with e-commerce data obtained from one e-commerce company in Indonesia to obtain poverty prediction data. The results of the experiments can be concluded that relevant to the results of the BPS estimate there are many features, if the 10-40 and 80-90 features are produced irrelevant.

Keywords: Poverty, BPS, *machine learning*, *decision tree*, *wrapper*.

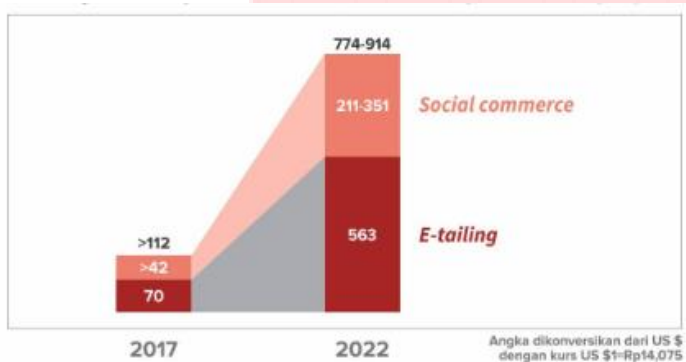
#### I. PENDAHULUAN

Kemiskinan merupakan masalah yang dihadapi hampir seluruh dunia, termasuk Indonesia dan ini menjadi tantangan besar supaya segera dituntaskan, adapun program pemerintah tentang penurunan angka kemiskinan belum mencapai target [1]. Badan Pusat Statistik Indonesia (BPS) mendeskripsikan kemiskinan berdasarkan kemampuan dalam memenuhi kebutuhan dasar dari sisi ekonomi [2]. Survei Sosial Ekonomi atau yang disingkat dengan Susenas diselenggarakan untuk mendapatkan data presentase kemiskinan yang dibutuhkan oleh pihak BPS. Susenas merupakan suatu aktivitas yang dilakukan oleh BPS untuk mengumpulkan data yang menyangkut mengenai kebutuhan rumah tangga, bidang pendidikan, kesehatan, dll. Pada tahun 2018, pengumpulan data Susenas Kor dilakukan pada setiap Bulan Maret dengan jumlah sampel total sebanyak 300.000 rumah tangga di Indonesia. Namun dengan melaksanakan kegiatan Susenas ini dapat diperkirakan memakan waktu yang tidak singkat karena dalam kegiatan tersebut diperlukannya tahap wawancara terhadap setiap kepala rumah tangga dengan menggunakan kuesioner oleh karena itu dapat dipastikan membutuhkan biaya

yang cukup besar serta para kepala rumah tangga tidak mudah untuk ditemui bahkan menghindari karena para kepala rumah tangga takut jika terjadi penipuan [3]. Di era digital saat ini dalam memperoleh data tidak harus melakukan wawancara untuk survei atau menggunakan kuesioner. Adapun metode untuk memperoleh data untuk memprediksi tingkat kemiskinan adalah dengan dataset *e-commerce*. Kaitannya data e-commerce dengan prediksi kemiskinan yaitu ketika BPS mengukur garis kemiskinan menggunakan konsep *basic needs approach* atau kemampuan memenuhi kebutuhan dasar (pangan dan non pangan) untuk mengukur tingkat kemiskinan disuatu daerah [2], dan dataset *e-commerce* lebih merepresentasikan pengeluaran belanja individu atau keluarga.

Indonesia merupakan negara yang berkontribusi 50% dari seluruh transaksi karena memiliki pasar e-commerce yang terbesar di Asia tenggara. Kontribusi ini akan diprediksikan meningkat lantaran penduduk Indonesia yang terus menggunakan internet guna mempermudah dalam melaksanakan semua kegiatan sehari - hari [4]. Pada tahun 2018, seorang dari perserikatan dagang konsultan manajemen McKinsey dan Company mempublikasikan hasil risetnya berupa status pasar e-commerce

Indonesia saat ini hingga masa yang akan datang. Hasil dari riset tersebut menunjukkan pertumbuhan nilai dan dampak pertumbuhan terhadap ekonomi dan sosial dari pasar e-commerce Indonesia saat ini sampai tahun 2022. Pasar e-commerce diperkirakan akan menghasilkan US\$65 miliar (Rp910 triliun). McKinsey menjelaskan tentang e-commerce sebagai tahapan transaksi jual beli barang online yang dibagi menjadi 2 jenis yaitu *Social Commerce* dan *E-tailing*. *Social Commerce* merupakan cara pemasaran secara online menggunakan sosial media serta pembayaran dan pengiriman yang menggunakan *platform* lain, sedangkan *E-tailing* adalah jual beli barang yang difasilitasi oleh beberapa perusahaan dagang online seperti shopee, Jd.Id, OLX, dll. Menurut McKinsey, pada tahun 2022 pasar e-commerce Indonesia diprediksikan akan meningkat delapan kali lipat [5]. Berikut adalah gambaran prediksi peningkatan data e-commerce di Indonesia :



Gambar I-1 Prediksi Peningkatan E-Commerce

Pada paparan diatas adapun metode lain yang digunakan melengkapi hasil survei yang diusulkan penulis untuk memprediksi tingkat kemiskinan disuatu daerah adalah menggunakan *machine learning* metode *decision tree* dan *wrapper feature selection* yang dibantu dengan data e-commerce yang diperoleh dari salah satu perusahaan e-commerce di Indonesia untuk mendapatkan data prediksi kemiskinan.

## II. METODE PENELITIAN

Metodologi yang digunakan dalam pengerjaan Aplikasi Pengolahan Data Prediksi Kemiskinan Berbasis E-commerce Menggunakan *Decision Tree* dan *Wrapper feature selection* berupa :

### 1. Penentuan Topik

Pada metode pengerjaan ini peneliti memilih topik yang nantinya dijadikan judul proyek akhir, setelah peneliti mempunyai berbagai macam topik akhirnya memutuskan topik tentang Aplikasi Pengolahan Data Prediksi Kemiskinan Berbasis E-commerce Menggunakan *Decision Tree* dan *Wrapper feature selection*.

### 2. Identifikasi Masalah

Pada metode pengerjaan ini peneliti mengidentifikasi masalah-masalah yang sering dihadapi oleh pemerintah Indonesia tentang prediksi tingkat kemiskinan.

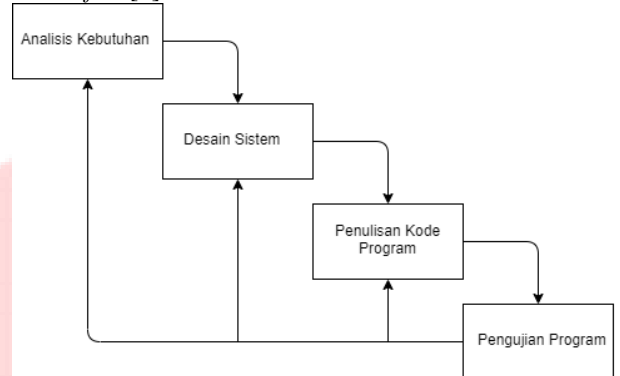
### 3. Studi Literatur

Pada pembuatan laporan ini peneliti menentukan fakta apa saja yang terkait dengan kemiskinan di Indonesia dan metode-metode untuk memprediksikan tingkat kemiskinan.

### 4. Perancangan Sistem

Perancangan sistem Aplikasi Pengolahan Data Prediksi

Kemiskinan Berbasis E-commerce menggunakan metode *machine learning decision tree* dan metode *wrapper feature selection* menggunakan metode pengembangan perangkat lunak SDLC (Software Development Life Cycle) dengan pendekatan *Waterfall*. Berikut adalah tahapan SDLC *Waterfall* [6]:



Gambar II-1 Metode Waterfall

Tahapan metode *waterfall* [7]:

#### 4.1 Analisis Kebutuhan

Dalam perancangan system, perlu adanya kebutuhan untuk membangun sebuah aplikasi. Dalam tahap ini, penulis harus mengetahui apa keinginan dari *user* dan batasan aplikasi. Beberapa hal yang harus dilakukan dalam pengumpulan data perangkat lunak pada proyek akhir ini, yaitu :

1. Melakukan wawancara dengan BPS Bandung untuk melihat proses sunsenas yang akan mempresentasikan tingkat kemiskinan.
2. Tinjauan pustaka yaitu dengan mencari referensi dari buku, jurnal dan *website* yang berhubungan dengan aplikasi yang akan dibangun.

#### 4.2 Desain Sistem

Dalam tahap desain sistem penulis melakukan pembuatan untuk memetakan proses bisnis dengan menggunakan tools BPMN, menentukan entitas dan atribut yang ada pada database dengan menggunakan ERD dan database yang digunakan yaitu MySQL untuk menyimpan semua data-data, menentukan fitur-fitur dan tampilan dengan menggunakan bahasa pemrograman *Python*.

#### 4.3 Penulisan Kode Program

Dalam tahap pengkodean, dilakukan pengkodean dari hasil perancangan dari aplikasi mengenai fungsionalitas yang akan diajukan kepada *user*

pengguna. Tools yang digunakan dalam pengkodean program yaitu Bahasa Pemrograman *Python* dan menggunakan *library sklearn.tree.DecisionTreeRegressor*.

#### 4.4 Pengujian Program

Setelah penulis melakukan pengkodean program, aplikasi akan diuji melalui pengujian *Black Box Testing* dan *User Acceptance Testing*. Pengujian ini bertujuan untuk memastikan aplikasi yang sudah dibangun terdapat *error* atau tidak, dan dapat mengetahui logika yang tidak sejalan dengan alur proses bisnis yang semestinya serta perangkat lunak diserahkan kepada pengguna untuk mengetahui apakah aplikasi yang dibangun sudah sesuai dengan harapan pengguna dan berfungsi sebagaimana layaknya.

### III. TINJAUAN PUSTAKA

Berikut merupakan beberapa teori pokok pembahasan yang sesuai dengan aplikasi yang dibangun dalam proyek akhir ini.

#### A. Machine Learning

*Machine learning* merupakan salah satu cabang dari disiplin ilmu Kecerdasan Buatan (*Artificial Intelligence*) yang membahas mengenai pembangunan sistem yang berdasarkan pada data. Banyak hal yang dipelajari, akan tetapi pada dasarnya ada 4 hal pokok yang dipelajari dalam *machine learning* [8].

##### 1. Pembelajaran Terarah (*Supervised Learning*).

Pada model ini sering disebut sebagai model terarah karena umumnya diberi instruksi yang jelas seperti apa saja yang perlu dipelajari dan bagaimana hal tersebut dapat dipelajari. Model ini umumnya digunakan untuk memprediksi masa depan berdasarkan data *historiys*. *Supervised learning* dibagi menjadi dua yakni [9]:

##### a. Classification

Pada metode ini paling umum untuk digunakan pada *data mining*. Dalam metode ini setiap atribut atau fitur harus diberikan label supaya komputer bisa mengetahui atau mengklasifikasikan sebuah objek dengan menggunakan label tersebut.

##### b. Regresion

Pada metode ini sebenarnya sama dengan metode *classification* namun pada metode ini diperuntukkan untuk membuat sebuah pola pada setiap atributnya. Pada metode ini bertujuan untuk mencari sebuah pola dan menentukan sebuah nilai numerik.

##### 2. Pembelajaran Tak Terarah (*Unsupervised Learning*).

Dalam metode ini tidak diberikan label, tetapi secara otomatis dibagi berdasarkan kemiripan dan struktur lain dari data tersebut.

#### B. Feature Selection

*Feature selection* atau seleksi fitur adalah salah satu teknik terpenting dan sering digunakan dalam *pre-processing*. Teknik ini mengurangi jumlah fitur yang terlibat dalam menentukan suatu nilai kelas target, mengurangi fitur *irelevan*, berlebihan dan data yang menyebabkan salah pengertian terhadap kelas target [10]. *Feature selection* dibagi menjadi dua kelompok yakni [10]:

##### 1. Rangking Selection

Pada kelompok ini, setiap fitur yang terdapat dalam data diberikan rangking dan mengesampingkan fitur yang tidak memenuhi syarat atau standar. *Rangking selection* bertujuan untuk menentukan tingkat rangking secara *independent* antara fitur yang satu dengan fitur yang lain.

##### 2. Subset Selection

Pada kelompok ini digunakan untuk mencari set dari fitur yang terdapat dalam data yang dianggap sebagai fitur yang optimal. Terdapat tiga jenis metode dalam *subset selection* yakni [10]:

##### a. Feature selection tipe wrapper

Pada tipe ini yang dilakukan adalah pemilihan secara bersamaan dengan pelaksanaan pemodelan.

##### b. Feature selection tipe filter

Pada tipe ini yang dilakukan adalah memanfaatkan salah satu fitur dari beberapa fitur yang terdapat dalam data.

##### c. Feature selection embedded

Pada tipe ini yang dilakukan adalah memanfaatkan suatu *machine learning* dalam proses *feature selection*. Fitur yang dianggap tidak berpengaruh dalam pengolahan data akan secara otomatis dihilangkan.

#### C. Decision Tree

Metode ini merupakan salah satu metode yang ada pada teknik klasifikasi dalam data mining. Metode *decision tree* ini untuk mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target [11]. Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan record. Atribut menyatakan suatu parameter yang disebut sebagai kriteria dalam pembentukan pohon. Misalkan untuk menentukan main tenis, kriteria yang diperhatikan adalah cuaca, angin, dan suhu. Salah satu atribut merupakan atribut yang menyatakan data solusi per item data yang disebut atribut hasil [12].

$$J(k, t_k) = \frac{m_{left}}{m} MSE_{left} + \frac{m_{right}}{m} MSE_{right}$$

$$\text{Where } \begin{cases} MSE_{node} = \sum_{i \in node} (\hat{y}_{node} - y^{(i)})^2 \\ \hat{y}_{node} = \frac{1}{m_{node}} \sum_{i \in node} y^{(i)} \end{cases} \quad (1).$$

#### D. Data Mining

Data mining merupakan proses yang memanfaatkan Teknik matematika, statistic dan kecerdasan buatan untuk mengidentifikasi informasi atau pola – pola yang valid, baru, memiliki potensi bermanfaat dan bisa dipahami dari sekumpulan data yang besar. Pola – pola tersebut bias dalam bentuk bisnis, kolerasi, trend atau model – model prediksi. Dalam proses data mining terdapat banyak langkah perulangan yang rumit yang dimana ada suatu digaan atau kesimpulan yang berbasis eksperimentasi yang dilibatkan [13].

#### E. Library Python Scikit-learn

*Scikit-Learn* merupakan suatu *library* untuk pengguna *python* pada *machine learning*. *Scikit-learn* ini merupakan free software yang dapat digunakan untuk melakukan berbagai pekerjaan dalam data *science* seperti regresi (*regression*), klasifikasi (*classification*), Pengelompokan (*clustering*), data *preprocessing*, *dimensionality reduction* dan model *selection* seperti perbandingan, validasi dan pemilihan parameter maupun model [14].

#### F. Library Python Scikit-Feature

*Scikit-feature* merupakan repositori *python* dari pemilihan fitur *open-resource* yang dikembangkan oleh Arizona State University. *Scikit-feature* ini memiliki 40 algoritma pemilihan suatu fitur, termasuk pemilihan fitur tradisional dan beberapa algoritma pemilihan fitur struktural dan *streaming*. *Scikit-feature* berfungsi sebagai *platform* yang memfasilitasi aplikasi pemilihan suatu fitur. Saat ini fitur *Scikit-feature* terdiri dari beberapa algoritma yakni *similarity based feature selection*, *information theoretical based feature selection*, *sparse learning based feature selection*, *statistical based feature selection*, *wrapper based feature selection*, *structural feature selection*, *streaming feature selection* [15].

#### G. Normalisasi

Normalisasi dalam kegiatan data mining merupakan proses penskalaan nilai atribut dari data sehingga bisa jatuh pada range tersebut. Ada beberapa metode yang digunakan untuk proses Normalisasi yaitu [16]:

##### 1. Min-Max

Metode *min-max* merupakan metode Normalisasi dengan melakukan transformasi *linier* terhadap data asli.

$$\text{MinMax} = \frac{x - \min(x)}{\max(x) - \min(x)} \times 10 \quad (2)$$

Keterangan :

$x$  = Value of each feature.

$\min(x)$  = The lowest value of each feature.

$\max(x)$  = The highest value of each feature

##### 2. Z-Score

Metode *Z-score* merupakan metode Normalisasi yang berdasarkan *mean* (nilai rata-rata) dan *standard deviation* (deviasi standar) dari data.

##### 3. Decimal Scaling

Metode *decimal scaling* merupakan metode Normalisasi dengan menggerakkan nilai *decimal* dari data kearah yang diinginkan.

##### 4. Sigmoidal

*Sigmoidal* merupakan metode *normalization* melakukan Normalisasi data secara *nonlinier* ke dalam range -1 - 1 dengan menggunakan fungsi *sigmoid*. Metode ini sangat berguna pada saat data-data yang ada melibatkan data-data *outlier*. Data *outlier* data yang keluar jauh dari jangkauan data lainnya.

#### H. Bahasa Pemrograman Python

*Python* adalah sebuah bahasa pemrograman *scripting* tingkat tinggi atau *high-level*, *interpreted*, *interactive*, dan *object-oriented*. *Python* dengan desain yang sangat mudah di baca dan dipahami, karena sama seperti bahasa pemrograman yang lainnya yaitu dengan menggunakan kata bahasa Inggris. Selain itu juga lebih sedikit dalam penggunaan rumus atau *syntac* [17].

#### I. R Squared

R-square merupakan ukuran statistik yang dimana antara 0 dan 1 yang berfungsi untuk menghitung seberapa mirip hasil prediksi dengan data aslinya yang ditandai dengan semakin dekatnya hasil prediksi dengan garis aslinya [18].

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^L (y_i - \hat{y}_i)^2}{\sum_{i=1}^L (y_i - \bar{y})^2} \quad (3)$$

#### J. RMSE (Root Mean Square Error)

RMSE merupakan suatu metode yang digunakan untuk mengevaluasi tingkat prediksi yang digunakan untuk mengukur tingkat akurasi pada suatu model. RMSE adalah nilai rata-rata dari jumlah kuadrat sebuah kesalahan atau mengukur tingkat kesalahan pada prediksi yang dihasilkan suatu model. Nilai RMSE rendah menunjukkan bahwa nilai prediksi yang dihasilkan mendekati nilai observasinya. Berikut merupakan rumus dari RMSE [18].

$$\text{RMSE}(y, \hat{y}) = \sqrt{\frac{\sum_{i=1}^L (y_i - \hat{y}_i)^2}{L}} \quad (4)$$

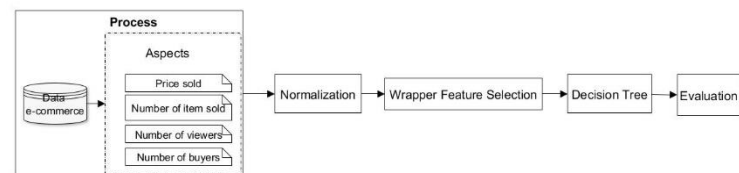
### IV. ANALISIS DAN PERANCANGAN

#### A. Gambaran Sistem Usulan

Berikut merupakan gambaran sistem usulan dari preprocessing hingga sistem usulan aplikasi :

##### 1. Gambaran Sistem Usulan Alur Pengembangan Model

Berikut merupakan gambaran alur pengembangan model.



Gambar IV-1 Sistem Usulan Alur Pengembangan Model

Pada gambaran diatas terdapat beberapa proses sebelum data *e-commerce* diolah ke dalam aplikasi berbasis *dataset e-commerce* untuk prediksi tingkat kemiskinan menggunakan algoritma *decision tree* dan *wrapper feature selection*.

##### a. Proses Data

Pada tahap proses data ini diperlukan karena pada tahap ini data akan dibersihkan dari data null atau *missing value* diubah menjadi 0 karena jika data yang digunakan masih terdapat *missing value* maka dapat menyebabkan data tidak siap untuk diproses kedalam proses lanjut. Pada data yang digunakan penulis masih terdapat *missing value* sehingga data yang *missing value* harus diganti dengan 0. Dalam proses data melihat aspek-aspek pada data *e-commerce* meliputi harga yang terjual, jumlah barang yang terjual, jumlah viewer, dan jumlah pembeli.

##### b. Normalisasi

Pada tahap kedua ini merupakan tahap normalisasi. Pada tahap normalisasi ini, data yang telah melalui tahap proses data akan dirasionalkan nilainya menjadi 0 – 10 menggunakan metode *min-max normalization* atau yang dikenal dengan *Rescaling*. Rumus *Rescaling* ditulis sebagai berikut.

$$\text{MinMax} = \frac{x - \min(x)}{\max(x) - \min(x)} \times 10 \quad (2)$$

Keterangan :

$x$  = nilai dari masing – masing fitur.

$\min(x)$  = nilai terendah dari setiap fitur.

$\max(x)$  = nilai tertinggi dari setiap fitur.

##### c. Wrapper Feature Selection

Metode *wrapper* membutuhkan satu algoritma *Machine learning* dan menggunakan kerjanya sebagai kriteria evaluasi. Metode ini

mencari fitur yang paling cocok untuk algoritma *machine learning* dan bertujuan untuk meningkatkan kinerja mining [15]. Langkah selanjutnya setelah normalisasi data yaitu ke proses seleksi fitur yakni *wrapper based feature selection* yang memiliki empat algoritma didalamnya seperti *svm\_forward*, *svm\_backward*, dan *decision\_tree\_forward*. Proses selanjutnya yaitu melihat skor dari masing – masing fitur, kemudian masuk kedalam proses berikutnya yakni proses *filtering* fitur, pada proses ini data akan diseleksi untuk menentukan fitur mana yang relevan.

d. *Decision Tree*

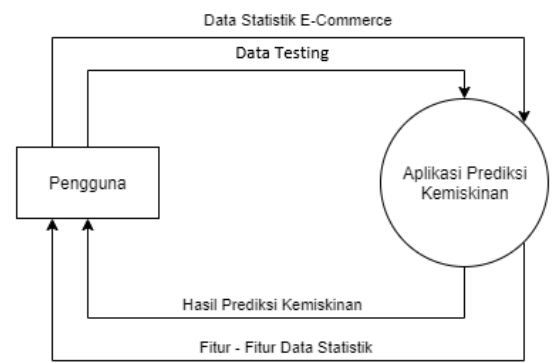
Tahapan selanjutnya adalah data yang sudah di filter melalui wrapper feature selection akan dimasukan kedalam *machine learning decision tree*. Machine learning ini merupakan salah satu metode yang ada pada teknik klasifikasi dan regresi dalam data mining. Machine learning *decision tree* ini untuk mengubah fakta yang sangat besar sebagai pohon keputusan yang menyajikan sebuah aturan serta untuk menggali data, mendapatkan ikatan tersembunyi antara sebuah variable target dengan sejumlah calon variable input. Pernyataan dalam bentuk tabel dengan atribut dan record umumnya digunakan dalam data decision tree. Atribut ini merupakan parameter dalam pembentukan pohon, salah satu contohnya adalah pada permainan tenis, parameter yang diperhatikan ialah cuaca, angin, dan suhu. Atribut hasil merupakan salah satu yang menyatakan solusi per item data [11].

e. Evaluasi

Tahapan terakhir adalah mengukur kesesuaian data menggunakan dua matriks yaitu RMSE dan  $R^2$  (*R-Square*). Fungsi RMSE adalah untuk mengukur kesalahan antara prediksi dengan *vector actual*, sedangkan fungsi  $R^2$  berupa bagian yang diwakili dari *varians vector* yang bisa diproyeksikan pada model regresi. Jika nilai  $R^2$  semakin besar maka hasil prediksi dapat disimpulkan nilai akurasi mendekati *vector actual*.

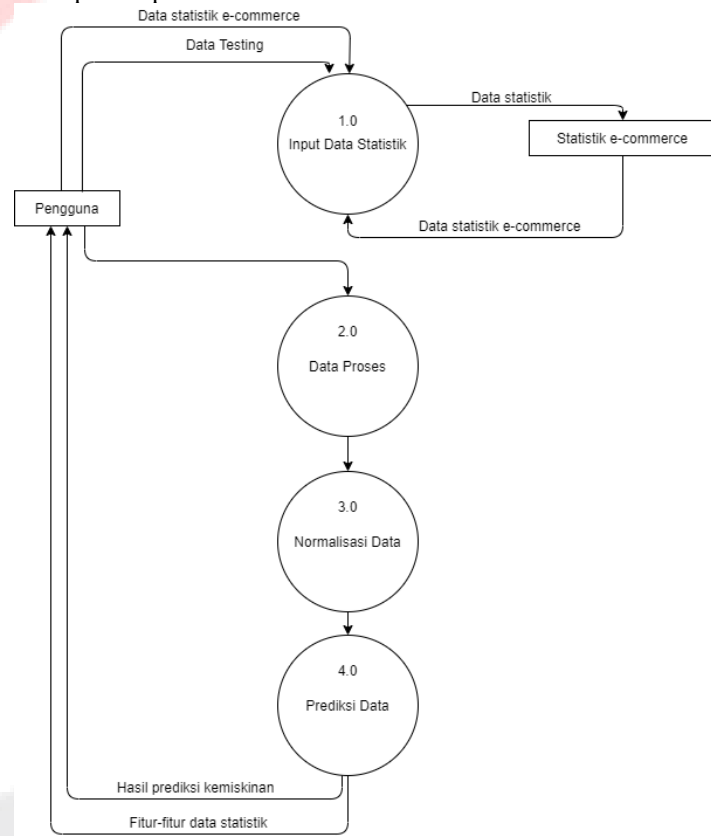
2. Gambaran Sistem Usulan Aplikasi

Pada sistem usulan ini, penulis mengusulkan menggunakan *dataset E-commerce* untuk menentukan *presentase* tingkat kemiskinan dengan aplikasi *Machine learning* menggunakan algoritma *Decision Tree* dan *Wrapper feature selection*. Alasan mengapa penulis menggambarkannya dengan data flow diagram karena pada aplikasi ini termasuk *structural programming* yang dimana pemrograman bertumpu pada pemanggilan *library* yang telah didefinisikan sebelumnya serta data flow diagram dapat menggambarkan fungsi, proses, penangkapan data, memanipulasi, menyimpan, mendistribusikan data antara suatu sistem pada lingkungannya serta antara komponen-komponen suatu sistem. Berikut adalah gambaran DFD yang diusulkan :



Gambar IV-2 Diagram Konteks

Pada gambaran ini, pengguna atau user yang menggunakan aplikasi prediksi kemiskinan dapat melihat grafik-grafik dan presentase tingkat kemiskinan disuatu daerah. Proses penggunaan aplikasi ini adalah user harus mempunyai data *ecommerce* sebagai data yang akan diolah. Berikut adalah alur dari aplikasi prediksi kemiskinan :



Gambar IV-3 DFD Level 1 Proses Prediksi Kemiskinan

Pada proses ini, data *e-commerce* yang akan diproses akan menuju proses input data statistik untuk melakukan pengolahan data, selanjutnya data tersebut akan menghasilkan data statistik yang disimpan kedalam statistik *e-commerce*, proses selanjutnya adalah kembali ke proses input data statistik supaya menghasilkan *dataset* yang nantinya akan digunakan user untuk melanjutkan proses selanjutnya berupa proses prediksi data. Pada proses prediksi data akan muncul grafik dan *presentase* tingkat kemiskinan disuatu daerah, dari hasil prediksi data maka kegiatan *user* yang telah menginputkan data *e-commerce* maka akan disimpan pada *history* prediksi.

V. IMPLEMENTASI DAN PENGUJIAN

A. Implementasi

Setelah tahap analisis dan perancangan, tahap selanjutnya adalah tahap implementasi dari aplikasi berbasis dataset *e-*

commerce untuk prediksi tingkat kemiskinan menggunakan algoritma decision tree dan wrapper feature selection. Berikut merupakan pengkodean dari beberapa fitur yang digunakan.

1) Halaman Registrasi.



Gambar V-1 Halaman Registrasi

Pada gambar ini, user akan melakukan registrasi dengan mengisi setiap form yang tersedia, jika ada salah satu kolom tidak diisi maka proses registrasi akan gagal.

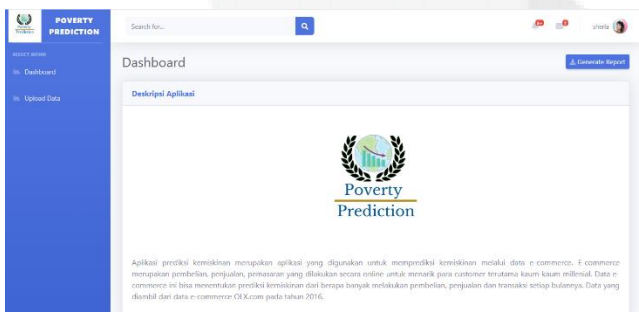
2) Halaman Login



Gambar V-2 Halaman Login

Setelah user memiliki akun, user dapat login ke aplikasi melalui halaman login, dalam halaman ini, user diharuskan mengisi kolom username dan password yang sudah dibuat sebelumnya. Jika user belum memiliki akun, maka user tersebut tidak bisa mengakses aplikasi ini.

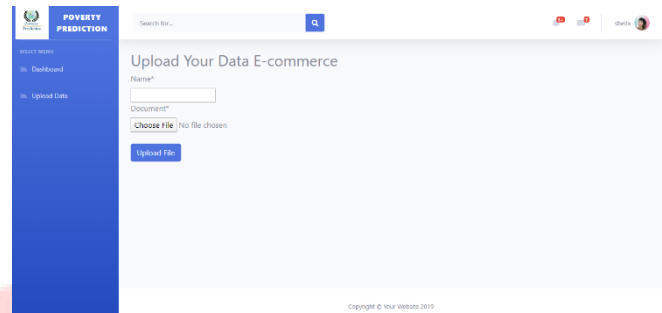
3) Halaman Dashboard



Gambar V-3 Halaman Dashboard

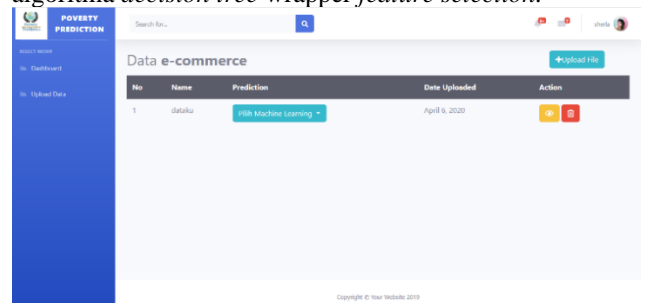
Pada gambar tersebut user mengisi username dan password yang sudah terdaftar untuk melakukan login.

4) Halaman Upload Data dan Halaman List Data



Gambar V-4 Halaman Upload Data

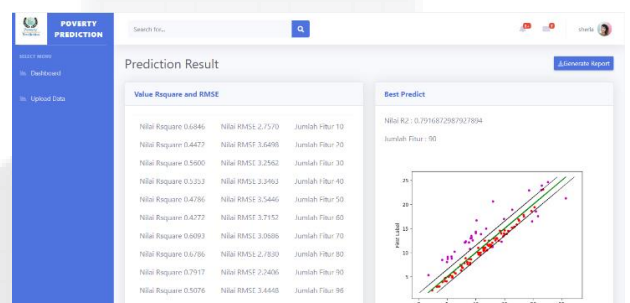
Pada halaman tersebut terdapat form upload data dan nama file untuk diprediksi. Format data yang diupload harus csv, karena template disediakan pada halaman list data. Berikut adalah halaman list data pada aplikasi berbasis dataset e-commerce untuk prediksi tingkat kemiskinan menggunakan algoritma decision tree wrapper feature selection.



Gambar V-5 Halaman List Data

Pada gambar tersebut user dapat menambah data dengan menekan button upload file. Button pilih machine learning merupakan dropdown untuk memilih algoritma machine learning yang tersedia berupa k-NN, Decision Tree, Naïve Bayes, Logistic Regression.

5) Halaman Hasil Prediksi



Gambar V-6 Halaman Hasil Prediksi

Pada gambar tersebut merupakan hasil prediksi data e-commerce yang diupload dan user dapat melihat nilai R2, RMSE. Kolom-kolom pada data tersebut akan ditampilkan data actual dan data prediksi.

B. Pengujian

Setelah melalui tahap implementasi maka aplikasi masuk kedalam proses selanjutnya yakni proses pengujian menggunakan black box testing. Berikut merupakan beberapa pengujian dengan menggunakan beberapa fitur algoritma yang dilakukan.

## 1) Pengujian Login

Pengujian dilakukan untuk menguji kesesuaian fungsionalitas login dengan spesifikasi kebutuhan pengguna. Berikut merupakan *scope of testing login*.

Tabel V-1 Scope of Testing Login

Perangkat Lunak	Aplikasi Pengolahan Data Prediksi Kemiskinan Berbasis E-Commerce Menggunakan Decision Tree Dan Wrapper Feature Selection.
Deskripsi	Aplikasi yang digunakan untuk memprediksi tingkat kemiskinan disuatu daerah.

Fungsi	Login
Aturan	<ol style="list-style-type: none"> <li>1. Username dan password harus diisi sesuai dengan data registrasi</li> <li>2. Username dan password tidak diisi sesuai dengan data registrasi</li> <li>3. Username dan password tidak diisi</li> </ol>

Tabel V-2 Test Case Matrix Function Login

No.	Function/Condition	Case No.	Test Case Description (Event)	Test Data (Input)	Expected Result	Actual Result / Comments	Conclusion/Kesimpulan
1.	Login	1.	Entry data login dengan mengikuti aturan 1 : 1. Username 2. Password	1. Username = pangestuade 2. Password = ade12345	Aplikasi menampilkan halaman <i>dashboard</i>	Aplikasi menampilkan halaman <i>dashboard</i>	Valid
		2.	Entry data login dengan mengikuti aturan 2 : 1. Username 2. Password	1. Username = pangestuade 2. Password = adepangestu	Aplikasi menampilkan Pesan error message "invalid creditials"	Aplikasi menampilkan Pesan error message "invalid creditials"	Valid
		3.	Entry data login dengan mengikuti aturan 3 : 1. Username 2. Password	1. Username = 2. Password =	Aplikasi menampilkan Pesan error message "invalid creditials"	Aplikasi menampilkan Pesan error message "invalid creditials"	Valid

## 2) Pengujian Registrasi

Pengujian dilakukan untuk menguji kesesuaian fungsionalitas registrasi dengan spesifikasi kebutuhan pengguna. Berikut merupakan *scope of testing registrasi*.

Tabel V-3 Scope of Testing Registrasi

Perangkat Lunak	Aplikasi Pengolahan Data Prediksi Kemiskinan Berbasis E-Commerce Menggunakan Decision Tree Dan Wrapper Feature Selection.
-----------------	---

Deskripsi	Aplikasi yang digunakan untuk memprediksi tingkat kemiskinan disuatu daerah.
Fungsi	Registrasi
Aturan	<ol style="list-style-type: none"> <li>1. First name, last name, email, username, password dan repeat password harus diisi</li> <li>2. First name, last name, email, username, password dan repeat password tidak diisi</li> </ol>

	3. Email tidak diisi.		4. Password dan repassword tidak sama
--	-----------------------	--	---------------------------------------

Tabel V-4 Test Case Matrix Function Registrasi

No.	Function/ Condition	Case No.	Test Case Description (Event)	Test Data (Input)	Expected Result	Actual Result / Comments	Conclution/ Kesimpulan
1.	Registrasi	1.	Entry data login dengan mengikuti aturan 1 : 1. First Name 2. Last Name 3. Username 4. Email 5. Password 6. Repeat Password	1. First Name = ade 2. Last Name = pangestu 3. Username = pangestuade 4. Email = pangestuade123@gmail.com 5. Password = ade12345 6. Repeat Password = ade12345	Aplikasi menampilkan halaman login	Aplikasi menampilkan halaman login	Valid
		2.	Entry data login dengan mengikuti aturan 2 : 1. First Name 2. Last Name 3. Username 4. Email 5. Password 6. Repeat Password	1. First Name = 2. Last Name = 3. Username = 4. Email = 5. Password = 6. Repeat Password =	Aplikasi menampilkan Pesan error message "all form must be set"	Aplikasi menampilkan Pesan error message "all form must be set"	Valid
		3.	Entry data login dengan mengikuti aturan 3 : 1. First Name 2. Last Name 3. Username 4. Email 5. Password	1. First Name = ade 2. Last Name = pangestu 3. Username = pangestuade 4. Email = 5. Password = ade12345 6. Repeat Password = ade12345	Aplikasi menampilkan Pesan error message "all form must be set"	Aplikasi menampilkan Pesan error message "all form must be set"	Valid



No.	Function/ Condition	Case No.	Test Case Description (Event)	Test Data (Input)	Expected Result	Actual Result / Comments	Conclution/ Kesimpulan
			6. Repeat Password				
		4.	Entry data login dengan mengikuti aturan 4 : 1. First Name 2. Last Name 3. Username 4. Email 5. Password 6. Repeat Password	1. First Name = ade 2. Last Name = pangestu 3. Username = pangestuade 4. Email = pangestuade123@gmail.com 5. Password = ade12345 6. Repeat Password = adepangestu	Aplikasi menampilkan Pesan error message "Password not maching"	Aplikasi menampilkan Pesan error message "Password not maching"	Valid

## 3) Pengujian Upload File

Pengujian dilakukan untuk menguji kesesuaian fungsionalitas *upload file* dengan spesifikasi kebutuhan pengguna. Berikut merupakan *scope of testing* dari *upload file*.

Tabel V-5 Scope of Testing Upload File

Perangkat Lunak	Aplikasi Pengolahan Data Prediksi Kemiskinan Berbasis E-Commerce Menggunakan Decision Tree Dan Wrapper Feature Selection.
-----------------	---

Deskripsi	Aplikasi yang digunakan untuk memprediksi tingkat kemiskinan disuatu daerah.
Fungsi	Upload file
Aturan	<ol style="list-style-type: none"> <li><b>Name dan document diisi</b></li> <li>Name dan document tidak diisi</li> </ol>

Tabel V-6 Test Case Matrix Function Upload File

No.	Function/ Condition	Case No.	Test Case Description (Event)	Test Data (Input)	Expected Result	Actual Result / Comments	Conclution/ Kesimpulan
1.	Upload File	1.	Entry data login dengan mengikuti aturan 1 : 1. Name 2. Document	1. Name = dataset 2. Document = testing.csv	Aplikasi menampilkan halaman <i>list</i> <i>document</i>	Aplikasi menampilkan halaman <i>list</i> <i>document</i>	Valid
		2.	Entry data login dengan mengikuti aturan 2 : 1. Name 2. Document	1. Name = 2. Document =	Aplikasi menampilkan Pesan error message	Aplikasi menampilkan Pesan error message	Valid

- 4) Pengujian Forgot Password  
Pengujian dilakukan untuk menguji kesesuaian fungsionalitas *forgot password* dengan spesifikasi kebutuhan pengguna. Berikut merupakan *scope of testing* dari *forgot password*.

Tabel V-7 Scope of Testing Forgot Password

Perangkat Lunak	Aplikasi Pengolahan Data Prediksi Kemiskinan Berbasis E-Commerce Menggunakan Decision Tree Dan Wrapper Feature Selection.
Deskripsi	Aplikasi yang digunakan untuk memprediksi tingkat kemiskinan disuatu daerah.
Fungsi	Forgot Password
Aturan	<ol style="list-style-type: none"> <li>1. <b>Email harus diisi</b></li> <li>2. Email dikosongkan</li> <li>3. <b><i>New Password dan new password confirmation</i> tidak boleh sama dengan data diri</b></li> </ol>

<ol style="list-style-type: none"> <li>4. <i>New Password dan new password confirmation</i> sama dengan data diri</li> <li>5. <b><i>New Password dan new password confirmation minimal 8 karakter</i></b></li> <li>6. <i>New Password dan new password confirmation</i> kurang dari 8 karakter</li> <li>7. <b><i>New Password dan new password confirmation</i> tidak boleh seluruhnya bersifat numerik</b></li> <li>8. <i>New Password dan new password confirmation</i> seluruhnya bersifat numerik</li> </ol>
--

Tabel V-8 Test Case Matrix Function Forgot Password

No.	Function/Condition	Case No.	Test Case Description (Event)	Test Data (Input)	Expected Result	Actual Result / Comments	Conclusion/ Kesimpulan
1.	Forgot Password	1.	Entry data <i>forgot password</i> dengan mengikuti aturan 1 :	1. Email = pangestuade123@gmail.com	Aplikasi menampilkan halaman <i>Password reset sent</i>	Aplikasi menampilkan halaman <i>Password reset sent</i>	Valid
		2.	Entry data <i>forgot password</i> dengan mengikuti aturan 2 :	1. Email =	Aplikasi menampilkan Pesan error message	Aplikasi menampilkan Pesan error message	Valid
		3.	Entry data <i>forgot password</i> dengan mengikuti aturan 3 :	1. <i>New Password</i> = naruto12345 2. <i>New Password Confirmation</i> = naruto12345	Aplikasi menampilkan halaman <i>Password reset Completed</i>	Aplikasi menampilkan halaman <i>Password reset Completed</i>	Valid

No.	Function/ Condition	Case No.	Test Case Description (Event)	Test Data (Input)	Expected Result	Actual Result / Comments	Conclusion/ Kesimpulan
			<i>New Password Confirmation</i>				
		4.	Entry data <i>forgot password</i> dengan mengikuti aturan 4 :  1. <i>New Password Confirmation</i>	1. <i>New Password</i> = adepangestu  2. <i>New Password Confirmation</i> = adepangestu	Aplikasi menampilkan Pesan error message	Aplikasi menampilkan Pesan error message	Valid
		5.	Entry data <i>forgot password</i> dengan mengikuti aturan 5 :  1. <i>New Password Confirmation</i>	1. <i>New Password</i> = sasuke12345  2. <i>New Password Confirmation</i> = sasuke12345	Aplikasi menampilkan halaman <i>Password reset Completed</i>	Aplikasi menampilkan halaman <i>Password reset Completed</i>	Valid
		6.	Entry data <i>forgot password</i> dengan mengikuti aturan 6 :  1. <i>New Password Confirmation</i>	1. <i>New Password</i> = naruto  2. <i>New Password Confirmation</i> = naruto	Aplikasi menampilkan Pesan error message	Aplikasi menampilkan Pesan error message	Valid
		7.	Entry data <i>forgot password</i> dengan mengikuti aturan 7 :  1. <i>New Password Confirmation</i>	1. <i>New Password</i> = naruto12345  2. <i>New Password Confirmation</i> = naruto12345	Aplikasi menampilkan halaman <i>Password reset Completed</i>	Aplikasi menampilkan halaman <i>Password reset Completed</i>	Valid
		8.	Entry data <i>forgot password</i> dengan mengikuti aturan 7 :  1. <i>New Password Confirmation</i>	1. <i>New Password</i> = 8080808080  2. <i>New Password Confirmation</i> = 8080808080	Aplikasi menampilkan Pesan error message	Aplikasi menampilkan Pesan error message	Valid

## VI. KESIMPULAN

Adapun kesimpulan yang di dapat oleh penulis adalah, sebagai berikut :

1. Aplikasi berhasil menampilkan fitur prediksi kemiskinan dengan data e-commerce berbasis decision tree dan algoritma wrapper feature selection.
2. Item yang berpengaruh dalam prediksi kemiskinan pada data e-commerce adalah motor dengan 13 fitur, mobil dengan 11 fitur, rumah dijual dengan 5 fitur, rumah disewa dengan 3 fitur, apartemen dijual dengan 5 fitur, apartemen disewa dengan 3 fitur, tanah dijual dengan 3 fitur, tanah disewa dengan 7 fitur.
3. Grafik yang ditampilkan merupakan perhitungan prediksi pada algoritma berupa nilai  $R^2$  dan RMSE.

Kelebihan fitur ini adalah akurasinya lebih tinggi daripada filter lain. Sedangkan Kelemahan dari metode ini adalah terlalu lama untuk memproses data. Ini karena pemilihan fitur pembungkus dalam pekerjaan mencoba untuk mengelola semua fitur yang ada.

## REFERENSI

- [1] M. A. Lingga, "Kemiskinan Masih Jadi Tantangan Besar Indonesia," *Kompas.com*, 2018. .
- [2] B. P. Statistik, "Badan Pusat Statistik," *Badan Pusat Statistik*, 2019. [Online]. Available: <https://www.bps.go.id/subject/23/kemiskinan-dan-ketimpangan.html>. [Accessed: 23-Sep-2019].
- [3] B. P. Statistik, *STATISTIK KESEJAHTERAAN RAKYAT WELFARE STATISTICS 2018*. 2018.
- [4] Statistica 2016, "Transaksi E-Commerce Indonesia Naik 500% dalam 5 Tahun," *databoks*, 2016. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2016/11/16/transaksi-e-commerce-indonesia-naik-500-dalam-5-tahun>. [Accessed: 19-Sep-2019].
- [5] D. Praditya, "Prediksi Perkembangan Industri E-commerce Indonesia pada Tahun 2022," *techinasia*, 2019. [Online]. Available: <https://id.techinasia.com/prediksi-e-commerce-indonesia>.
- [6] Fitho Galandi, "Metode Waterfall : Definisi, Tahapan, Kelebihan dan Kekurangan," *Pengetahuan dan Teknologi*, 2016. [Online]. Available: <http://www.pengetahuandanteknologi.com/2016/09/metode-waterfall-definisi-tahapan.html>.
- [7] I. Fahrurrozi and A. SN, "Proses pemodelan software dengan metode waterfall dan extreme programming : Studi kasus perbandingan," *Univ. Gajah Mada*, pp. 1–10, 2015.
- [8] Mr BigsmiLe, "Mengenal Teknologi Machine Learning (Pembelajaran Mesin)," *CODEPOLITAN*, 2016. [Online]. Available: <https://www.codepolitan.com/mengenal-teknologi-machine-learning-pembelajaran-mesin>. [Accessed: 19-Sep-2019].
- [9] B. Lisam, "Apa Itu Machine Learning? Berikut Penjelasannya," *PodFeeder*, 2019. [Online]. Available: <http://www.podfeeder.com/teknologi/apa-itu-machine-learning-berikut-penjasannya/>. [Accessed: 23-Sep-2019].
- [10] P. S. R. Rajaleximi, M. S. I. Ahmed, and A. Alenezi, "Feature selection using optimized multiple rank score model for credit scoring," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 2, pp. 74–84, 2019.
- [11] Rismayanti, "Decision Tree Penentuan Masa Studi Mahasiswa Prodi Teknik Informatika ( Studi Kasus : Fakultas Teknik dan Komputer Universitas Harapan Medan )," *Query*, vol. 5341, no. April, pp. 16–24, 2018.
- [12] A. Géron, *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems*. 2017.
- [13] M. K. Albert Verasius Dian Sano, S.T., "DEFINISI, KARAKTERISTIK, DAN MANFAAT DATA MINING -SERI DATA MINING FOR BUSINESS INTELLIGENCE (2)," *BINUS UNIVERSITY*, 2019. [Online]. Available: <https://binus.ac.id/malang/2019/01/definisi-karakteristik-dan-manfaat-data-mining-seri-data-mining-for-business-intelligence-2/>.
- [14] H. Azizul, "Berkenalan dengan scikit-learn (Part 1) – Preparations," *hkaLabs*, 2018. [Online]. Available: <https://hakim-azizul.com/berkenalan-dengan-scikit-learn/>.
- [15] J. Li, "Data Mining, Data Science, Feature Extraction, Feature Selection, Machine Learning, Python," *kdnuggets.com*, 2016. [Online]. Available: <https://www.kdnuggets.com/2016/03/scikit-feature-open-source-feature-selection-python.html>.
- [16] A. Chandra, "Beberapa Cara untuk Preprocessing Data dalam Machine Learning," *Medium.com*, 2018. [Online]. Available: <https://medium.com/warung-pintar/beberapa-cara-untuk-preprocessing-data-dalam-machine-learning-13cef4294614>.
- [17] R. Amaliyah, "Pengertian Python Beserta Kelebihan dan Kekurangan Python, Sudah Tahu?," *NESABEMEDIA*, 2019. [Online]. Available: <https://www.nesabamedia.com/pengertian-python/>. [Accessed: 19-Sep-2019].
- [18] D. R. Wijaya, R. Sarno, and E. Zulaika, "Noise filtering framework for electronic nose signals: An application for beef quality monitoring," *Comput. Electron. Agric.*, vol. 157, no. January, pp. 305–321, 2019.