APLIKASI BERBASIS DATASET E-COMMERCE UNTUK PREDIKSI KEMISKINAN MENGGUNAKAN ALGORITMA NAÏVE BAYES, XGBOOST DAN SIMILARITY BASED FEATURE SELECTION

APLICATION BASED ON E-COMMERCE DATASET FOR POVERTY PREDICTION USING NAÏVE BAYES ALGORITHM, XGBOOST AND SIMILARITY BASED FEATURE SELECTION

Sherla Yualinda[1], Dr. Dedy Rahman Wijaya, S.T., M.T.[2], Elis Hernawati, S.T., M.Kom.[3] [123]Program Studi D3 Sistem Informasi, Fakultas Ilmu Terapan Universitas Telkom
sherla@student.telkomuniversity.ac.id[1], dedyrw@tass.telkomuniversity.ac.id[2], elishernawati@tass.telkomuniversity.ac.id[3]

Abstrak

Kemiskinan menginterpretasikan salah suatu keadaan seseorang tidak mampu untuk memenuhi kebutuhan dasar mereka seperti halnya sandang, papan, pangan, kesehatan, dalam menuntut ilmu, dll. Badan Pusat Statistik atau lebih dikenal dengan sebutan BPS menggunakan konsep kemampuan untuk dapat memenuhi kebutuhan (*basic needs approach*) guna mengukur tingkat kemiskinan di Indonesia. Dengan menggunakan konsep ini, pengeluaran menjadi tolak ukur dari kemiskinan yang dipandang sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan pangan dan *non* pangan, sehingga penduduk yang tidak mampu (miskin) adalah penduduk yang memiliki pengeluaran perkapita perbulan dibawah garis kemiskinan. Metode lain yang diusulkan penulis untuk melengkapi hasil survei dan sensus guna memprediksi kemiskinan di suatu daerah di Indonesia adalah menggunakan *naive bayes* dengan metode *XGBoost* dan *Similarity Based berbasis e-commerce*. Dalam percobaan yang telah dilakukan, nilainya cukup relevan antara fitur dan nilai asli. Banyaknya fitur yang terlalu sedikit tidak selalu menghasilkan nilai akurasi yang juga kecil, demikian juga sebaliknya, di mana penggunaan sejumlah besar fitur tidak selalu menghasilkan akurasi yang tinggi.

Kata Kunci: Kemiskinan, BPS, *Naive Bayes*, *XGBoost*, *Similarity Based*, data *e- commerce*.

Abstract

Poverty interprets one of the conditions a person is unable to meet their basic needs such as clothing, shelter, food, health, in studying, etc. The Central Statistics Agency or better known as BPS uses the concept of ability to be able to meet needs (basic needs approach) to measure poverty levels in Indonesia. By using this concept, expenditure becomes a benchmark of poverty which is seen as an inability from the economic side to meet food and non-food needs, so that the poor can be those who have per capita expenditure per month below the poverty line. Another method proposed by the author to supplement survey and census results to predict poverty in an area in Indonesia is to use Naive Bayes with XGBoost and Similarity Based e-commerce methods. In the experiments that have been carried out, the value is quite relevant between the features and the original values. The number of features that are too little does not always produce a value of accuracy that is also small, as well as vice versa, where the use of a large number of features does not always produce high accuracy.

Keywords: Poverty, BPS, Naive Bayes, XGBoost, Similarity Based, e-commerce data.

I. PENDAHULUAN

Negara Indonesia adalah negara yang memiliki jumlah penduduk terpadat ke-4 setelah China, India dan Amerika Serikat. Jumlah penduduk saat ini mencapai kurang lebih 255,5 juta jiwa. Jumlah penduduk Indonesia akan terus meningkat dari tahun ke tahun karena laju pertumbuhan di Indonesia tergolong tinggi yakni 1,49 persen. Pertumbuhan penduduk yang tinggi dapat menimbulkan berbagai masalah dan dampak negatif bagi masyarakat dan negara jika tidak segera diatasi. Salah satu dampak negatif dari laju pertumbuhan yang tinggi adalah tingginya angka kemiskinan. Angka kemiskinan di Indonesia berdasarkan Badan Pusat Statistik (BPS) mencapai 27,77 juta jiwa atau 10,6 persen dari total jumlah penduduk Indonesia. Tingginya angka kemiskinan di Indonesia menjadi tugas utama pemerintah karena angka kemiskinan menjadi indikator perekonomian sebuah negara. Angka kemiskinan di Indonesia didasarkan pada tingkat ketidakmampuan masyarakat untuk memenuhi kebutuhan pokok. Beberapa faktor yang menjadi penyebab masalah kemiskinan di Indonesia adalah Pendidikan yang rendah, kemampuan (*skill*) yang rendah, tingat pertumbuhan yang tinggi, serta tidak meratanya pembangunan infrastruktur[1].

Badan Pusat Statistik atau lebih dikenal dengan sebutan BPS menggunakan konsep kemampuan untuk dapat memenuhi kebutuhan (*basic needs approach*) guna mengukur tingkat kemiskinan di Indonesia. Dengan menggunakan konsep ini, pengeluaran menjadi tolak ukur dari kemiskinan yang dipandang sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan pangan dan *non* pangan, sehingga penduduk yang tidak mampu (miskin) adalah penduduk yang memiliki pengeluaran perkapita perbulan dibawah garis kemiskinan[2]. Konsep garis kemiskinan[2] :

1. Garis Kemiskinan (GK) adalah hasil dari penjumlahan antara Garis Kemiskinan Makan (GKM) dengan Garis Kemiskinan *Non* Makan (GKNM). Kategori penduduk miskin ialah penduduk yang pengeluaran perbulan rata-ratanya yakni dibawah Garis Kemiskinan.
2. Garis Kemiskinan Makan (GKM) adalah nilai atau hasil pengeluaran kebutuhan minimum makan yang disertakan dengan 2100 kilokalori perkapita perhari. Paket komoditi kebutuhan dasar makanan diwakili oleh 52 jenis komoditi (

umbi-umbian, padi-padian, daging, telur, susu, sayuran, kacang-kacangan, buah-buahan , minyak , dll).

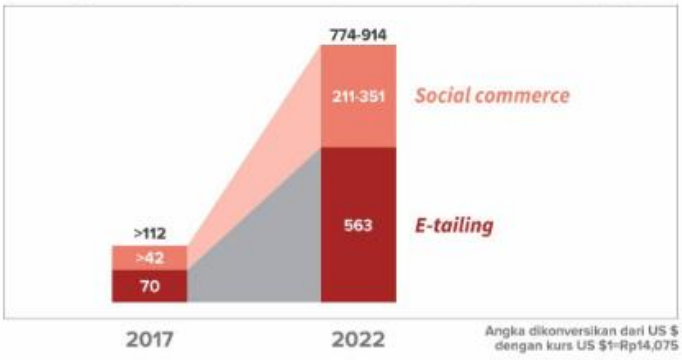
3. Garis Kemiskinan Non Makan (GKNM) adalah suatu kebutuhan minimum untuk sandang, papan, kesehatan dan juga kebutuhan dalam menuntut ilmu. Kebutuhan dasar *non* pangan diwakili oleh 47 jenis di pedesaan dan 51 jenis di perkotaan.

Salah satu aspek yang telah diberikan oleh Garis Kemiskinan Makanan (GKM) sebanyak 73,48% seperti halnya beras dan rokok ini merupakan salah satu kontribusi yang paling besar terhadap total Garis Kemiskinan (GK) di daerah perkotaan maupun pedesaan[3].

Badan Pusat Statistik atau yang disingkat dengan BPS khususnya BPS Bandung menyelenggarakan Survei Sosial Ekonomi Nasional (Susenas) yang bertujuan untuk mendapatkan data sebagai gambaran mengenai kondisi sosial ekonomi. Mulai tanggal 2015, pengumpulan data Survei Sosial Ekonomi Nasional dilaksanakan pada bulan Maret dan pada tahun 2018 Susenas dilaksanakan di seluruh provinsi di Indonesia (34 Provinsi) dengan ukuran sempel mencapai 300.000 rumah tangga yang tersebar di 514 kabupaten /kota di Indonesia. Sample tersebut tidak termasuk rumah tangga yang tinggal dalam rumah tangga khusus maupun *blok* khusus seperti asrama, penjara dll[4].

Pengumpulan data dari rumah tangga yang dipilih akan dilakukan wawancara langsung antara pencacah dengan responden (kepala rumah tangga, suami/istri maupun anggota keluarga yang lain) yang terpilih. Proses pengolahan yang dilakukan petugas dari BPS untuk mendapatkan data diawali dengan tahapan penyimpanan data melalui rekaman data, mengecek kesesuaian isi data dengan hasil kuesioner hingga tahap tabulasi dengan bantuan komputer. Sebelum dilaksanakannya bebrapa tahap diatas, terlebih dahulu petugas akan mengcek kelengkapan isian daftar pertanyaan, pengeditan terhadap isian data yang dianggap tidak wajar termasuk keterkaitan atau konsistensi antara jawaban satu dengan jawaban yang lainnya[4].

Pada beberapa pernyataan diatas dapat disimpulakan bahwasannya masalah dalam menentukan tingkat kemiskinan disuatu daerah mengkaitkan banyak hal, oleh karena itu dibutuhkan pandangan lain sehingga kemiskinan dapat terlihat lebih dalam dan akurat. Pengumpulan data yang dilakukan BPS akan membutuhkan waktu yang tidak singkat dan tahapan yang dilakukannya pun cukup rumit, selain itu pada proses wawancara yang dilangkukan kepada kepala rumah tangga umumnya sulit dilakukan karena kepala rumah tangga tersebut sulit untuk ditemui atau menghindar karena takut adanya penipuan. Sehingga metode lain yang diusulkan peneliti untuk melengkapi hasil survei dalam memprediksi kemiskinan di suatu daerah adalah menggunakan naive bayes dengan metode XGBoost dan Similarity Based berbasis e-*commerce*. Alasan penulis memilih menggunakan data *e-commerce* sebagai data yang nantinya akan diolah sehingga mengghasilkan presentase tingkat kemiskinan disuatu daerah karena di Asia Tenggara khususnya di negara Indonesia memiliki pasar *e-commerce* terbesar yang memberikan kontribusi sekitar 50% dari seluruh transaksi yang berjalan. Kontribusi ini akan terus meningkat lantaran biasanya penduduk Indonesia sering menggunakan internet untuk beraktivitas sehari-hari. Tahun 2018, hasil riset yang diungkapkan oleh firma konsultan manajemen McKinsey dan Company adalah temuan – temuan yang meliputi pertumbuhan nilai pasar *e-commerce* Indonesia hingga tahun 2022 dan potensi dampak pertumbuhan terhadap ekonomi dan sosial Indonesia. Berikut merupakan gambaran prediksi peningkatan data *e-commerce* di Indonesia.



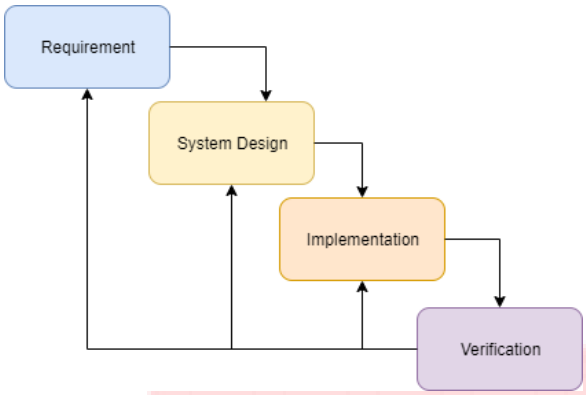
Gambar I-1 Prediksi Peningkatan *E-Commerce*

Pertumbuhan pasar *E-commerce* di Indonesia diprediksikan akan menghasilkan sekitar US$65 miliar atau Rp910 triliun. E-Commerce merupakan alur jual beli barang secara *online* yang diungkapkan oleh McKinsey yang dibagi menjadi 2 kategori, diantaranya yakni *E-tailing merupakan* jual beli formal yang menggunakan media *platform online* untuk memfasilitasi transaksi, selain *E-tailing* terdapat juga *Social Commerce* yang memenfaatkan media sosial seperti Facebook atau Instagram sebagai media perdagangan barang dengan pembayaran serta pengirimannya melalui *platform* lain. Pada tahun 2022 pasar *e-commerce* akan diprediksikan meningkat sebesar 8 kali lipat oleh McKinsey[5].

II. METODE PENELITIAN

Berikut merupakan metode penelitian aplikasi berbasis dataset *e-commerce* untuk prediksi tingkat kemiskinan menggunakan algoritma naive bayes, *xgboost dan similarity based feature selection*.

1. Penentuan Topik
   Pada tahap ini penulis pertama-tama menentukan topik yang nantinya akan dibuat sebuah aplikasi untuk menyelesaikan proyek akhir di semester 6 mendatang dengan mengangkat judul Aplikasi untuk memprediksi kemiskinan menggunakan *machine learning* naive bayes dengan metode *xgboost* dan *similarity based feature selection* yang nantinya akan menampilkan hasil berupa grafik maupun nilai presentase prediksi kemiskinan sesuai dengan data *e-commerce* yang diinputkan.
2. Identifikasi Masalah
   Pada tahap ini penulis mengidentifikasi masalah – masalah yang dihadapi oleh pemerintah Indonesia mengenai prediksi tingkat kemiskinan.
3. Studi Literatur
   Pada pembuatan laporan proyek akhir ini untuk menentukan fakta-fakta apa saja yang terkait dengan kemiskinan di Indonesia dan metode – metode yang digunakan untuk menentukan prediksi tingkat kemiskinan.
4. Perancangan Sistem
   Pada tahap perancangan sistem, penulis menggunakan metode *waterfall* yang dimana menggambarkan struktur pendekatan yang sistematis dan berurutan pada pengembangan perakat lunak yang dimulai dari tahap *Requirement, Design, Implementation, Verification, Maintenance*, seperti gambar dibawah ini[6].



Gambar II-1 Metode Waterfall

4.1 Requirement Analysis

Pada tahap ini bertujuan untuk memahami perangkat lunak yang akan dibangun dengan mencari data-data yang diperlukan oleh penulis dengan cara wawancara, serta tinjauan pustaka atau studi literatur dengan mencari referensi buku, web, jurnal yang berhubungan dengan perangkat lunak yang akan dibangun.

4.2 System Design

Pada tahapan ini penulis menggunakan bahasa pemograman *python* untuk membangun perangkat lunak atau aplikasi, menggunakan *tools* BPMN untuk menggambaran proses bisnis yang sedang berjalan (*As-Is*) atau yang akan berjalan (*To-Be*) dan merancang database mnggunakan ERD untuk menentukan *atribut* serta *entitas*-nya.

4.3 Implementation

Pada perancangan ini, penulis menggunakan metode *naïve bayes* dimana Tahapan dalam proses algoritma *naïve bayes* adalah sebagai berikut[7]:

a) Menghitung jumlah kelas / label

b) Menghitung jumlah kasus per kelas

c) Kalikan semua variable kelas

d) Bandingkan hasil perkelas

4.4 Verification

Setelah tahap *implementation*, semua unit dikembangkan dan diintegrasikan kedalam sistem pengujian guna mengecek setiap kegagalan atau kesalahan yang terdapat dalam perangkat lunak atau aplikasi. Pada tahap ini penulis menggunakan metode pengujian *Blackbox Testing* untuk memastikan sistem berjalan dengan lancar dan sesuai dengan yang diharapkan. Selain menggunakan *blackbox testing*, penulis juga menggunakan UAT (User Acceptance Testing) guna memastikan apakah aplikasi sudah sesuai dengan keinginan user.

III. TINJAUAN PUSTAKA

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

A. Machine Learning

Machine learning belajar serta memperbaiki diri dari pengalaman tanpa deprogram secara *eksplisit*. mesin berfokus dengan pembelajaran pada pengembangan program komputer yang bisa mengakses data dami menggunakannya untuk belajar sendiri[8]. Sistem pembelajaran mesin dibagi menjadi tiga yakni[9] :

a) Model adalah sistem yang membentuk prediksi atau identifikasi
b) Parameter adalah faktor yang digunakan model untuk membentuk keputusannya
c) Pembelajaran adalah sistem yang menyesuaikan parameter serta model dalam prediksi versus hasil *actual*

Pada definisinya , *Machine learning* adalah cabang aplikasi dari AI (*Artifical Intelligence*) atau suatu kecerdasan buatan. *Machine learning* dalam belajar harus membutuhkan data sebagai acuan untuk belajar dalam mengeluarkan sebuah *output*, tanpa data tersebut *machine learning* tidak dapat bekerja atau berfungsi dengan baik. *Machine learning* dibagi menjadi tiga model yakni[10] :

1. Model *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 historis. *Supervised learning* dibagi menjadi dua yakni[10] :
   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 diperuntuhkan untuk membuat sebuah pola pada setiap *atribut*-nya. Pada metode ini bertujuan untuk mencari sebuah pola dan menentukan sebuah nilai numerik.
2. Model *Unsupervised Learning*

   Dalam metode ini tidak diberikan label , tetapi secara otomatis dibagi berdasarkan kemiripan dan struktur lain dari data tersebut[10].

B. *Feature Selection*

Feature selection merupakan suatu kegiatan preprocessing dan bertujuan untuk memilih *feature* yang berpengaruh maupun tidak berpengaruh dalam penganalisasian suatu data. *Feature selection* dibagi menjadi dua kelompok yakni[11]:

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 dari fitur yang terdapat dalam data yang dianggap sebagai fitur yang optimal. Terdapat tiga jenis metode dalam *subnet selection* yakni[11]:

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. *Naïve Bayes*

Naïve Bayes merupakan suatu metode klasifikasi yang menggunakan metode probabilitas dan statistik. Algoritma *Naïve Bayes* dapat memprediksi peluang di masa yang akan datang dari pengalaman sebelumnya yang dikenal dengan *Teorema Bayes*[7]. Keuntungan menggunakan *Naïve Bayes* adalah metode ini hanya menggunakan jumlah data pelatihan (*Training Data*) yang kecil guna menentukan rentang parameter yang diperlukan dalam proses pengklasifikasian. Berikut merupakan persamaan dari *teorema bayes*[12] :

$$P(H \mid X) = \frac{P(X \mid H).P(H)}{P(X)}$$

Keterangan :

X : Data dengan *class* yang belum diketahui

H : Hipotesis data merupakan suatu *class* yang spesifik

P(H | X) : Probabilitas hipotesis H berdasar kondisi X (*posterior probabilitas*)

P(H) : Probabilitas hipotesis H (*prior probabilitas*)

P(X|H) : Probabilitas hipotesis X berdasar kondisi pada hipotesis H

P(X) : Probabilitas hipotesis X.

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 bisa 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 *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 pembandingan, validasi dan pemilihan parameter maupun model[14].

F. *Library Pyhon 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 memfalisitasi 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 di dalam proses *data mining* adalah proses penskalaan nilai *atribut* atau fitur dari data yang akan dinormalisasi sehingga data tersebut bisa memiliki skala atau *range* yang telah ditetapkan sebelumnya. Ada beberapa metode yang digunakan untuk proses normalisasi yaitu *min-max*, *z-score*, *decimal scaling*, *sigmoidal* , dll. Metode *min-max* dilakukan untuk transformasi linier terhadap data asli. Metode *z-score* adalah normalisasi yang berdasarkan nilai rata-rata atau biasanya disebut dengan mean dan *standart deviation* (deviasi standar) dari data. Metode *decimal scaling* adalah normalisasi yang menggerakkan nilai desimal dari suatu data. Metode *sigmoidal* adalah normalisasi yang secara *non*-linier kedalam *range* -1-1 dengan menggunakan fungsi *sigmoid*, dalam metode ini berguna bagi data yang melibatkan data *outlier* (data yang jauh dari jangkauan data lainnya)[16].

H. Bahasa Pemrograman *Python*

Python adalah salah satu bahasa pemrograman yang popular di dunia kerja. *Python* secara default telah terpasang di berbagai sistem operasi berbasis Linux seperti Ubuntu, Linux Mint, dan Fedora. Selain itu *Python* memiliki sebuah *package manager* yang popular dan unggul bernama PIP. Dengan PIP pengguna dapat menghapus atau memasang pustaka *Python*[17].

I. 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].

J. R Squared

R square merupakan ukuran statik yang dimana antara 0 dan 1 yang berfungsi untuk mehitunug seberapa mirip hasil prediksi dengan data aslinya yang ditandai dengan semakin dengkatnya hasil prediksi dengan garis aslinya[19].

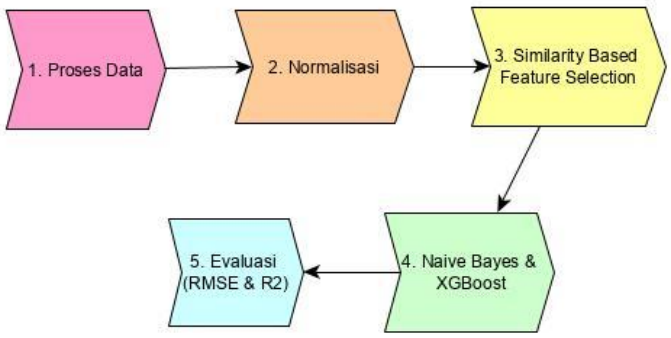
IV. ANALISIS DAN PERANCANGAN

A. Gambaran Sistem Usulan

Dibawah ini adalah gambaran sistem usulan dari alur pengembangan model 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

Gambar diatas merupakan beberapa tahapan sebelum data *e-commerce* diolah ke dalam aplikasi berbasis dataset *e-commerce* untuk prediksi tingkat kemiskinan menggunakan algoritma *naive bayes, xgboost dan similarity based feature selection*.

a. Proses Data
Pertama data akan masuk ke dalam *preprocessing* data atau *data cleaning*. Tahap ini adalah tahap yang penting karena pada tahap ini bertujuan untuk membersihkan data yang tidak lengkap atau *missing value* yang dapat mengganggu jalannya data sehingga data tidak dapat diproses kedalam tahap selanjutnya dikarenakan data dinilai tidak siap. Di dalam data *e-commerce* dari salah satu perusahaan e-commerce di Indonesia yang diperoleh sebelumnya masih terdapat *missing value* atau data bernilai *null* sehingga data yang bernilai *null* nantinya akan diganti dengan 0, setelah itu data akan ditentukan mana yang merupakan fitur serta data mana yang merupakan label.

b. Normalisasi
Setelah memalui tahapan pertama yakni proses data, kemudian masuk ke dalam proses normalisasi yang dimana data yang telah ditentukan diskalakan nilai datanya dari 0-10. Alasan mensklalakan nilai data dari 0-10 karena jika range nya 0-1, maka nilai dibelakang koma semakin panjang dan machine learningnya akan semakin sulit untuk di training karena nilainya terlalu kecil. Pada proses normalisasi penulis menggunakan metode *Rescaling atau yang disebut dengan min-max normalization*. Berikut merupakan rumus dasarnya :

$$MinMax = \frac{x - \min(x)}{\max(x) - \min(x)} * 10$$

Keterangan :

x = nilai dari masing – masing fitur.

min(x) = nilai terendah dari setiap fitur.

max(x) = nilai tertinggi dari setiap fitur.

c. *Similarity Based Feature Selection*
Setelah data selesai dinormalisasi, data akan masuk ke dalam proses selanjutnya yakni seleksi fitur *XGBoost* dan *Similarity Based Feature Selection*. Pada algoritma *Similarity Based Feature Selection* mengekploitasi berbagai jenis fitur untuk menentukan fitur mana yang cocok. Pada *supervised feature selection*, *similarity based* data dapat diturunkan dari informasi label sedangkan pada *unsupervised feature*, sebagian besar memanfaatkan langkah-langkah yang metrik yang berbeda untuk mendapatkan data. Pada *Similarity Based Feature Selection* terdapat beberapa metode pemilihan fitur yakni *fisher_score, reliefF, trace_ratio*. Fisher Score adalah algoritma pemilihan fitur yang diawasi di mana nilai fitur di kelas yang sama adalah sama dan nilai fitur di kelas yang berbeda tidak sama. Berikut ini adalah rumusnya.

$$\text{Fisher\_score}(\boldsymbol{f_i}) = \frac{\sum_{j=1}^{c} \boldsymbol{n_j}(\boldsymbol{\mu_{ij}} - \boldsymbol{\mu_i})^2}{\sum_{j=1}^{c} \boldsymbol{n_j}\boldsymbol{\sigma_{ij}^2}}$$

Secara umum ditetapkan bahwa n_j adalah jumlah sampel yang tersedia di kelas j, μ_i adalah nilai rata-rata dalam fitur f_i, μ_{ij} adalah nilai fitur f_i untuk sampel di kelas j dan akhirnya, σ_{ij}^2 adalah varian dari fitur fi untuk sampel di kelas j. Sementara ReliefF memilih fitur yang digunakan untuk dipisahkan. Berikut ini adalah rumus dari ReliefF.

$$\text{ReliefF\_score}(\boldsymbol{f_i}) = \frac{1}{c}\sum_{j=1}^{I}\left(-\frac{1}{m_j}\sum_{x,\epsilon NH(j)} \boldsymbol{d}\big(\boldsymbol{X(j,i)} - \boldsymbol{X(r,i)}\big) + \sum_{y \neq y_j} \frac{1}{h_{jy}} \frac{p(y)}{1-p(y)} \sum_{x,\epsilon NM(j,y)} \boldsymbol{d}\big(\boldsymbol{X(j,i)} - \boldsymbol{X(r,i)}\big)\right)$$

$NH(j)$ dan $NM(j,y)$ adalah contoh terdekat dengan x_j dari kelas yang sama dan di kelas y, ukuran masing-masing adalah m_j dan h_{jy}. $p(y)$adalah rasio instance di kelas y. Dan berikutnya adalah trace_ratio yang secara langsung memilih secara global fitur subnet berdasarkan skor yang sesuai. Berikut ini adalah rumus untuk trace_ratio.

$$\text{Trace\_ratio } (\boldsymbol{S}) = \frac{\boldsymbol{tr(W'X'L_bXW)}}{\boldsymbol{tr(W'X'L_wXW)}}$$

di mana L_b dan L_w adalah matriks Laplac ian dari S_a dan S_b.

d. *Naïve Bayes* dan *XGBoost*
Setelah data masuk kedalam proses *Similarity Based Feature Selection*, kemudian data akan masuk kedalam *machine learning naïve bayes* atau XGBoost yang dimana *naïve bayes* adalah suatu metode pengklasifikasian yang menggunakan metode probabilitas dan statistic. Algoritma *Naïve Bayes* dapat memprediksi peluang di masa yang akan datang dari pengalaman sebelumnya yang dikenal dengan *Teorema Bayes*. Berikut merupakan *Teorema Bayes* :

$$\boldsymbol{P(H \mid X)} = \frac{\boldsymbol{P(X \mid H)}.\boldsymbol{P(H)}}{\boldsymbol{P(X)}}$$

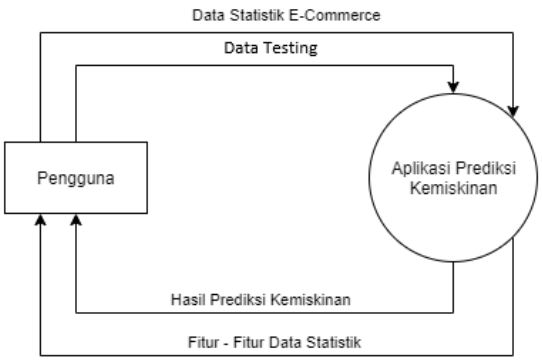
Keterangan :
X : Data dengan *class* yang belum diketahui
H : Hipotesis data merupakan suatu *class* yang spesifik
P(H | X) : Probabilitas hipotesis H berdasar kondisi X (*posterior probabilitas*)
P(H) : Probabilitas hipotesis H (*prior probabilitas*)
P(X|H) : Probabilitas hipotesis X berdasar kondisi pada hipotesis H
P(X) : Probabilitas hipotesis X

Sedangkan XGBoost atau *Extreme Gradient Boosting* merupakan algoritma suatu mesin yang kuat dan cepat. *XGBoost* menyediakan dua tingkatan dalam pengambilan sampel pada kolom yakni *colsample_bytree* dan *colsample_bylevel*. XGBoost

perlu menghitung *hessian*, sehingga membutuhkan fungsi objektif dua kali lipat.

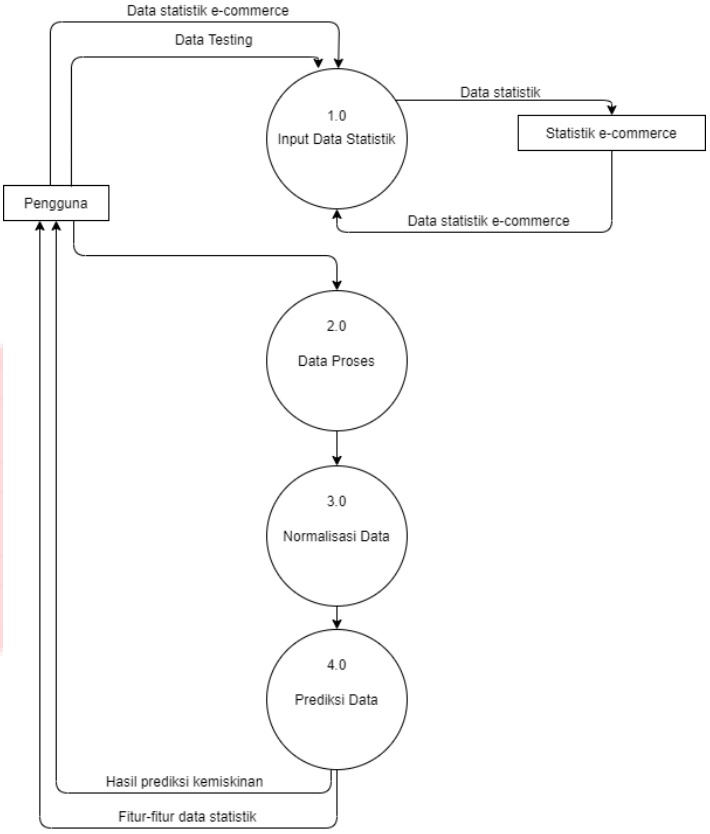
e. *Evaluasi*
Dalam tahapan terakhir ini data akan dievaluasi menggunakan 2 matrik guna regresi. Dua matrik tersebut yakni R^2 (*R-Square*) dan RMSE. RMSE digunakan untuk mengukur perbedaan maupun kesalahan antara *vector actual* dengan prediksi. Jika nilai RMSE tinggi maka hasil yang dikeluarkan memiliki banyak perbedaan antara nilai sebenarnya dengan hasil prediksi, sedangkan R^2 digunakan untuk mewakili bagian dari *varians vector* yang dapat diprediksi oleh model regresi. Jika R^2 bernilai 1 maka dapat dikatakan sesuai dalam memprediksi nilai, jika sebaliknaya R^2 bernilai negative atau kurang dari 1 maka dapat dikatakan salah dalam memprediksi nilai.

2. Gambaran Sistem Usulan Aplikasi
Berikut merupakan gambaran sistem usulan pada aplikasi berbasis dataset *e-commerce* untuk prediksi tingkat kemiskinan menggunakan algoritma *naive bayes*, *xgboost* dan *similarity-based feature selection* yang digambarkan dengan menggunakan data flow diagram. Alasan mengapa penulis menggambarkannya dengan *data flow diagaram* karena pada aplikasi ini termasuk *structural programing* 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.



Gambar IV-2 Diagram Konteks

Pada gambaran proses diatas, pengguna yang nantinya akan menggunakan aplikasi prediksi kemiskinan dengan tujuan ingin mengetahui presentase tingkat kemiskinan, menggunakan data *ecommerce* sebagai data yang akan diolah kedalam aplikasi. Berikut merupakan alur dari aplikasi prediksi kemiskinan.



Gambar IV-3 DFD Level 1 Proses Prediksi Kemiskinan

Pada proses ini, data *e-commerce* yang akan diproses masuk kedalam proses input data statistik untuk diolah, kemudian akan menghasilkan data statistik yang akan disimpan kedalam satistik *e-commerce*, setelah itu data akan masuk kedalam proses input data statistik lagi untuk menghasilkan dataset yang nantinya akan digunakan pengguna untuk masuk kealam proses selanjutnya yakni proses normalisasi data kemudian data akan masuk ke dalam proses prediksi data. Pada proses ini akan menghasilkan hasil 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 naive bayes, *xgboost dan similarity-based feature selection*. Berikut merupakan implementasi tampilan.

1) Halaman Registrasi.

Gambar V-1 Halaman Registrasi

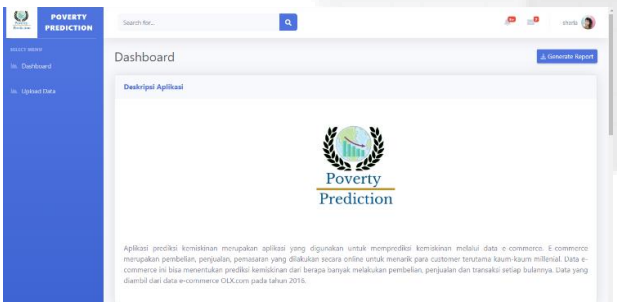
Pada halaman ini, user dapat membuat akun jika user tersebut belum memiliki akun. Dalam halaman ini user diharuskan mengisi kolom-kolom seperti first_name, last_name, email, username, password dan repeat password.

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.

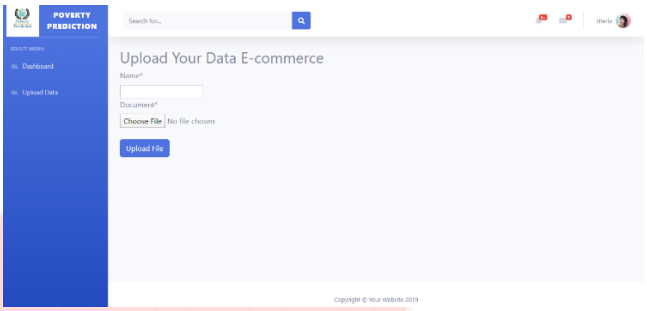
3) Halaman Dashboard



Gambar V-3 Halaman Dashboard

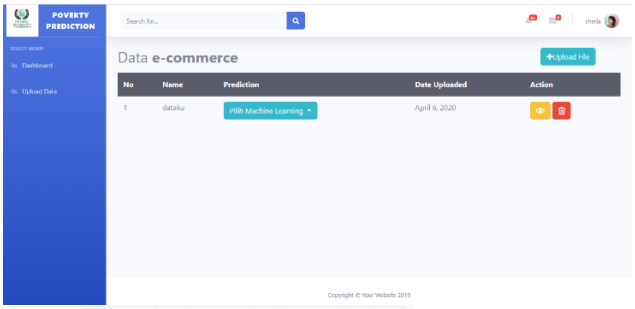
Pada halaman ini, user harus login terlebih dahulu untuk dapat mengakses halaman *dashboard* ini. Dalam halaman utama ini terdapat menu dashboard dan upload data. Di dalam menu dashboard terdapat deskripsi dari masing – masing algoritma yang disediakan oleh aplikasi.

4) Halaman Upload Data dan Halaman List Data



Gambar V-4 Halaman Upload Data

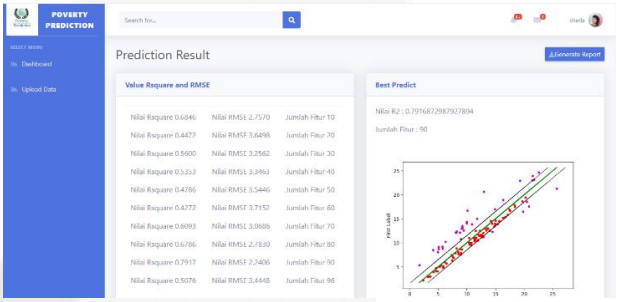
Pada halaman ini, user dapat mengupload data yang akan diprediksi. Data tersebut harus sesuai dengan template yang telah ditentukan oleh aplikasi dan data tersebut harus berupa csv.



Gambar V-5 Halaman List Data

Selain itu pada menu upload data, terdapat *list data* jika user tersebut sudah meng-upload data.

5) Halaman Hasil Prediksi



Gambar V-6 Halaman Hasil Prediksi

Pada tampilan hasil prediksi ini akan tampil jika user terlah memilih algoritma *machine learning* yang dipilih untuk memprediksi kemiskinan.

B. Pengujian

Setelah melalui tahap implementasi maka aplikasi masuk kedalam proses selanjutnya yakni proses pengujian. Berikut merupakan pengujian menggunakan *black box testing*.

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 Berbasis Dataset E-Commerce Untuk Prediksi Kemiskinan Menggunakan Algoritma Naïve Bayes, Xgboost Dan Similarity Based Feature Selection.
Deskripsi	Aplikasi yang digunakan untuk memprediksi tingkat kemiskinan disuatu daerah.
Fungsi	Login
Aturan	**1. Username dan password harus diisi sesuai dengan data registrasi** 2. Username dan password tidak diisi sesuai dengan data registrasi 3. Username dan password tidak diisi

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	Conclution/ Kesimpulan
1.	Login	1.	Entry data login dengan mengikuti aturan 1 : 1. Username 2. Password	Username dan Password sesuai dengan data registrasi	Aplikasi menampilkan halaman *dashboard*	Aplikasi menampilkan halaman *dashboard*	Valid
		2.	Entry data login dengan mengikuti aturan 2 : 1. Username 2. Password	Username dan Password tidak sesuai dengan data registrasi	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	Username dan Password dikosongkan	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.*

Perangkat Lunak	Aplikasi Berbasis Dataset E-Commerce Untuk Prediksi Kemiskinan Menggunakan Algoritma Naïve Bayes, Xgboost Dan Similarity Based Feature Selection.
Deskripsi	Aplikasi yang digunakan untuk memprediksi tingkat kemiskinan disuatu daerah.
Fungsi	Registrasi

Tabel V-3 Scope of Testing Registrasi

Aturan	1. **First name, last name, email, username, password dan repeat password harus diisi** 2. First name, last name, email, username, password dan repeat password tidak diisi		3. **Password dan repassword sama** 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	Semua form diisi	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	Semua form tidak diisi	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 6. Repeat Password	Password : Sherla123 Repeat Password : Sherla123	Aplikasi menampilkan halaman login	Aplikasi menampilkan halaman login	Valid
		4.	Entry data login dengan mengikuti aturan 4 : 1. First Name	Password : Sherla123 Repeat Password	Aplikasi menampilkan Pesan error	Aplikasi menampilkan Pesan error	Valid

No.	Function/ Condition	Case No.	Test Case Description (Event)	Test Data (Input)	Expected Result	Actual Result / Comments	Conclution/ Kesimpulan
			2. Last Name 3. Username 4. Email 5. Password 6. Repeat Password	: Sherla12345	message "Password not maching"	message "Password not maching"	

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 Berbasis Dataset E-Commerce Untuk Prediksi Kemiskinan Menggunakan Algoritma Naïve Bayes, Xgboost Dan Similarity Based Feature Selection.
Deskripsi	Aplikasi yang digunakan untuk memprediksi tingkat kemiskinan disuatu daerah.
Fungsi	Upload file
Aturan	**1. *Name* dan *document* diisi** 2. *Name* dan *document* tidak diisi

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 *upload file* dengan mengikuti aturan 1 : 1. Name 2. Document	Semua form diisi	Aplikasi menampilkan halaman *list document*	Aplikasi menampilkan halaman *list document*	Valid
		2.	Entry data *upload file* dengan mengikuti aturan 2 : 1. Name 2. Document	Semua form tidak diisi	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 Berbasis Dataset E-Commerce Untuk Prediksi Kemiskinan Menggunakan Algoritma Naïve Bayes, Xgboost

	Dan Similarity Based Feature Selection.
Deskripsi	Aplikasi yang digunakan untuk memprediksi tingkat kemiskinan disuatu daerah.
Fungsi	Forgot Password
Aturan	1. **Email harus diisi** 2. Email dikosongkan 3. ***New Password dan new password confirmation* tidak boleh sama dengan data diri** 4. *New Password dan new password confirmation* sama dengan data diri 5. ***New Password dan new password confirmation* minimal 8 karakter** 6. *New Password dan new password confirmation* kurang dari 8 karakter 7. ***New Password dan new password confirmation* tidak boleh seluruhnya bersifat numerik** 8. *New Password dan new password confirmation* seluruhnya bersifat numerik 9. **New password dengan new password confirmation harus sama** 10. New password dengan new password confirmation tidak sama

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	Conclution / Kesimpulan
1.	Forgot Password	1.	Entry data *forgot password* dengan mengikuti aturan 1 : 1. Email	Email diisi sesuai dengan akun registrasi, contoh : yualindasherli@gmail.com	Aplikasi menampilkan halaman *Password reset sent*	Aplikasi menampilkan halaman *Password reset sent*	Valid
		2.	Entry data *forgot password* dengan mengikuti aturan 2 : 1. Email	Email dikosongkan	Aplikasi menampilkan Pesan error message	Aplikasi menampilkan Pesan error message	Valid
		3.	Entry data *forgot password* dengan mengikuti aturan 3 : 1. *New Password* 2. *New Password*	*New Password* dan *New Password Confirmation* tidak boleh sama dengan data diri	Aplikasi menampilkan halaman *Password reset Completed*	Aplikasi menampilkan halaman *Password reset Completed*	Valid

No.	Function/ Condition	Case No.	Test Case Description (Event)	Test Data (Input)	Expected Result	Actual Result / Comments	Conclution / Kesimpulan
			Confirmation				
		4.	Entry data *forgot password* dengan mengikuti aturan 4 : 1. *New Password* 2. *New Password Confirmation*	*New Password* dan *New Password Confirmation* sama dengan data diri	Aplikasi menampilkan Pesan error message	Aplikasi menampilkan Pesan error message	Valid
		5.	Entry data *forgot password* dengan mengikuti aturan 5 : 1. *New Password* 2. *New Password Confirmation*	*New Password* dan *New Password Confirmation* minimal 8 karakter, contoh : uuiiojhg123	Aplikasi menampilkan halaman *Password reset Completed*	Aplikasi menampilkan halaman *Password reset Completed*	Valid
		6.	Entry data *forgot password* dengan mengikuti aturan 6 : 1. *New Password* 2. *New Password Confirmation*	*New Password* dan *New Password Confirmation* kurang dari 8 karakter, contoh : yuhj	Aplikasi menampilkan Pesan error message	Aplikasi menampilkan Pesan error message	Valid
		7.	Entry data *forgot password* dengan mengikuti aturan 7 : 1. *New Password* 2. *New Password Confirmation*	*New Password* dan *New Password Confirmation* tidak boleh seluruhnya bersifat numerik, contoh: mnhghj345i	Aplikasi menampilkan halaman *Password reset Completed*	Aplikasi menampilkan halaman *Password reset Completed*	Valid
		8.	Entry data *forgot password* dengan mengikuti aturan 7 :	*New Password* dan *New Password Confirmation* seluruhnya bersifat	Aplikasi menampilkan Pesan error	Aplikasi menampilkan Pesan error	Valid

No.	Function/ Condition	Case No.	Test Case Description (Event)	Test Data (Input)	Expected Result	Actual Result / Comments	Conclution / Kesimpulan
			1. *New Password* 2. *New Password Confirmation*	numerik, contoh: 89977667	message	message	
		9.	Entry data *forgot password* dengan mengikuti aturan 7 : 1. *New Password* 2. *New Password Confirmation*	*New password* dengan *new password confirmation* harus sama, contoh : *New Password* : syualinda123 *New Password Confirmation* : syualinda123	Aplikasi menampilkan halaman *Password reset Completed*	Aplikasi menampilkan halaman *Password reset Completed*	Valid
		10.	Entry data *forgot password* dengan mengikuti aturan 7 : 1. *New Password* 2. *New Password Confirmation*	*New password* dengan *new password confirmation* tidak sama, contoh : *New Password* : syualinda1234 *New Password Confirmation* : syualinda123	Aplikasi menampilkan Pesan error message	Aplikasi menampilkan Pesan error message	Valid

VI. KESIMPULAN

Dari implementasi dan pengujian yang telah dilakukan dapat disimpulkan bahwa:

1. Aplikasi dapat menampilkan hasil prediksi kemiskinan dengan data *e-commerce* berbasis Naïve bayes dan XGBoost dengan algoritma *similarity based feature selection*.

2. Item yang berpengaruh dalam prediksi kemiskinan pada data e-commerce pada aplikasi adalah mobil, motor, rumah, dst.

3. Dapat mengembangkan aplikasi berbasis web yang dapat menampilkan grafik rsuqre dan rmse.

REFERENSI

[1] R. Ilmugeografi, "13 Alasan Indonesia Termasuk Negara Berkembang dan Solusinya," *Ilmugeografi.com*. [Online]. Available: https://ilmugeografi.com/ilmu-sosial/alasan-indonesia-termasuk-negara-berkembang. [Accessed: 23-Sep-2019].

[2] B. P. Statistik, "Kemiskinan dan Ketimpangan," *Badan Pusat Statistik*. [Online]. Available: https://www.bps.go.id/subject/23/kemiskinan-dan-ketimpangan.html. [Accessed: 23-Sep-2019].

[3] F. M. FARUK, "Berkenalan Dengan Kemiskinan," *GEOTIMES*. [Online]. Available: https://geotimes.co.id/opini/berkenalan-dengan-kemiskinan/. [Accessed: 23-Sep-2019].

[4] B. P. Statistik, "Statistik Kesejahteraan Rakyat 2018," *Badan Pusat Statistik*, 2018. .

[5] D. Praditya, "Prediksi Perkembangan Industri E-commerce Indonesia pada Tahun 2022," *techinasia*, 2019. [Online]. Available: https://id.techinasia.com/prediksi-ecommerce-indonesia.

[6] F. Galandi, "Metode Waterfall : Definisi, Tahapan, Kelebihan dan Kekurangan," *pengetahuan dan teknologi.com*, 2018. [Online]. Available: http://www.pengetahuandanteknologi.com/2016/09/metode-waterfall-definisi-tahapan.html.

[7] INFORMATIKALOGI, "Algoritma Naive Bayes," *NFORMATIKALOGI.COM*, 2017. [Online]. Available: https://informatikalogi.com/algoritma-naive-bayes/. [Accessed: 04-Jan-2020].

[8] F. Burstein and C. W. Holsapple, *Handbook on Information Systems 1: Basis Themes*. 2008.

[9] Podfeeder, "Apa Itu Machine Learning," *podfeeder.com*. [Online]. Available: http://www.podfeeder.com/teknologi/apa-itu-machine-learning-berikut-penjelasannya/. [Accessed: 23-Sep-2019].

[10] V. N. Drozdov, V. A. Kim, and L. B. Lazebnik, *Modern approach to the prevention and treatment of NSAID-gastropathy.*, no. 2. 2011.

[11] G. Forman, "An extensive empirical study of feature selection metrics for text classification," *J. Mach. Learn. Res.*, vol. 3, pp. 1289–1305, 2003.

[12] V. Ratnasari, "Pengoptimalan Naïve Bayes Dan Regresi Logistik Menggunakan Algoritma Genetika Untuk Data Klasifikasi," p. 86, 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*. [Online]. Available: https://binus.ac.id/malang/2019/01/definisi-karakteristik-dan-manfaat-data-mining-seri-data-mining-for-business-intelligence-2/.

[14] Hakim-azizul, "Berkenalan dengan scikit-learn (Part 1) – Preparations," *hkaLabs*. [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] Noviandi, "Data Mining," 2018.

[17] R. Fajar, "Memulai Pemrograman dengan Python," *codepolitan*, 2016. [Online]. Available: https://www.codepolitan.com/memulai-pemrograman-python. [Accessed: 23-Sep-2019].

[18] Kuliahkomputer, "Training dan Testing Ilmu Komputer 'Root Mean Square Error,'" *Kuliahkomputer*, 2018. [Online]. Available: http://www.kuliahkomputer.com/2018/07/training-dan-testing-ilmu-komputer-root.html.

[19] A. Hershy, "Calculating R-squared from scratch (using python)," *towardsdatascience*, 2019. [Online]. Available: https://towardsdatascience.com/r-squared-recipe-5814995fa39a.