FEATURE SELECTION MENGGUNAKAN *GENETIC ALGORITHM* DALAM MELAKUKAN PREDIKSI *REPUCHASE INTENTION*

SKRIPSI

Diajukan untuk memenuhi salah satu syarat guna mencapai gelar
Sarjana dalam bidang ilmu Teknik Industri

Disusun oleh :

Nama : Wellington
NPM : 2016610036

PROGRAM STUDI SARJANA TEKNIK INDUSTRI
JURUSAN TEKNIK INDUSTRI
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS KATOLIK PARAHYANGAN
BANDUNG

2020

FEATURE SELECTION MENGGUNAKAN *GENETIC ALGORITHM* DALAM MELAKUKAN PREDIKSI *REPUCHASE INTENTION*

SKRIPSI

Diajukan untuk memenuhi salah satu syarat guna mencapai gelar
Sarjana dalam bidang ilmu Teknik Industri

Disusun oleh :

Nama : Wellington

NPM : 2016610036

PROGRAM STUDI SARJANA TEKNIK INDUSTRI
JURUSAN TEKNIK INDUSTRI
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS KATOLIK PARAHYANGAN
BANDUNG

2020

FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS KATOLIK PARAHYANGAN
BANDUNG

Nama	: Wellington
NPM	: 2016610036
Program Studi	: Teknik Industri
Judul Skripsi	: *FEATURE SELECTION* MENGGUNAKAN *GENETIC ALGORITHM* DALAM MELAKUKAN PREDIKSI *REPUCHASE INTENTION*

TANDA PERSETUJUAN SKRIPSI

Bandung, Agustus 2020
Ketua Program Studi Sarjana
Teknik Industri

(Romy Loice, S.T., M.T)

Pembimbing Tunggal

(Dedy Suryadi, S.T., M.S., Ph.D)

PERNYATAAN TIDAK MENCONTEK ATAU MELAKUKAN PLAGIAT

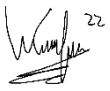
Saya yang bertanda tangan di bawah ini,
Nama : Wellington
NPM : 2016610036

dengan ini menyatakan bahwa Skripsi dengan Judul:
FEATURE SELECTION MENGGUNAKAN *GENETIC ALGORITHM* DALAM MELAKUKAN PREDIKSI *REPUCHASE INTENTION*

adalah hasil pekerjaan saya dan seluruh ide, pendapat atau materi dari sumber lain telah dikutip dengan cara penulisan referensi yang sesuai.

Pernyataan ini saya buat dengan sebenar-benarnya dan jika pernyataan ini tidak sesuai dengan kenyataan, maka saya bersedia menanggung sanksi yang akan dikenakan kepada saya.

Bandung, 11 Agustus 2020



Wellington
NPM : 2016610036

ABSTRAK

Permasalahan *feature selection* merupakan permasalahan pada pemilihan *features* yang dianggap paling relevan dalam melakukan prediksi suatu keluaran. Dalam *feature selection* dilakukan pemilihan sekumpulan *features* yang paling berpengaruh atau yang disebut *feature subset*, dimana pada konteks ini, *features* dianggap sebagai kata-kata penting dalam ulasan. Banyaknya jumlah kata dalam ulasan membuat pencarian solusi optimal untuk pemilihan *feature subset* menjadi sangat sulit dan memakan waktu lama.

Untuk menyelesaikan permasalahan *feature selection* pada penelitian ini, metode pendekatan yang digunakan adalah algoritma metaheuristik. Salah satu algoritma metaheuristik adalah *Genetic Algorithm* (GA) yang digunakan dalam penelitian ini. Algoritma ini terinspirasi dari teori evolusi yang dikemukakan oleh Charles Darwin. Proses evolusi terjadi dengan kegiatan pertukaran gen dalam kromosom, mutasi nilai gen, hingga seleksi alam dari kromosom terbaik. Nilai gen dalam kromosom berisi nilai indeks dari *feature* unik, dimana setiap kromosom mewakili kombinasi *feature subset*. Proses crossover dibantu dengan pemakaian *feature importance* sedangkan proses mutasi dibantu dengan penerapan sebuah ukuran yang diusulkan dalam penelitian ini, yaitu nilai proporsi kecenderungan. GA memerlukan metode untuk memeriksa keakuratan *feature subset*. Metode yang digunakan adalah algoritma *decision tree*.

Pada studi kasus terhadap data dari website sociolla.com, diterapkan 24 kombinasi parameter GA dalam mendapatkan *feature subset*. Model prediksi menggunakan *feature subset* terpilih memiliki keakuratan yang lebih baik dibandingkan model serupa yang menggunakan seluruh *features* maupun model prediksi dengan *features subset* acak. Berdasarkan *two-sample t-test*, didapatkan kesimpulan bahwa model prediksi dengan *feature subset* terpilih memiliki rata-rata akurasi yang lebih tinggi secara signifikan dibandingkan model prediksi lainnya (nilai $\alpha = 0,05$), dengan rata-rata kenaikan akurasi model untuk kasus “cleanser” sebesar 7,1 persen dan rata-rata kenaikan akurasi model untuk kasus “treatment” sebesar 11,1 persen.

ABSTRACT

Feature selection problem is a problem in selecting features that are considered the most relevant in predicting an output. In feature selection, a set of features which have the most influence or so called a feature subset is chosen, where in this context, features are considered as important words in the review. The large number of words in the review make the search of optimal solution for feature subset is very difficult and time consuming.

To solve the feature selection problem in this study, the approach method used is a metaheuristic algorithm. This study apply one of the metaheuristic algorithms, which is Genetic Algorithm (GA). Genetic Algorithm is inspired by the theory of evolution, founded by Charles Darwin. The process of evolution occurs by exchanging genes in chromosomes, mutations in gene values, to natural selection from the best chromosomes. The gene value in the chromosome contains the index value of unique features, where each chromosome represents a combination of feature subsets. The crossover process is assisted by the application of feature importance while the mutation process is assisted by the application of a measure proposed in this study, i.e. the value of the tendency proportion. GA requires a method for checking the accuracy of feature subsets. The method used is decision tree algorithm.

In the case study where data taken from the sociolla.com website, 24 combinations of GA parameters were applied to obtain the feature subset. Prediction models using selected feature subsets have better accuracy than similar models that use all features and prediction models with random features subsets. Based on the two-sample t-test, it is concluded that prediction model with selected feature subset has a significantly higher average accuracy than the other prediction models ($\alpha = 0.05$), with the average increase in the accuracy of the model for the case "cleanser" by 7.1 percent and the average increase in model accuracy for "treatment cases by 11.1 percent.

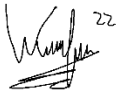
KATA PENGANTAR

Puji dan syukur penulis panjatkan karena atas berkat dan rahmatNya, penulis dapat menyelesaikan penelitian skripsi dengan judul "*Feature Selection* Menggunakan *Genetic Algorithm* Dalam Melakukan Prediksi *Repuchase Intention*" sebagai syarat kelulusan. Penulisan skripsi ini tentu tidak lepas dari dukungan beberapa pihak. Penulis mengucapkan banyak terima kasih kepada:

1. Bapak Dedy Suryadi, S.T., M.S., Ph.D. selaku dosen pembimbing penulis yang selalu meluangkan waktu untuk senantiasa memberikan masukan dan dukungan terkait penelitian dan penulisan skripsi penulis.
2. Bapak Alfian Tan, S.T., M.T. selaku dosen wali penulis di UNPAR yang banyak memberikan masukan dan bantuan terkait kegiatan perkuliahan dan kebimbangan penulis selama penelitian skripsi.
3. Bapak Romy Loice, S.T., M.T. selaku koordinator skripsi penulis yang sudah menyusun kegiatan skripsi dengan baik.
4. Orang tua dan keluarga penulis yang selalu menemani dan memberikan dukungan kepada penulis dalam menyelesaikan penelitian dan penulisan skripsi.
5. Orang terdekat penulis yang tidak bisa disebutkan satu per satu yang senantiasa menemani penulis terkait penyusunan skripsi dan menemani penulis selama kegiatan perkuliahan di Unpar.

Penulis ingin meminta maaf apabila terdapat kesalahan, sehingga penulis terbuka terhadap saran dan kritik terkait laporan ini. Semoga hasil karya penulis dapat memberikan manfaat bagi banyak pihak. Demikian penyusunan skripsi inim saya ucapkan terima kasih.

Bandung, 11 Agustus 2020



Wellington

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

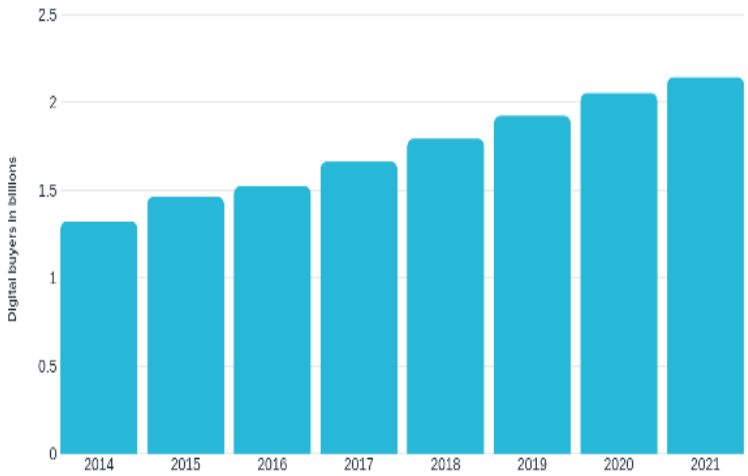
DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

Subbab ini menjelaskan tentang latar belakang masalah, identifikasi dan rumusan masalah, pembatasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian dan sistematika penulisan yang dilakukan pada penelitian terkait *feature selection* menggunakan *Genetic Algorithm* dalam memprediksi *repurchase intention* dari pelanggan.

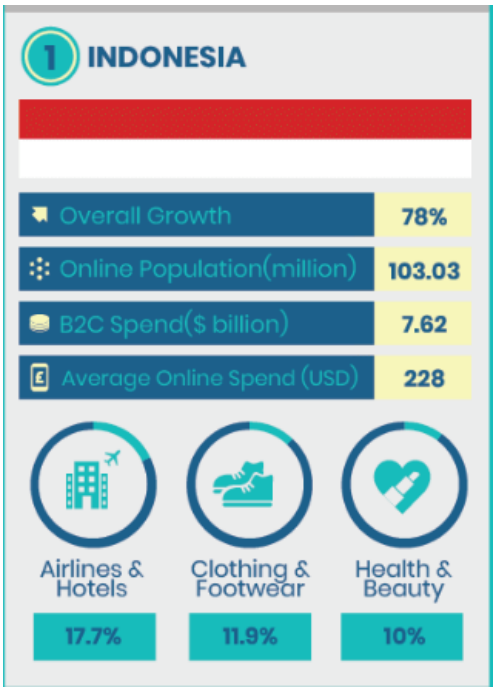
I.1 Latar Belakang Masalah

Di era globalisasi, perkembangan teknologi menuntut beberapa aspek dalam kehidupan untuk bergerak dinamis. Salah satunya terkait dengan cara perolehan informasi. Segala macam jenis informasi saat ini dapat dengan mudahnya diakses melalui internet. Bahkan, saat ini dapat melakukan transaksi secara *online*. Tren transaksi secara *online* ini dikenal dengan istilah *e-commerce.* Atas kemudahan dalam bertransaksi ini, semakin besar peluang adanya konsumen baru yang mampu melakukan kegiatan transaksi *online.* Hal ini juga didukung dengan data terkait perkembangan konsumen *online* yang bertambah banyak dari tahun ke tahun seperti dapat dilihat pada Gambar I.1.



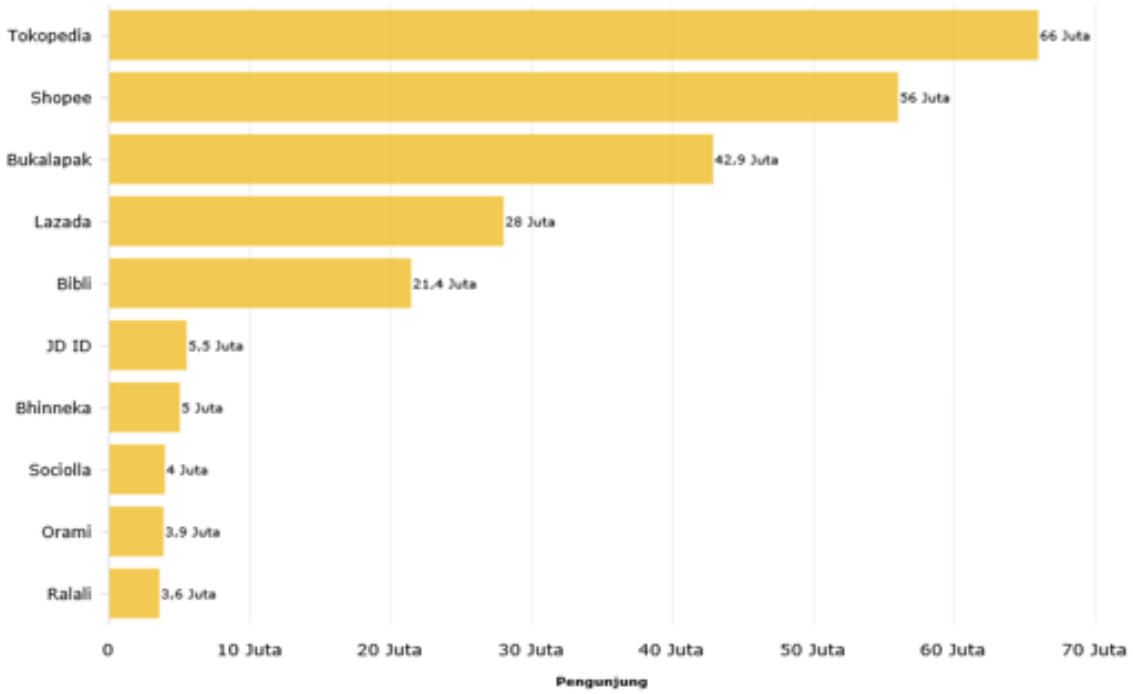
Gambar I.1 Grafik Perkembangan *Digital Buyers* Per Tahun
(Sumber : https://www.oberlo.com/blog/ecommerce-statistics-guide-your-strategy)

Pada Gambar I.1, dapat dilihat bahwa setiap tahunnya terjadi peningkatan jumlah *digital buyer*. Selain itu, terdapat statistik bahwa Indonesia adalah salah satu negara dengan pertumbuhan *e-commerce* tercepat di dunia. Gambaran terkait statistik *e-commerce* Indonesia dapat dilihat pada Gambar I.2.



Gambar I.2 *Summary E-commerce Indonesia*
(Sumber : https://merchantmachine.co.uk/saturated-sectors)

Dalam *e-commerce* terdiri dari banyak sekali sektor sesuai dengan kebutuhan konsumen. Terdapat beberapa sektor dimana masyarakat Indonesia sering menghabiskan uangnya dalam berbelanja secara *online*, diantaranya para konsumen paling sering menghabiskan uang pada sektor *Airline & Hotel* (17.7%), *Clothing & Footwear* (11.9 %), dan *Health & Beauty* (10%). Terkait data tersebut, situs *e-commerce* tentu mengalami pertumbuhan yang cukup positif, baik dari segi pendapatan maupun banyaknya pengunjung. Dari sekian banyak situs *e-commerce* yang ada di Indonesia, terdapat beberapa situs yang telah bergerak menjadi perusahaan raksasa di bidang *e-commerce*. Hal ini dapat ditandai dari banyaknya pengunjung yang datang ke situs *e-commerce* terkait. Salah satu dari beberapa situs *e-commerce* yang memiliki pengunjung terbanyak adalah Sociolla. Grafik terkait statistik *e-commerce* Indonesia dengan pengunjung terbesar selama kuartal III pada tahun 2019 dapat dilihat pada Gambar I.3.



Gambar I.3. Grafik *E-commerce* Indonesia Pengunjung Terbesar Kuartal III 2019 (Sumber : https://databoks.katadata.co.id/datapublish/2019/10/10/tren-pengguna-*e-commerce*-2017-2023)

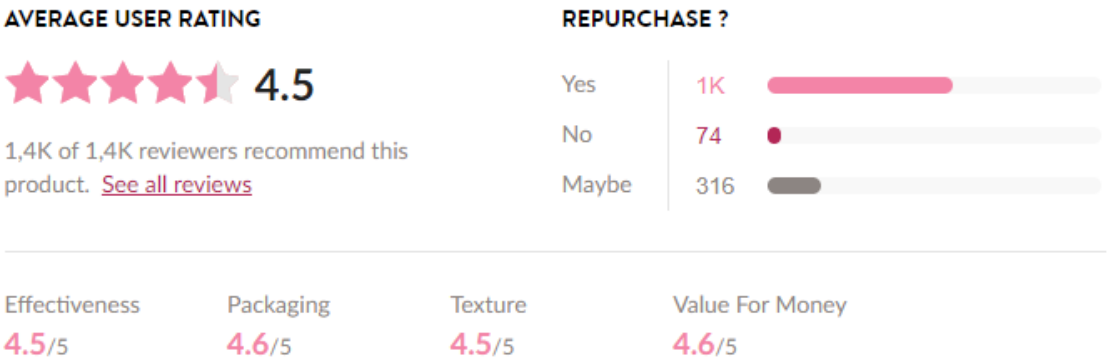
Sociolla merupakan satu dari banyak situs yang bergerak di bidang kecantikan. SOCO by Sociolla adalah situs resmi dari Sociolla yang menampung para pengguna dalam satu cakupan komunitas. Salah satu fitur yang terdapat dalam SOCO adalah para pengguna dapat memberikan aspirasi melalui ulasan secara lengkap terkait kosmetik maupun produk kecantikan lainnya. Ulasan yang diberikan kepada suatu produk nantinya ditulis dengan berbagai poin-poin penilaian. Salah satu poin penilaian yang penting adalah ingin membeli kembali produk tersebut atau tidak. Dengan mengetahui apakah produk dibeli kembali atau tidak menyatakan suatu produk sukses dan diterima oleh masyarakat. Hal ini dapat disebut dengan *repurchase intention.* Untuk mengetahui *repurchase intention*, dapat dibuat sebuah model guna memprediksi keputusan tersebut dengan menggunakan bantuan dari ulasan. Dengan mengetahui sukses tidaknya suatu produk diterima oleh masyarakat berpotensi dalam mengetahui perihal kenaikan maupun penurunan produk. Hal ini penting untuk diketahui oleh para produsen agar produsen dapat memperbaiki kesalahan yang ada pada produknya ataupun membuat strategi yang tepat untuk produk terkait. Kemampuan untuk memprediksi pembeli akan membeli kembali produk terkait atau tidak tentu menjadi suatu

keunggulan bagi suatu perusahaan. Yang terjadi saat ini adalah, belum ada perusahaan yang memprediksi keinginan konsumen membeli kembali produk (*repurchasing).* Padahal dengan mengetahui keinginan *repurchasing*, perusahaan dapat menyusun langkah yang tepat dalam menjalankan strategi pemasaran.

I.2 Identifikasi dan Rumusan Masalah

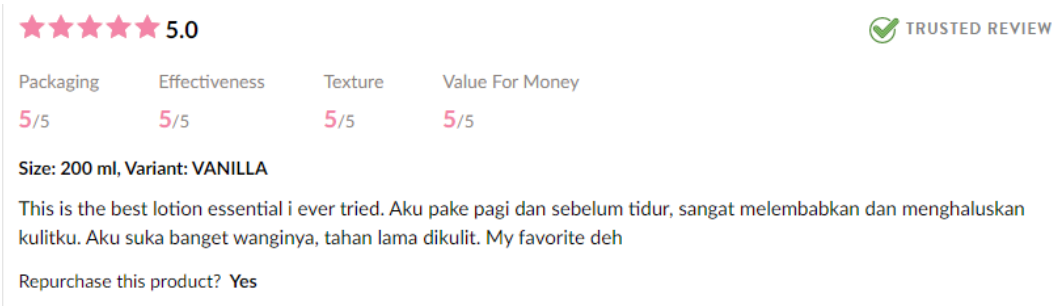
Secara mendasar, meramalkan keinginan *repurchasing* dapat dilakukan dengan melihat dari *ulasan* yang ada. Semakin positif sebuah ulasan, maka semakin besar peluang seorang konsumen untuk membeli kembali produk serupa.

Ulasan yang ditemukan dalam SOCO by Sociolla sangat banyak sehingga perlu adanya pengolahan data ulasan produk kosmetik agar mudah untuk dibaca. Gambar terkait *summary* ulasan produk yang ada dalam SOCO by Sociolla dapat dilihat pada Gambar I.4



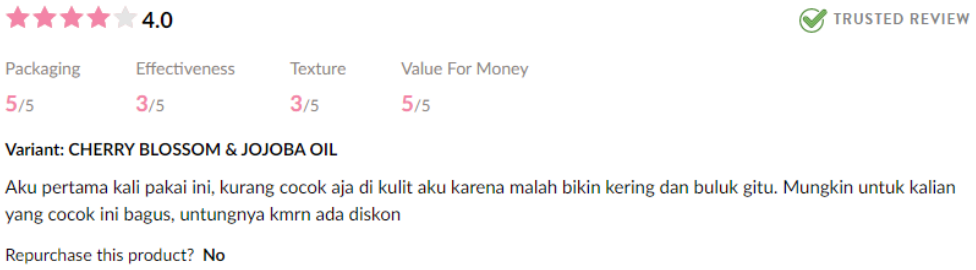
Gambar I.4 *Summary* Ulasan Produk
(Sumber : https://review.soco.id/product/body-lotion/20799-body-essential)

Setiap ulasan yang ditulis oleh konsumen memiliki poin penilaian *repurchase* . Penilaian terkait produk tersebut cocok atau tidak sehingga dapat timbul keinginan *repurchase* dituliskan dengan jelas dalam ulasan terkait. Gambar I.5 merupakan gambar *review* salah seorang konsumen melakukan keputusan *repurchase* terhadap produk tersebut.



Gambar I.5 *Review Repurchase* "Yes" Konsumen
(Sumber : https://review.soco.id/product/body-lotion/20799-body-essential?page=18)

Setiap produk tentu tidak dapat diterima dengan baik oleh seluruh pasar dalam masyarakat. Tidak heran bila ditemukan ulasan konsumen dimana terdapat keinginan untuk tidak melakukan keputusan *repurchase* dengan berbagai sebab. Gambar terkait ketidakinginan melakukan *repurchase* seorang konsumen dapat dilihat pada Gambar I.6.



Gambar I.6 *Review Repurchase* "No" Konsumen
(Sumber : https://review.soco.id/product/body-lotion/20799-body-essential?page=7)

Setiap kata dalam ulasan dapat dianggap sebagai sebuah *feature*. Semakin banyak kata dalam sebuah ulasan akan memiliki *features* yang banyak pula. Menurut Ma & Xia (2017), banyaknya *features* membuat hasil prediksi model yang diperoleh menjadi tidak akurat. Hal ini disebabkan karena belum diketahui apakah seluruh *features* yang ada adalah *features* yang relevan. Semakin banyak *features*, waktu yang dibutuhkan untuk membuat model *repurchase* yang akurat juga semakin lama. Menurut Homsapaya & Sornil (2017) dan Shaharanee & Hadzig (2015), semakin banyak *feature* yang tidak relevan, akan membuat performansi sistem dari segi akurasi dan interpretasi model yang dibuat semakin buruk. Hal ini dapat menimbulkan *curse of the dimensionality,* dimana semakin besar dimensi, dalam kasus ini adalah *features,* maka semakin besar tingkat *error* yang dihasilkan (Pavlenko, 2003). Oleh karena itu, akan dilakukan pengurangan

features menjadi *feature subset* yang hanya terdiri dari feature yang dianggap penting.

Permasalahan pemilihan *feature* yang tepat dapat diselesaikan dengan metode *feature selection*. *Feature selection* merupakan metode untuk mengurangi jumlah variabel input menjadi variabel yang diyakini paling berguna bagi model untuk memprediksi variabel target (Brownlee, 2019). Dalam *feature selection,* perlu mempertimbangkan dua hal yaitu generasi *subset* dan evaluasi *subset.* Generasi *subset* menentukan kandidat *feature subset* dan evaluasi subset untuk mengukur kualitas dari *subset*.

Dengan keterbatasan waktu dalam memilih *feature subset* terbaik secara tepat, perlu dilakukan metode pendekatan. Disaat inilah, algoritma metaheuristik diperlukan untuk mendapat *feature subset* hampir terbaik dalam waktu relatif singkat. Algoritma metaheuristik berguna untuk optimisasi permasalahan kompleks, berbeda dengan algoritma heuristik yang hanya memberikan solusi terkait suatu permasalahan. Algoritma metaheuristik untuk kasus ini berguna untuk mengoptimisasi *feature subset* untuk model prediksi keputusan *repurchase* konsumen. Beberapa metode algoritma metaheuristik seperti *Genetic Algorithm* (GA), *Genetic Programming* (GP), *Ant Colony Optimization* (ACO) dan *Particle Swarm Optimization* (PSO), telah membuktikan performansi yang baik dalam mencari solusi untuk permasalahan kompleks sehingga cocok untuk digunakan dalam menyelesaikan permasalahan *feature selection* (Ma & Xia, 2017). Dalam studi yang dilakukan oleh Santana et. al. (2010) dalam Ma & Xia (2017) menyatakan bahwa *Ant Colony Optimization* cocok digunakan jika variabel yang diperhitungkan berjumlah sedikit, namun jika variabel target berjumlah banyak, maka *Genetic Algorithm* lebih tepat untuk digunakan.

Hasil dari *Genetic Algorithm* ini memiliki menghasilkan *feature subset* yang berkualitas dan memberikan jawaban yang menjanjikan. Dalam melakukan GA, terdapat beberapa parameter yang harus ditentukan, seperti *population size, mutation rate* dan *crossover rate*. Kombinasi dari parameter yang berbeda menghasilkan *feature subsets* yang berbeda pula. Dengan pilihan *feature subses* yang beragam, tidak semua *feature subsets* yang dihasilkan memiliki features terbaik sehingga perlu dilakukan evaluasi terhadap akurasi *feature subset*.

Akurasi dari *feature subset* mempengaruhi akurasi dari prediksi *repurchase intention*. *Tool* yang dimaksud untuk mengukur keakuratan *feature*

subset tersebut adalah menggunakan *Machine Learning*. Algoritma *Machine learning* akan digunakan untuk memprediksi keputusan *repurchasing* dengan mempelajari *pattern* dari *feature.*

Dalam melakukan peramalan *repurchase intention*, perlu 2 kegiatan yang dilakukan. Pertama, perlu melakukan *feature selection* guna mereduksi dimensi *feature* ulasan menjadi *feature subset* terbaik. Setelah mendapat *feature subset*, hasil *feature selection* tersebut digunakan sebagai input data dalam model algoritma *machine learning*. Model algoritma akan digunakan untuk memprediksi *repurchase intention*, dimana semakin banyak jawaban (*label)* yang berhasil diprediksi maka akurasi model semakin tinggi. Dalam penelitian ini, akan diterapkan 2 buah usulan dalam algoritma Genetic Algorithm untuk mencari solusi yang lebih baik. Kedua usulan ini, yaitu penerapan proporsi kecenderungan dan *feature importance,* membedakan penelitian ini dibandingkan penelitian terdahulu. Penjelasan lebih rinci terkait penerapan usulan akan dipaparkan pada bab-bab berikutnya. Dengan teridentifikasinya alur permasalahan tersebut, ditemukan beberapa pokok permasalahan. Berikut ini merupakan rumusan masalah yang dibuat untuk penelitian ini :

1. Bagaimana penerapan *Genetic Algorithm* untuk melakukan *feature selection* dalam mendapatkan *feature subset* yang baik*?*
2. Apa saja *features* yang mempengaruhi keputusan *repurchase* konsumen terhadap produk kosmetik?
3. Bagaimana perbandingan performansi *Genetic Algorithm* dengan berbagai kombinasi parameter dalam mendapatkan *feature subset* terbaik?

I.3 Pembatasan Masalah

Berdasarkan identifikasi masalah yang telah dirumuskan, dilakukan pembatasan masalah yang digunakan selama penelitian ini. Terdapat beberapa batasan masalah yang ditetapkan. Pembatasan masalah tersebut adalah sebagai berikut.

1. Kategori produk yang digunakan dalam penelitian ini terbatas pada kategori produk *cleanser* dan *treatment.*
2. Data ulasan konsumen yang digunakan dalam penelitian ini diambil pada tanggal 18 Maret 2019 hingga 26 Agustus 2019

3. Label atau kelas dari data ulasan yang digunakan bersifat *binary* dengan menggunakan keputusan *repurchase* "yes" dan "no"
4. Jumlah data ulasan yang digunakan selama penelitian per kategori produk adalah 6250 ulasan

I.4 Tujuan Penelitian

Setelah mengetahui pembatasan masalah, dilanjutkan dengan tujuan penelitian. Tujuan penelitian berguna untuk menjawab rumusan masalah. Berikut ini adalah tujuan dari penelitian :

1. Menerapkan *Genetic Algorithm* melakukan *feature selection* dalam mengoptimisasi *feature subset*.
2. Mengetahui *feature* yang mempengaruhi keputusan *repurchase* konsumen terhadap produk kosmetik.
3. Mendapatkan perbandingan performansi *Genetic Algorithm* dengan berbagai kombinasi *parameter* dalam mendapatkan *feature subset* terbaik.

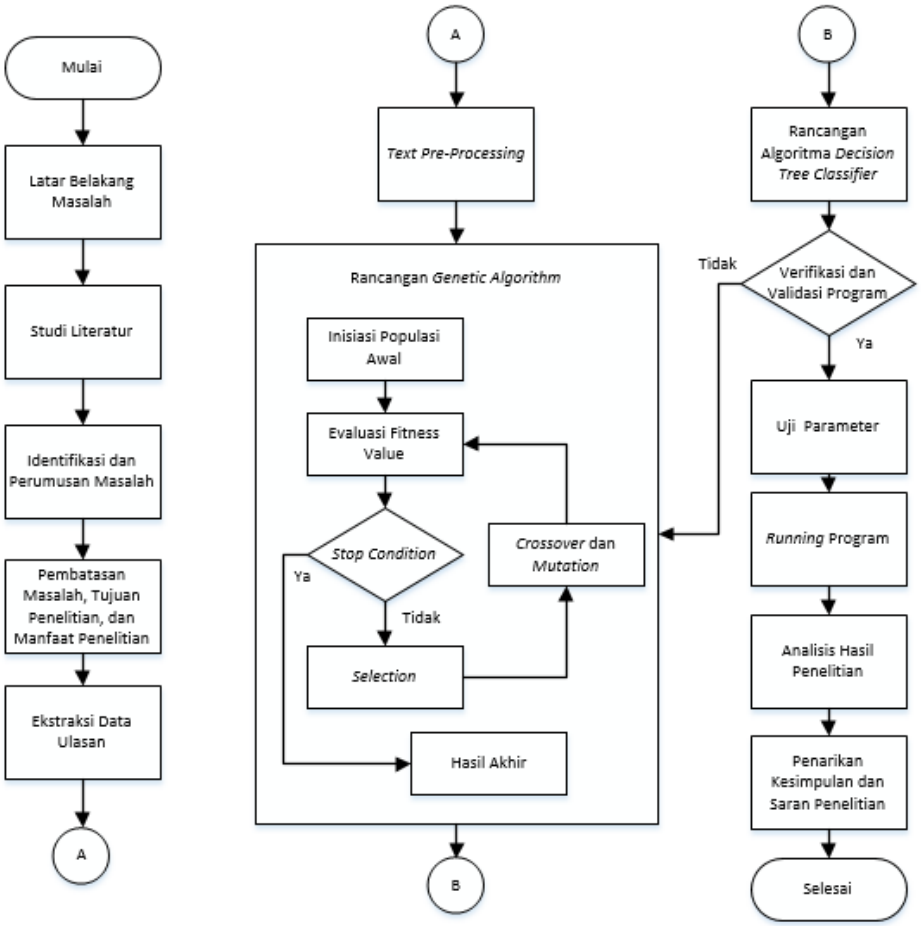
I.5 Manfaat Penelitian

Manfaat penelitian berguna untuk mengetahui dampak yang diberikan dari dilakukan suatu penelitian. Terdapat beberapa manfaat yang dapat diberikan dalam penelitian ini. Berikut ini merupakan manfaat dari penelitian yang dilakukan:

1. Memberikan alternatif penyelesaian untuk melakukan *Feature Selection.*
2. Menambahkan pengetahuan dan wawasan pembaca mengenai penerapan *Genetic Algorithm* dalam *Feature Selection.*
3. Menambahkan pengetahuan dan wawasan pembaca mengenai penerapan algoritma *Machine learning*.
4. Menambah referensi untuk penelitian yang berkaitan dengan *Genetic Algorithm, Machine Learning Algorithm,* dan *Feature Selection.*

I.6 Metodologi Penelitian

Metodologi penelitian merupakan langkah-langkah penelitian yang dilakukan. Metodologi penelitian terdiri dari beberapa tahapan. Gambar I.7 merupakan alur metodologi penelitian yang akan dilakukan.



Gambar I.7 Alur Metodologi Penelitian

Pada Gambar I.7 terdapat beberapa langkah yang dilakukan dalam penelitian ini. Metodologi bertujuan untuk mengetahui langkah apa saja yang harus dilakukan dalam penelitian. Berikut ini merupakan penjelasan dari metodologi penelitian yang digambarkan dalam bentuk *flowchart.*

1. Latar Belakang Masalah

Dalam memulai suatu penelitian, diawali dengan mencari latar belakang masalah. Latar belakang masalah menjadi suatu acuan dalam melakukan penelitian dan diperlukan guna mengetahui bagaiman suatu masalah bisa terjadi. Setelah mengetahui latar belakang suatu masalah, dilanjutkan dengan tahap berikutnya

2. Studi Literatur

Studi literatur dilakukan untuk membantu memahami masalah dan penyelesaian yang dapat digunakan. Pada tahap ini dilakukan pengumpulan

sumber atau referensi terkait dengan *Feature Selection, Machine Learning Algorithm* dan *Genetic Algorithm*.

3. Identifikasi dan Perumusan Masalah

Identifikasi dan perumusan masalah merupakan dasar dari dilakukan penelitian. Pada identifikasi masalah dijelaskan mengenai permasalahan yang akan diteliti yaitu *feature selection*, hal-hal yang akan diteliti, dan penentuan metode yang digunakan untuk menyelesaikan masalah yaitu GA. lalu dilakukan penyelesaian masalah yang lebih besar yakni meramalkan kemampuan repurchasing konsumen dengan menggunakan metode algoritma *machine learning*. Berdasarkan identifikasi masalah, dibuat rumusan masalah dari penelitian yang dilakukan.

4. Pembatasan Masalah, Penentuan Tujuan dan Manfaat Penelitian

Pembatasan masalah bertujuan agar penelitian yang dilakukan menjadi lebih terfokus. Penelitian yang lebih terfokus membantu menghasilkan kesimpulan yang baik dan benar. Setelah melakukan pembatasan masalah akan dilanjutkan dengan penentuan tujuan dan manfaat penelitian. Penentuan tujuan dan manfaat penelitian ditentukan agar penelitian yang dilakukan memiliki tujuan yang jelas dan dapat memberikan manfaat untuk pembaca.

5. Ekstraksi Data Ulasan

Pada tahap ini akan dilakukan ekstraksi data yang dibantu menggunakan program PyCharm. Ekstraksi data dilakukan sesuai dengan objek penelitian yaitu Review SOCO by Sociolla. Setelah melalui tahap ekstraksi, akan dilanjutkan dengan tahap berikutnya

6. *Text Pre-Processing*

Tahap ini melakukan pengolahan data mentah menjadi dataset yang siap untuk digunakan dalam pembuatan model prediksi. Tahap ini mengolah teks data mentah dalam beberapa tahap yang berguna untuk menghilangkan kata yang dianggap tidak memberikan makna yang dibutuhkan sebagai sebuah features penting. Setelah melalui tahap ini akan dilanjutkan pada tahap rancangan *Genetic Algorithm*.

7. Merancang *Genetic Algorithm*

Pada tahap ini akan dilakukan perancangan *Genetic Algorithm.* Awalnya akan dilakukan inisiasi populasi awal. Populasi awal dibuat dengan membuat

features menjadi suatu kromosom. Sebuah populasi akan diisi dengan beberapa sampel kromosom acak. Kemudian akan dilakukan proses *mutasi, crossover* dan *seleksi* untuk mencari nilai *fitness value* (FV) terbaik.

Setelah mendapat nilai *fitness value* terbaik, akan dilakukan perbandingan *fitness value* terbaik dalam beberapa kombinasi parameter. Hal ini dilakukan agar menentukan *feature subset* mana yang akan digunakan dalam pembuatan model dalam tahap berikutnya. Selanjutkan akan dilakukan perancngan algoritma decision tree.

8. Merancang Algoritma *Decision Tree*

Pada tahap ini akan dilakukan perancangan *Decision Tree.* Tahap ini akan menyusun algorithma *machine learning* yang digunakan untuk melakukan prediksi *repurchase intention* dan mengevaluasi feature subset. Hasil akhir dari kegiatan prediksi ini adalah akurasi dari berhasil atau tidaknya *machine learning* dalam menebak suatu review dalam *repurchase intention*. Setelah membuat rancangan algoritma, akan dilanjutkan verifikasi dan validasi program.

9. Verifikasi dan Validasi Program

Program yang telah dibuat berdasarkan rancangan algoritma akan diverifikasi dan validasi. Hal ini bertujuan untuk memastikan model program yang telah dibuat dapat diaplikasikan ditanpa adanya kesalahan. Setelah melakukan verifikasi dan validasi, akan dilanjutkan dengan melakukan uji parameter.

10. Uji Parameter

Pada tahap ini akan dilakukan dibuat kombinasi dari nilai jumlah kromosom, *mutation rate*, *crossover rate* dan *population size*. Setelah dibuat kombinasi, akan dilakukan pemilihan kombinasi parameter. Setelah melalui tahap ini, akan dilanjutkan dengan tahap berikutnya

11. *Running* Program

Berdasarkan nilai parameter terpilih akan dilakukan running sebanyak beberapa replikasi guna membuat model *repurchase intention*. Setelah memalui tahap ini akan dilanjutkan ke analisis hasil penelitian.

12. Analisis Hasil Penelitian

Analisis dilakukan untuk melihat dan membandingkan antara hasil penerapan metode algoritma *Decision Tree* menggunakan *feature subset* yang didapatkan melalui *Genetic Algorithm* dengan hasil penerapan metode algoritma

machine learning menggunakan *random feature subset*. Analisis juga dilakukan untuk membandingkan hasil *fitness value* yang dihasilkan oleh *feature subset* hasil penerapan GA dan *Decision Tree* dengan hasil *fitness value* menggunakan seluruh *features*. Setelah melakukan analisis, akan dilakukan penarikan kesimpulan dan saran.

13. Kesimpulan dan Saran

Tahap terakhir pada penelitian yang dilakukan adalah penarikan kesimpulan berdasarkan penelitian yang dilakukan. Selain itu, akan diberikan juga saran untuk penelitian selanjutnya.

I.7 Sistematika Penulisan

Penulisan hasil penelitian ini disusuh dalam lima bab yang dilakukan secara berurutan dan jelas sehingga hasil penelitian ini dapat dengan mudah dibaca dan dipahami. Berikut merupakan sistematika penulisan hasil penelitian ini.

BAB I PENDAHULUAN

Bab I berisi mengenai latar belakang masalah, identifikasi dan perumusan masalah, pembatasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan yang dilakukan pada penelitian terkait *feature selection* menggunakan algoritma *Binary Particle Swarm Optimization* (BPSO) dalam membuat model prediksi *repurchase intention*.

BAB II TINJAUAN PUSTAKA

Bab II berisi mengenai tinjauan pustaka, teori, serta studi literatur yang digunakan terkait dengan penelitian yang dilakukan. Pada bab ini akan dijabarkan mengenai teori-teori, formula atau rumus, serta metode yang digunakan sebagai dasar pemecahan masalah, pengolahan data, serta analisis terkait *feature selection* menggunakan algoritma *Genetic Algorithm* dalam membuat model prediksi *repurchase intention*.

BAB III PENGOLAHAN DATA

Bab III berisi mengenai proses pengolahan data ulasan mentah menjadi data yang siap untuk digunakan pada tahap berikutnya. Pada bab ini dilakukan

proses penghilangan data sesuai dengan batasan penelitian, penyortiran data yang tidak digunakan, pengubahan bentuk ulasan menjadi bentuk tabel.

Pada bab ini juga dilakukan perancangan algoritma *Genetic Algorithm (*GA*) dan* perancangan algoritma *Decision Tree*. Selanjutnya adalah menerapkan algoritma dan mengoperasikan algoritma ini untuk menyelesaikan kasus *feature selection* dalam membuat model prediksi *repurchase intention* pada program *PyCharm* dengan menggunakan bahasa pemrograman *Python*. Dalam bab ini pula dilakukan pengujian parameter untuk setiap parameter pada kedua algoritma yang digunakan. Terakhir bab ini melakukan pencarian fitur yang berpengaruh pada model prediksi *repurchase intention* dengan mempertimbangkan performansi model.

BAB IV ANALISIS HASIL

Bab IV berisi mengenai analisis hasil penerapan algoritma *Genetic Algorithm* (GA) dalam menyelesaikan kasus *feature selection* untuk membangun model prediksi *repurchase intention*. Analisis ini meliputi analisis pemilihan parameter, analisis algoritma, analisis hasil pemilihan fitur, dan analisis performansi model prediksi *repurchase intention*.

BAB V KESIMPULAN DAN SARAN

Bab V berisi mengenai kesimpulan dari penelitian *feature selection* menggunakan algoritma *Genetic Algorithm* (GA) dalam membuat model prediksi *repurchase intention* serta memberikan saran untuk penelitian selanjutnya yang memiliki topik serupa.