

***FEATURE SELECTION MENGGUNAKAN  
ALGORITMA BINARY PARTICLE SWARM  
OPTIMIZATION (BPSO) UNTUK MEMPREDIKSI  
REPURCHASE INTENTION***

**SKRIPSI**

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

Disusun oleh :

Nama : Dimas Adrianto  
NPM : 2016610166



**PROGRAM STUDI SARJANA TEKNIK INDUSTRI  
JURUSAN TEKNIK INDUSTRI  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS KATOLIK PARAHYANGAN  
BANDUNG  
2020**

***FEATURE SELECTION MENGGUNAKAN  
ALGORITMA BINARY PARTICLE SWARM  
OPTIMIZATION (BPSO) UNTUK MEMPREDIKSI  
REPURCHASE INTENTION***

**SKRIPSI**

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

Disusun oleh :

Nama : Dimas Adrianto  
NPM : 2016610166



**PROGRAM STUDI SARJANA TEKNIK INDUSTRI  
JURUSAN TEKNIK INDUSTRI  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS KATOLIK PARAHYANGAN  
BANDUNG  
2020**

**FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS KATOLIK PARAHYANGAN  
BANDUNG**



Nama : Dimas Adrianto  
NPM : 2016610166  
Jurusan : Teknik Industri  
Judul Skripsi : *FEATURE SELECTION MENGGUNAKAN ALGORITMA BINARY PARTICLE SWARM OPTIMIZATION (BPSO) UNTUK MEMPREDIKSI REPURCHASE INTENTION*

**TANDA PERSETUJUAN SKRIPSI**

Bandung, 26 Agustus 2020

**Ketua Program Studi Sarjana Teknik Industri**

Romy Lince, S.T., M.T.

**Pembimbing Tunggal**

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



Jurusan Teknik Industri  
Fakultas Teknologi Industri  
Universitas Katolik Parahyangan

## **Pernyataan Tidak Mencontek atau Melakukan Tindakan Plagiat**

Saya, yang bertanda tangan di bawah ini,

Nama : Dimas Adrianto

NPM : 2016610166

dengan ini menyatakan bahwa skripsi dengan judul :

***“ FEATURE SELECTION MENGGUNAKAN ALGORITMA BINARY PARTICLE SWARM OPTIMIZATION (BPSO) UNTUK MEMREDIKSI REPURCHASE 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, 26 Agustus 2020

Dimas Adrianto  
2016610166

## ABSTRAK

*Feature selection* merupakan proses untuk memilih sejumlah fitur relevan yang berpengaruh terhadap *output* yang diinginkan dari sebuah model. Di dalam kasus model prediksi *repurchase intention* pada penelitian ini, ulasan pelanggan merupakan dokumen yang direpresentasikan sebagai kumpulan kata-kata. Kata dalam konteks *feature selection* merupakan sebuah fitur. Banyaknya jumlah ulasan pelanggan menyebabkan meningkatnya jumlah kata (fitur) yang perlu diseleksi sehingga *feature selection* merupakan masalah yang cukup kompleks untuk dilakukan dan diperlukan sebuah metode pendekatan untuk membantu melakukan *feature selection* dalam membangun model prediksi *repurchase intention* dari ulasan pelanggan.

Penelitian ini bertujuan untuk menerapkan algoritma *Binary Particle Swarm Optimization* (BPSO) untuk melakukan *feature selection* dalam membangun sebuah model prediksi *repurchase intention* dari ulasan pelanggan serta mengetahui fitur (kata) mana yang berpengaruh di dalam model prediksi *repurchase intention*. Algoritma BPSO merupakan modifikasi dari algoritma *Particle Swarm Optimization* (PSO) yang meniru proses mencari makanan pada kehidupan populasi burung. Algoritma *k-Nearest Neighbors* (*k-NN*) digunakan untuk mengukur performansi model prediksi *repurchase intention*. Pada penelitian ini juga dilakukan sebuah usulan untuk melakukan pengurangan kata (fitur) awal berdasarkan frekuensi terbesar dengan menggunakan prinsip *Pareto* (*80/20 rule*) serta mengurutkan kata (fitur) berdasarkan kecenderungan sentimen kata tersebut dengan menggunakan metode *Sentiment Orientation-Pointwise Mutual Information* (SO-PMI).

Berdasarkan hasil *feature selection* pada dua *dataset* ulasan pelanggan yang diambil dari *website sociolla.com* (*dataset moisturizer 1* dengan 2.614 ulasan dan *dataset moisturizer 2* dengan 5.334 ulasan), algoritma BPSO mampu mengurangi dari 341 fitur menjadi 229 fitur serta meningkatkan akurasi model prediksi dari 78,44% menjadi 81,31% pada *dataset moisturizer 1* dan mampu mengurangi dari 605 fitur menjadi 389 fitur serta meningkatkan akurasi model prediksi dari 78,66% menjadi 79,87% pada *dataset moisturizer 2*. Dengan menggunakan *two-sample t-test* ( $\alpha = 5\%$ ), dapat ditarik kesimpulan bahwa terjadi perbedaan rata-rata yang signifikan secara statistik terhadap peningkatan akurasi sebelum dan setelah *feature selection* pada kedua *dataset moisturizer*.

## **ABSTRACT**

*Feature selection is a process for selecting a number of relevant features that affect the desired output of a model. In this study, customer reviews are documents that are represented as collections of words. Words in the context of feature selection is a feature. The large number of customer reviews causes an increase in the number of words (features) that need to be selected. Feature selection is quite a complex problem and an approach method is needed to help perform feature selection in building a predictive model based on customer reviews.*

*This study aims to apply the Binary Particle Swarm Optimization (BPSO) algorithm to perform feature selection in building a repurchase intention prediction model based on customer reviews and determine which features (words) affect the repurchase intention prediction model. The BPSO algorithm is a modification of the Particle Swarm Optimization (PSO) algorithm that mimics the movement of organisms in a bird flock. k-Nearest Neighbors (k-NN) algorithm is used to measure the performance of the repurchase intention prediction model. In this study a proposal was made to reduce initial words (features) based on the largest frequency by using the Pareto principle (80/20 rule) and words (features) are sorted based on the sentiment tendencies of the words using the Sentiment Orientation-Pointwise Mutual Information method (SO-PMI).*

*Based on the feature selection results on two datasets of customer reviews taken from the sociolla.com website (dataset moisturizer 1 that contains 2,614 reviews and dataset moisturizer 2 that contains 5,334 reviews), the proposed BPSO algorithm is able to reduce from 341 features to 229 features and increase the accuracy of prediction models from 78, 44% to 81.31% in the moisturizer 1 dataset and was able to reduce from 605 features to 389 features and increase the accuracy of the prediction model from 78.66% to 79.87% in the moisturizer 2 dataset. Using the two-sample t-test ( $\alpha = 5\%$ ), it can be concluded that there is a statistically significant average difference to the increase in accuracy before and after feature selection on the two datasets.*

## KATA PENGANTAR

Puji dan Syukur saya ucapkan kepada Tuhan Yang Maha Esa atas berkat dan rahmat-Nya, saya dapat menyelesaikan karya ilmiah saya yang dalam bentuk skripsi untuk mencapai gelar Sarjana dalam bidang Teknik Industri dengan judul “*FEATURE SELECTION MENGGUNAKAN ALGORITMA BINARY PARTICLE SWARM OPTMIZATION (BPSO) UNTUK MEMPREDIKSI REPURCHASE INTENTION*”. Dalam proses penelitian dan penyusunan skripsi ini, saya mendapat banyak bantuan dan bimbingan dari berbagai pihak. Saya menyadari bahwa masih terdapat kekurangan dalam penyusunan skripsi ini, sehingga saya terbuka atas saran serta masukan dari berbagai pihak untuk menyempurnakan skripsi ini.

Skripsi ini saya dedikasikan secara khusus kepada ayah saya, Alm. Bapak Dr. Rudianto Ekawan, S.T., M.T. yang telah menjadi penyemangat dan inspirasi besar bagi pribadi saya. Saya juga ingin mengucapkan banyak terima kasih kepada semua pihak yang telah membantu, membimbing, mendukung, dari seluruh rangkaian penelitian hingga penyusunan skripsi ini, terutama saya ucapkan terima kasih kepada:

1. Ibu saya Wieke Widayani, kedua adik saya Achmad Dinand W. dan M. Reza Ardianto serta seluruh keluarga besar yang telah memberikan dukungan secara moral serta materi.
2. Bapak Dedy Suryadi, S.T., M.S., Ph.D sebagai dosen pembimbing skripsi saya yang telah membimbing dan banyak membantu saya dalam penelitian dan penyusunan laporan skripsi.
3. Bapak Alfian, S.T., M.T. dan Ibu Cynthia Prithadevi Juwono, Ir., M.S. sebagai dosen penguji proposal yang telah memberi masukan dan arahan untuk penelitian saya ini.
4. Bapak Hanky Fransiscus, S.T., M.T. dan Ibu Cynthia Prithadevi Juwono, Ir., M.S. sebagai dosen penguji siding skripsi yang telah memberi masukan dan arahan untuk penelitian saya ini.
5. Sahabat serta teman dekat (terutama Rizka Ayudia, Yudhistira Ari Kurniaputra, Difa Aulia G., M. Adhitya D.A.Y., Rahadian Agung H., M. Haseemy G., Naufaldhy O.S., Marcus Alexander, Reynaldi Pierera G.,

Cecilia Carissa, Dian Anggraini S., Adinda Oktaviani R., dan Wellington) yang telah menemani dan menyemangati perjalanan kehidupan perkuliahan serta skripsi saya di Teknik Industri Unpar.

6. Seluruh rekan-rekan Teknik Industri Unpar (terutama kelas A Teknik Industri Unpar angkatan 2016) yang telah menemani perjalanan kehidupan perkuliahan saya di Teknik Industri Unpar.

Saya ingin berterima kasih serta meminta maaf apabila terdapat pihak-pihak yang belum saya sebutkan dan apabila terdapat kesalahan pada penulisan laporan skripsi ini. Demikian penyusunan laporan skripsi ini, semoga laporan skripsi saya ini dapat bermanfaat bagi khalayak umum dan khususnya dalam ilmu teknik industri.

Bandung, 9 Agustus 2020

Dimas Adrianto



## DAFTAR ISI

<b>ABSTRAK</b> .....	i
<b>ABSTRACT</b> .....	ii
<b>KATA PENGANTAR</b> .....	iii
<b>DAFTAR ISI</b> .....	v
<b>DAFTAR TABEL</b> .....	vii
<b>DAFTAR GAMBAR</b> .....	ix
<b>DAFTAR LAMPIRAN</b> .....	xi
<b>BAB I PENDAHULUAN</b> .....	I-1
I.1 Latar Belakang Masalah .....	I-1
I.2 Identifikasi dan Rumusan Masalah.....	I-4
I.3 Pembatasan Masalah dan Pengambilan Asumsi Penelitian.....	I-7
I.4 Tujuan Penelitian .....	I-7
I.5 Manfaat Penelitian .....	I-8
I.6 Metodologi Penelitian.....	I-8
I.7 Sistematika Penulisan.....	I-12
<b>BAB II TINJAUAN PUSTAKA</b> .....	II-1
II.1 <i>Repurchase Intention</i> .....	II-1
II.2 <i>Feature Selection</i> .....	II-1
II.3 Algoritma Metaheuristik.....	II-2
II.4 Algoritma <i>Particle Swarm Optimization</i> (BPSO) .....	II-3
II.5 Algoritma <i>Binary Particle Swarm Optimization</i> (BPSO) .....	II-6
II.6 <i>Supervised Learning</i> .....	II-7
II.7 Algoritma <i>k-Nearest Neighbors</i> .....	II-8
II.8 <i>Confusion Matrix</i> .....	II-10
II.9 <i>Sentiment Orientation – Pointwise Mutual Information</i> (SO-PMI) .....	II-11
II.10 <i>t-Test</i> .....	II-12
II.11 <i>Text Pre-processing</i> .....	II-13
<b>BAB III PENGOLAHAN DATA</b> .....	III-1
III.1 Pemilihan Data Ulasan .....	III-1
III.2 <i>Text Pre-processing</i> .....	III-1
III.2.1 <i>Tokenizing</i> .....	III-2

III.2.2 <i>Stemming</i> .....	III-2
III.2.2 <i>Stopwords Removal</i> .....	III-2
III.3 Pengambilan Fitur Awal .....	III-3
III.3.1 Menghitung Frekuensi Kata yang Muncul.....	III-4
III.3.2 Penentuan Urutan Fitur Awal .....	III-5
III.4 Perancangan Algoritma <i>Binary Particle Swarm Optimization</i> .....	III-7
III.4.1 <i>Encoding</i> dan <i>Decoding</i> Partikel.....	III-8
III.4.2 Notasi Algoritma BPSO pada Kasus <i>Feature Selection</i> .....	III-9
III.4.3 Algoritma Utama <i>Binary Particle Swarm Optimization</i> .....	III-10
III.4.4 Algoritma Utama <i>Update</i> Posisi Partikel .....	III-12
III.4.4 Algoritma Evaluasi Posisi Partikel .....	III-13
III.5 Implementasi Algoritma pada Kasus <i>Feature Selection</i> .....	III-14
III.5.1 Verifikasi dan Validasi Algoritma .....	III-14
III.5.2 Penentuan Nilai Parameter .....	III-21
III.6 Implementasi Algoritma.....	III-27
III.6.1 Implementasi Algoritma pada <i>Dataset Moisturizer 1</i> .....	III-27
III.6.2 Implementasi Algoritma pada <i>Dataset Moisturizer 2</i> .....	III-33
III.7 Perbandingan Performansi Model Prediksi .....	III-39
III.8 Fitur-fitur Terpilih dari Hasil <i>Feature Selection</i> .....	III-41
<b>BAB IV ANALISIS</b> .....	IV-1
IV.1 Analisis Proses <i>Text Pre-processing</i> .....	IV-1
IV.2 Analisis Pemilihan Parameter .....	IV-1
IV.3 Analisis Hasil <i>Feature Selection</i> .....	IV-4
IV.4 Analisis Kata Terpilih .....	IV-5
<b>BAB V KESIMPULAN DAN SARAN</b> .....	V-1
V.1 Kesimpulan.....	V-1
V.2 Saran.....	V-2
<b>DAFTAR PUSTAKA</b>	
<b>LAMPIRAN</b>	
<b>DAFTAR RIWAYAT HIDUP</b>	

## DAFTAR TABEL

Tabel II.1	<i>Confusion Matrix 2x2</i> .....	II-10
Tabel III.1	Contoh Tahap <i>Tokenizing</i> .....	III-2
Tabel III.2	Contoh Tahap <i>Stemming</i> .....	III-2
Tabel III.3	Contoh Tahap <i>Stopwords Removal</i> .....	III-3
Tabel III.4	Sepuluh Kata dengan Frekuensi Terbesar <i>Dataset Moisturizer 1</i> ..	III-4
Tabel III.5	Duapuluh Kata Berdasarkan Nilai SO-PMI <i>Dataset Moisturizer 1</i> ..	III-7
Tabel III.6	Frekuensi per Kata Studi Kasus Sederhana .....	III-8
Tabel III.7	<i>Encoding</i> Partikel Kasus Sederhana .....	III-9
Tabel III.8	Matriks $r_2$ Partikel Studi Kasus Sederhana .....	III-15
Tabel III.9	Kecepatan Partikel Studi Kasus Sederhana.....	III-16
Tabel III.10	Nilai Sigmoid Partikel Studi Kasus Sederhana .....	III-16
Tabel III.11	Rekapitulasi Akurasi Percobaan Nilai $c$ .....	III-23
Tabel III.12	Rekapitulasi Parameter.....	III-26
Tabel III.13	Rekapitulasi Hasil <i>Dataset Moisturizer 1</i> .....	III-29
Tabel III.14	Jumlah Fitur Hasil <i>Dataset Moisturizer 1</i> .....	III-29
Tabel III.15	Akurasi Hasil <i>Dataset Moisturizer 1</i> .....	III-30
Tabel III.16	<i>Precision</i> Hasil <i>Dataset Moisturizer 1</i> .....	III-30
Tabel III.17	<i>Recall</i> Hasil <i>Dataset Moisturizer 1</i> .....	III-31
Tabel III.18	<i>F-score</i> Hasil <i>Dataset Moisturizer 1</i> .....	III-31
Tabel III.19	Rekapitulasi Hasil <i>Dataset Moisturizer 2</i> .....	III-35
Tabel III.20	Jumlah Fitur Hasil <i>Dataset Moisturizer 2</i> .....	III-35
Tabel III.21	Akurasi Hasil <i>Dataset Moisturizer 2</i> .....	III-36
Tabel III.22	<i>Precision</i> Hasil <i>Dataset Moisturizer 2</i> .....	III-36
Tabel III.23	<i>Recall</i> Hasil <i>Dataset Moisturizer 2</i> .....	III-37
Tabel III.24	<i>F-score</i> Hasil <i>Dataset Moisturizer 2</i> .....	III-37
Tabel III.25	Performansi Model Prediksi .....	III-40
Tabel III.26	Lima belas Fitur Terpilih <i>Dataset Moisturizer 1</i> .....	III-41
Tabel III.27	Lima belas Fitur Terpilih <i>Dataset Moisturizer 2</i> .....	III-41



## DAFTAR GAMBAR

Gambar I.1	Grafik Perkembangan <i>E-commerce</i> di Asia Tenggara 2019.....	I-1
Gambar I.2	Peningkatan per Kategori Produk <i>E-commerce</i> di Indonesia .....	I-2
Gambar I.3	Produk <i>Derma Angel Acne Patch Day</i> .....	I-3
Gambar I.4	Contoh Ulasan Pelanggan .....	I-6
Gambar I.5	Metodologi Penelitian.....	I-11
Gambar II.1	Urutan Model <i>Post-purchase</i> .....	II-1
Gambar II.2	Ilustrasi <i>Particle Swarm Optimization</i> .....	II-4
Gambar II.3	Ilustrasi Pergerakan Partikel .....	II-5
Gambar II.4	Model <i>Supervised Learning</i> .....	II-8
Gambar II.5	Cara Kerja Algoritma <i>k-Nearest Neighbors</i> .....	II-9
Gambar III.1	<i>Pareto Chart</i> Fitur Awal <i>Dataset Moisturizer 1</i> .....	III-5
Gambar III.2	Ilustrasi <i>Flowchart</i> Algoritma Utama .....	III-11
Gambar III.3	Ilustrasi <i>Flowchart</i> Algoritma <i>Update</i> Posisi Partikel.....	III-12
Gambar III.4	Ilustrasi <i>Flowchart</i> Algoritma <i>Evaluasi</i> Posisi Partikel.....	III-13
Gambar III.5	Program Membuka <i>Dataset</i> .....	III-17
Gambar III.6	Program <i>Update</i> dan <i>Evaluasi</i> Posisi .....	III-18
Gambar III.7	Contoh Program <i>Backend PySwarms</i> .....	III-19
Gambar III.8	Hasil <i>Program Input Dataset</i> .....	III-20
Gambar III.9	Hasil <i>Program</i> Algoritma BPSO .....	III-21
Gambar III.10	Grafik Percobaan Penentuan Parameter <i>Iters</i> .....	III-22
Gambar III.11	Probability Plot Percobaan nilai $c_1 >$ nilai $c_2$ .....	III-23
Gambar III.12	Probability Plot Percobaan nilai $c_2 >$ nilai $c_1$ .....	III-24
Gambar III.13	Hasil <i>Two Sample t-Test</i> Pengujian Parameter .....	III-24
Gambar III.14	Grafik Percobaan Penentuan Parameter $w = 0,75$ .....	III-25
Gambar III.15	Grafik Percobaan Penentuan Parameter $w = 1,5$ .....	III-25
Gambar III.16	Grafik Percobaan Penentuan Parameter $k$ .....	III-26
Gambar III.17	<i>Confusion Matrix</i> Sebelum Seleksi <i>Dataset Moisturizer 1</i> .....	III-28
Gambar III.18	<i>Confusion Matrix</i> Setelah Seleksi <i>Dataset Moisturizer 1</i> .....	III-28
Gambar III.19	Uji Distribusi Akurasi Sebelum Seleksi <i>Dataset Moisturizer 1</i> ..	III-32
Gambar III.20	Uji Distribusi Akurasi Setelah Seleksi <i>Dataset Moisturizer 1</i> ....	III-32
Gambar III.21	<i>One-tailed t-test</i> Akurasi <i>Dataset Moisturizer 1</i> .....	III-33

Gambar III.22 <i>Confusion Matrix</i> Sebelum Seleksi <i>Dataset Moisturizer 2</i> .....	III-34
Gambar III.23 <i>Confusion Matrix</i> Setelah Seleksi <i>Dataset Moisturizer 2</i> .....	III-35
Gambar III.24 Uji Distribusi Akurasi Sebelum Seleksi <i>Dataset Moisturizer 2</i> ..	III-38
Gambar III.25 Uji Distribusi Akurasi Setelah Seleksi <i>Dataset Moisturizer 2</i> ....	III-38
Gambar III.26 <i>One-tailed t-test</i> Akurasi <i>Dataset Moisturizer 2</i> .....	III-39

## DAFTAR LAMPIRAN

LAMPIRAN A CONTOH ULASAN PELANGGAN .....	A-1
LAMPIRAN B <i>SYNTAX PROGRAM</i> .....	B-1
LAMPIRAN C <i>LIST FITUR TERPILIH</i> .....	C-1





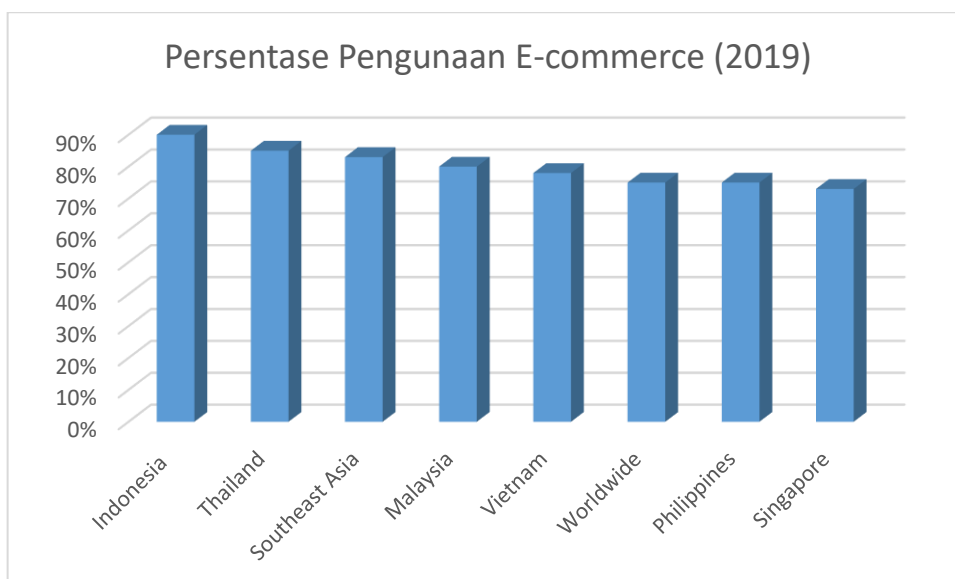
# BAB I

## PENDAHULUAN

Pada bab ini akan membahas 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 membangun model prediksi *repurchase intention* dari ulasan pelanggan.

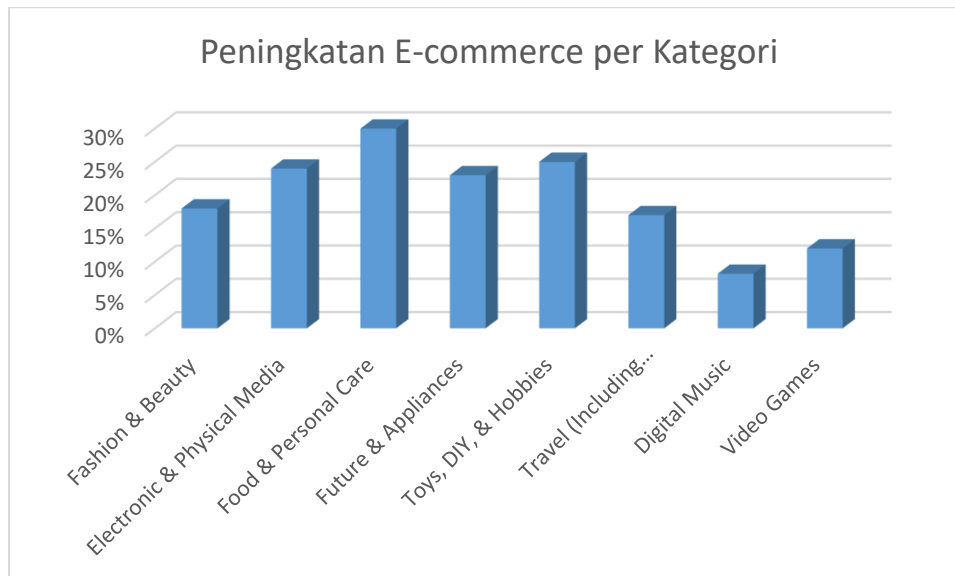
### I.1 Latar Belakang Masalah

Meningkatnya konsumsi belanja melalui perangkat digital di Indonesia di latar belakangnya oleh meningkatnya penggunaan internet di Indonesia. Menurut penelitian yang dilakukan oleh *We Are Social* pada tahun 2019 dalam *Digital 2019 Spotlight E-commerce in Southeast Asia*, penetrasi internet di Indonesia mencapai 65% dari populasi penduduk di Indonesia. Pada kuartal ketiga tahun 2019, 90% pengguna internet di Indonesia dalam rentang umur 16 – 64 tahun, pernah menggunakan internet untuk berbelanja online dalam kurun waktu satu bulan, lebih tinggi dibandingkan dengan Thailand dan bahkan Asia Tenggara secara keseluruhan.



Gambar I.1 Grafik Perkembangan *E-commerce* di Asia Tenggara 2019  
(Sumber: *Hootsuite We Are Social Global Digital Report 2019*)

Dengan peningkatan ini, Indonesia merupakan negara dengan potensi pasar yang besar terutama untuk pasar *e-commerce*. Pada tahun 2019, total pengeluaran pelanggan *e-commerce* untuk kategori produk makanan dan produk kebutuhan pribadi di Indonesia meningkat 30% dari tahun 2018. Peningkatan untuk kategori produk makanan dan produk kebutuhan pribadi merupakan yang tertinggi dibandingkan dengan kategori produk lain yang tersedia di *e-commerce*.



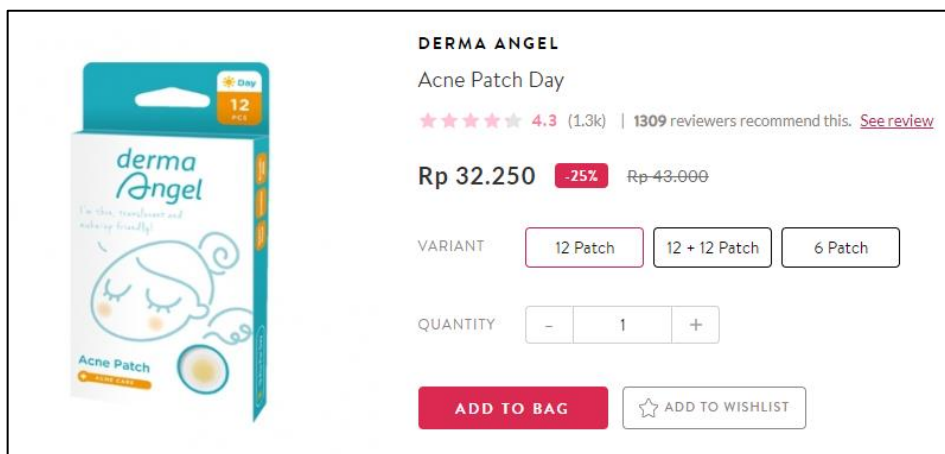
Gambar 1.2 Peningkatan per Kategori Produk *E-commerce* di Indonesia  
(Sumber: *Hootsuite We Are Social Global Digital Report 2019*)

Kebutuhan pelanggan *e-commerce* semakin berkembang dan meningkat, sehingga para pelaku usaha *e-commerce* perlu meningkatkan kualitas pelayanan maupun kualitas produk yang disediakan. Kepuasan pelanggan dihasilkan dari pengalaman pelanggan terhadap kualitas pelayanan serta kualitas produk yang ditawarkan (Hellier, Geursen, Carr, & Rickard, 2002). Kepuasan pelanggan ini memiliki pengaruh besar terhadap niat beli ulang pelanggan atau *repurchase intention*.

Salah satu cara yang dapat digunakan oleh pelaku usaha *e-commerce* adalah melalui ulasan yang diberikan oleh pelanggan atau *online customer review* yang telah membeli produk atau menggunakan jasa yang ditawarkan pada *e-commerce*. Meningkatkan kepuasan pelanggan terhadap pelayanan atau produk serta menimbulkan kepercayaan pelanggan saat berbelanja secara online di *e-commerce* merupakan hal penting dalam kegiatan *e-commerce*, terutama untuk

menarik minat pelanggan dan menimbulkan niat berbelanja online kembali di tempat atau *e-commerce* yang sama (Saragih & Rizky, 2012).

*Sociolla.com* merupakan salah satu situs *e-commerce* di Indonesia yang menawarkan kategori produk kebutuhan pribadi, yaitu produk kecantikan dan kosmetik. Pada platform *sociolla.com* terdapat kolom ulasan yang dapat diisi oleh pelanggan berdasarkan pengalamannya berbelanja pada *sociolla.com* serta pelanggan dapat memberikan informasi *repurchase intention*. Dengan tersedianya kolom ulasan dan informasi *repurchase intention* yang diberikan pelanggan, diperlukan sebuah metode untuk membantu *sociolla.com* dalam menafsirkan ulasan ini dan dapat memprediksi *repurchase intention* dari pelanggan. Akan tetapi, dengan banyaknya produk yang ditawarkan dan banyaknya ulasan yang diberikan oleh pelanggan, usaha untuk menafsirkan ulasan ini dan memprediksi *repurchase intention* menjadi hal yang cukup menyulitkan. Salah satu contoh yang diambil adalah produk *Derma Angel Acne Patch Day* seperti pada Gambar 1.3. Produk ini telah diulas oleh 1309 pelanggan.



Gambar 1.3 Produk *Derma Angel Acne Patch Day*  
(Sumber: <https://www.sociolla.com/skin-care/10023-acne-patch-day-12.html>)

Ulasan pada produk yang dijual di *sociolla.com* seperti pada contoh produk di atas merupakan data yang berbentuk teks. Salah satu metode untuk mengolah data dengan bentuk teks adalah dengan menggunakan metode *text mining*. *Text mining* dapat membantu produsen yang memasarkan produknya di *sociolla.com* untuk menafsirkan ulasan pelanggan serta dapat memprediksi *repurchase intention* dari pelanggan dengan menemukan pola dalam ulasan yang diberikan pelanggan.

## I.2 Identifikasi dan Rumusan Masalah

Perkembangan zaman pada lingkungan yang serba cepat maupun instan diiringi dengan perkembangan internet pada saat ini, memungkinkan para pelaku usaha *e-commerce* salah satunya adalah *sociolla.com* untuk mengakses sejumlah data yang besar secara mudah, salah satu contohnya adalah ulasan pelanggan. Dengan meningkatnya jumlah ulasan yang diberikan pelanggan, informasi yang dapat diolah pun meningkat. Untuk mengambil informasi dari ulasan pelanggan tersebut, *sociolla.com* perlu membaca banyaknya ulasan yang diberikan pelanggannya untuk mengambil informasi. Membaca secara manual ulasan ini merupakan hal yang memakan waktu dan usaha yang besar sehingga diperlukan sebuah metode untuk membantu *sociolla.com* untuk mengambil informasi dan menganalisa informasi tersebut dari ulasan pelanggan yang jumlahnya sangat banyak.

Besarnya jumlah data yang digunakan dari ulasan pelanggan menyebabkan proses menarik informasi ini menjadi sulit. Semakin besar jumlah data yang digunakan berbanding lurus secara eksponensial terhadap ruang pencarian solusi optimal (Sathesh, 2018). Fenomena ini disebut juga sebagai "*Curse of Dimensionality*". Selain itu, besarnya jumlah data yang digunakan juga dapat meningkatkan biaya komputasi salah satunya adalah waktu komputasi yang panjang.

Sebelum melakukan penafsiran dan menemukan pola serta memprediksi *repurchase intention* dalam data ulasan pelanggan yang berbentuk teks, diperlukan sebuah metode untuk mengurangi besarnya data tersebut. Salah satunya adalah *feature selection*. *Feature selection* merupakan proses untuk memilih fitur yang berpengaruh terhadap output yang diinginkan. Di dalam konteks *text mining*, dokumen direpresentasikan sebagai kumpulan kata-kata. Dokumen yang berisi kata-kata tersebut merupakan vektor yang memiliki dimensi besar, dimana banyaknya dimensi ini bergantung pada jumlah fitur yang digunakan (Nedjah, Mourelle, Kacprzyk, Franca, & De Souza, 2009) dan pada kasus ulasan pelanggan *sociolla.com* ini, fitur merupakan kata.

Fitur dalam konteks *machine learning* adalah variabel input yang berisi informasi dalam membangun sebuah model prediksi. Pada kasus *feature selection* ulasan pelanggan, dokumen merupakan kumpulan ulasan pelanggan. Setiap ulasan pelanggan dapat direpresentasikan sebagai sebuah baris yang memiliki

banyak kolom. Setiap kolom merupakan representasi dari kata yang muncul atau terdapat di dalam ulasan pelanggan tersebut. Dengan kata lain, setiap kolom (yang berisi kata) adalah fitur di dalam model prediksi.

*Feature selection* dalam ulasan pelanggan merupakan sebuah metode yang digunakan dalam *machine learning* untuk mengurangi fitur (kata) yang tidak relevan dan berlebihan untuk mengurangi waktu komputasi tanpa harus mengurangi akurasi dari hasil (Liu, Lu, & Song, 2014). Penggunaan *feature selection* juga memiliki beberapa alasan, yaitu (James, Witten, Hastie, & Tibshirani, 2017):

1. Membuat model menjadi lebih sederhana sehingga dapat lebih mudah untuk diinterpretasikan oleh pengguna.
2. Waktu komputasi yang lebih singkat.
3. Menghindari terjadinya *curse of dimensionality*.
4. Meningkatkan generalisasi model dengan menggunakan *overfitting* pada model.

Pada *sociolla.com*, bukan hanya ulasan saja yang dapat diberikan oleh pelanggan. Informasi mengenai *repurchase intention* juga dapat diberikan oleh pelanggan. Dengan berdasarkan data yang tersedia (ulasan pelanggan serta informasi *repurchase intention* dari pelanggan), *sociolla.com* dapat membangun sebuah model prediksi *repurchase intention* pelanggan menggunakan ulasan pelanggan. Namun, dengan banyaknya ulasan pelanggan yang tersedia, jumlah fitur (kata) yang akan digunakan pada model prediksi *repurchase intention* meningkat pula. Salah satu contoh ulasan pelanggan dapat dilihat pada Gambar I.4.

Ulasan tersebut memiliki 77 kata yang berbeda, sehingga dari contoh ulasan di bawah ini sudah terdapat 77 fitur yang digunakan di dalam model prediksi *repurchase intention*. Apabila terdapat 1308 ulasan seperti pada contoh produk pada Gambar I.3, akan terdapat banyak sekali fitur yang digunakan pada model prediksi. Oleh karena itu, *feature selection* dapat digunakan untuk mengurangi sejumlah fitur (kata) yang kurang bahkan tidak relevan digunakan dalam model prediksi *repurchase intention*.



Gambar I.4 Contoh Ulasan Pelanggan  
(Sumber: <https://www.sociolla.com/skin-care/10023-acne-patch-day-12.html>)

Memilih fitur dalam *feature selection* dari sejumlah fitur yang tersedia merupakan masalah yang cukup kompleks untuk dilakukan sehingga diperlukan sebuah metode pendekatan untuk membantu menyelesaikan permasalahan *feature selection*. Salah satu metode pendekatan yang diusulkan untuk membantu menyelesaikan kasus *feature selection* adalah dengan menggunakan metode algoritma metaheuristik. Metode algoritma metaheuristik yang diusulkan pada penelitian ini adalah algoritma *Binary Particle Swarm Optimization* (BPSO).

Di dalam kasus memilih fitur, algoritma *Binary Particle Swarm Optimization* (BPSO) digunakan dibandingkan hanya dengan menggunakan algoritma *Particle Swarm Optimization* (PSO) umum. Hal ini dikarenakan pada kasus pemilihan fitur merupakan kasus diskrit sehingga penggunaan algoritma *Binary Particle Swarm Optimization* (BPSO) yang merupakan modifikasi dari algoritma *Particle Swarm Optimization* (PSO) untuk menyelesaikan masalah diskrit. Algoritma *Binary Particle Swarm Optimization* (BPSO) digunakan untuk memilih fitur (kata) apa saja yang menjadi fitur paling berpengaruh dalam model prediksi *repurchase intention* dari ulasan pelanggan.

Berdasarkan identifikasi masalah yang telah dijabarkan di atas, maka dapat diperoleh rumusan masalah sebagai berikut:

1. Bagaimana penerapan algoritma *Binary Particle Swarm Optimization* (BPSO) yang dirancang dengan *encoding* partikel dibantu oleh perpaduan prinsip *Pareto (80/20 rule)* dan metode *Sentiment Orientation-Pointwise Mutual Information* (SO-PMI) untuk menyelesaikan *feature selection* dalam membangun model prediksi *repurchase intention* dari ulasan pelanggan?
2. Fitur apa saja yang paling berpengaruh dalam model prediksi *repurchase intention* dari ulasan pelanggan?

### **I.3 Pembatasan Masalah dan Pengambilan Asumsi Penelitian**

Berdasarkan identifikasi dan rumusan masalah di atas, maka diambil beberapa batasan masalah yang dapat menjadi acuan dalam penelitian ini. Berikut merupakan batasan masalah yang diambil dalam penelitian ini:

1. Kategori produk yang digunakan dalam penelitian ini terbatas pada kategori produk *moisturizer*.
2. Data ulasan pelanggan yang digunakan pada penelitian ini merupakan ulasan yang diberikan oleh pelanggan dari tanggal 18 Maret 2019 hingga 26 Agustus 2019.
3. Label atau kelas yang digunakan pada keputusan *repurchase* hanya "yes" dan "no".

Setelah menentukan batasan yang digunakan pada penelitian ini, asumsi ditentukan untuk membantu melaksanakan penelitian ini agar lingkup penelitian dapat lebih spesifik. Berikut merupakan asumsi masalah yang digunakan di dalam penelitian ini:

1. Model yang dibentuk dengan menggunakan *dataset* sebagian produk pada kategori produk *moisturizer* diasumsikan dapat digunakan dan menggambarkan seluruh kategori produk *moisturizer*.
2. Tidak terjadi perubahan yang besar pada produk di kategori *moisturizer* sehingga model yang dibentuk dengan menggunakan *dataset moisturizer* yang diambil dari tanggal 18 Maret 2019 hingga 26 Agustus 2019 diasumsikan dapat digunakan sekarang dan kedepannya.

#### **I.4 Tujuan Penelitian**

Berdasarkan identifikasi dan rumusan masalah serta pembatasan masalah di atas, maka penelitian ini bertujuan untuk:

1. Menerapkan algoritma *Binary Particle Swarm Optimization* (BPSO) yang dirancang dengan *encoding* partikel dibantu oleh perpaduan prinsip *Pareto (80/20 rule)* dan metode *Sentiment Orientation-Pointwise Mutual Information* (SO-PMI) untuk menyelesaikan masalah *feature selection* dalam membuat model prediksi *repurchase intention* dari ulasan pelanggan.
2. Mengetahui fitur apa saja yang paling berpengaruh dalam model prediksi *repurchase intention* dari ulasan pelanggan.

#### **I.5 Manfaat Penelitian**

Berdasarkan tujuan dari penelitian ini di atas, Berikut merupakan beberapa manfaat yang didapatkan dari penelitian ini baik bagi peneliti maupun bagi pembaca lain:

1. Pembaca dapat menambah pengetahuan mengenai *feature selection* beserta metode penyelesaiannya dengan menggunakan algoritma *Binary Particle Swarm Optimization* (BPSO).
2. Pembaca dapat menjadikan hasil penelitian ini sebagai referensi bagi penelitian selanjutnya yang memiliki tujuan atau topik penelitian yang serupa.
3. Memberikan metode alternatif yang dapat digunakan untuk menyelesaikan kasus *feature selection*.

#### **I.6 Metodologi Penelitian**

Penelitian ini memiliki tahapan-tahapan dalam pelaksanaannya. Pada sub bab ini akan dijelaskan mengenai tahapan yang digunakan pada penelitian ini. Dimulai dari studi literatur sebelum dilakukannya penelitian hingga penarikan kesimpulan serta saran yang didapatkan dari hasil penelitian. Metodologi penelitian ini akan dijelaskan dalam bentuk tulisan dan diagram alir (*flow chart*) penelitian yang digambarkan pada Gambar I.5. Berikut ini merupakan metodologi penelitian *feature selection* menggunakan algoritma *Binary Particle Swarm*



*Optimization* (BPSO) untuk memprediksi *repurchase intention* dari ulasan pelanggan.

1. Studi Literatur

Pada tahap ini, dilakukan studi awal terhadap objek, pemahaman konsep, serta pemilihan metode penyelesaian kasus *feature selection*. Pada tahap ini juga, dikumpulkan berbagai referensi yang dapat digunakan untuk menunjang penelitian. Literatur yang digunakan berasal dari buku, jurnal, maupun artikel.

2. Identifikasi dan Perumusan Masalah

Pada tahap ini, dilakukan identifikasi serta merumuskan masalah pada *feature selection* untuk memprediksi *repurchase intention* dari ulasan pelangga sebagai objek utama penelitian.

3. Pembatasan Masalah

Pada tahap ini, dilakukan pembatasan masalah terhadap masalah yang terdapat pada penelitian agar lingkup penelitian ini tidak terlalu luas dan dilakukan pembatasan akibat keterbatasan dari penelitian.

4. Penentuan Tujuan dan Manfaat Penelitian

Pada tahap ini, dilakukan penentuan tujuan serta manfaat dilakukannya penelitian ini. Tujuan penelitian ini ditentukan sehingga penelitian ini memiliki arah dan apabila tujuan dari penelitian ini tercapai, maka penelitian ini diharapkan mampu memberikan manfaat baik bagi peneliti dan pembaca dari penelitian ini.

5. Ekstraksi Data Ulasan Pelanggan

Pada tahap ini dilakukan ekstraksi atau pengambilan data ulasan pelanggan dari *sociolla.com* dengan menggunakan program *PyCharm*. Ekstraksi data ini dilakukan untuk mengambil data yang cocok digunakan dalam penelitian ini.

6. *Data Pre-processing*

Pada tahap ini, dilakukan proses pembersihan data yang telah diekstrak pada tahap sebelumnya. Pembersihan data ini dilakukan agar data yang digunakan pada tahap selanjutnya lebih terstruktur dan menghilangkan *noise* pada data yang dapat mengganggu tahap penelitian selanjutnya. Tahap ini terbagi menjadi beberapa sub-tahap, yaitu *Tokenizing*,

*Stopwords Removal*, *Stemming*, dan Konversi Bentuk Data. Selain itu, pada tahap ini dilakukan pembagian data menjadi *train data* dan *test data* yang akan digunakan untuk membantu proses klasifikasi dengan menggunakan algoritma *k-Nearest Neighbors* (*k-NN*). Tahapan ini dilakukan dengan menggunakan program *PyCharm* dengan bantuan *package PySastrawi* dan *Microsoft Excel*.

7. Perancangan Algoritma *Binary Particle Swarm Optimization* (BPSO)  
Pada tahap ini, dilakukan perancangan algoritma *Binary Particle Swarm Optimization* (BPSO) untuk membantu menyelesaikan kasus *feature selection*. Perancangan ini dimulai dari *encoding* partikel yang digunakan, inisiasi posisi awal partikel, evaluasi posisi partikel, menghitung nilai  $p_{best}$  dan  $g_{best}$ , evaluasi posisi terbaik menggunakan *fitness value*, serta menentukan posisi terbaik partikel.
8. Perancangan Algoritma *k-Nearest Neighbors* (*k-NN*)  
Pada tahap ini, dilakukan perancangan algoritma *k-NN* menggunakan *PyCharm* dengan bantuan *package sklearn*.
9. Verifikasi dan Validasi Program  
Pada tahap ini, dilakukan verifikasi dan validasi terhadap program yang dirancang untuk penelitian *feature selection* ini. Tahap ini untuk memastikan bahwa program yang dirancang telah sesuai dengan algoritma yang telah dirancang. Perancangan program algoritma *Binary Particle Swarm Optimization* (BPSO) dilakukan pada program *PyCharm* dengan bantuan *package PySwarms*.
10. Uji Parameter  
Pada tahap ini, dilakukan pengujian parameter yang dapat dikendalikan dari algoritma *Binary Particle Swarm Optimization* (BPSO) yang berisi  $w$ ,  $c_1$ , dan  $c_2$  serta parameter  $k$  pada algoritma *k-Nearest Neighbors* (*k-NN*). Pengujian parameter ini dilakukan untuk mencari nilai parameter terbaik yang dapat digunakan pada penelitian ini agar menghasilkan hasil yang paling sesuai.
11. *Running Program*

Pada tahap ini, dilakukan *running program* berulang kali untuk mendapatkan hasil *feature selection* yang paling sesuai untuk membangun model prediksi *repurchase intention* dari ulasan pelanggan.

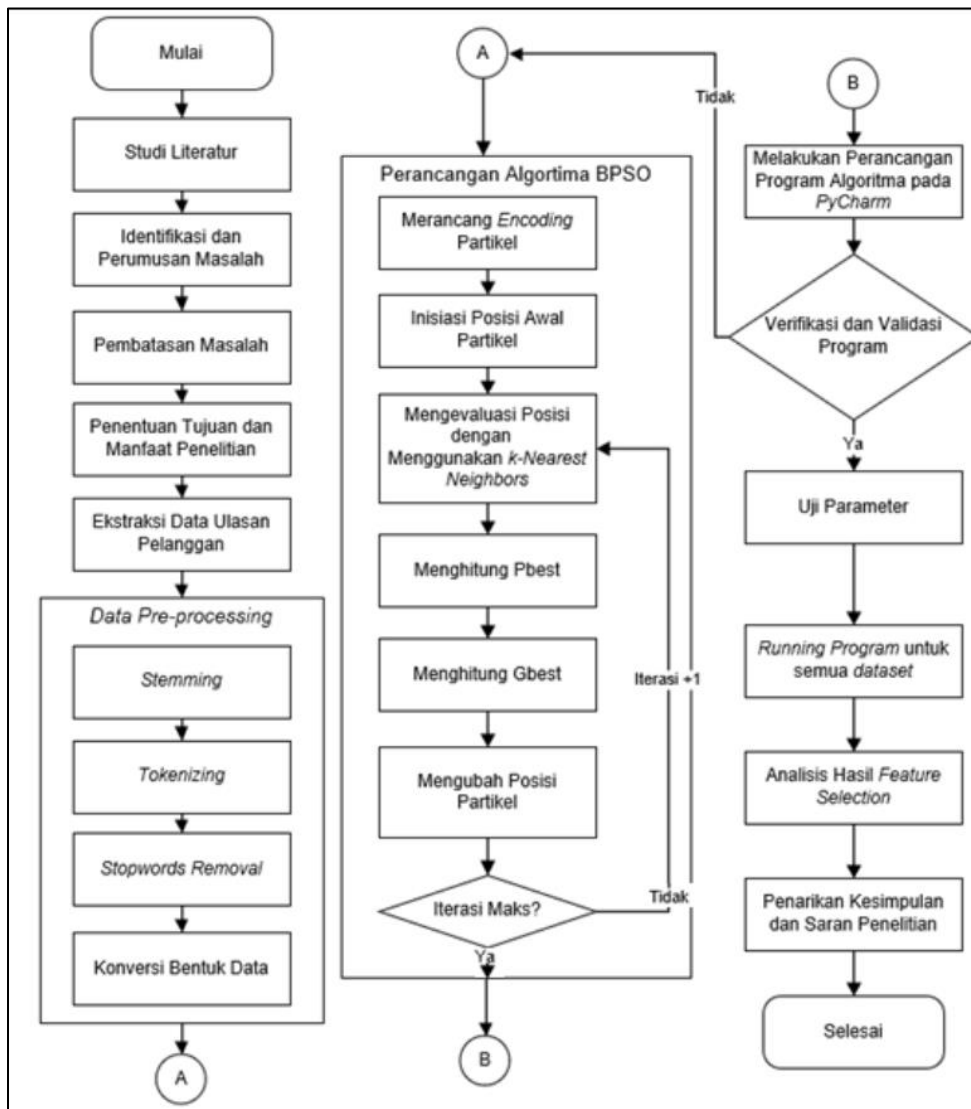
12. Analisis Hasil *Feature Selection*

Pada tahap ini, dilakukan analisis terhadap hasil dari penelitian *feature selection* dengan menggunakan algoritma *Binary Particle Swarm Optimization* (BPSO) untuk membangun model prediksi *repurchase intention* dari ulasan pelanggan. Analisis ini dilakukan dengan menggunakan nilai performansi klasifikasi yang dilakukan dengan menggunakan algoritma *k-Nearest Neighbors* (*k*-NN).

13. Penarikan Kesimpulan dan Saran Penelitian

Pada tahap ini, dilakukan penarikan kesimpulan yang menjawab rumusan masalah sesuai dengan hasil penelitian yang dilakukan pada tahap-tahap sebelumnya serta memberikan saran atas penelitian bagi penelitian selanjutnya yang memiliki topik yang serupa.

Berikut di bawah ini merupakan gambar diagram alir (*flowchart*) dari metodologi penelitian yang digunakan di dalam penelitian *feature selection* menggunakan algoritma *Binary Particle Swarm Optimization* (BPSO) untuk memprediksi *repurchase intention* dari ulasan pelanggan.



Gambar I.5 Metodologi Penelitian

## I.7 Sistematika Penulisan

Penulisan hasil penelitian ini disusun dalam enam 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 *Binary Particle Swarm Optimization* (BPSO) dalam membuat model prediksi *repurchase intention*.

## **BAB III PENGOLAHAN DATA**

Bab III berisi mengenai proses pengubahan data ulasan mentah yang diekstrak sebelumnya menjadi bentuk data yang lebih sesuai untuk tahap selanjutnya. Pada bab ini dilakukan proses penghilangan data sesuai dengan batasan penelitian, pengubahan bentuk ulasan menjadi bentuk tabel, dan pemilihan fitur awal yang digunakan pada tahap selanjutnya dengan menggunakan metode SO-PMI. Bab ini juga berisi mengenai perancangan algoritma *Binary Particle Swarm Optimization* (BPSO) yang dimulai dari encoding partikel hingga perubahan posisi partikel. Perancangan algoritma *k-Nearest Neighbors* (*k*-NN) juga dilakukan pada bab ini. 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*. Pada bab ini pula dilakukan pengujian parameter untuk setiap parameter pada kedua algoritma yang digunakan. Pada bab ini juga 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 *Binary Particle Swarm Optimization* (BPSO) dalam menyelesaikan kasus *feature selection* untuk membangun model prediksi *repurchase intention*. Analisis ini meliputi analisis pemilihan parameter, 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 *Binary Particle Swarm Optimization* (BPSO) dalam membuat model prediksi *repurchase intention* serta memberikan saran untuk penelitian selanjutnya yang memiliki topik serupa.