

**PERANCANGAN MODEL PREDIKSI *REPURCHASE INTENTION* BERDASARKAN ULASAN KONSUMEN  
PADA *WEBSITE X***

**SKRIPSI**

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

Disusun oleh :

Nama : Joshua Budidharma Immanuel  
NPM : 2017610200



**PROGRAM STUDI SARJANA TEKNIK INDUSTRI  
JURUSAN TEKNIK INDUSTRI  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS KATOLIK PARAHYANGAN  
BANDUNG  
2021**

**PERANCANGAN MODEL PREDIKSI *REPURCHASE INTENTION* BERDASARKAN ULASAN KONSUMEN  
PADA *WEBSITE X***

**SKRIPSI**

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

Disusun oleh :

Nama : Joshua Budidharma Immanuel  
NPM : 2017610200



**PROGRAM STUDI SARJANA TEKNIK INDUSTRI  
JURUSAN TEKNIK INDUSTRI  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS KATOLIK PARAHYANGAN  
BANDUNG  
2021**

**FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS KATOLIK PARAHYANGAN  
BANDUNG**



Nama : Joshua Budidharma Immanuel  
NPM : 2017610200  
Jurusan : Teknik Industri  
Judul Skripsi : PERANCANGAN MODEL PREDIKSI *REPURCHASE INTENTION* BERDASARKAN ULASAN KONSUMEN PADA *WEBSITE X*

**TANDA PERSETUJUAN SKRIPSI**

Bandung, Agustus 2021

**Ketua Program Studi Sarjana  
Teknik Industri**

3 Sept  
2021

(Dr. Ceicalia Tesavrita, 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 : Joshua Budidharma Immanuel  
NPM : 2017610200

dengan ini menyatakan bahwa Skripsi dengan Judul:  
PERANCANGAN MODEL PREDIKSI *REPURCHASE INTENTION*  
BERDASARKAN ULASAN KONSUMEN PADA *WEBSITE X*

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 Juli 2021

Joshua Budidharma Immanuel  
NPM : 2017610200

## ABSTRAK

Dalam pasar *e-commerce* Indonesia yang terus berkembang, *Website X* merupakan *website* penjual produk kosmetik dan kecantikan terbesar di Indonesia. Sebagai perusahaan yang bergerak di bidang kecantikan, diperlukan *customer feedback* yang memaparkan *repurchase intention* konsumen melalui media digital. *Website X* menyediakan informasi yang berkelimpahan mengenai *repurchase intention* dan faktor lain yang mempengaruhinya. Data ulasan konsumen yang berkelimpahan mempersulit pemrosesan. Oleh karena itu, dipakai *feature selection* untuk memilih fitur. Model prediksi *repurchase intention* dibuat dengan algoritma *decision tree classifier* untuk membuat model yang dapat memprediksi *repurchase intention* konsumen.

Data dibagi menjadi dua bagian, data *training* dan data *test* dengan rasio 80:20. Dari berbagai metode *feature selection*, diuji metode *term frequency*, TF-IDF, Fisher Score, dan *Distinguishing Feature Selector*. Setelah membagi data *training* secara acak menjadi 10 *balanced* sampel, 50 fitur yang paling relevan diambil untuk dibuat menjadi model prediksi. Kemudian, dibuatlah model prediksi dengan algoritma *decision tree*. Confusion matrix menyediakan informasi bagi perhitungan akurasi sampel.

Proses pembuatan model prediksi dilakukan untuk setiap sampel dari tiga kategori produk, yakni *cleanser*, *moisturizer*, dan *treatment*. Dari hasil perbandingan antar model, didapatkan bahwa model prediksi Fisher Score memiliki tingkat akurasi tertinggi sekitar 55%. Setelah dilakukan uji signifikansi *paired t test*, terbukti bahwa model dari Fisher Score memiliki rata-rata akurasi yang berbeda secara signifikan dengan kebanyakan metode lainnya. Didapati pula bahwa model Fisher Score juga dapat menggambarkan *repurchase intention* dengan lebih baik dibandingkan model yang hanya menggunakan *rating*. Maka dari itu, model prediksi *repurchase intention* terbaik adalah model yang menggunakan metode Fisher Score.

## **ABSTRACT**

*In Indonesia's ever-growing e-commerce market, Website X is the largest website for purchasing cosmetics and beauty products in Indonesia. As a company engaged in the beauty sector, customer feedback is crucial to identify consumers' repurchase intentions through digital media. Website X provides abundant information regarding repurchase intentions and other factors that influence it. Massive amounts of consumer reviews complicates processing prowess. Aside from that, there are 3 instances of repurchase intention, "Yes", "No", and "Maybe". Therefore, a decision tree classifier algorithm is used to create a model that can predict consumer repurchase intentions. However, the reviews has an excess amount of factors/features that makes prediction inaccurate. Therefore, feature selection is used to reduce the features involved.*

*The data is divided into two parts, training data and test data with a ratio of 80:20. From various feature selection methods, term frequency, TF-IDF, Fisher Score, and Distinguishing Feature Selector methods were tested. After randomly dividing the training data into 10 balanced samples, the 50 most relevant features were taken to be made into a predictive model. Then, a prediction model is made using a decision tree algorithm. The confusion matrix provides information for calculating sample accuracy.*

*The process of making prediction models is carried out for each sample from three product categories, namely cleanser, moisturizer, and treatment. It is found that the Fisher Score prediction model has the highest accuracy rate of about 55%. After testing significance with paired t-test, it is evident that the Fisher Score model has an average accuracy that is significantly different from most other methods. The Fisher Score model is able to describe customer repurchase intention better than the rating model. Therefore, the best repurchase intention prediction model is the model that uses the Fisher Score method.*

## KATA PENGANTAR

Puji dan syukur pada Tuhan Yang Maha-Esa atas berkat dan anugerah-Nya, penulis dapat menyelesaikan penelitian skripsi dengan judul “Perancangan Model *Repurchase Intention* Berdasarkan Ulasan Konsumen Pada *Website X*”. Penulis juga mengucapkan terima kasih sebesar-besarnya kepada pihak-pihak yang telah membantu dan membimbing proses penelitian skripsi dari awal sampai selesai. Penulis sangat bersyukur kepada pihak-pihak tersebut, antara lain:

1. Bapak Dedy Suryadi, S.T., M.S., Ph.D, selaku dosen pembimbing penulis yang selalu memberikan waktu, tenaga, dan masukan sehingga penulis dapat menyelesaikan skripsi dengan baik,
2. Dr. Ceicalia Tesavrita, S.T., M.T., selaku koordinator skripsi yang telah membantu penulis dalam proses administrasi skripsi dan telah menyusun kegiatan mata kuliah skripsi dengan sebaik-baiknya,
3. Orang tua dan keluarga terkasih penulis yang senantiasa menemani dan menyemangati penulis di waktu senang maupun sedih dalam menyelesaikan penelitian skripsi, dan
4. Teman-teman terdekat penulis yang namanya tidak bisa disebutkan satu per satu yang telah senantiasa mendukung dan menyemangati penulis dalam penelitian skripsi sehingga penulis dapat .

Penulis menyadari bahwa penelitian ini masih jauh dari sempurna. Maka dari itu, penulis terbuka untuk menerima saran dan kritik terkait dengan penelitian ini. Sebagai penutup, penulis berharap penelitian ini dapat bermanfaat bagi pihak-pihak yang memerlukan. Sekian kata pengantar dari penulis, penulis mengucapkan terima kasih.

Bandung, 13 Agustus 2021



Joshua Budidharma Immanuel

## DAFTAR ISI

<b>ABSTRAK</b> .....	<b>i</b>
<b>ABSTRACT</b> .....	<b>ii</b>
<b>KATA PENGANTAR</b> .....	<b>iii</b>
<b>DAFTAR ISI</b> .....	<b>v</b>
<b>DAFTAR TABEL</b> .....	<b>vii</b>
<b>DAFTAR GAMBAR</b> .....	<b>ix</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>xi</b>
<b>BAB I PENDAHULUAN</b> .....	<b>I-1</b>
I.1 Latar Belakang Masalah.....	I-1
I.2 Identifikasi dan Perumusan Masalah .....	I-5
I.3 Pembatasan Masalah dan Asumsi Penelitian .....	I-8
I.4 Tujuan Penelitian.....	I-9
I.5 Manfaat Penelitian.....	I-9
I.6 Metodologi Penelitian .....	I-10
I.7 Sistematika Penulisan .....	I-13
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>II-1</b>
II.1 Repurchase Intention .....	II-1
II.2 <i>Machine Learning</i> .....	II-1
II.3 <i>Text Pre-processing</i> .....	II-3
II.4 Feature Selection .....	II-4
II.4.1 <i>Frequency-based Feature Selection</i> .....	II-5
II.4.2 Term Frequency – Inverse Document Frequency (TF-IDF).....	II-5
II.4.3 Fisher Score .....	II-6
II.4.4 DFS .....	II-7
II.5 <i>Decision Tree</i> .....	II-8
II.6 <i>Algoritma Decision Tree Multiclass</i> .....	II-10
II.7 <i>Cross Validation</i> .....	II-11
II.8 <i>Confusion Matrix</i> .....	II-14
II.9 <i>Paired T-Test</i> .....	II-16
<b>BAB III PENGOLAHAN DATA</b> .....	<b>III-1</b>

III.1	Pembuatan Data Sampel.....	III-1
III.2	<i>Text Pre-processing</i> .....	III-2
III.2.1	<i>Noise Removal</i> .....	III-2
III.2.2	<i>Tokenizing</i> .....	III-4
III.2.3	<i>Lowercase Conversion</i> .....	III-4
III.2.4	<i>Stopword Removal</i> .....	III-5
III.2.5	<i>Stemming</i> .....	III-5
III.3	Perhitungan Metode <i>Feature Selection</i> .....	III-6
III.3.1	Perhitungan Frekuensi Kata Ulasan.....	III-6
III.3.2	Perhitungan Nilai TF-IDF.....	III-7
III.3.3	Perhitungan Fisher Score.....	III-8
III.3.4	Perhitungan DFS.....	III-10
III.4	<i>Feature Selection</i> .....	III-11
III.5	<i>Cross Validation of Training Sample</i> .....	III-18
III.6	<i>Decision Tree Classification</i> .....	III-20
III.7	Pengujian Model Prediksi tanpa <i>Feature Selection</i> .....	III-24
III.8	Uji Signifikansi.....	III-27
<b>BAB IV ANALISIS.....</b>		<b>IV-1</b>
IV.1	Analisis Penggunaan <i>Text Pre-processing</i> .....	IV-1
IV.2	Analisis Pemilihan Metode <i>Feature Selection</i> .....	IV-2
IV.3	Analisis Data <i>Balancing</i> .....	IV-5
IV.4	Analisis Fitur yang Digunakan.....	IV-7
IV.5	Analisis Uji Signifikansi.....	IV-9
IV.6	Analisis Interpretasi Hasil.....	IV-10
<b>BAB V KESIMPULAN DAN SARAN.....</b>		<b>V-1</b>
V.1	Kesimpulan.....	V-1
V.2	Saran.....	V-1
<b>DAFTAR PUSTAKA</b>		
<b>LAMPIRAN</b>		
<b>RIWAYAT HIDUP PENULIS</b>		

## DAFTAR TABEL

Tabel II.1 Algoritma Decision Tree dan Karakteristiknya .....	II-10
Tabel II.2 Contoh Algoritma <i>Decision Tree</i> .....	II-10
Tabel II.3 Metrik <i>Decision Tree</i> .....	II-11
Tabel II.4 <i>Confusion Matrix Two Class Classifier</i> .....	II-14
Tabel II.5 Contoh <i>Confusion Matrix Multiclass</i> .....	II-15
Tabel III.1 Contoh Ulasan Sebelum dan Sesudah Text Pre-processing .....	III-3
Tabel III.2 Contoh Ulasan Sebelum dan Sesudah <i>Tokenization</i> .....	III-4
Tabel III.3 Contoh Ulasan Sebelum dan Sesudah <i>Lowercase Conversion</i> .....	III-4
Tabel III.4 Contoh Ulasan Sebelum dan Sesudah <i>Stopword Removal</i> .....	III-5
Tabel III.5 Contoh Ulasan Sebelum dan Sesudah <i>Stemming</i> .....	III-6
Tabel III.6 Contoh Data Frekuensi TF-IDF pada Ulasan Sampel.....	III-7
Tabel III.7 Contoh Data Rata-rata TF-IDF .....	III-9
Tabel III.8 Tabel Data Document Frequency Fitur Contoh .....	III-10
Tabel III.9 Contoh Matriks <i>Word Count</i> .....	III-17
Tabel III.10 Contoh <i>Confusion Matrix</i> dan Hasil Algoritma <i>Decision Tree Classifier</i> .....	III-21
Tabel III.11 Akurasi Model Prediksi Metode Frekuensi.....	III-21
Tabel III.12 Akurasi Model Prediksi Metode TF-IDF .....	III-22
Tabel III.13 Akurasi Model Prediksi Metode Fisher Score .....	III-22
Tabel III.14 Akurasi Model Prediksi Metode DFS .....	III-23
Tabel III.15 Tabel Perbandingan Rata-rata Akurasi Metode <i>Feature Selection</i> .....	III-24
Tabel III.16 Akurasi Model Prediksi <i>Rating</i> .....	III-24
Tabel III.17 Akurasi Model Prediksi <i>Words</i> .....	III-25
Tabel III.18 Akurasi Model Prediksi <i>Rating</i> dan <i>Words</i> .....	III-26
Tabel III.19 Tabel Perbandingan Rata-rata Akurasi Fisher Score dengan Metode Lain.....	III-26
Tabel III.20 Hasil <i>Paired T-Test</i> Perbandingan Fisher Score Kategori <i>Cleanser</i> .....	III-28

Tabel III.21 Hasil <i>Paired T-Test</i> Perbandingan Fisher Score Kategori <i>Moisturizer</i> .....	III-29
Tabel III.22 Hasil <i>Paired T-Test</i> Perbandingan Fisher Score Kategori <i>Treatment</i> .....	III-29
Tabel III.23 <i>Paired T-Test</i> Fisher Score Sebelum <i>Feature Selection</i> Kategori <i>Cleanser</i> .....	III-30
Tabel III.24 <i>Paired T-Test</i> Fisher Score Sebelum <i>Feature Selection</i> Kategori <i>Moisturizer</i> .....	III-30
Tabel III.25 <i>Paired T-Test</i> Fisher Score Sebelum <i>Feature Selection</i> Kategori <i>Treatment</i> .....	III-31
Tabel IV.1 <i>Feature Importance Cleanser</i> Sampel ke-1 .....	IV-12
Tabel IV.2 <i>Feature Importance Moisturizer</i> Sampel ke-1 .....	IV-13
Tabel IV.3 <i>Feature Importance Treatment</i> Sampel ke-1 .....	IV-14

## DAFTAR GAMBAR

Gambar I.1 Pengguna dan Tingkat Penetrasi <i>E-Commerce</i> per Tahun di Indonesia .....	I-3
Gambar I.2 Statistika Web Traffic <i>Website X</i> 2019-2020 .....	I-3
Gambar I.3 <i>Review</i> Produk <i>Repurchase</i> “Yes” .....	I-6
Gambar I.4 <i>Review</i> Produk <i>Repurchase</i> “No” .....	I-6
Gambar I.5 Metodologi Penelitian .....	I-10
Gambar II.1 <i>Machine Learning Based on Type of Data</i> .....	II-2
Gambar II.2 <i>Four Key Steps of Feature Selection</i> .....	II-4
Gambar II.3 <i>Decision Tree</i> .....	II-9
Gambar II.4 <i>10-fold Cross Validation</i> .....	II-13
Gambar III.1 Contoh <i>Plot</i> Frekuensi Sampel.....	III-7
Gambar III.2 Contoh <i>Plot</i> Metode Frekuensi Sampel .....	III-11
Gambar III.3 Contoh Hasil <i>Feature Selection</i> Frekuensi Sampel.....	III-13
Gambar III.4 Contoh <i>Plot</i> Metode TF-IDF .....	III-14
Gambar III.5 Contoh Hasil <i>Feature Selection</i> Frekuensi Sampel.....	III-14
Gambar III.6 Contoh <i>Plot</i> Metode Fisher Score .....	III-15
Gambar III.7 Contoh Hasil <i>Feature Selection</i> Fisher Score.....	III-16
Gambar III.8 Contoh <i>Plot</i> Metode DFS.....	III-16
Gambar III.9 Contoh Hasil <i>Feature Selection</i> DFS .....	III-17
Gambar III.10 Algoritma <i>Cross Validation</i> .....	III-19
Gambar III.11 Algoritma Hasil <i>Cross Validation</i> .....	III-19
Gambar III.12 Contoh <i>Decision Tree</i> .....	III-20
Gambar III.13 <i>Paired T-Test</i> Fisher Score dan Frekuensi.....	III-28
Gambar IV.1 Contoh <i>Memory Error</i> pada <i>Python Console</i> .....	IV-7

## DAFTAR LAMPIRAN

LAMPIRAN A SYNTAX ALGORITMA .....	A-1
LAMPIRAN B <i>DECISION TREE</i> .....	B-1
LAMPIRAN C <i>FLOW CHART DECISION TREE</i> .....	C-1
LAMPIRAN D CONTOH <i>PLOTTING</i> .....	D-1
LAMPIRAN E HASIL UJI SIGNIFIKANSI .....	E-1

# BAB I

## PENDAHULUAN

Sebagai pendahuluan untuk penelitian ini, bab ini akan menjelaskan mengenai dasar, tujuan pelaksanaan, dan metodologi penelitian. Pendahuluan penelitian juga akan meliputi permasalahan yang diteliti, batasan dan asumsi yang dipakai dalam penelitian, manfaat penelitian, dan sistematika penulisan penelitian. Berikut merupakan pendahuluan penelitian.

### I.1 Latar Belakang Masalah

Dalam dunia perdagangan, suatu perusahaan pastinya ingin meningkatkan produk maupun layanan yang diberikan agar dapat memenuhi kebutuhan konsumen dengan sebaik mungkin. Salah satu cara dalam meningkatkan kualitas produk atau layanan tersebut adalah melalui pengambilan data *user experience* dari konsumen. Data tersebut dapat diambil dengan berbagai macam cara, yakni cara-cara langsung atau konvensional seperti survey, observasi pasar, analisis data historis, dan cara tidak langsung, yakni dengan memakai *customer review* yang ada pada *website* ataupun *digital marketplace* yang menawarkan produk ataupun layanan yang diberikan perusahaan.

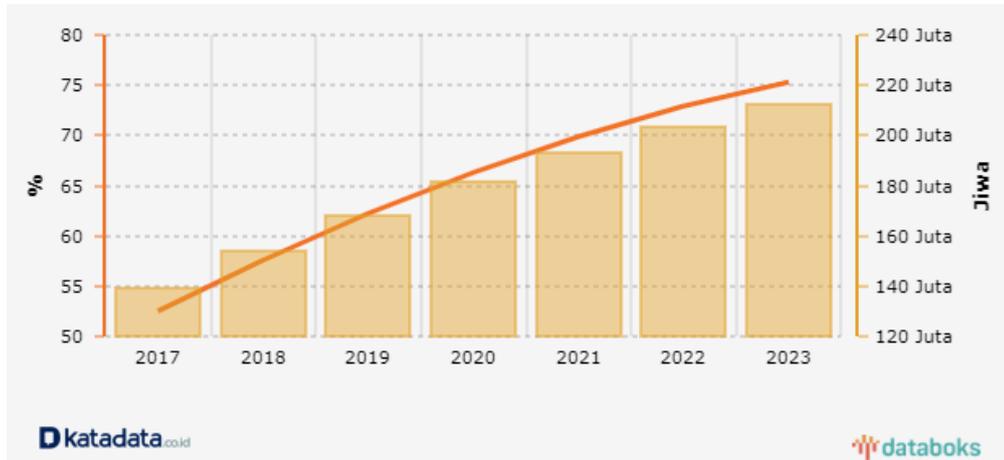
Pada tahun 2020, Indonesia mengalami pembatasan interaksi sosial atau PSBB akibat adanya pandemi yang berlangsung sepanjang tahun. Dengan adanya PSBB tersebut, toko-toko mengalami penurunan jumlah pengunjung yang datang dengan drastis. Hal ini dibuktikan lebih lanjut oleh Liputan6 (2021) yang mengemukakan bahwa berdasarkan APPBI, pengunjung pada mal di Jakarta, salah satu kota metropolitan terbesar di Indonesia, tidak pernah melebihi 40% dari tahun-tahun sebelum terlandanya pandemi. Hal ini tidak hanya berdampak besar pada jumlah transaksi secara langsung, tetapi juga pada aspek-aspek lain seperti penentuan strategi dan riset dari perusahaan atau *product owner*.

Dengan adanya pandemi yang berlangsung, perusahaan atau *product owner* mengalami pengurangan yang cukup drastis mengenai *feedback* dari

konsumen. Hal ini akan berdampak cukup signifikan terhadap bagian *decision making* dari perusahaan, terutama pada perusahaan yang bergantung pada *feedback* konsumen untuk mengembangkan kualitas produk atau jasanya. Maka dari itu, diperlukan cara bagi perusahaan atau *product owner* untuk menjual produk-produknya dan mengumpulkan informasi mengenai *feedback* dari konsumen yang memadai melalui media yang menghindari kontak sosial. Salah satu cara tersebut adalah perdagangan melalui media *online* atau perdagangan dengan pasar *e-commerce*.

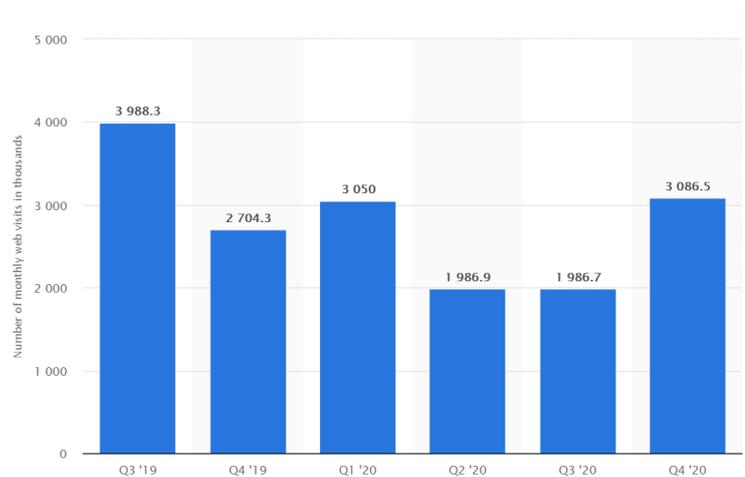
Sebagai negara yang berkembang, Indonesia akan terus mengalami peningkatan pengguna media *online* dari tahun ke tahun. Indonesia yang termasuk ke dalam golongan negara yang berkembang akan mengalami dampak yang cukup drastis dalam globalisasi atau peralihan masyarakat dari tradisional menjadi modern. Dengan kenaikan luas pasar tersebut, potensial untuk mendapatkan keuntungan dalam media *digital* menjadi lebih besar dari tahun ke tahun. Maka dari itu, dapat dikatakan bahwa Indonesia mulai beralih ke arah *online shopping* dengan membuat *website* tersendiri maupun dengan mendaftarkan produk-produknya di media *digital marketplace*. Perusahaan pada masa kini tentunya harus memiliki sarana untuk menjual-belikan produk secara *online*. Sarana perdagangan tersebut dapat dibuat oleh pihak perusahaan tersendiri, atau memakai *platform digital marketplace* yang disediakan oleh pihak lainnya, seperti Shopee, Tokopedia, dan lain-lain. Alhasil, sudah jarang sekali terdapat toko yang tidak memiliki cara untuk melakukan transaksi jual-beli secara *online*. Dengan adanya media yang bersifat *digital* untuk berdagang produk, konsumen dapat mengakses produk yang dibutuhkan dengan mudah dan instan.

Dengan adanya berbagai alternatif pilihan untuk cara berdagang secara *online*, perusahaan atau *product owner* dapat membuat atau memakai suatu sistem untuk mengakses dan memenuhi kebutuhan pasar yang lebih luas, terutama pada zaman yang mengedepankan sarana teknologi *digital* sebagai landasan kebutuhan manusia. Hal ini berkenaan dengan meningkatnya penetrasi pengguna *platform e-commerce* di Indonesia. Gambar I.1 menunjukkan grafik peningkatan pengguna yang diwakilkan dengan histogram, dan tingkat penetrasi pasar *e-commerce* yang diwakilkan dengan garis berwarna oranye, di Indonesia.



Gambar I.1 Pengguna dan Tingkat Penetrasi *E-Commerce* per Tahun di Indonesia (Sumber: <https://databoks.katadata.co.id/>)

Dari grafik di atas, didapati bahwa persentase pengguna (*bar*) dan tingkat penetrasi pasar (*garis*) memiliki peningkatan yang cukup drastis sampai dengan tahun 2023. Maka dari itu, perusahaan atau *product owner* yang bergerak di bidang kecantikan memerlukan suatu media *e-commerce* seperti *website*, *digital marketplace*, dan media *e-commerce* lainnya untuk memasarkan produk di pasar *online*. Salah satu *digital marketplace* terdepan pada bidang kecantikan adalah *Website X*. *Website X* adalah suatu pasar *e-commerce* yang mendagangkan berbagai jenis produk kosmetik dan kecantikan seperti *make up*, produk *skin care*, produk perawatan rambut sampai dengan aksesoris. Gambar I.2 menunjukkan *web traffic* dari *Website X* dari kuartal ke-3 tahun 2019 sampai dengan akhir 2020.



Gambar I.2 Statistika *Web Traffic Website X* 2019-2020 (Sumber: <https://www.statista.com>)

Dari grafik yang telah ditunjukkan, didapati bahwa sekitar 8 juta pengunjung telah mengakses atau menelusuri *Website X* pada tahun 2020. Selain sebagai media perdagangan *online*, *Website X* dapat digunakan oleh perusahaan atau *product owner* untuk melakukan penelitian mengenai produk-produknya, ataupun menyebarkan program promosi dari produk. Dengan *web traffic* yang cukup padat per tahun, perusahaan atau *product owner* yang bergerak dalam bidang kecantikan akan kehilangan potensial pasar yang cukup luas apabila tidak memiliki kehadiran dalam *Website X*.

Menurut Dean dan Bowen (1994), menjaga kepuasan konsumen adalah aspek yang penting dalam dunia bisnis. Konsumen yang puas dapat menghasilkan *customer loyalty*, sedangkan konsumen yang kurang puas dapat memberikan tanggapan atau kritik membangun sebagai patokan untuk mengembangkan produk. Maka dari itu, perusahaan atau *product owner* memerlukan suatu solusi yang dapat memproses data *customer feedback* yang tersedia pada *Website X* agar dapat diproses lebih lanjut menjadi cara menjaga kepuasan konsumen. Dengan kata lain, perusahaan memerlukan suatu cara untuk mengumpulkan data *user experience* agar perusahaan dapat meningkatkan kualitas produk sesuai dengan kebutuhan konsumen yang memakainya.

Umumnya, *user experience* biasanya dilihat dari reaksi konsumen yang datang ke dalam toko dan berbelanja ataupun melalui metode konvensional seperti survey yang diberikan terkait dengan layanan yang diberikan di toko secara langsung. Berbeda dengan hal tersebut, toko atau perusahaan yang berdagang melalui media *e-commerce* tidak dapat menangkap *user experience* secara langsung. Seorang *user* dapat meninggalkan suatu *review* ataupun nilai *rating* yang dapat menggambarkan pengalaman berbelanja di layanan *e-commerce* terkait.

*Website X* menyediakan berbagai jenis sarana untuk memberikan penilaian pada ulasan, antara lain *rating*, ulasan, profil kulit, gambar produk, dan lain-lain. Salah satu dari sarana tersebut adalah indikator *status repurchase intention* dimana konsumen dapat menyatakan apakah konsumen akan membeli kembali produk yang diberi ulasan tersebut. Karena berkaitan erat dengan konsep loyalitas konsumen, status pembelian kembali ini dapat menjadi faktor

yang terkait dengan *user experience* dari konsumen terhadap produk yang dibelinya.

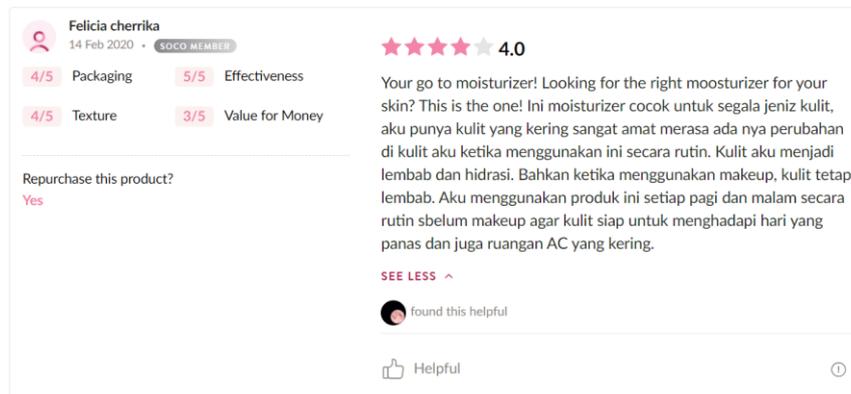
Jadi, untuk menangkap informasi *user experience*, perusahaan atau *product owner* dapat memakai komponen dalam ulasan pada *Website X* untuk memprediksi keinginan pelanggan membeli kembali produk yang dipasarkannya. Dengan memakai *repurchase intention* sebagai variabel prediktor, perusahaan dapat mengumpulkan *feedback* dari konsumen untuk mengevaluasi *user experience* dari *platform e-commerce* yang dilakukan, dan menetapkan strategi bisnis berdasarkan karakteristik produk dan hasil prediksi. Alhasil, perusahaan terkait dapat menyesuaikan kebutuhan atau prioritas yang ditetapkan oleh konsumen dan memanfaatkan hal-hal yang mempertahankan loyalitas konsumen seefektif mungkin.

## **I.2 Identifikasi dan Perumusan Masalah**

*Website X* adalah *platform e-commerce* yang memperjual-belian produk-produk kosmetik dan kecantikan. *Website X* memiliki layanan *customer review* dimana konsumen dapat melihat *review* dari konsumen lain yang sudah pernah membeli produk yang sedang ditinjau. Sebagai konteks lebih lanjut, *customer review* tersebut memiliki beberapa aspek tersendiri, antara lain: *rating total*, *rating long wear*, *rating effectiveness*, *rating pigmentation*, *rating packaging*, *rating texture*, *rating value for money*, *review*, dan status *repurchase*. Dalam penelitian ini, aspek status *repurchase* menjadi aspek yang ingin diprediksi. Jadi, aspek *repurchase* ini akan menjadi basis dari penelitian dimana model penelitian ditujukan untuk dapat memprediksi opsi *repurchase* tersebut.

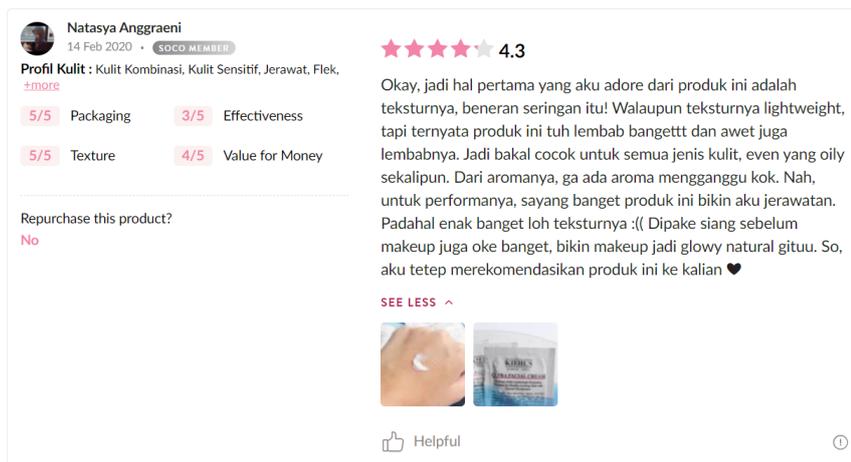
Dari aspek-aspek yang telah disebutkan, dapat dikatakan bahwa aspek *customer review* selain status *repurchase* dapat menjadi faktor-faktor penentu dari keinginan konsumen untuk membeli produk kembali. Hal ini berguna bagi perusahaan untuk mengetahui mengapa konsumen ingin membeli kembali produk dan mengapa konsumen tidak ingin membeli produk kembali.

Secara logis, tentunya konsumen akan membeli kembali produk yang disukainya. Artinya, konsumen akan membeli produk apabila ia memberikan *rating* yang tinggi. Akan tetapi, terdapat beberapa *review* yang berkontradiksi dari pernyataan tersebut. Gambar I.3 menunjukkan contoh *review* dari konsumen yang ingin membeli kembali produk.



Gambar I.3 Review Produk *Repurchase* “Yes”  
(Sumber: <https://review.soco.id/>)

Dari gambar di atas, didapati bahwa contoh *customer review* di atas memiliki beberapa fitur utama, yakni *rating total*, *rating per aspek*, *review*, dan *repurchase intention*. Pada *review*, konsumen menjelaskan lebih lanjut mengenai alasan pemberian fitur *rating* dan *repurchase intention*. Didapati pula bahwa konsumen memberikan skor total sebesar 4,0. Selanjutnya, terdapat pula *review* produk yang sama, namun *status repurchase intention* pada produk tersebut adalah “No” yang ditunjukkan pada Gambar I.4.



Gambar I.4 Review Produk *Repurchase* “No”  
(Sumber: <https://review.soco.id/>)

Dari gambar di atas, didapati bahwa *customer review* memberikan nilai yang lebih besar dari konsumen sebelumnya, yakni 4,3. Dapat dikatakan bahwa *rating* dari *review* konsumen dengan *repurchase intention* “No” lebih tinggi dibandingkan dengan *rating* dari *review* konsumen dengan *repurchase intention*

“Yes”. Maka dari itu, diperlukan pendalaman lebih lanjut mengenai dasar-dasar mengapa seorang konsumen atau *customer* ingin membeli kembali suatu produk.

Sebagai rekapitulasi hasil observasi singkat, didapati beberapa kontradiksi dari *rating* yang diberikan dan *review* yang ditulis. Seperti yang telah dikatakan, hal ini bertentangan dengan logika bahwa konsumen yang memberikan nilai *rating* yang tinggi pasti akan membeli kembali produknya. Didapati pula bahwa peninjauan *rating* atau bintang saja tidak pasti akan menghasilkan *repurchase intention*. Dapat dilihat pula bahwa *review* memiliki dampak terhadap *repurchase intention* dari konsumen di luar *rating* yang diberikan konsumen.

Beranjak dari pernyataan tersebut, *review* dari konsumen dapat menggambarkan *user experience* konsumen secara lebih detail. Artinya, *review* dapat menangkap poin-poin atau alasan lebih lanjut mengapa konsumen akan membeli kembali produk atau tidak. Jadi, diperlukan penelitian lebih lanjut mengenai *review* yang disertakan konsumen pada produk yang ditinjau.

Akan tetapi, *review* tersendiri memerlukan pengupasan. Membuat *review* menjadi suatu fitur berarti meninjau kata per kata dari *review* tersebut dan didata frekuensinya sebagai suatu fitur. Dengan adanya jumlah *review* yang mencapai ribuan, peninjauan *review* dengan cara manual bukan metode yang efisien. Maka dari itu, dipakailah suatu langkah yang dapat mengurangi faktor yang perlu ditinjau, yakni *feature selection*. Dengan memanfaatkan basis bahasa pemrograman Python, fitur yang bersifat *redundant* atau kurang relevan dapat dihilangkan. Hasil dari *feature selection* akan bisa memprediksi *repurchase intention* dengan *workload* dan waktu yang cukup kecil.

Beranjak dari peninjauan *review*, penelitian ini bertujuan untuk membuat suatu model prediksi *repurchase intention* dari *Website X*. Dari puluhan ribu *review* yang dinilai terpercaya terhadap produk yang terpilih, proses *feature selection* akan memangkas fitur/kata yang kurang relevan dan meningkatkan akurasi dari prediksi *repurchase intention*. Diharapkan bahwa dengan adanya *feature selection*, hasil prediksi *repurchase intention* menjadi lebih akurat dibandingkan dengan penebakan secara manual atau model prediksi yang tidak menggunakan *feature selection*.

Berbeda dengan penelitian yang telah dilakukan sebelumnya (Adrianto (2020), dan Wellington (2020)), penelitian ini tidak menggunakan metode yang

bersifat *cost-heavy* dan iteratif. Dengan kata lain, penelitian ini akan memanfaatkan metode non-metaheuristik seperti TF-IDF, Fisher Score, atau DFS untuk melakukan perhitungan nilai fitur atau kata yang terlibat dalam *customer review*. Hal ini dikarenakan adanya peninjauan *label multiclass* yang merumitkan formulasi prediksi. Maka dari itu, dipakailah pemrosesan metode-metode yang relatif simpel dalam penelitian ini. Selain untuk meliputi metode-metode yang lebih banyak, penelitian ini juga diharapkan menjadi *baseline* atau dasar untuk penelitian selanjutnya yang memakai algoritma yang lebih rumit.

Dari masalah yang sudah diidentifikasi, dapat dirumuskan beberapa poin masalah yang menjadi acuan untuk penelitian ini. Masalah tersebut menjadi dasar dari tujuan dan metodologi penelitian yang dipakai. Perumusan masalah yang ditinjau pada penelitian ini adalah sebagai berikut.

1. Apa saja fitur yang berpengaruh pada akurasi model prediksi *repurchase intention* dari konsumen *Website X*?
2. Bagaimanakah model yang terbaik dari *data set* yang dimiliki untuk memprediksi *repurchase intention* dari konsumen *Website X*?

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

Penelitian yang dilakukan memerlukan suatu pembatas agar ranah penelitian dapat diarahkan sesuai dengan penyelesaian masalah yang telah ditentukan dan memenuhi tujuan penelitian. Maka dari itu, dibuatlah batasan penelitian untuk mengurangi jumlah variabel yang diuji. Berikut adalah batasan yang dipakai untuk penelitian ini:

1. Faktor yang dipakai dalam penelitian terbatas pada faktor yang tersedia pada ulasan produk
2. Data yang dipakai untuk membuat model prediksi *repurchase intention* berupa data yang diambil sampai dengan tanggal 30 Agustus 2019 dengan kategori *cleanser*, *moisturizer*, dan *treatment*.
3. Hasil yang didapatkan dari penelitian hanya berlaku pada ulasan kategori sejenis.
4. Data *customer review* yang akan diteliti terbatas pada *review* yang memuat isi ulasan, *rating* yang ada pada semua jenis produk (*overall*, *packaging*, *texture*, *value for money*), dan *repurchase intention* konsumen.

Selain batasan dari penelitian, ditetapkan pula asumsi penelitian untuk mengurangi faktor-faktor yang tidak terukur maupun variabel yang tidak relevan dengan penelitian. Asumsi yang ditetapkan juga harus sesuai dengan rumusan masalah dan tujuan penelitian yang telah ditetapkan. Asumsi yang diterapkan pada penelitian ini adalah:

1. Data yang tersedia di *customer review* dapat menggambarkan perilaku konsumen untuk tiap kategori yang ditinjau, terutama *repurchase intention*.
2. Setiap *product owner* dari kategori masing-masing memiliki permasalahan yang sama, yakni keperluan peningkatan informasi *customer feedback* melalui media *online*.

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

Dalam menjalankan penelitian, diperlukan suatu tujuan penelitian yang dapat mengarahkan penelitian. Maka dari itu, diperlukan tujuan sebagai acuan dari penelitian yang dilakukan. Tujuan dari penelitian ini adalah sebagai berikut.

1. Dapat mengidentifikasi fitur-fitur yang berpengaruh pada akurasi model prediksi *repurchase intention* dari konsumen dengan *feature selection*.
2. Membuat model yang dapat memprediksi *repurchase intention* dari konsumen pada *Website X* dengan akurasi terbaik dari antara *data set* yang dipakai.

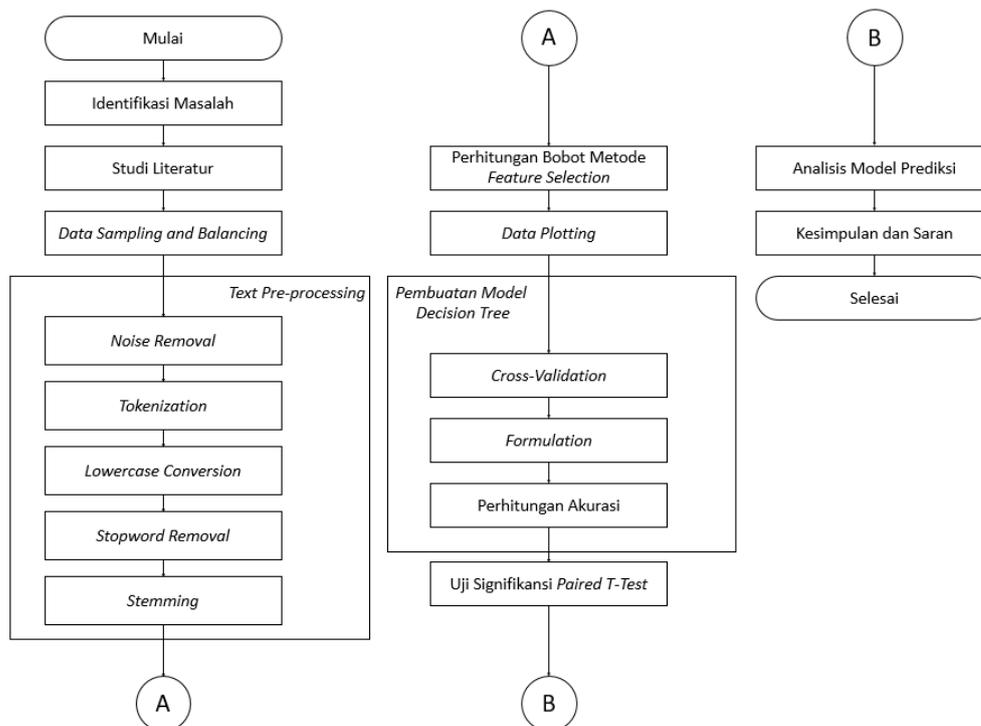
#### **I.5 Manfaat Penelitian**

Dalam penelitian ini, terdapat berbagai manfaat yang dapat dirumuskan. Manfaat tersebut adalah dampak dari hasil yang telah didapati untuk diaplikasikan oleh ilmu pengetahuan, pemilik masalah, maupun peneliti ke depannya agar dapat berkembang lebih lanjut. Manfaat dari penelitian ini adalah sebagai berikut.

1. Dapat mengetahui indikasi dari apa yang membuat seorang konsumen ingin membeli kembali produk
2. Bagi perusahaan agar dapat menggunakan model prediksi *repurchase intention* untuk mengedepankan riset pasar *e-commerce*

## I.6 Metodologi Penelitian

Dalam sebuah penelitian, diperlukan arahan yang bersifat sistematis agar penelitian tersebut dapat dilakukan dengan jelas dan lancar. Maka dari itu, diperlukan suatu metodologi penelitian sebagai pemberi konteks dari arah penelitian. Gambar I.5 adalah *flow chart* yang menggambarkan metodologi penelitian yang akan dilakukan.



Gambar I.5 Metodologi Penelitian

Dari *flow chart* yang telah dibuat, dapat dilihat bahwa alur penelitian cukup *straightforward*, dengan pengecualian pada pengolahan data yang dapat dibagi menjadi sub-langkah tersendiri. Metodologi penelitian ini menggambarkan alur serta menjadi acuan dari penelitian yang akan dilakukan. Deskripsi dari langkah-langkah metodologi penelitian adalah sebagai berikut.

### 1. Identifikasi masalah

Beranjak dari latar belakang masalah penelitian, dilakukan proses perumusan masalah yang ada pada penelitian. Dari identifikasi tersebut, dibuat pula batasan, asumsi, tujuan, serta penjelasan manfaat dari penelitian ini.

2. Studi literatur

Setelah dilakukan identifikasi masalah, dilakukan studi literatur mengenai objek penelitian, informasi umum mengenai *Website X*, *feature selection* yang digunakan, alat pembuat model prediksi dari hasil *feature selection*. Studi literatur dilakukan dengan mencari sumber referensi resmi yang relevan dengan topik penelitian, yakni *feature selection*, *repurchase intention*, dan *decision tree classifier*.

3. *Data Sampling and Balancing*

Dari data yang sudah tersedia, data tersebut dibagi menjadi 2 bagian, yakni data untuk *training set*, dan data untuk *test set*. Data untuk *training set* akan dibuat sampel data acak sebanyak 30 buah sampel dengan *sample size* sebesar 6000 *review* per sampel. Isi dari setiap sampel akan memiliki jumlah jawaban *status repurchase intention* yang sama, yakni 2000 per status. Proses ini dilakukan untuk data dari masing-masing kategori produk.

4. *Text Pre-processing*

Setelah dilakukan pembuatan sampel, sampel data tersebut diproses agar menjadi format yang sesuai untuk diteliti. Kata-kata yang bersifat kurang relevan dalam proses pembuatan model prediksi akan dihilangkan, dan kalimat *review* disesuaikan untuk dapat menangkap sisa kata-kata yang ada. Penyesuaian meliputi penghilangan *noise* seperti tanda baca, kurung, tanda kutip, simbol, kode gambar, serta nomor, imbuhan, dan pergantian huruf menjadi huruf non-kapital.

5. Perhitungan Nilai (DFS, Frekuensi, TF-IDF, Fisher Score)

Setelah didapatkan data yang sudah bersih, dilakukan *feature selection* dengan 4 metode, yakni Frequency-based, DFS, TF-IDF, dan Fisher Score. Keempat metode ini berguna untuk memberikan nilai kepada masing-masing kata atau fitur yang ada pada *review*. Frequency based *feature selection* menggunakan frekuensi kata secara total dalam korpus sebagai nilai fitur, sedangkan TF-IDF, Fisher Score, dan DFS menggunakan rumus masing-masing untuk mencari nilai fitur tersebut. Dari hasil *feature selection*, fitur yang memiliki nilai-nilai sampai dengan batasan yang telah ditentukan akan dipilih untuk menjadi basis tahap selanjutnya.

6. *Data Plotting*

Niali bobot tiap fitur pada *sample* akan disusun dari nilai tertinggi ke nilai terendah. Setelah itu, data tersebut di-*plot* menjadi grafik untuk menunjukkan persebaran bobot fitur. Pada penelitian ini, dipilih sebanyak 50 fitur dengan bobot terbesar pada setiap sampel. Data *plotting* dilakukan untuk semua sampel *training* yang ditinjau pada tiap kategori dan setiap metode.

7. *Feature Selection*

Setelah mendapatkan 50 fitur yang dibutuhkan untuk membuat model prediksi, fitur yang terpilih akan disimpan. Data *training* dan test akan di-*fit* ke dalam matriks yang meninjau *word count* fitur terpilih dari ulasan ke-*i* ( $i = 1, 2, \dots, m$ ; dimana  $m = 6000$  untuk data *training*,  $m = 3000$  untuk data *test*) dengan menggunakan package *CountVectorizer*.

8. Pembuatan Model *Decision Tree*

Dari hasil pemilihan fitur yang telah dilakukan, dibuatlah model yang menggambarkan prediksi sesuai dengan fitur yang dibuat. Model tersebut akan dibuat dengan meninjau hasil evaluasi *subset* dari metode *feature selection*. Dengan melihat hasil tersebut, dapat dibuat model *repurchase intention* konsumen berdasarkan fitur atau faktor yang paling relevan dengan keinginan membeli ulang dari konsumen. Model tersebut akan dibuat *confusion matrix*-nya untuk dihitung akurasi dari model prediksi yang dibuat. Akurasi model prediksi dihitung untuk setiap sampel per metode pada ketiga kategori yang ditinjau. Akurasi tersebut akan dirata-rata untuk setiap metode yang ditinjau. Dibuat juga model prediksi *repurchase intention* untuk data yang hanya berupa *rating*, data kata-kata dari ulasan, dan data *rating* dan ulasan tanpa *feature selection*. Akurasi model prediksi dari data-data tersebut dihitung untuk setiap sampel dari ketiga kategori.

9. Uji Signifikansi

Setelah mendapatkan akurasi model prediksi terbaik dari antara metode *feature selection*, akurasi terbaik tersebut dilakukan uji signifikansi terhadap hasil akurasi model dengan metode lainnya. Dilakukan uji signifikansi juga dari akurasi model prediksi dengan metode *feature selection* terbaik terhadap akurasi model-model yang tidak

menggunakan *feature selection*. Hasil yang didapatkan akan dibahas pada analisis.

10. Analisis Model Prediksi

Dari hasil yang diteliti, dilakukan analisis mendalam mengenai fitur yang menjadi *key factor*, fitur yang tidak berpengaruh, maupun adanya kemungkinan memakai kombinasi fitur dalam memprediksi *repurchase intention* dari konsumen. Analisis tersebut melihat hasil yang telah didapatkan dari *training* maupun *test set* untuk keseluruhan *subset*, serta kategori yang ditinjau untuk menjadi evaluasi pengembangan algoritma. Dibahas pula mengenai hasil uji signifikansi, kegunaannya, dan implikasinya terhadap model prediksi yang telah dibuat.

11. Kesimpulan dan Saran

Dari hasil yang telah didapatkan tersebut, dibuatlah suatu kesimpulan yang berkenaan dengan tujuan penelitian. Berdasarkan hasilnya pula dibuatlah saran penelitian untuk penelitian ke depannya.

### **I.7 Sistematika Penulisan**

Dalam bab ini, dikemukakan runtutan penelitian dari awal sampai akhir. Penelitian ini akan dibagi menjadi lima bab utama, dan setiap bab akan dijelaskan secara singkat dan jelas. Berikut adalah sistematika penulisan penelitian ini.

## **BAB I PENDAHULUAN**

Bab I meliputi latar belakang masalah, identifikasi dan perumusan masalah, pembatasan masalah, asumsi penelitian, tujuan penelitian, manfaat penelitian, metodologi yang dipakai dalam penelitian, dan sistematika penulisan dari penelitian

## **BAB II TINJAUAN PUSTAKA**

Bab II meliputi informasi yang berkaitan dengan penelitian yang dilakukan berdasarkan sumber referensi yang telah dicari. Informasi yang terdapat pada bab ini menjelaskan mengenai teori-teori, rumus-rumus, jenis-jenis metode penelitian, dan jenis sarana pembuatan model yang bersifat relevan dengan penelitian yang dilakukan.

### **BAB III PENGUMPULAN DAN PENGOLAHAN DATA**

Bab III meliputi proses pengumpulan data, dan pengolahan data ulasan dari *Website X*. Secara ringkas, bagian penelitian ini berisi mengenai pembagian data menjadi sampel yang bersifat *balanced*, pemrosesan data menjadi bentuk teks yang dapat diterima program, pembuangan kata-kata yang kurang relevan dengan data secara keseluruhan, dan pembuatan model prediksi *repurchase intention* berdasarkan algoritma *decision tree classifier*. Selain itu, bab ini juga berisi mengenai uji signifikansi antar metode yang ditinjau untuk melihat adanya perbedaan yang signifikan dari rata-rata akurasi model prediksi yang didapatkan.

### **BAB IV ANALISIS**

Bab IV meliputi analisis proses dan hasil pengolahan data yang telah dilakukan. Analisis meliputi analisis pemilihan variabel terikat dan bebas, analisis *data sampling*, analisis *data balancing*, analisis pemilihan alat pembuat model prediksi, analisis interpretasi hasil, dan analisis fitur yang terpilih.

### **BAB V KESIMPULAN DAN SARAN**

Bab V berisi mengenai kesimpulan dari awal proses sampai analisis hasil penelitian yang telah dilakukan. Dari peninjauan proses dan kesimpulan, dibuat pula saran bagi penelitian ke depannya yang merupakan pengembangan ataupun mengambil referensi dari penelitian ini.