# LEMBAR JUDUL SKRIPSI

**SISTEM REKOMENDASI PRODUK *SKIN CARE* BERDASARKAN PERMASALAHAN KULIT WAJAH DENGAN METODE *CONTENT BASED FILTERING***

****

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat kelulusan Strata Satu (S1)

**DWI AYU NOUVALINA**

**16180009**

**Program Studi Sistem Informasi**

**Sekolah Tinggi Manajemen Informatika dan Komputer**

**ANTAR BANGSA**

**Tangerang**

**2022**

# LEMBAR PERSEMBAHAN

Dengan mengucap puji syukur kepada Allah subhanahu wa ta’ala, skripsi ini saya persembahkan untuk :

1. Kedua orangtua dan keluarga tercinta yang dengan doa-doa, motivasi, bimbingannya tak pernah putus untuk kesuksesan saya.
2. Ibu Kusuma Hati, M.M., M.Kom., yang selalu sabar dalam membimbing saya menyelesaikan skripsi.
3. Semua dosen yang selalu memberi ilmu dan semangat kepada mahasiswa-mahasiswanya.
4. Teman-teman seperjuangan yang selalu mendukung.
5. Diri saya sendiri yang telah berjuang dengan sungguh-sungguh.

# LEMBAR PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan di bawah ini :

Nama : Dwi Ayu Nouvalina

NIM : 16180009

Program Studi : Sistem Informasi

Perguruan Tinggi : STMIK Antar Bangsa Tangerang

Dengan ini menyatakan bahwa skripsi yang telah saya buat dengan judul “**Sistem Rekomendasi Produk *Skin Care* Berdasarkan Permasalahan Kulit Wajah Dengan Metode *Content Based Filtering***” adalah asli (orisinil) atau tidak plagiat (menjiplak) dan belum pernah diterbitkan/dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah lembar pernyataan ini saya buat dengan sebenar-benarnya tanpa ada paksaan dari pihak manapun juga. Apabila di kemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain mengklaim bahwa skripsi yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari **Sekolah Tinggi Manajemen Informatika & Komputer** **Antar Bangsa** dicabut/dibatalkan.

Dibuat di : Tangerang

Pada tanggal :

Yang menyatakan,

Dwi Ayu Nouvalina

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH**

Yang bertanda tangan di bawah ini :

Nama : Dwi Ayu Nouvalina

NIM : 16180009

Program Studi : Sistem Informasi

Perguruan Tinggi : STMIK Antar Bangsa Tangerang

Dengan ini menyutujui untuk memberikan ijin kepada pihak **Sekolah Tinggi Manajemen Informatika & Komputer** **Antar Bangsa**, Hak Bebas Royalti Non- Eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah kami yang berjudul: “**Sistem Rekomendasi Produk *Skin Care* Berdasarkan Permasalahan Kulit Wajah Dengan Metode *Content Based Filtering***”, beserta perangkat yang diperlukan (apabila ada).

Dengan **Hak Bebas Royalti Non-Eksklusif** ini, pihak **Sekolah Tinggi Manajemen Informatika & Komputer** **Antar Bangsa** berhak menyimpan, mengalih-media atau *format*-kan, mengelolakannya dalam pangkalan data (*database*), mendistribusikannya dan menampilkan atau mempublikasikannya di *internet* atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari kami selama tetap mencantumkan nama kami sebagai penulis/pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak **Sekolah Tinggi Manajemen Informatika & Komputer** **Antar Bangsa**, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya..

Dibuat di : Tangerang

Pada tanggal :

Yang menyatakan,

Dwi Ayu Nouvalina

# LEMBAR PERSETUJUAN DAN PENGESAHAN SKRIPSI

Skripsi ini diajukan oleh :

Nama : Dwi Ayu Nouvalina

NIM : 16180009

Program Studi : Sistem Informasi

Jenjang : Strata Satu (S1)

Judul Skripsi : **Sistem Rekomendasi Produk *Skin Care* Berdasarkan Permasalahan Kulit Wajah Dengan Metode *Content Based Filtering***

Telah dipertahankan pada Periode I-2022 di hadapan penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh Sarjana Ilmu Komputer (S.Kom) pada Program Strata Satu (S1) Program Studi Sistem Inforrmasi di Sekolah Tinggi Manajemen Informatika & Komputer Antar Bangsa.

Tangerang, 2022

**PEMBIMBING SKRIPSI**

**Kusuma Hati, M.M., M.Kom**

**D E W A N P E N G U J I**

Dosen Penguji I : ....................................... ............................................

Dosen Penguji II : ....................................... ............................................

# LEMBAR PANDUAN PENGGUNAAN HAK CIPTA

Skripsi sarjana yang berjudul “Sistem Rekomendasi Produk *Skin Care* Berdasarkan Permasalahan Kulit Wajah Dengan Metode *Content Based Filtering*” adalah karya tulis asli DWI AYU NOUVALINA dan bukan hasil terbitan sehingga peredaran karya tulis hanya berlaku di lingkungan akademik saja, serta memiliki hak cipta. Oleh karena itu dilarang keras untuk menggandakan baik sebagian maupun seluruhnya karya tulis ini, tanpa seizin penulis.

Referensi kepustakaan diperkenankan unutk dicatat tetapi pengutipan atau peringkasan isi tulisan hanya dapat dilakukan dengan seizin penulis dan disertai ketentuan pengutipan secara ilmiah dengan menyebutkan sumbernya.

Untuk keperluan perizinan pada pemilik dapat menghubungi informasi yang tertera di bawah ini:

Nama : DWI AYU NOUVALINA

Alamat : Jl. Pipa, Lr.Cendana, Gedubang Jawa, Kecamatan Langsa Baro, Kota Langsa, Aceh

No. HP : 082310955119 (*WhatsApp*)

Email : han.yunanouv@gmail.com

# KATA PENGANTAR

Puji syukur alhamdullillah, penulis panjatkan kehadirat Allah yang telah melimpahkan rahmat dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan skripsi ini dengan baik. Di mana skripsi ini penulis sajikan dalam bentuk buku yang sederhana. Adapun judul penulisan skripsi, yang penulis ambil adalah sebagai berikut : “**Sistem Rekomendasi Produk *Skin Care* Berdasarkan Permasalahan Kulit Wajah Dengan Metode *Content Based Filtering*”.**

Tujuan penulisan skripsi ini dibuat sebagai salah satu syarat kelulusan Program Strata Satu(S1) STMIK Antar Bangsa. Sebagai bahan penulisan diambil berdasarkan hasil penelitian(eksperimen), observasi dan beberapa sumber literatur yang mendukung penulisan ini. Penulismenyadari bahwa tanpa bimbingan dan dorongan dari semua pihak, maka penulisan skripsi initidak akan lancar. Oleh karena itu pada kesempatan ini, izinkanlah penulis menyampaikan ucapanterima kasih kepada :

1. Ketua STMIK Antar Bangsa.
2. Wakil Ketua I STMIK Antar Bangsa.
3. Ketua Program Studi Sistem Informasi STMIK Antar Bangsa.
4. Bu Kusuma Hati, M.M., M.Kom selaku Dosen Pembimbing skripsi.
5. Staff / karyawan / dosen di lingkungan STMIK Antar Bangsa.
6. Orang tua tercinta yang telah memberikan dukungan moral maupun spritual.
7. Rekan-rekan mahasiswa kelas 16.8B.

Serta semua pihak yang terlalu banyak untuk disebut satu persatu sehingga terwujudnya penulisan ini. Penulis menyadari bahwa penulisan skripsi ini masih jauh sekali dari sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan dimasa yang akan datang.

# ABSTRAKSI

**Dwi Ayu Nouvalina (16180009), Sistem Rekomendasi Produk *Skin Care* Berdasarkan Permasalahan Kulit Wajah Dengan Metode *Content Based Filtering*.**

Pada tulisan ini, penulis menggunakan metode *waterfall*. Aplikasi yang dibuat menggunakan MySQL untuk *database* dan pemrograman PHP sebagai alat bantuk pembuatan sistem informasi tersebut. Sistem ini dapat membantu dalam pembuatan laporan barang masuk, barang keluar, dan persediaan barang tersampaikan dengan baik.

**Kata Kunci :**

# ABSTRACT

***Dwi Ayu Nouvalina (16180009),* Sistem Rekomendasi Produk *Skin Care* Berdasarkan Permasalahan Kulit Wajah Dengan Metode *Content Based Filtering*.**

*At this writing, the author uses waterfall method. Application are created using MySQL for the databases and PHP Programming as a tool for making information system. This system can help to make reports well.*

***Keywords :***

# DAFTAR ISI

[LEMBAR JUDUL SKRIPSI i](#_Toc46091505)

[LEMBAR PERSEMBAHAN ii](#_Toc46091506)

[LEMBAR PERNYATAAN KEASLIAN SKRIPSI iii](#_Toc46091507)

[LEMBAR PERSETUJUAN DAN PENGESAHAN SKRIPSI v](#_Toc46091508)

[LEMBAR PANDUAN PENGGUNAAN HAK CIPTA vi](#_Toc46091509)

[KATA PENGANTAR vii](#_Toc46091510)

[ABSTRAKSI ix](#_Toc46091511)

[ABSTRACT x](#_Toc46091512)

[DAFTAR ISI xi](#_Toc46091513)

[DAFTAR SIMBOL xiii](#_Toc46091514)

[DAFTAR GAMBAR xviii](#_Toc46091515)

[DAFTAR TABEL xix](#_Toc46091516)

[DAFTAR LAMPIRAN xx](#_Toc46091517)

[BAB I PENDAHULUAN 1](#_Toc46091518)

[1.1. Latar Belakang Masalah 1](#_Toc46091519)

[1.2. Identifikasi Permasalahan 2](#_Toc46091520)

[1.3. Perumusan Masalah 2](#_Toc46091521)

[1.4. Maksud Dan Tujuan 2](#_Toc46091522)

[1.5. Metode Penelitian 3](#_Toc46091523)

[1.5.1 Teknik Pengumpulan Data 3](#_Toc46091524)

[1.5.2 Model Pengembangan Sistem 4](#_Toc46091525)

[1.5.2.1 Pengembangan *Machine Learning* 4](#_Toc46091525)

[1.5.2.2 Pengembangan *Software* 5](#_Toc46091525)

[1.6. Ruang Lingkup 7](#_Toc46091526)

[BAB II LANDASAN TEORI 8](#_Toc46091527)

[2.1. Tinjauan Pustaka 8](#_Toc46091528)

[2.1.1 Konsep Dasar Sistem Rekomendasi 8](#_Toc46091529)

[2.1.2 *Machine Learning* 9](#_Toc46091530)

[2.1.3 *Content Based Filtering* 12](#_Toc46091531)

[2.1.4 *Term Frequency (TF) – Inverse Document Frequency (IDF)* 12](#_Toc46091532)

[2.1.5 *Cosine Similarity* 14](#_Toc46091533)

[2.1.6 Bahasa Pemrograman Python 15](#_Toc46091534)

[2.1.7 *Web Scrapping* 15](#_Toc46091534)

[2.1.8 *Framework Streamlit* 16](#_Toc46091534)

[2.2. Penelitian Terkait 1](#_Toc46091537)6

[BAB III ANALISA SISTEM BERJALAN 14](#_Toc46091538)

[3.1. Tinjauan Institusi/Perusahaan 14](#_Toc46091539)

[3.1.1 Tinjauan Perusahaan 14](#_Toc46091540)

[3.1.2 Struktur Organisasi 14](#_Toc46091541)

[3.2. Pengumpulan Data Pakar 15](#_Toc46091542)

[3.2.1 Objek Pakar 14](#_Toc46091541)

[3.2.2 Hasil Wawancara Pakar 14](#_Toc46091541)

[3.3. Algoritma Sistem Pakar 17](#_Toc46091543)

[3.4. Basis Pengetahuan 17](#_Toc46091543)

[3.4.1 Tabel Pakar 17](#_Toc46091544)

[3.4.2 Rule-rule Pada Pakar 17](#_Toc46091545)

[3.4.3 Pohon Keputusan Pakar 17](#_Toc46091545)

[BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM 19](#_Toc46091546)

[4.1. Analisa Kebutuhan *Software* 19](#_Toc46091547)

[4.2. Desain 27](#_Toc46091548)

[4.2.1 Database 27](#_Toc46091549)

[4.2.2 Software Architecture 33](#_Toc46091550)

[4.2.3 User Interface 42](#_Toc46091551)

[4.3. Code Generation 45](#_Toc46091552)

[4.4. Testing 52](#_Toc46091553)

[4.5. Support 54](#_Toc46091554)

[4.5.1 Publikasi Web 54](#_Toc46091555)

[4.5.2 Spesifikasi Hardware Dan Software 56](#_Toc46091556)

[BAB V PENUTUP 58](#_Toc46091558)

[5.1. Kesimpulan 58](#_Toc46091559)

[5.2. Saran 58](#_Toc46091560)

[DAFTAR PUSTAKA 59](#_Toc46091561)

[DAFTAR RIWAYAT HIDUP 60](#_Toc46091562)

[LEMBAR KONSULTASI BIMBINGAN 61](#_Toc46091563)

[SURAT KETERANGAN RISET 62](#_Toc46091564)

[LAMPIRAN 63](#_Toc46091565)

[A. Daftar Pertanyaan/Kuesioner riset (Wawancara Pakar) 58](#_Toc46091560)

[B. Daftar Pertanyaan/Kuesioner riset (Post-Test) 58](#_Toc46091560)

# DAFTAR SIMBOL

# DAFTAR GAMBAR

Gambar II.1 Cabang Ilmu *Artificial Intelligence (AI)* 9

Gambar II.2 Perbedaan Pemrograman Normal vs *Machine Learning* 10

Gambar II.3 Tipe-tipe Pembelajaran *Machine Learning* 11

Gambar III.1 Bagan Alur Penelitian 19

Gambar III.2 Perancangan Sistem 21

Gambar III.3 Tahapan Pembuatan Dataset 23

Gambar III.4 *Machine Learning Workflow* 26

Gambar III.5 Detail Konsep Sistem Rekomendasi 28

Gambar III.6 Ilustrasi Model *Deployment* 28

# DAFTAR TABEL

Tabel III.1 Fitur-fitur Pada Dataset 25

Tabel IV.2

# DAFTAR LAMPIRAN

Lampiran A.1 Surat 3

Lampiran A.2

# PENDAHULUAN

## Latar Belakang Masalah

Seiring berkembangnya zaman, semakin banyak juga perubahan yang terjadi pada gaya hidup manusia, salah satunya adalah di bidang kecantikan, khususnya wajah. Untuk mempertahankan penampilan agar selalu terlihat bagus, banyak sekali jenis produk perawatan kulit wajah yang beredar di kalangan masyrakat. Selain itu, tak jarang bagi kita melihat fenomena dimana banyak masyarakat membuat konten ulasan produk *skin care* ataupun cara perawatan kulit menggunakan *skin care*. Hal tersebut membuktikan bahwa perkembangan *skin care* sangatlah pesat dan seolah menjadi kebutuhan primer bagi banyak orang.

Menurut data statistik Badan Pengawas Obat dan Makanan (BPOM) RI hingga tahun 2022 saat ini terdapat sejumlah 218.344 produk kosmetika. Hal itu mengindikasikan banyaknya produk *skin care* yang beredar di kalangan masyarakat. Di satu sisi hal itu merupakan hal yang bagus karena jumlah inovasi dan variasi produk yang beragam, namun di sisi lain hal itu membuat kita kesulitan menemukan produk yang tepat diantara banyaknya produk. Hal tersebut juga dibuktikan dari riset terhadap responden mengenai permasalahan kulit yang mereka alami serta produk *skin care* yang mereka gunakan.

Sebagaimana yang diketahui juga bahwa jika salah dalam menggunakan produk *skin care*, alih-alih merawat kulit menjadi lebih baik sebaliknya malah menyebabkan kondisi kulit semakin parah. Melihat adanya celah masalah pada pemilihan produk perawatan kulit tersebut, muncul sebuah pemecahan masalah yaitu dengan memanfaatkan kecanggihan teknologi dalam membuat sebuah sistem yang mampu merekomendasikan produk *skin care* berdasarkan kondisi dan kebutuhan kulit wajah masing-masing.

Sistem rekomendasi memiliki tiga metode pendekatan yang umum digunakan yaitu *content based filtering*, *collaborative filtering* dan *hybrid filtering*. Dalam kasus pemilihan produk *skin care* berdasarkan kondisi dan kebutuhan kulit masing masing konsumen atau pengguna, maka metode yang dipilih ialah *Content Based Filtering* yang melihat kecocokan atribut *item* dengan preferensi pengguna.

Dari uraian di atas penulis tertarik untuk mengusulkan penerapan sistem informasi yang berjudul “Sistem Rekomendasi Produk *Skin Care* Berdasarkan Permasalahan Kulit Wajah dengan Metode *Content Based Filtering*”.

## Identifikasi Permasalahan

Berdasarkan latar belakang di atas, maka penulis mencoba untuk mengidentifikasi masalah yaitu, banyaknya produk *skin care* yang beredar, menyebabkan banyak masyarakat masih bingung memilih produk *skin care* yang tepat berdasarkan permasalahan kulit wajah yang mereka alami.

## Perumusan Masalah

Berdasarkan uraian latar belakang dan identifikasi masalah di atas, perumusan masalahnya yaitu bagaimana membangun sistem yang dapat memberikan rekomendasi produk *skin care* berdasarkan kondisi dan permasalahan kulit wajah pengguna.

## Maksud Dan Tujuan

Maksud dari penulisan skripsi ini adalah :

* + - * 1. Memperkecil lingkup produk skin care yang terlalu banyak.
        2. Memecahkan masalah dan memberikan solusi dari pemilihan produk *skin care* yang tepat.
        3. Menghemat waktu dan biaya pengguna dalam memilih produk karena rekomendasi ditinjau dari permasalahan kulit wajah yang dialami pengguna.
        4. Membuat rancangan sistem rekomendasi produk *skin care* berdasarkan preferensi pengguna.

Adapun tujuan dari penulisan skripsi ini adalah untuk memenuhi salah satu syarat kelulusan Strata Satu (S1) pada program studi Sistem Informasi di STMIK Antar Bangsa.

## Metode Penelitian

### Teknik Pengumpulan Data

* + - 1. Survei

Survei ini adalah salah satu metode pengumpulan data yang penulis lakukan untuk mengamati dan mengumpulkan informasi dari responden dalam hal penggunaan *skin care*. Pertanyaan berupa permasalahan kulit wajah, produk yang saat ini sedang digunakan, dan produk yang ingin digunakan sehingga bisa dijadikan acuan dalam membuat sistem.

* + - 1. *Web Scrapping*

Metode ini adalah proses pengambilan data atau ekstraksi dari *website-website* penjualan produk *skin care* sesuai dengan kebutuhan dan disimpan dalam format tertentu untuk kemudian dijadikan sebagai dataset.

* + - 1. Studi Pustaka

Selain kegiatan di atas, penulis juga melakukan studi pustaka guna mendukung dan memperlengkap data yang didapat, berupa referensi dari buku dan jurnal agar lebih relevan dengan topik atau pokok masalah yang sedang penulis teliti.

### Model Pengembangan Sistem

**1.5.2.1 Pengembangan Pakar (Pengembangan *Machine Learning*)**

Pengembangan model *machine learning* yang penulis gunakan dalam skripsi adalah sebagai berikut :

*Unsupervised Learning*

*Unsupervised learning* adalah salah satu tipe pembelajaran yang dilakukan oleh *machine learning* dalam pembuatan *Artifical Intelligence*. Dalam pendekatan ini, kita tidak perlu melatih algoritma (melakukan *training data*) untuk mengenali pola penyusun AI, karena *unsupervised learning* mampu mempelajari pola data tanpa target data. Salah satu penerapan metode *unsupervised* *learning* adalah sistem rekomendasi. Menggunakan data sebelumnya, *unsupervised learning* dapat membantu menemukan tren data yang dapat digunakan untuk memberikan rekomendasi produk, sehingga konsumen dapat tertarik untuk melakukan pembelian.

*Content Based Filtering*

Dalam pembuatan sistem rekomendasi, terdapat beberapa pilihan metode yang bisa digunakan untuk melakukan *filtering* (penyaringan data sesuai tujuan rekomendasi). Penulis menggunakan teknik rekomendasi *content based filtering*, yaitu sistem rekomendasi yang dirancang berdasarkan preferensi pengguna*.*

*Cosine Similarity*

*Cosine similarity* adalah salah satu algoritma di dalam text mining yang berfungsi untuk mengelompokkan suatu teks atau dokumen dengan cara menormalisasi panjang vector dengan membandingkan kemiripan teks satu dengan yang lain. *Cosine similarity* digunakan dalam menghitung kemiripan kategori fitur pada produk.

*TF-IDF (Term Frequency-Inverse Document Frequency)*

Pengembangan model *machine leaning* pada skripsi ini menggunakan algoritma *TF-IDF (Term Frequency-Inverse Document Frequency).* Algoritma *TF-IDF* bekerja dengan melakukan pemberian bobot pada setiap kata kunci disetiap kategori untuk mencari kemiripan kata kunci dengan kategori yang tersedia.

**1.5.2.2 Pengembangan Software**

Model pengembangan perangkat lunak yang digunakan dalam skripsi ini adalah dengan menggunakan model pengembangan *waterfall*.

1. Analisa Kebutuhan Software

Dalam tahapan ini dilakukan analisa kebutuhan sistem, meliputi bagaimana pengumpulan data, perancangan sistem yang akan dibuat, hingga hasil studi pustaka.

1. Desain

Desain merupakan gambaran usulan perangkat lunak seperti, metode dan teknik usulan, arsitektur perangkat lunak, representasi *interface*, dan prosedur pengkodean. Desain inilah yang akan digunakan untuk merancang sistem rekomendasi. Pada penelitian ini, desain sistem yang digunakan bukanlah pemrograman biasa, melainkan menggunakan algoritma dalam *Machine Learning*.

1. *Code Generation*

Untuk mengimplementasikan desain yang telah dirancang sebelumnya, penulis menggunakan bahasa pemrograman Python beserta *library-library* nya dan *framework machine learning* Tensorflow dalam membuat sistem rekomendasi ini.

1. *Testing*

*Testing* atau pengujian merupakan tahapan yang dilakukan untuk mengecek kembali apakah ada kesalahan dan memastikan keluaran yang dihasilkan sesuai dengan yang diharapkan. Model *machine learning* yang telah dibuat akan dicoba di dalam server sebelum di-*deploy*. Pada tahap deployment, penulis menggunakan Streamlit, yaitu *framework* berbasis Python untuk membangun aplikasi web dari model *machine learning*.

1. Support

Support adalah tahapan terakhir dari metode pengembangan waterfall. Software yang telah jadi akan dideploy untuk digunakan oleh penggunanya. Selain itu disini juga termasuk tahapan untu memastikan apakah sistem yang dibuat telah berjalan dengan baik.

## Ruang Lingkup

Untuk memudahkan dalam penyusunan skripsi ini, maka penulis membatasi penulisan ini yaitu membuat Sistem Rekomendasi Produk *Skin Care* yang terdiri dari :

1. Pembahasan terpusat hanya pada permasalahan kulit wajah dan produk *skin care.*

2. Produk yang direkomendasikan hanyalah produk *skin care* dasar (*facial wash, toner, serum, moisturizer, sunscreen*) berdasarkan tipe kulit normal/kering/berminyak/kombinasi/sensitive.

3. Rekomendasi yang diberikan berdasarkan dari preferensi pengguna dan tanpa terpengaruh oleh rating pengguna lain.

# LANDASAN TEORI

## Tinjauan Pustaka

### Konsep Dasar Sistem Rekomendasi

Sistem rekomendasi merupakan sebuah alat atau teknik perangkat lunak yang digunakan untuk melakukan prediksi terhadap suatu objek. Sistem rekomendasi dapat memberikan saran-saran item yang bermanfaat bagi pengguna, seperti misalnya membantu dalam menentukan item yang akan dibeli, musik yang akan didengarkan, film yang akan ditontokan, dsb. [1]

Sistem rekomendasi merupakan suatu aplikasi untuk menyediakan dan merekomendasikan suatu item dalam membuat suatu keputusan yang diinginkan oleh pengguna*.* [2]

Sehingga, bisa disimpulkan bahwa sistem rekomendasi merupakan sistem yang memudahkan manusia dalam menentukan pilihan dari banyaknya pilihan yang tersedia. Sistem rekomendasi juga bisa dirancang sesuai dengan preferensi pengguna agar lebih tepat sasaran.

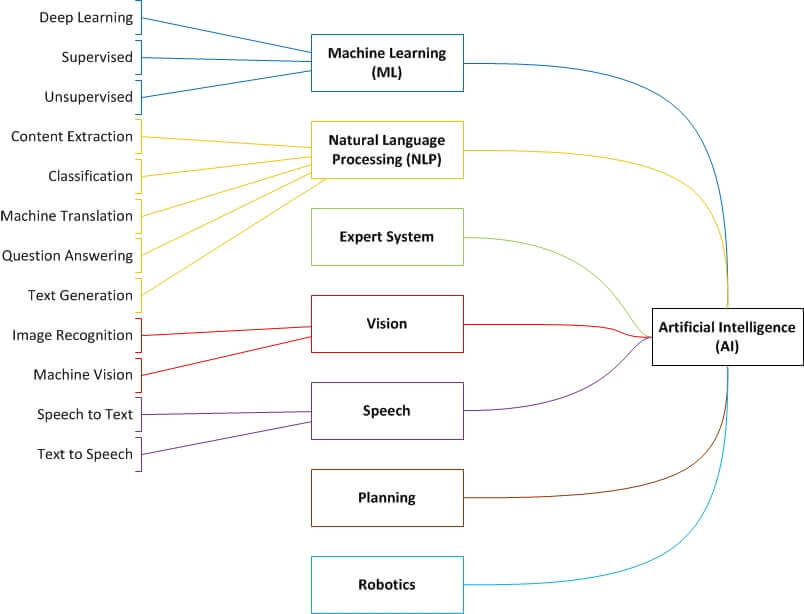
Sistem rekomendasi terbagi menjadi beberapa metode pendekatan, antara lain : [3]

1. *Collaborative Filtering*, yaitu sistem rekomendasi yang didasarkan pada pemberian rating, ulasan, atau *feedback* yang diberikan oleh pengguna.
2. *Content-based Filtering*, yaitu tipe sistem rekomendasi yang didasarkan pada preferensi pengguna itu sendiri tanpa melibatkan pengguna lain. *Content based filtering* memberikan rekomendasi dengan membangun profil pengguna.
3. *Hybrid*, yaitu sistem rekomendasi yang menggabungkan dua atau lebih metode. Hal tersebut dilakukan dengan tujuan salinge melengkapi kekurangan dari metode yang digunakan.

Beberapa peneliti menambahkan metode baru, Knowledge Based, yaitu metode yang menggunakan pola pengetahuan untuk memberikan hasil rekomendasi.

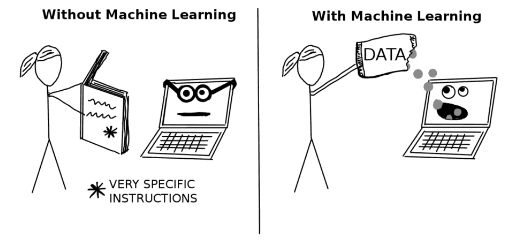
### *Machine Learning*

Istilah *machine learning* pertama kali didefiniskan pada tahun 1959 oleh Arthur Samuel. Menurut Arthur, machine learning adalah suatu bidang ilmu komputer yang memberikan kemampuan pembelajaran kepada komputer untuk mengetahui sesuatu tanpa pemrograman yang jelas. Machine learning merupakan salah satu cabang ilmu dari *Artificial Intelligence* *(AI)* yang menarik karena *machine learning* dapat belajar layaknya manusia.



Gambar 2.1. Cabang Ilmu *Artificial Intelligence (AI)*

*Machine learning* adalah seperangkat metode yang memungkinkan komputer belajar dari data untuk membuat dan meningkatkan prediksi (misalnya kanker, penjualan mingguan, cicilan tetap). *Machine learning* merupakan perubahan paradigma dari “pemrograman normal” dimana semua intruksi harus diberikan secara eksplisit dan jelas kepada komputer, menjadi sebuah “perograman tidak langsung” dengan cara mempresentasikan data-data tersebut dengan algoritma pembelajaran. [4]

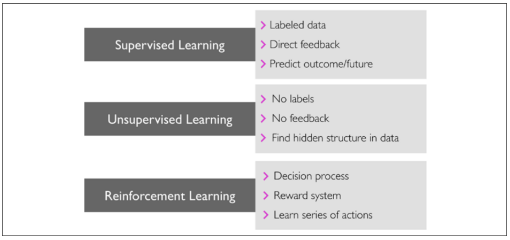


Gambar 2.2. Perbedaan Pemrograman Normal vs *Machine Learning*

Gambar diatas menunjukkan perbandingan pemrograman normal dan machine learning dimana algoritma yang digunakan tentu berbeda, dimana tanpa machine learning, programmer harus memberikan instruksi yang sangat spesifik dalam membuat program.

Algoritma adalah seperangkat aturan yang diikuti mesin untuk mencapai tujuan tertentu. Algoritma dapat dianggap sebagai resep yang mendefinisikan input, output, dan semua langkah yang diperlukan untuk mendapatkan dari input ke output. Algoritma *machine learning* adalah program yang digunakan untuk mempelajari model pembelajaran mesin dari data. [4]

Dalam membuat model *machine learning*, terdapat perbedaan mendasar antara tipe-tipe pembelajaran *machine learning* bergantung pada maksud dan tujuannya: [5]



Gambar 2.3. Tipe-tipe pembelajaran *Machine Leaning*

*Supervised Learning*

Penggunaan skenario ini merupakan pembelajaran yang menggunakan masukan data pembelajaran yang telah diberi label. Setelah itu membuat prediksi dari data yang telah diberi label.

*Unsupervised Learning*

Pembelajaran ini merupakan pembelajaran yang menggunakan masukan data pembelajaran yang tidak diberi label. Setelah itu mencoba untuk mengelompokkan data berdasarkan karakteristik-karakteristik yang ditemui.

*Reinforcement Learning*

Pembelajaran *reinforcement learning* merupakan pembelajaran yang mencampur fase pembelajaran dan fase tes atau dengan kata lain disaat melakukan tes, maka disaat itulah belajar.

### *Content Based Filtering*

*Content-based filtering* adalah salah satu metode dalam sistem rekomendasi. *Content-based filtering* tidak melibatkan pengguna lain dalam menentukan rekomendasi, namun hanya pengguna itu sendiri. Algoritma ini hanya akan memilih item dengan konten yang mirip berdasarkan apa yang dicari *user* untuk direkomendasikan. Hal ini akan membuat keragaman rekomendasi lebih sedikit, tetapi akan berfungsi baik walaupun pengguna tidak memberikan penilaian (*rating* atau *feedback*). [1]

Teknik – teknik yang biasa digunakan dalam *content-based filtering* seperti *TF-IDF, Bayesian Classifiers, Cluster Analysis, Decision Trees* dan *Artificial Neural Networks*. Metode ini memiliki kelebihan yaitu dapat memberikan rekomendasi tanpa diperlukan adanya *rating* oleh pengguna, melainkan memberikan rekomendasi berdasarkan informasi item.

Dengan kata lain, algoritma ini mencoba merekomendasikan item yang sama dengan yang disukai pengguna di masa lalu atau memeriksa tren saat ini. Secara khusus, berbagai item kandidat dibandingkan dengan item yang sebelumnya dipilih oleh pengguna, dan item yang paling cocok kemudian direkomendasikan. Pendekatan ini didasarkan pada riwayat interaksi pengguna dan penyaringan informasi. [6]

### *Term Frequency* (TF) - Inverse Document Frequency (IDF)

Metode TF/IDF merupakan salah satu metode populer yang digunakan dalam menentukan bobot setiap kata. Bobot tersebut dapaat mencerminkan tingkat kepentingannya dalam sebuah dokumen. [7]

Metode ini menggabungkan dua konsep untuk perhitungan bobot yaitu, frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu yang disebut *Term Frequency* (*TF*) dan inverse frekuensi dokumen yang mengandung kata yang disebut *Inverse Document Frequency (IDF*). Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata tersebut di dalam dokumen. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut akan rendah pada kumpulan dokumen. [8]

Rumus umum untuk TF-IDF adalah :

**Wdt = TFdt \* IDF*t***

Dimana:

t = kata ke-t dari kata kunci

d = dokumen ke-d

W = bobot dokumen ke-d terhadap kata ke-t

TF = banyaknya kata yang dicari pada sebuah dokumen

IDF = Inverse Document Frequency

*Term* yang sering muncul pada dokumen tapi jarang muncul pada kumpulan dokumen memberikan nilai bobot yang tinggi. W akan meningkat dengan jumlah kemunculan term pada dokumen dan berkurang dengan jumlah term yang muncul pada dokumen. Setelah bobot (W) masing-masing dokumen diketahui, maka dilakukan proses pengurutan dimana semakin besar nilai W, semakin besar tingkat kecocokan dokumen tersebut terhadap kata kunci, demikian sebaliknya. [8]

### *Cosine Similarity (Kesamaan Cosinus)*

*Cosine similarity* adalah ukuran kesamaan antara dua vektor bukan nol yang memperkirakan sudut kosinus di antara mereka. Jika orientasi sudut cosinus antara dua vektor adalah sama maka mereka memiliki kesamaan cosinus 1 dan juga dengan orientasi yang berbeda kesamaan cosinus akan menjadi 0 atau di antara 0-1. Kesamaan kosinus terutama digunakan dalam ruang positif, di mana hasilnya dibatasi dengan rapi di [1,0].

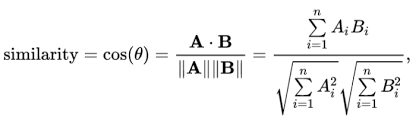
Dalam pencarian informasi, *cosine similarity* merupakan algoritma yang umum digunakan di berbagai bidang; dalam kasus ini, penulis menggunakan *cosine similarity* untuk memeriksa kesamaan antara fitur produk yang disediakan dan fitur produk yang nantinya akan dipilih pengguna. Ada banyak aplikasi kesamaan kosinus yang bisa diaplikasikan, misalnya:

Kesamaan teks *anchor*

Kesamaan dokumen dengan dokumen

Kesamaan kata kunci antara situs anda dan situs pesaing [9]

Diberikan dua buah vektor atribut, yaitu A dan B, *cosine similarity*, cos (θ), direpresentasikan sebagai berikut:



Dimana *Ai* dan *Bi* merupakan komponen vektor A dan B.

Untuk pencocokan teks, vektor atribut A dan B biasanya merupakan vektor frekuensi istilah dari dokumen. Kemiripan kosinus dari dua dokumen akan berkisar dari 0 hingga 1, karena istilah frekuensi tidak boleh negatif. Sudut antara vektor frekuensi dua suku tidak boleh lebih besar dari 90o. Kami menggunakan kesamaan cosinus untuk mengukur sudut cosinus dari konten dua situs web apakah konten situs serupa atau tidak jika itu maka berapa banyak. Rentang yang kami sukai adalah [ 0.3-0.5]. [9]

### Bahasa Pemrograman Python

Python sangat kuat, namun sangat mudah diakses. Python telah menjadi bahasa pemrograman paling populer untuk ilmu data karena memungkinkan kita melupakan bagian pemrograman yang membosankan dan menawarkan lingkungan di mana kita dapat dengan cepat menuliskan ide-ide kita dan menerapkan konsep secara langsung ke dalam tindakan. [5]

Python adalah sebuah bahasa multi paradigma dan bisa digunakan untuk membuat script, bahasa pemrograman prosedural yang berbasis object oriented, dan sekaligus bahasa fungsional. Python diketik secara longgar atau dinamis dan interaktif. Tidak ada kompiler terpisah, melainkan perintah yang diketik ke dalam interpreter secara otomatis akan dikompilasi, ditautkan ke sebuah link(sesuai kebutuhan), dan dieksekusi. [10]

### Web Scrapping

Secara teori, web scraping adalah praktik pengumpulan data melalui cara apa pun selain dari program yang berinteraksi dengan API (atau, tentu saja, melalui manusia yang menggunakan browser web). Ini paling sering dilakukan dengan menulis program otomatis yang menanyakan server web, meminta Nata (biasanya dalam bentuk HTML dan file lain yang terdiri dari halaman web), dan kemudian mem-*parsing* data tersebut untuk mengekstrak informasi yang diperlukan.

Dalam praktiknya, *web scraping* mencakup berbagai teknik dan teknologi pemrograman, seperti analisis data dan keamanan informasi. [11]

### Framework Streamlit

Streamlit adalah sebuah framework berbasis Python yang bersifat *open-source* yang dibuat untuk memudahkan dalam membangun apikasi *web* di bidang *machine learning* yang interaktif dan ilmu data. Salah satu hal menarik dari framework ini adalah kita tidak perlu mengetahui banyak hal tentang teknologi *web development*.

Streamlit cara paling andal untuk membangun aplikasi data, termasuk kemampuan untuk menampilkan dan menata data, menggambar bagan dan peta, menambahkan widget interaktif, menyesuaikan tata letak aplikasi, menghitung cache, dan menentukan tema . [12]

## Penelitian Terkait

Pada penulisan skripsi ini penulis mengambil beberapa jurnal sebagai referensi, pendukung dan perbandingan, jurnal yang didapat yaitu sebagai berikut:

1. Jurnal Manajemen Informatika & Sistem Informasi Volume 4, No, 1 “Sistem Rekomendasi Product Emina Cosmetics Dengan Menggunakan Metode *Content-Based Filtering*” karya Fatoni Batari Larasati dan Herny Februariyanti. Jurnal ini membahas mengenai sistem rekomendasi produk *skin care* dari salah satu brand kosmetik “Emina Cosmetic” dengan menggunakan metode *content-based filtering*. Berbeda dengan penulis yang mengambil data dari banyak brand *skin care*, jurnal ini hanya memberikan hasil rekomendasi khusus produk dari Emina Cosmetic, sehingga dari proses pengumpulan data nya sudah berbeda. Selain itu, dalam hal memberikan rekomendasi, pengguna memasukkan kata kunci terlebih dahulu, kemudian muncul produk-produk yang direkomendasikan. Berbeda dengan penulis yang lebih spesifik terhadap permasalahan kulit wajah pengguna, kategori produk, manfaat yang ingin didapatkan, serta memberikan rekomendasi lanjutan.
2. Jurnal IJITEE Volume 3, No.2 “*Product Recommendation System Design Using Cosine Similarity and Content-based Filtering Methods*” karya Cut Fiarni, Herastia Maharani, dan Nathania Calista. Jurnal ini membahas mengenai rekomendasi produk fiber mulai dari bahan baku, cara pencucian, printing, dan lain-lain berdasarkan segmentasi customer di sebuah perusahaan. Perbedaan tentu terletak pada jenis produk yang ingin direkomendasikan dimana penulis merekomendasikan produk *skin care* sedangkan dalam jurnal ini produk yang direkomendasikan adalah produk fiber. Selain itu, jurnal ini menggunakan algoritma dan metode lain selain *Cosine Similarity* dan *Content-Based Filtering*, yaitu *Profile Matching* dan *K-*

*Means Algorithm*.

1. Jurnal LINK Vol. 27, No. 1 “Pemanfaatan Algoritma TF/IDF Pada Sistem Informasi Ecomplaint Handling” membahas mengenai sebuah aplikasi complain yang dapat mengelompokkan complain berdasarkan kategori dengan memanfaatkan algoritma TF-IDF. Perbedaan yang sangat mendasar adalah pemanfaatan dari algoritma TF-IDF. Meskipun sama-sama menggunakan algoritma TF-IDF, pada jurnal ini, algoritma tersebut digunakan dalam proses penentuan tingkat similaritas dengan cara mengukur tingkat kesamaan antar dokumen dengan membandingkan suatu *keyword* dengan *template* komplain yang sudah ada sebelumnya. Sedangkan penulis memanfaatkan algoritma TF-IDF dalam proses penentuan tingkat similaritas antar dokumen yang disediakan dengan kata kunci pada fitur yang disediakan, sehingga bisa memberikan rekomendasi yang tepat berdasarkan tingkat kemiripan yang paling tinggi.

# ANALISA DAN RANCANGAN SISTEM BERJALAN ?

## Tinjauan Institusi/Perusahaan => (Alur Penelitian)

Alur penelitian yang dilakukan dalam penelitian ini terdiri dari beberapa tahapan yang digambarkan pada gambar 3.1. berikut.

Studi Pustaka

Pengambilan Data

Survei Pendahuluan

Pengujian Sistem

Perancangan & Implementasi Sistem

Analisa Hasil

Penarikan Kesimpulan dan Saran

Gambar 3.1. Bagan Alur Penelitian

### Sejarah Institusi/Perusahaan => (Survei Penelitian)

Pada tahap ini, penulis melakukan observasi kepada para wanita pengguna skin care untuk mengetahui permasalahan kulit wajah serta produk-produk yang sedang dan ingin digunakan. Penulis melakukan observasi dengan membagikan link Google Form kepada para responden secara daring untuk mendapatkan insight dalam melakukan perancangan sistem.

### Struktur Organisasi dan Fungsi => (Studi Pustaka)

Penulis mencari sumber referensi baik dari jurnal maupun buku terkait topik sistem rekomendasi maupun algoritma yang digunakan. Topik yang terkait dengan penelitian seperti pemahaman mengenai sistem rekomendasi, *machine learning, cosine similarity*, dan algoritma *TF-IDF*.

### => (Pengumpulan Data)

Data yang digunakan dalam penelitian ini adalah data rincian produk skin care yang diekstrak dari beberapa *website e-commerce* dan *marketplace* untuk diambil fitur-fitur yang dibutuhkan. Fitur-fitur tersebut nantinya akan dijadikan sebuah dataset yang akan digunakan dalam proses pembuatan *machine learning*.

### => (Perancangan dan Implementasi Sistem)

### Pada tahap ini, penulis membangun rancangan sistem rekomendasi dengan menggunakan konsep *machine learning* berbasis web dan dibangun menggunakan bahasa pemrograman Python dan framework Tensorflow. Dalam merancang sistem rekomendasi, penulis mengaplikasikan metode *content-based filtering* dan menggunakan *cosine similarity* dan algoritma TF-IDF. Tahapan perancangan sistem dengan konsep *machine learning* dapat dilihat pada gambar 3.2 berikut.

Pemrosesan Data

Pemodelan

Dataset

*Deployment*

*Get Recommendation*

Gambar 3.2. Perancangan Sistem

### => (Pengujian Sistem)

Pengujian terhadap model *machine learning* bukanlah pengujian benar atau salah, karena *machine learning* merupakan teknologi yang digunakan untuk memprediksi suatu hal, dalam kasus ini adalah memprediksi produk yang tepat untuk dijadikan rekomendasi. Dalam memberikan sebuah rekomendasi, tidak ada kata benar atau salah karena rekomendasi tersebut bisa saja digunakan bisa juga tidak. Sehingga, dalam hal ini pengujian sistem dilakukan dalam rangka menentukan seberapa akurat sistem yang dibangun dengan menggunakan *content-based filtering*. Selama sistem mampu memberikan rekomendasi yang masuk akal, maka sistem sudah dikatakan bagus atau bisa digunakan. Misalnya seperti, ketika mencari rekomendasi produk *facial wash*, namun yang keluar adalah produk toner, maka sistem dikatakan gagal. Namun, jika mencari rekomendasi produk *facial wash* dan yang keluar juga *facial wash*, maka sistem dikatakan berhasil.

### => (Analisa Hasil)

Pada tahap analisa hasil, penulis melakukan analisa terhadap sistem yang sudah diuji. Analisa hasil dilakukan dengan melakukan pengujian terhadap pengguna yang menggunakan sistem dalam mencari produk yang diinginkan, melihat apakah produk yang direkomendasikan masuk akal, ataupun menganalisa keakuratan produk rekomendasi.

### => (Penarikan Kesimpulan dan Saran)

Tahap ini merupakan tahap terakhir dari serangkaian tahapan penelitian. Penulis menarik kesimpulan terhadap seberapa akurat dan bermanfaat sistem yang dibangun. Penarikan kesimpulan tersebut dapat dijadikan saran untuk peneliti lain yang ingin melanjutkan penelitian.

## Pengumpulan Data Pakar (*Dataset*)

### => (Rancangan Dataset)

Dalam membuat model machine learning, *dataset* merupakan hal yang paling penting karena merupakan sumber data yang akan diolah. Dataset biasanya bisa didapatkan dengan mudah di berbagai *platform repository machine learning* seperti Kaggle, UCI Machine Learning, dll. Namun, dataset yang tersedia tidak menjamin bisa memenuhi semua kebutuhan dataset yang dibutuhkan *programmer,* ada beberapa dataset yang sulit untuk didapatkan dikarenakan penelitian yang jarang terjadi. Untuk itu, dikarenakan keterbatasan persediaan *database* dan fitur yang dibutuhkan dalam penelitian ini, penulis membuat dataset mandiri yang dibutuhkan melalui proses pengumpulan data *web scrapping* sehingga fitur yang didapatkan lebih tepat sasaran.

*Web scrapping* merupakan metode yang sangat cocok sekali untuk membuat *dataset* dikarenakan fitur yang akan diambil merupakan fitur yang akan diolah. Sehingga, hal ini memberikan efisiensi pada proses pengumpulan *dataset*. Gambar 3.3. dibawah menunjukkan tahapan pembuatan dataset.

*Data Organizing with Excel*

*Data Cleaning with Excel*

*Web Scrapping*

*Dataset*

Gambar 3.3. Tahapan Pembuatan Dataset

Berbeda dengan *database* yang merupakan wadah dari data yang sudah terstruktur, *dataset* merupakan data yang belum terstruktur dan hanya mengacu pada kumpulan data yang dipilih dan disusun dalam baris dan kolom. Meskipun begitu, salah satu keunggulan dataset adalah tidak hanya bisa digunakan untuk satu model, namun juga dapat digunakan dalam membuat berbagai macam model *machine learning* sesuai dengan yang diinginkan.

***Web Scrapping***

Web scrapping merupakan salah satu proses pengumpulan data primer melalui situs-situs web yang tersedia dengan memanfaatkan *extension* pada Google Chrome. Pada penelitian ini, penulis melakukan *scrape* terhadap beberapa situs web yang menjual produk-produk *skin care*, dengan total ±1200 data. Beberapa diantaranya yaitu :

*Marketplace* Shopee

*Official Website* Sephora

*Official Website* Sociolla

*Official Website* Zalora

*Official Website* The Face Shop

*Ofiicial Website* Beauty Haul

*Official Website* Wardah Beauty

dan situs web lainnya.

### Objek Pakar => (Fitur pada Dataset)

Dataset yang digunakan pada penulisan ini merupakan hasil *web scraping* sehingga fitur yang diinginkan lebih spesifik sesuai dengan yang dibutuhkan. Adapun fitur-fitur yang ada pada dataset tersajkan pada tabel 3.1.

|  |  |
| --- | --- |
| **Fitur** | **Keterangan** |
| Link produk | Berisi tautan produk yang mengarah ke *website* |
| Nama produk | Nama produk *skin care* |
| Harga | Tertera harga produk *skin care* |
| Deskripsi | Deskripsi produk *skin care* |
| *Notable Effect* | Berupa data kategorikal dimana pengguna bisa memilih manfaat yang dibutuhkan dari tipe produk *skin care* yang dipilih. Misalnya, *hydrating* (melembapkan), *pore-care* (perawatan pori-pori), *brightening* (mencerahkan), *acne-free* (bebas jerawat), dll. |
| *Skin type* (tipe kulit) | Berupa data kategorikal dimana pengguna bisa memilih tipe kulit wajah nya, yaitu normal, *dry* (kering), *oily* (berminyak), *combination* (kombinasi), *sensitive* (sensitive) |
| *Picture* | Berisi tautan gambar produk |
| Tipe produk | Berupa data kategorikal dimana pengguna bisa memilih tipe produk apa yang dicari, yaitu *facial wash* (sabun cuci muka), toner, serum, *moisturizer* (pelembap), *sunscreen*. |

Tabel 3.1. Fitur-fitur Pada Dataset

### Hasil Wawancara Pakar => () ???

## Algoritma Sistem Pakar => (*Machine Learning Workflow*)

*Machine learning* memiliki kemampuan independen karena mesin dirancang untuk belajar dan membuat sistem secara mandiri. Dalam membuat *machine learning project*, ada tahapan yang harus dilalui sebelum *project* dapat diimplementasikan. Tahapan yang dimaksud dapat digambarkan sebagai berikut.

*Data Collection Process*

*Exploratory Data Analysis*

*Data Preprocessing*

*Model Selection*

*Get Top-N Recommendation*

*Deployment*

Gambar 3.4. *Machine Learning Workflow*

Pada dasarnya, alur kerja *machine learning* disesuaikan dengan *project* yang akan dibuat. Artinya, tahapan tersebut bersifat iteratif dan tahapan nya bisa saja berbeda antara satu project dengan yang lainnya. Penjelasan pada tahapan berikut adalah sebagai berikut.

* + 1. ***Data Collection Process* (Proses Pengumpulan Data)**

Tahap ini merupakan tahap awal dalam mengerjakan *project machine learning* dimana *dataset* yang digunakan memberikan pengaruh sekitar 50% terhadap performa model. Artinya, sekalipun menggunakan algoritma yang *powerful,* jika *dataset* nya tidak bagus maka model machine learning yang dihasilkan juga tidak akan bagus. *Dataset* yang sudah jadi bisa didapatkan dari *public repository* seperti Kaggle atau membuat dataset mandiri kemudian mengolahnya. Dalam kasus ini, dataset yang digunakan adalah dataset mandiri yang didapatkan dari hasil *web scrapping* di beberapa situs *web* penjualan produk *skin care*.

* + 1. ***Exploratory Data Analysis (EDA)***

*Exploratory Data Analysis* atau EDA berfungsi untuk memberikan analisa awal terhadap eksplorasi data yang dilakukan untuk melihat kualitas data, meminimalkan potensi kesalahan yang bisa mempengaruhi performa *model*, atau melihat gambaran secara keseluruhan dari fitur-fitur pada dataset. Eksplorasi data yang dimaksud bisa berupa melihat persentase tiap fitur, mengecek data terduplikat, membuat diagram data, dan sebagainya.

* + 1. ***Data Preprocessing* (Pemrosesan Data)**

*Data preprocessing* merupakan tahapan lanjut setelah melakukan EDA, dimana data memasuki tahapan untuk diolah agar siap pakai dalam pembuatan model. Pada tahap ini terjadi proses *data cleaning* dan *data preparati*o*n. Data cleaning* merupakan tahap pembersihan data seperti menghapus baris kosong, menangani missing value, data terduplikat, dll. Sedangkan pada *data preparation* merupakan proses untuk mempersiapkan data sebelum masuk ke bagian pemodelan. Hal yang bisa dilakukan misalnya seperti merubah skala data, normalisasi, standarisasi, atau hal lain bergantung pada model yang akan dibuat.

* + 1. ***Model Selection* (Seleksi Model)**

Tahap ini sering disebut sebagai tahap *modeling* (pemodelan) dimana setelah melalui tahapan sebelumnya, maka akan dilakukan seleksi model. Seleksi model dapat berupa pemilihan algoritma ataupun pemilihan *hyperparamtaer* terbaik untuk metode yang dipilih. Pada kasus ini, penulis menggunakan *TF-IDF Vectorizer* yang bisa digunakan untuk menemukan representasi fitur penting dari setiap data unik pada fitur, lalu ditransformasikan ke bentuk matriks. Kemudian menghitung *cosine similarity* (kesamaan kosinus) pada matriks *TF-IDF*.

* + 1. ***Get Recommendation* (Mendapatkan Rekomendasi)**

Pada sistem rekomendasi, biasanya hasil rekomendasi didapatkan dengan memanfaatkan *rating* atau *feedback* pada suatu hal, misalnya memberikan *rating* pada film yang telah ditonton, maka film yang serupa akan direkomendasikan kepada pengguna. Namun, bagaimana jika sistem tidak menyimpan riwayat aktivitas pengguna atau itu adalah pengalaman pertama pengguna sehingga dia tidak memiliki riwayat pemberian rating sebelumnya, apakah tidak ada rekomendasi yang akan diberikan? Tentu hal itu menimbulkan pertanyaan.

Terkadang ada beberapa kasus dimana sistem rekomendasi dibuat tanpa menggunakan *rating* ataupun tanpa riwayat aktivitas, namun sangat jarang penelitian itu terjadi. Hal yang serupa juga sedang dialami oleh penulis, dikarenakan fokus permasalahan adalah memberikan rekomendasi produk terhadap permasalahan masing-masing individu. Lalu, untuk mengatasi masalah tersebut penulis menggabungkan dua konsep untuk menghasilkan rekomendasi. Ilustrasi detail konsep sistem rekomendasi adalah seperti yang terlihat pada gambar 3.5 dibawah.

Data yang sudah siap pakai

Memanfaatkan *TF-IDF* dan *Cosine Similarity* dalam menemukan kesamaan fitur

Mencari tingkat kemiripan fitur yang dipilih oleh pengguna dengan produk yang tersedia

Produk dengan tingkat kemiripan yang tinggi akan direkomendasikan

Pengguna bisa memilih salah satu dari produk yang direkomendasikan untuk mendapat rekomendasi yang lain

Diberikan rekomendasi produk lain yang serupa

Gambar 3.5. Detail Konsep Sistem Rekomendasi

* + 1. *Deployment*

Tahap *deployment* adalah tahap siap pakai dimana model yang sudah dibuat dan diuji coba (dalam hal ini uji coba mendapatkan rekomendasi) akan disimpan kemudian di *deploy*. Ilustrasi model *deployment* dapat dilihat pada gambar 3.6.

*User*

*Web App*

*Web service*

***Model***

*Web*

*Inputs*

*Predictions*

Gambar 3.6. Ilustrasi Model *Deployment*

## Basis Pengetahuan => ()

*Menjelaskan proses terbentuknya model machine learning*

### Tabel Pakar => ()

*Menjelaskan proses melakukan pemrosesan data*

### Rule-rule Pada Pakar => ()

### Pohon Keputusan Pakar => ()

# BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM

## Analisa Kebutuhan *Software*

## Desain

### Database

### 

### Software Architecture

### User Interface

## Code Generation

## Testing

## Support

### Publikasi Web

### Spesifikasi Hardware Dan Software

Tabel IV.13 Spesifikasi *Hardware* dan *Software*

|  |  |
| --- | --- |
| Kebutuhan | Keterangan |
| Sistem Operasi |  |
| Processor |  |
| RAM |  |
| Hardisk |  |
| Pegkodean |  |
| Aplikasi *database* |  |
| *Web Browser* |  |
| Perancangan Sistem |  |

# PENUTUP

## Kesimpulan

## Saran

# DAFTAR PUSTAKA

|  |  |  |
| --- | --- | --- |
| [1] |  | F. B. A. Larasati and H. Februariyanti, "Sistem Rekomendasi Product Emina Cosmetics Dengan Menggunakan Metode Content - Based Filtering," Jurnal Manajemen Informatika & Sistem Informasi, vol. Vol.3 No.1, pp.45-54, 2021. [Online] Available: http://e-journal.stmik.lombok.ac.id/index.php/misi. [Accessed April 1, 2022] |
| [2] |  | U. Ungkawa, D.Rosmala and F.Aryanti, "PEMBANGUNAN APLIKASI TRAVEL RECOMMENDER DENGAN METODE CASE BASE REASONING ," *Jurnal Informatika Itenas,* vol.4 No.2, pp.3, 2013. |
| [3] |  | A. Y. Leonardo, "PEMILIHAN KERJA UNTUK MAHASISWA ATMAJAYA YOGYAKARTA MENGGUNAKAN METODE CONTENT-BASED FILTERING ," *Universitas Atma Jaya Yogyakarta*, 2015. |
| [4] |  | C. Molnar, "Interpretable Machine Learning A Guide for Making Black Box Models Interpretable," 2020. [E-book] Available: www.lulu.com. |
| [5]  [6] |  | S. Raschka, and V. Mirjalili, Python Machine Learning | Machine Learning and Deep Learning with Python, Scikit-Learn, and TensorFlow 2, 3rd Edition, Birmingham, UK: Packt Publishing, 2019.  C. Fiarni, H. Maharani and N. Calista, "Product Recommendation System Design Using Cosine Similarity and Content-based Filtering Methods," IJITEE, vol.3 No.2, pp. 43-44, 2019. |
| [7] |  | M. N. Saadah, et al.*,* "Sistem Temu Kembali Dokumen Teks dengan Pembobotan TF-IDF dan LCS", Jurnal Ilmiah Teknologi Informasi,vol. 11 No. 1, pp. 17-20, 2013. |
| [8] |  | R. A. Sasmita and A. Z. Falani, "PEMANFAATAN ALGORITMA TF/IDF PADA SISTEM INFORMASI ECOMPLAINT HANDLING," JURNAL LINK, vol. 27 No.1, p.30, 2018. |
| [9] |  | T. Banik, Step-By-Step Guide for AI-Powered Advanced SEO Secrets Finally Revealed! Proven Data-Driven Strategies Any Beginner Can Implement, Chennai, India: Notion Press, 2019. |
| [10] |  | J. W. B. Lin, A Hands-On Introduction to Using Python In The Atmospheric and Oceanic Science, 2012. [E-book] Available: www.lulu.com. |
| [11] |  | R. Mitchell, Web Scraping with Python | Collecting Data From the Modern Web*,* Sebastopol, California: O'Reilly Media, 2015. |
| [12] |  | Nurhayati, TEKNIK ENSEMBLE LEARNING UNTUK PENINGKATAN PERFORMA AKURASI MODEL PREDIKSI (SELEKSI MAHASISWA PENERIMA BEASISWA): Pascal Books, 2022. [E-book]. |

# DAFTAR RIWAYAT HIDUP

1. **Biodata Mahasiswa**

NIM : 16180009

Nama : Dwi Ayu Nouvalina

Tempat Tanggal Lahir : Langsa, 14 November 1999

Alamat : Jl. Pipa, Lr. Cendana, Gedubang Jawa, Kecamatan Langsa Baro, Kota Langsa, Aceh

1. **Riwayat Pendidikan Formal**

Tamatan : SDN 1 Langsa, lulus tahun 2011

Tamatan : SMPN 3 Langsa, lulus tahun 2014

Tamatan : SMAN Modal Bangsa Aceh, lulus tahun 2017

Tangerang, Agustus 2022

Foto

3x4

Dwi Ayu Nouvalina

# LEMBAR KONSULTASI BIMBINGAN

|  |  |
| --- | --- |
|  | **LEMBAR KONSULTASI BIMBINGAN SKRIPSI** |
| **SEKOLAH TINGGI MANAJEMEN INFORMATIKA & KOMPUTER**  **ANTAR BANGSA** |

* NIM : 16180009
* Nama Lengkap : Dwi Ayu Nouvalina
* Dosen Pembimbing : Kusuma Hati, M.M., M.Kom
* Judul Skripsi : Sistem Rekomendasi Produk *Skin Care* Berdasarkan Permasalahan Kulit Wajah Dengan Metode *Content Based Filtering*

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Tanggal Bimbingan** | **Pokok Pembahasan** | **Paraf Dosen Pembimbing** |
| 1. |  |  |  |
| 2. |  |  |  |
| 3. |  |  |  |
| 4. |  |  |  |
| 5. |  |  |  |
| 6. |  |  |  |
| 7. |  |  |  |
| 8. |  |  |  |

Catatan untuk Dosen Pembimbing:

Bimbingan Skripsi

* Dimulai pada tanggal :
* Diakhiri pada tanggal :
* Jumlah pertemuan bimbingan :

Disetujui oleh,

Dosen Pembimbing

**()**

# SURAT KETERANGAN RISET

# LAMPIRAN