

LAPORAN TUGAS BESAR
PERANCANGAN APLIKASI SAINS DATA



Disusun Oleh:

Kelompok 6

Syifa Putri Fadhilla (1305210084)
Rosalie Naurah Husna (1305210101)
Nakhwa Azizah (1305213023)

DS 45 02
PROGRAM STUDI SAINS DATA
FAKULTAS INFORMATIKA
UNIVERSITAS TELKOM
BANDUNG
2023

BAB 1

PENDAHULUAN

1.1 Latar belakang

Dalam era digital yang semakin berkembang pesat, banyak orang mengandalkan pencarian online untuk mencari informasi dan ulasan tentang produk kecantikan sebelum membeli dan mencobanya. Mengetahui pengaruh besar ulasan produk dalam pengambilan keputusan pembelian, pengembangan web aplikasi yang memfasilitasi pengguna dalam menemukan dan mengakses ulasan produk kecantikan akan menjadi solusi yang sangat berguna.

Web aplikasi ini akan memberikan akses cepat dan mudah bagi pengguna untuk mencari produk kecantikan yang diminati sesuai dengan kriteria yang diinginkan, serta membaca ulasan yang informatif untuk membandingkan produk dengan produk lainnya. Dengan adanya web aplikasi ini, pengguna akan lebih menemukan informasi yang mereka butuhkan dan membuat keputusan pembelian yang lebih informatif.

Dalam proyek pembuatan web ini, data yang digunakan adalah “Sephora Products and Skincare Reviews” yang kami peroleh dari Kaggle. Data ini menjadi sumber informasi utama yang akan diolah untuk membangun sistem yang memberikan akses cepat dan mudah kepada pengguna. Dengan menggunakan data ini, kami akan mengembangkan fitur-fitur yang membantu pengguna dalam mencari produk yang diminati, seperti sistem pencarian yang efisien berdasarkan kriteria tertentu. Selain itu, ulasan dan peringkat yang diberikan oleh pengguna sebelumnya akan memberikan wawasan yang berharga bagi pengguna baru yang ingin mengetahui pengalaman pengguna lain dengan produk tersebut.

Melalui penggunaan data "Sephora Products and Skincare Reviews" dan pengembangan web aplikasi yang inovatif, diharapkan pengguna akan mendapatkan pengalaman dalam memilih produk kecantikan yang sesuai dengan kebutuhan dan preferensi mereka.

1.2 Rumusan masalah

1. Kebingungan dalam memilih produk yang tepat
2. Ketidakpastian kecocokan produk dengan kulit
3. Informasi produk yang terbatas
4. Rekomendasi personal

1.3 Tujuan

Membantu pengguna dalam menemukan produk kecantikan sesuai yang sesuai dengan menyediakan informasi-informasi tentang produk kecantikan serta menampilkan rekomendasi personal sesuai dengan kebutuhan pengguna.

1.4 Arsitektur

Arsitektur yang kamu rancang akan digunakan sebagai panduan dalam mengerjakan Tugas Besar ini. Berikut merupakan gambaran dari arsitektur serta tools dan resource yang kami gunakan.



Data yang kami dapat dari kaggle kami eksplorasi menggunakan python agar nantinya akan lebih mudah untuk diolah. Setelah itu, 2 data yang sudah kami dapat akan dilakukan preprocessing dalam microsoft machine learning studio (classic). Pada Github kami melakukan perancangan web app dengan menggunakan model machine learning dan diakhir mendeployment melalui streamlit.

1.5 Proses agile

1.5.1 Pembentukan team(kerjanya ngapain)

PIC	Tugas Bersama	Tugas Individu	Keterangan
Syifa Putri Fadhilla	Preprocessing Data di Machine Learning Studio (Classic)	Implementasi Model Machine Learning	Tuntas
Rosalie Naurah Husna		Menyusun Laporan (Agile process, EDA)	Tuntas
Nakhwa Azizah		Streamlit (Front-End)	Tuntas

1.5.2 Perencanaan Sprint

Sprint	Deskripsi	Waktu
1	Perencanaan, Analisis, dan Eksplorasi data	5 Juni 2023
2	Preprocessing Data di Azure Machine Learning Studio (Classic)	8 Juni 2023
3	Implementasi Model	11 Juni 2023

4	Deploy Web App dan Monitoring	12 Juni 2023
5	Penyusunan Laporan	12 Juni 2023

1.5.3 Perancangan

Web app yang akan kami rancang adalah web app yang akan membantu pengguna dalam menemukan produk kecantikan sesuai yang sesuai dengan menyediakan informasi-informasi tentang produk kecantikan serta menampilkan rekomendasi personal sesuai dengan kebutuhan pengguna. Web app tersebut akan menampilkan visualisasi berdasarkan kategori. Selain itu, Web app tersebut terdapat fitur pencarian dan penyaringan untuk mempermudah pengguna menemukan produk kecantikan yang mereka inginkan berdasarkan kriteria yang telah dipilih seperti kategori produk, tipe kulit, dan warna kulit.

BAB 2

ANALISIS

2.1 Dataset

2.1.1 Penjelasan Dataset

Dataset "Sephora Products and Skincare Reviews" merupakan kumpulan data yang didapatkan pada bulan march 2023. Pada dataset "Sephora Products and Skincare Reviews" terdapat 2 data berbeda yaitu product_info dan review.

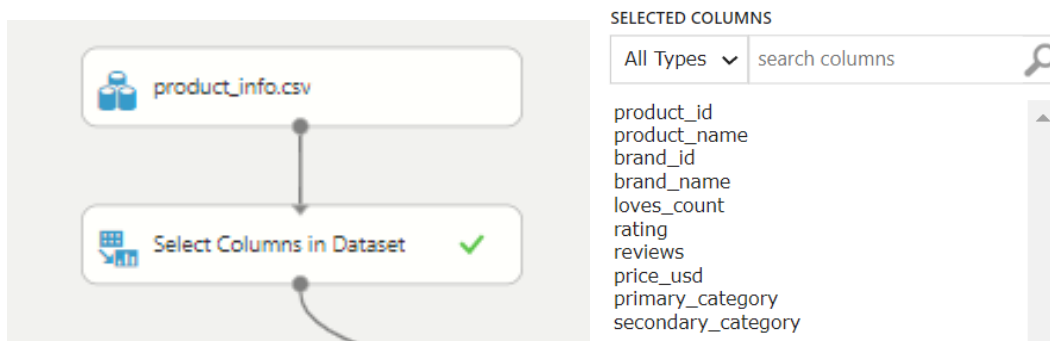
2.1.2 Eksplorasi

Terdapat 2 dataset yang kami gunakan yaitu data product_info yang berisi 8.000 data dengan atribut product_id, product_name, brand_id, brand_name, loves_name, rating, reviews, size, variation_value, variation_desc, ingredients, price_usd, value_price_usd, sale_price_usd, limited_edition, new, online_only, out_of_stock, sephora_exclusive, highlights, primary_category, secondary_category, tertiary_category, child_count, child_max_price, dan child_min_price.

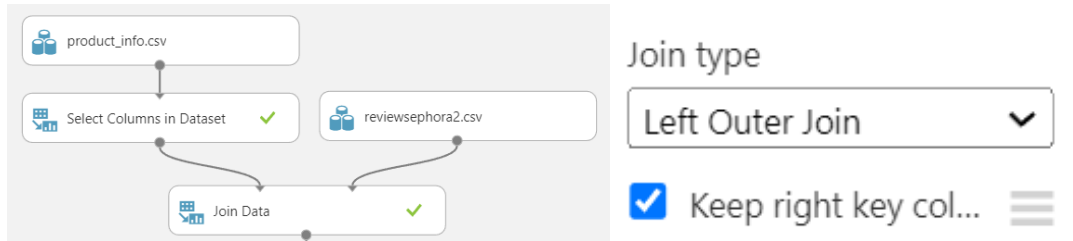
Pada data Review terdapat ulasan pengguna lebih dari 1 juta pada lebih dari 2.000 produk dari semua produk kategori skincare yang berisikan atribut author_id, rating, is_recomended, helpfulness, total_feedback_count, total_pos_feddback_count, submission_time, review_title, skin_tone, eye_color, skin_type, hair_color, dan product_id

2.1.3 Preprocessing

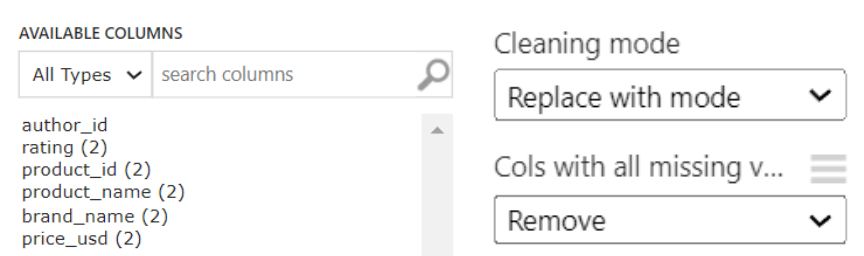
Pertama kami melakukan Drop Select Column in Dataset untuk memilih kolom mana saja yang akan digunakan. Pada data info kami menggunakan column berikut ini:



Selanjutnya kami menggabungkan dua data yaitu dataset product_info dengan dataset review sephora yang sudah kami preprocessing pada python. Kami melakukan join data produk_id pada kedua dataset tersebut dengan join type left outer join dimana setiap baris dari tabel product_info akan diulangi sebanyak jumlah review yang sesuai dari tabel data review.



Dikarenakan masih terdapat kolom yang nantinya tidak terpakai, kami akan menghapus kolom - kolom tersebut menggunakan select column. Selanjutnya kami melakukan replace missing value menggunakan mode pada data yang telah digabungkan.



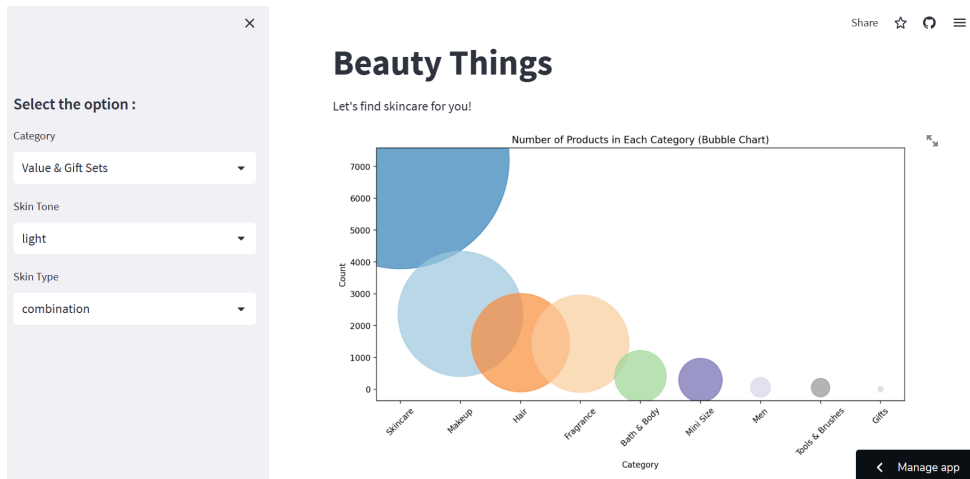
2.2 Implementasi (model yang digunakan, pengimplementasian model tsb)

Pada tahap ini kami melakukan pembangunan model menggunakan pendekatan content based filtering yaitu metode dalam sistem rekomendasi yang menggunakan analisis terhadap konten atau atribut dari item yang direkomendasikan. Selanjutnya kami mempersiapkan dataset untuk pelatihan. Pada pelatihan model, kami membuat objek random forest classifier dan melatih menggunakan fitur target pada set pelatihan. Selanjutnya, kami akan mengevaluasi model dan menghitung akurasi model.

2.3 Fungsionalitas web app

- Visualisasi

Visualisasi yang akan kami tampilkan berupa bubble chart yang berisikan jumlah kategori dari produk kami. Guna dari adanya visualisasi yang nanti ditampilkan untuk membantu para pengguna mengetahui berapa banyak jumlah berdasarkan kategori. Berikut Visualisasi yang kami tampilkan



- Pencarian dan Penyaringan

Dengan adanya fitur pencarian dan penyaringan dalam web app, pengguna dapat dengan mudah menemukan produk kecantikan yang mereka inginkan berdasarkan kriteria yang telah dipilih seperti kategori produk, tipe kulit, dan warna kulit.

The screenshot shows the same web application interface as above, but the main area displays a table titled 'Hasil Pencarian:'. The table has columns: 'product_name', 'brand_name', 'review_title', 'loves_count', 'price_usd', and 'review_text'. The first row shows a product named 'Fragrance Discovery Set' by '19-69' with an 'Amazing' review title, 6320 loves, and a price of 35.0000. The second row shows 'African Beauty' by '19-69' with a 'Love this serum!' review title, 7777 loves, and a price of 35.0000. A 'Manage app' button is at the bottom right.

	product_name	brand_name	review_title	loves_count	price_usd	review_text
0	Fragrance Discovery Set	19-69	Amazing	6320	35.0000	Love this serum! I' only used it a couf days, but the difference in smoothness on m skin is obvious. I h constant issues wi texture as I am old This absorbs quicl and sits well unde makeup. NNo obnoxious smell
1	African Beauty	19-69	Love this serum!	7777	35.0000	Love this serum! I' only used it a couf days, but the difference in smooth skin is o

2.4 Deploy Web App

Pada tahap ini kami melakukan deployment menggunakan github yang dihubungkan dengan stremlit untuk membuat web app yang telah kami rancang. Penggunaan streamlit pada proyek ini memungkinkan kami untuk menghadirkan web app yang interaktif dan bermanfaat pada pengguna kami.

Deploy an app

Repository

[Paste GitHub URL](#)

nakhwaazizah/Sephora

Branch

main

Main file path

WebApp.py

[Advanced settings...](#)

Deploy!

2.5 Monitoring

Pada tahap monitoring ini kami melakukan pemantauan kesalahan (Error monitoring) terhadap web app yang kami rancang. Pemantauan kesalahan yang kami lakukan berupa pemeriksaan apakah terjadi kesalahan yang terjadi di dalam web app kami. Selain itu, kami juga memonitor waktu respon web app kami untuk mengukur seberapa cepat aplikasi merespon pengguna. Dengan melakukan monitoring tersebut web app kami dapat mengoptimalkan kinerja aplikasi, mengukur keberhasilan dan dapat mengidentifikasi masalah yang terjadi di dalam web app kami.

BAB 3

Penutup

3.1 Kesimpulan

Kesimpulan dari proyek tugas besar mata kuliah perancangan aplikasi sains data kami adalah kami dapat merancang Web app melalui github dengan menghubungkan streamlit yang akan menampilkan visualisasi berdasarkan kategori dan menampilkan fitur pencarian dan penyaringan untuk mempermudah pengguna menemukan produk kecantikan yang mereka inginkan berdasarkan kriteria yang telah dipilih seperti kategori produk, tipe kulit, dan warna kulit. Dengan adanya web app sains data yang kami rancang ini diharapkan dapat membantu para pengguna dalam menentukan produk apa saja yang ia butuhkan sesuai dengan kriteria yang diinginkan dan memenuhi tugas besar matakuliah Perancangan aplikasi sains data.

3.2 Lampiran

- **Dataset :** [Sephora Products and Skincare Reviews | Kaggle](#)
- **Web App :** <https://nakhwaazizah-sephora-webapp-fg48b3.streamlit.app/>