

**LAPORAN PROYEK AKHIR  
PRAKTIKUM DATA SCIENCE**

**ANALISIS SENTIMEN TERHADAP  
PAKAIAN WANITA**



AHMAD ABDULLAH AZZAM  
ASWIRA YUNiar SYARIF

123190013  
123190009

**PROGRAM STUDI INFORMATIKA  
JURUSAN TEKNIK INFORMATIKA  
FAKULTAS TEKNIK INDUSTRI  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
YOGYAKARTA  
2021**

## PENDAHULUAN

Sejak mewabahnya pandemi Covid-19 dan masyarakat dihimbau untuk tetap berada di rumah sehingga kegiatan transaksi belanja online meningkat pesat. Berbagai macam toko online mulai memanfaatkan momentum ini untuk mendapatkan banyak pelanggan baru, salah satu contohnya adalah pakaian wanita. Selama masa pandemi, belanja online menjadi alternatif utama yang banyak dipilih masyarakat. 92 persen mencoba metode belanja baru, 57 persen masyarakat yang melakukan pembelian secara digital dan 48 persen layanan grocery pick up & aplikasi pengiriman (Ferry Kusnowo, 2020). Banyak pelanggan baru yang belum mengetahui kualitas produk yang dijual di toko online sehingga membuat pelanggan baru melihat review produk terlebih dahulu sebelum membelinya. Oleh karena itu, supaya memudahkan pelanggan baru mengetahui kesimpulan dari review yang dilihat, maka dilakukan sebuah pengujian dengan mengambil data-mining berupa komentar atau review dari pakaian wanita.

### 1. Analisis Sentimen

Sentiment Analysis (Analisis Sentimen) atau sering disebut juga dengan opinion mining (penambangan opini) adalah studi komputasi untuk mengenali dan mengekspresikan opini, sentiment, evaluasi, sikap, emosi, subjektivitas, penilaian atau pandangan yang terdapat dalam suatu teks (Liu, 2010, 629). Analisis sentimen juga merupakan sebuah metode yang digunakan untuk mengekstrak data opini, memahami serta mengolah tekstual data secara otomatis untuk melihat sentimen yang terkandung dalam sebuah opini. (Sari & Wibowo, 2019).

Ulasan merupakan bagian dalam aspek kecerdasan manusia, dan mempunyai peran yang penting dalam pengambilan keputusan yang dilakukan oleh manusia. In computer human interaction largely ignores emotion where as it has a biasing role in human-to-human communication in our everyday life (Danisman & Alpkocak, 2008).

### 2. Pre-Processing

Preprocessing merupakan salah satu tahapan yang penting untuk data pada proses mining. Data yang digunakan dalam proses mining tidak selamanya dalam kondisi yang ideal untuk diproses. Terkadang pada data tersebut terdapat berbagai permasalahan yang dapat mengganggu hasil dari proses mining itu sendiri seperti diantaranya adalah missing value, data redundant, outliers, ataupun format data yang tidak sesuai dengan sistem. Oleh karenanya untuk mengatasi permasalahan tersebut dibutuhkan tahap Preprocessing. Preprocessing merupakan salah satu tahapan menghilangkan permasalahan-permasalahan yang dapat mengganggu hasil daripada proses data. Dalam kasus klasifikasi dokumen yang menggunakan data bertipe teks, terdapat beberapa macam proses yang dilakukan

umumnya diantaranya case folding, filtering (remove punctuation), stopword removal, stemming, tokenization dan sebagainya. (Samir, 2019).

### 3. Shiny

Shiny merupakan sebuah paket dalam R yang memungkinkan penggunanya membangun web apps yang interaktif. Shiny menggabungkan antara kekuatan komputasi statistika R dan interaksinya dengan web modern interaktif. Struktur shiny terdiri dari 4 chunk atau komponen yaitu global, ui, server, dan run-app.

Pada chunk Global berisi paket-paket (library) yang diperlukan sehingga web dapat berjalan dengan baik. UI merupakan fungsi yang mendefinisikan tampilan web dari aplikasi yang akan dijalankan. Fungsinya memuat seluruh input dan output yang akan ditampilkan dalam app. Server merupakan fungsi yang mendefinisikan logika kerja analysis dari sisi server pada aplikasi. Run-App merupakan fungsi dari aplikasi yang memanggil UI dan Server untuk menjalankan aplikasi.

### 4. Algoritma Latent Dirichlet Allocation (LDA)

Analisis sentimen dapat membantu dalam mengekstrasi pendapat yang berasal dari sebuah dokumen, komentar, review produk, dan data-data lainnya sehingga akan lebih mudah dalam melakukan pemantauan. Salah satu metode untuk melakukan analisis sentimen adalah Latent Dirichlet Allocation (LDA) yang mampu digunakan untuk mengekstraksi topik dari kumpulan dokumen komentar, dimana topik tersebut direpresentasikan sebagai kemunculan kata-kata dengan probabilitas topik yang berbeda. Oleh karena itu diperlukan representasi data dalam bentuk visual yang mudah dipahami daripada teks dan tabel. Salah satu bentuk visualisasi data adalah wordcloud yang memberikan gambaran visual frekuensi kemunculan kata.

### 5. Word cloud

Word cloud (disebut juga text cloud atau tag cloud) merupakan salah satu metode untuk menampilkan data teks secara visual. Grafik ini populer dalam text mining karena mudah dipahami. Dengan menggunakan word cloud, gambaran frekuensi kata-kata dapat ditampilkan dalam bentuk yang menarik namun tetap informatif. Semakin sering satu kata digunakan, maka semakin besar pula ukuran kata tersebut ditampilkan dalam word cloud.

## METODE

Proses penelitian menggunakan teknik unsupervised learning yang dijelaskan oleh (Vishwanathan dan Smola, 2010). unsupervised learning merupakan salah satu tipe algoritma machine learning yang digunakan untuk menarik kesimpulan dari dataset. Metode ini hanya akan mempelajari suatu data berdasarkan kedekatannya saja atau yang biasa disebut dengan clustering. Proses clustering menggunakan metode unsupervised learning untuk menghasilkan kluster positif dan negatif dari review pakaian wanita.

### 1. Sumber Dataset

Pengumpulan data training review women's clothing e-commerce menggunakan parameter yang sudah tersedia dari website <https://www.kaggle.com> untuk mendapatkan review kategori positif atau negatif. Data yang telah diambil disimpan dalam file data womensclothing.csv dan akan dilanjutkan proses preprocessing data.

### 2. Preprocessing Data

Data yang telah kita simpan akan memasuki proses preprocessing data. Pemrosesan data ini dilakukan dengan beberapa tahap. Adapun tahapan preprocessing nya adalah :

#### a. Normalisasi Data

Ada beberapa komponen khas yang biasa ada di kolom ulasan review yakni karakter khusus seperti tanda baca (.,?! dll), karakter lainnya (\$,%,\* , dll). Karena karakter khusus, tanda baca dan karakter lainnya tidak memiliki pengaruh apapun terhadap nilai sentimen, maka komponen-komponen khas diatas akan dibersihkan dengan cara membuang atribut dalam sebuah ulasan review. Berikut langkah-langkah pada tahap normalisasi fitur:

1. Data akan diperiksa apakah terdapat karakter khusus seperti tanda baca (.,?! dll), karakter lainnya (\$,%,\* , dll).
2. Jika terdapat karakter khusus, tanda baca dan karakter lainnya maka akan dihilangkan.

Selanjutnya yaitu melakukan Case Folding. Tahap ini akan dilakukan perubahan semua karakter menjadi lowercase atau huruf kecil. Langkah-langkah dari case folding adalah :

1. Memeriksa ukuran setiap karakter dari awal hingga akhir
2. Jika ditemukan karakter dengan huruf kapital (uppercase), maka karakter tersebut akan diubah menjadi huruf kecil (lowercase).

### 3. Topic Modeling Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) digunakan untuk menentukan topik secara otomatis dari sekumpulan ulasan review. Ulasan review yang diteliti memiliki struktur tersembunyi (hidden structure) berupa topik, distribusi topik per ulasan review, dan penentuan topik per kata dalam setiap ulasan review. LDA

menggunakan kumpulan ulasan review tersebut untuk inference struktur topik tersembunyi. Jumlah topik yang akan dihasilkan telah ditentukan sebelum proses LDA dilaksanakan.

Setelah model topik LDA dibuat, sebuah ulasan review dapat ditentukan distribusi topiknya yang mendeskripsikan kumpulan kata dalam ulasan review. LDA menggunakan asumsi bag of words, yaitu urutan kemunculan kata dalam dokumen diabaikan.

#### 4. Word Cloud

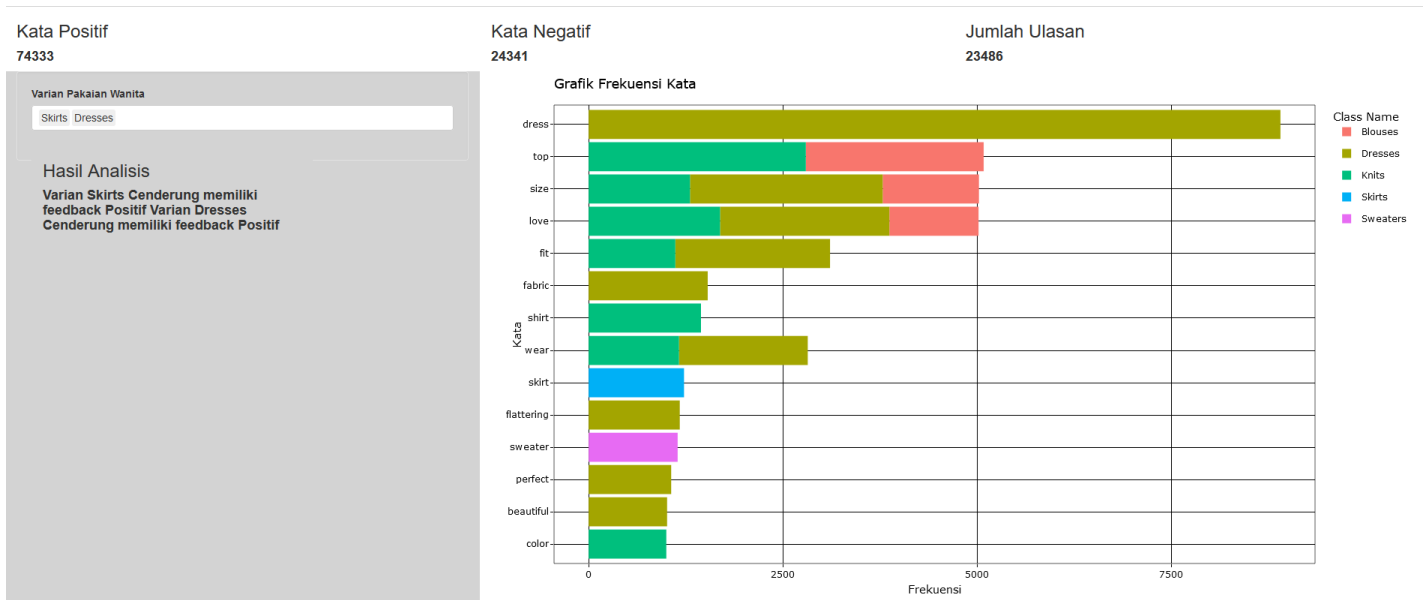
Pembuatan word cloud menggunakan package wordcloud dan dapat diinstal dalam R-studio dari repositori CRAN kemudian dikerjakan dalam dua tahap, yaitu persiapan data dan pembuatan word cloud. Data yang diambil berupa kumpulan kata negatif dan positif yang sudah diklasifikasikan. Kemudian tampilan word cloud kata negatif berwarna hitam dan berada di bagian atas sedangkan kata positif berwarna biru yang berada dibagian bawah. Total kata yang ditampilkan dalam word cloud berisi 200 kata negatif dan positif.

## HASIL DAN PEMBAHASAN

Setelah dilakukan pengumpulan data dengan menggunakan library(vroom) dengan aplikasi RStudio, diperoleh 23486 ulasan review.

Selanjutnya dilakukan preprocessing atau tahap persiapan data. Preprocessing atau tahap persiapan data merupakan tahapan yang paling penting dan krusial dalam sebuah penelitian yang menggunakan metode text mining. Pada penelitian ini, tahap preprocessing dilakukan dengan membuang tanda baca dan karakter lainnya yang terdapat pada data. Selanjutnya data akan diubah semua ukuran huruf menjadi huruf kecil (lowercase). Hal ini sangat penting untuk dilakukan karena kata yang terdiri dari huruf-huruf yang sama akan dibaca oleh sistem sebagai kata yang berbeda apabila dari huruf-hurufnya memiliki ukuran huruf yang berbeda. Misalnya seperti kata 'Recommend dan 'recommend. Apabila hal tersebut dibiarkan, akan memberikan dampak kepada bertambahnya waktu komputer untuk mengolah data. Namun yang terpenting adalah, hasil yang akan didapat akan menjadi rancu atau tidak jelas.

Setelah Seluruh data sudah dibersihkan dan difilter, kita akan mendapatkan grafik interface hasil Sentiment Analysis pakaian wanita yang sudah dikategorikan menjadi beberapa variasi seperti (blouses, dresses, knits, skirts, sweaters) dan kesimpulan hasil analisis sesuai dengan varian yang dipilih, yang dapat dilihat pada gambar dibawah :

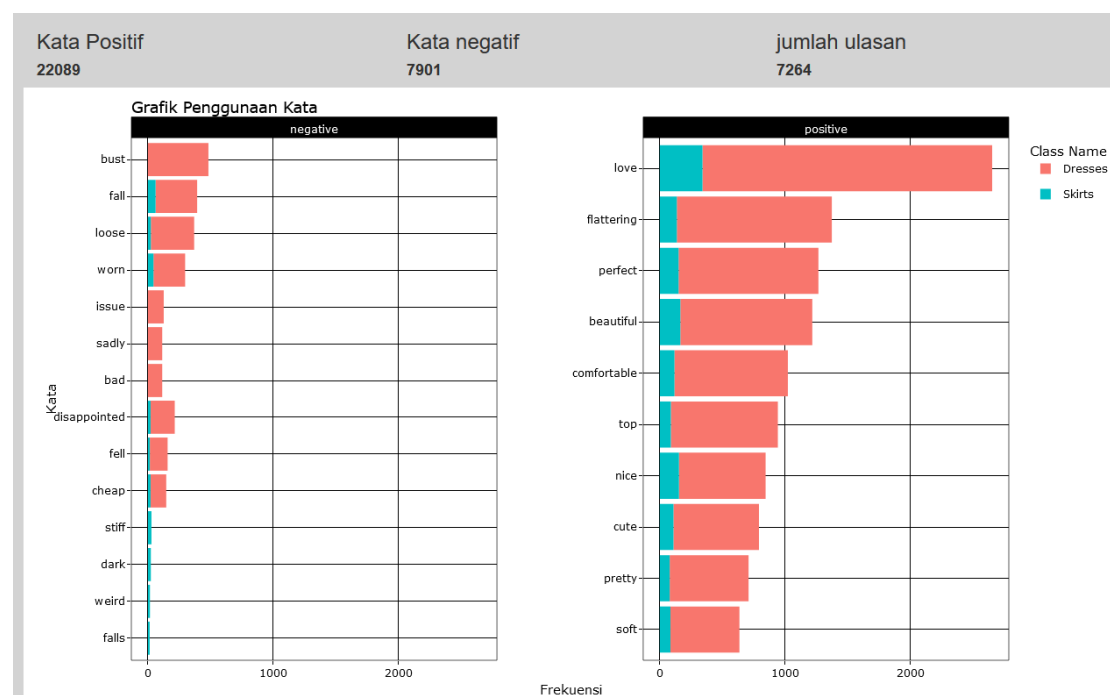


Gambar 1 Word Frequency Graphic

Hasil dari polaritas yang diambil berdasarkan ulasan review dikategorikan menjadi dua bentuk analisis polaritas seperti: Positif dan Negatif. Sehingga dapat dilihat pada hasil tabel 1. Menunjukkan jumlah masing-masing kategori polaritas. Dari hasil yang didapatkan sebelumnya disajikan ke dalam bentuk grafik interface yang terlihat pada Gambar 2 untuk mempermudah proses pembacaan hasil akhir.

**Tabel 1. Jumlah kata pada pakaian wanita berdasarkan sentiment untuk kategori dresses dan skirts**

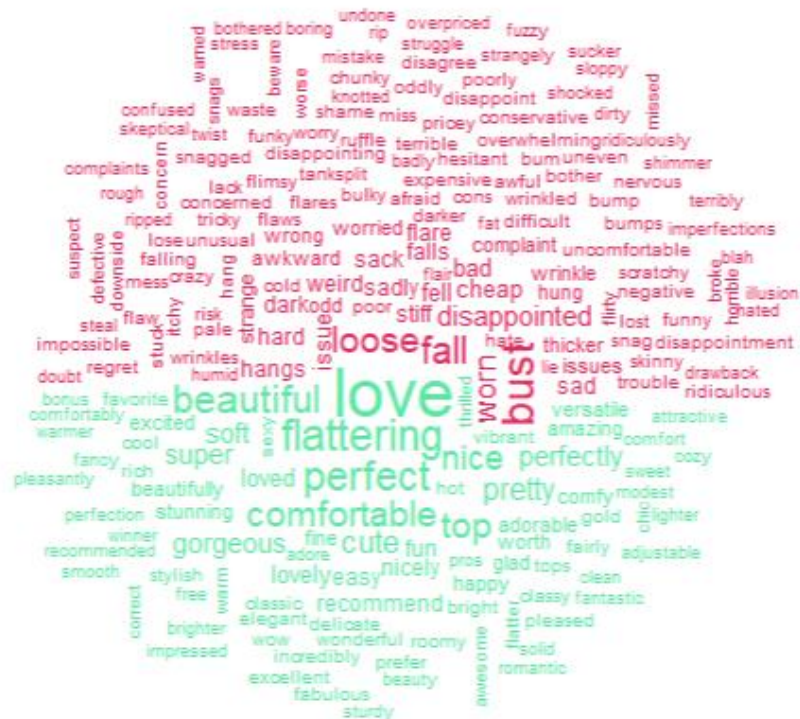
Sentimen	Jumlah
Positif	22089
Negatif	7901



Gambar 2 Word Usage Graohic

Hasil dari word cloud diambil berdasarkan ulasan review yang dikategorikan menjadi dua bentuk analisis yaitu kata negatif dan positif seperti gambar 3. Semakin besar kata dalam word could yang ditampilkan maka kata tersebut adalah kata yang paling banyak atau sering digunakan dalam ulasan review.

NEGATIF



POSITIF

*Gambar 3 Word Cloud*

## **KESIMPULAN**

Metode Latent Dirichlet Allocation (LDA) dapat digunakan untuk proses pengklasifikasian suatu data berupa text-mining. Penggunaan LDA seringkali dipergunakan untuk mengelola atau mengklasifikasi data-data berdasarkan topik modeling. Tentunya dengan menghasilkan keputusan yang benar.

Berdasarkan hasil dari analisis sentimen yang telah dilakukan, dapat ditarik kesimpulan bahwa masyarakat lebih banyak memberikan respon yang bersentimen positif dibandingkan yang bersentimen negatif. Berdasarkan hasil word cloud semakin besar kata yang berada dalam word cloud maka kata tersebut adalah kata yang paling banyak atau sering digunakan dalam ulasan review.