

# **LAPORAN *REVIEW* DUA TUGAS AKHIR**

*Disusun untuk memenuhi tugas mata kuliah Metodologi Penelitian*

## **TUGAS INDIVIDU**



**Oleh :**

Angelia Charlotte Francis

(55201122013)

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER DAN INFORMATIKA  
UNIVERSITAS NURTANIO  
BANDUNG  
2025**

# **1. ANALISIS PEMODELAN SENTIMEN BERBASIS ASPEK TERHADAP BRAND LOKAL KECANTIKAN PADA TWITTER MENGGUNAKAN METODE DECISION TREE CART.**

- Latar belakang dari judul tersebut ialah

Twitter merupakan salah satu media sosial yang berkembang pesat di Indonesia dan sering digunakan masyarakat, khususnya remaja, untuk membagikan opini, termasuk tentang produk kecantikan. Produk kecantikan terdiri dari skincare dan make up, dan semakin banyaknya brand lokal membuat ulasan konsumen berperan penting dalam membentuk persepsi terhadap kualitas, harga, dan kecocokan produk. Namun, opini-opini yang tersebar di media sosial sering kali masih bersifat asumsi dan belum dianalisis secara ilmiah, sehingga prediksi perusahaan terhadap preferensi konsumen kerap tidak akurat. Oleh karena itu, analisis pemodelan sentimen dengan pendekatan machine learning, seperti Decision Tree CART, menjadi penting untuk mengklasifikasikan opini menjadi sentimen positif atau negatif secara otomatis. Dengan penerapan pemodelan sentimen yang tepat, perusahaan dapat memahami respons konsumen secara lebih akurat, meningkatkan kualitas produk, serta merespons umpan balik pasar secara lebih efektif.

- Hal-hal yang melatarbelakangi penelitian ini antara lain:
  - a. Pesatnya penggunaan Twitter
  - b. Meningkatnya minat masyarakat terhadap produk kecantikan
  - c. Pengaruh opini konsumen dalam keputusan pembelian
  - d. Keterbatasan analisis opini secara ilmiah
  - e. Kebutuhan akan pemodelan sentimen yang sistematis

- Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka dalam penelitian ini permasalahan yang akan dirumuskan adalah:

1. Bagaimana implementasi dalam klasifikasi analisis pemodelan sentimen berbasis aspek terhadap *brand* lokal kecantikan pada Twitter menggunakan metode *Decision Tree CART*?
2. Bagaimana hasil akurasi yang dihasilkan dari klasifikasi analisis pemodelan sentimen berbasis aspek menggunakan metode *Decision Tree CART*?

- Tujuan Penelitian

Berdasarkan rumusan masalah diatas maka didapat tujuan dari penelitian ini sebagai berikut:

1. Mengimplementasikan klasifikasi analisis pemodelan sentimen berbasis aspek terhadap *brand* lokal kecantikan pada Twitter menggunakan metode *Decision Tree CART*.
2. Mengetahui hasil akurasi dari penerapan klasifikasi analisis pemodelan sentimen berbasis aspek menggunakan metode *Decision Tree CART*

- Teori yang digunakan pada judul ini ialah:

1. Analisis Sentimen
2. Text Mining
3. NLP
4. Ketidak Seimbangan Data (Data Imbalance)
5. Word embedding
6. *Decisin Tree CART*
7. Model Confusion Matrix

- Metodologi yang digunakan ialah:

1. Metode Pengumpulan data

- a. Data menurut sifat

Dalam penelitian ini menggunakan data yang bersifat kuantitatif. Data kuantitatif diperoleh melalui survei dan memberikan jawaban dalam bentuk angka. Ciri utamanya adalah bersifat objektif dan interpretasi data cenderung tidak bervariasi.

- b. Data menurut sumber

Dalam penelitian ini menggunakan sumber data eksternal. Data eksternal diperoleh dari sumber di luar organisasi atau lokasi penelitian. Jenis data ini sering digunakan untuk membandingkan satu tempat dengan tempat lainnya, seperti data kependudukan, penjualan produk dari perusahaan lain, atau jumlah siswa dari sekolah lain.

2. Metode Implementasi

- a. Pengumpulan Data

Proses pengumpulan data dilakukan dengan mencari data di media soaial. Data tersebut kemudian dianalisis untuk diolah dan diproses.

b. Preprocessing dan Labelling

Setelah pengumpulan data, dilakukan dengan beberapa tahapan, seperti Preprocessing untuk membersihkan dan mempersiapkan data serta melakukan label sentiment positif dan negatif.

c. Implementasi dan Pengujian

Proses implementasi dengan membagi data train dan data test, merepresentasikan kata menggunakan *fasttext*, dan mengklasifikasikan menggunakan metode *decision tree*. Pengujian dilakukan untuk mendapatkan hasil akurasi dari model *Decision Tree CART* dalam melakukan klasifikasi menggunakan *confusion matrix*.

- Hasil Penelitian

1. Dari kedua produk Somethinc dan Wardah menunjukkan lebih banyak sentimen positif dibandingkan sentimen negatif. Untuk produk Somethinc dari total data sebanyak 2215 data dengan sentimen positif berjumlah 1461 dan sentimen negatif berjumlah 754, sedangkan untuk produk Wardah sebanyak 2098 data terdapat sentimen positif berjumlah 1638 data dan sentimen negatif berjumlah 460 data. Sehingga menunjukkan bahwa kedua brand ini memiliki penerimaan yang baik di kalangan konsumen.
2. Untuk mengatasi ketidakseimbangan kelas sentimen, upsampling dilakukan pada data latih, memungkinkan model belajar lebih efektif. Metode Decision Tree dengan upsampling menghasilkan akurasi 80,13%, presisi 81,90%, recall 90,20%, dan F1-Score 85,84% untuk produk Somethinc, sedangkan untuk produk Wardah menghasilkan akurasi 86,19%, presisi 89,41%, recall 93,25%, dan F1-Score 91,28%, hasil tersebut menunjukkan kinerja model yang dibangun sudah baik dalam mengklasifikasikan sentimen pada kedua produk.

## **2. PEMODELAN ANALISIS SENTIMEN K-POPERS DI TWITTER (X) TERHADAP HARGA TIKET KONSER MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (STUDI KASUS : GOLDEN DISK AWARDS 2024).**

- Latar Belakang dari judul tersebut ialah,

Fenomena Korean Pop (K-Pop) di Indonesia telah mengalami perkembangan pesat dalam beberapa tahun terakhir. Hal ini dipengaruhi oleh kemajuan teknologi digital dan kemudahan akses terhadap internet, yang memungkinkan masyarakat Indonesia, khususnya generasi muda, untuk mengakses berbagai konten hiburan dari Korea Selatan secara cepat dan luas. K-Pop bukan hanya menghadirkan musik, tetapi juga budaya, gaya hidup, dan visual yang menarik perhatian. Daya tarik ini diperkuat oleh kehadiran komunitas penggemar aktif yang menyebarkan informasi melalui media sosial, serta penyelenggaraan konser dan acara bertema Korea yang semakin marak di Indonesia.

- Hal-hal yang melatarbelakangi penelitian ini antara lain:

- a. Kemajuan Teknologi dan Akses Internet
- b. Daya Tarik Visual dan Koreografi
- c. Kehadiran Komunitas Penggemar yang Solid
- d. Strategi Globalisasi Industri Hiburan Korea
- e. Kebutuhan Ekspresi Diri dan Identitas Anak Muda

- Rumusan masalah

Berdasarkan latar belakang yang telah dijabarkan diatas, rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana analisis sentimen K-Popers di media sosial Twitter (X) terhadap harga tiket konser pada studi kasus Golden Disk Awards 2024?
2. Bagaimana pemodelan analisis sentimen K-Popers di media sosial Twitter (X) dengan menggunakan metode Support Vector Machine dalam mengklasifikasikan harga tiket konser pada studi kasus Golden Disk Awards 2024?
3. Bagaimana mengukur model berdasarkan tingkat akurasi hasil pemodelan analisis sentimen K-Popers di media sosial Twitter (X) dalam mengklasifikasikan harga tiket konser pada studi kasus Golden Disk Awards 2024 dengan menggunakan Confusion Matrix?

- Tujuan

Berdasarkan rumusan masalah yang telah dijabarkan diatas, tujuan pada penelitian ini adalah sebagai berikut:

1. Menganalisis sentimen K-Popers di media sosial Twitter (X) terhadap harga tiket konser pada studi kasus Golden Disk Awards 2024.
2. Memodelkan analisis sentimen K-Popers di media sosial Twitter (X) dengan menggunakan metode Support Vector Machine dalam mengklasifikasikan harga tiket konser pada studi kasus Golden Disk Awards 2024.
3. Mengukur model berdasarkan tingkat akurasi hasil pemodelan analisis sentimen K-Popers di media sosial Twitter (X) dalam mengklasifikasikan harga tiket konser pada studi kasus Golden Disk Awards 2024 dengan menggunakan Confusion Matrix

- Teori yang digunakan

1. Text Mining

Text Mining adalah metode yang dapat digunakan untuk memenuhi kebutuhan untuk proses klasifikasi dokumen dengan konten apapun. Metode ini bekerja dengan menghasilkan variasi dari kumpulan data dalam jumlah besar dan kemudian mencoba menemukan pola yang sesuai dengan apa yang diharapkan dari kumpulan data teks yang ada

- a. Preprocessing

Preprocessing merupakan tahap menyiapkan data untuk dianalisis atau proses dengan menghapus data yang tidak relevan, guna meningkatkan efisiensi system, dalam preprocessing terdapat

1. Case folding
2. Cleansing
3. Tokenizing
4. Normalization
5. Stopword

2. Natural Language Processing (NLP)

Pemrosesan Bahasa Alami (Natural Language Processing/NLP) membuat program yang memiliki kemampuan untuk memahami bahasa manusia.

### 3. Analisis Sentimen

Analisis sentimen merupakan sebuah metode untuk mengekstrak data opini dan memahami serta mengolah tekstual data secara otomatis untuk melihat atau mengidentifikasi sentimen yang terkandung dalam sebuah opini

### 4. Crawling Data

Crawling Data merupakan proses otomatis untuk mengumpulkan data dari berbagai sumber seperti situs web, database, atau dokumen dengan menggunakan software atau aplikasi khusus untuk mengakses sumber data dan mengambil informasi yang dibutuhkan. Data yang dikumpulkan melalui proses Crawling Data kemudian dapat diproses dan digunakan untuk berbagai tujuan, seperti analisis data, penelitian, atau pengembangan sistem informasi

### 5. Balanced Dataset

Balanced dataset adalah data yang perbandingan setiap data dalam suatu kelas seimbang, sebaliknya data yang setiap kelasnya memiliki jumlah yang berbeda secara signifikan, maka dataset tersebut disebut imbalance. Ketidakseimbangan kelas data adalah masalah umum dalam klasifikasi Machine Learning dimana terdapat rasio yang tidak proporsional di setiap kelas

### 6. Word2vec

Word2vec merupakan algoritma dalam bidang NLP yang digunakan untuk membuat representasi vektor kata (word embedding) dari teks. Tujuan dari Word2vec yaitu untuk menghubungkan setiap kata dalam teks ke dalam ruang vektor, dimana kata-kata dengan makna yang mirip atau sering kali muncul bersama-sama akan memiliki representasi vektor yang berdekatan satu sama lain, yang memungkinkan komputer untuk mengenali dan memahami hubungan semantic antara kata-kata tersebut.

### 7. Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan metode klasifikasi menerapkan konsep dalam mengklasifikasikan data, yaitu menentukan hyperplane terbaik untuk memberi jarak atau pemisah antar kelas. Dalam memprediksi suatu kelas dari data, SVM akan melabelinya berdasarkan daerah kelas mana yang merupakan tempat dari data tersebut. Prinsip SVM adalah membangun hyperplane yang memiliki ukuran margin yang sama dan tidak cenderung mendekati daerah dari salah satu kelas, usaha pada pencarian hyperplane terbaik sebagai pemisah antar kelas merupakan inti dari metode SVM

## 8. Confusion Matrix

Confusion Matrix merupakan suatu metode yang digunakan sebagai evaluasi klasifikasi untuk dapat mengetahui seberapa akurat metode yang di evaluasi.

- Metodologi penelitian

### 1. Metodologi Pengumpulan Data

#### a. Data Berdasarkan Sifat

Data dari penelitian ini bersifat Data Kuantitatif. Data kuantitatif merupakan data yang diperoleh melalui sebuah survei yang menghasilkan jawaban yang disajikan dalam bentuk angka, sehingga data tersebut cenderung bersifat objektif. Oleh karena itu, ketika membaca data tersebut penafsiran atau pengartiannya tidak akan berbeda.

#### b. Data Berdasarkan Sumber

Sumber data dari penelitian ini menggunakan Data Eksternal. Data eksternal merupakan data yang diperoleh dari situasi dan kondisi di luar organisasi atau tempat penelitian dilakukan. Jenis data ini sering digunakan sebagai pembanding suatu tempat I-7 dengan tempat lainnya. Contohnya adalah data tingkat kepuasan konsumen, data kependudukan, data penjualan produk dan sebagainya.

#### c. Data Berdasarkan Cara Memperoleh

Cara memperoleh data dalam penelitian ini menggunakan Data Primer. Data primer merupakan data yang dikumpulkan secara langsung dari objek penelitian, baik oleh seorang peneliti maupun sebuah organisasi. Pengumpulan data dapat berupa wawancara, kuesioner, survei atau observasi

#### d. Data Berdasarkan Waktu

Dalam penelitian ini waktu pengumpulan data termasuk dalam data cross-section. Data cross-section merupakan data yang dikumpulkan dalam satu waktu sehingga data menunjukkan titik waktu tertentu tanpa mempertimbangkan perubahan dari waktu ke waktu. Misalnya adalah data pengangguran januari 2023

### 2. Metodologi Implementasi

#### a. Pengumpulan Data

Proses pengumpulan data dilakukan dengan mencari data dari media sosial untuk mendapatkan data mentah, yang kemudian diolah dan diproses



b. Preprocessing dan Labeling

Setelah proses pengumpulan data, tahap Preprocessing melibatkan pengolahan awal data mentah dengan membersihkan dan mempersiapkan data sebelum digunakan dalam model serta melakukan label sentimen positif dan negatif.

c. Implementasi dan Pengujian

Proses implementasi dengan membagi data menjadi data train dan data test, merepresentasikan kata sentimen menggunakan Word2vec dan mengklasifikasikan menggunakan metode Support Vector Machine. Pengujian dilakukan untuk mendapatkan hasil akurasi dari hasil klasifikasi model menggunakan algoritma yang telah diimplementasikan menggunakan Confusion Matrix.

- Hasil penelitian

1. Sentimen K-Popers terhadap harga tiket konser Golden Disk Awards 2024 di media sosial Twitter (X) cenderung negatif. Dari total data sampel sebanyak 2698 data tweet, ditemukan sentimen positif sebanyak 1071 data (37,69%) dan sentimen negatif sebanyak 1681 data (62,31%). Hal ini mengindikasikan bahwa kebijakan harga tiket yang ditetapkan perlu dievaluasi kembali oleh pihak promotor atau penyelenggara acara.
2. Data train yang digunakan pada penelitian ini dilakukan upsampling untuk mengatasi ketidakseimbangan kelas dengan menyeimbangkan kelas positif terhadap kelas negatif sebagai kelas mayoritas sehingga model dapat belajar lebih baik dari masing-masing kelas.
3. Hasil klasifikasi model dengan metode Support Vector Machine menunjukkan akurasi tertinggi dengan proporsi data latih dan uji 90:10 menggunakan kernel polynomial. Persamaan kernel polynomial yang digunakan adalah  $K(x_i, x_j) = ((x_i \cdot x_j) + 1)^4$  dengan parameter  $C = 10$ ,  $\gamma = 1$ ,  $\text{coef0} = 1$ ,  $\text{degree} = 4$ . Model ini menghasilkan akurasi sebesar 85,19%, presisi 83,04%, recall 81,58%, dan F1-Score 82,30%. Berdasarkan hasil tersebut, rasio pembagian dataset disimpulkan mempengaruhi performa model dan berdasarkan akurasi yang didapat maka dapat dikatakan bahwa kinerja model yang dibangun sudah baik dalam mengklasifikasikan sentimen K-Popers terhadap harga tiket konser.

### **3. PENERAPAN MACHINE LEARNING, DEEP LEARNING, DAN DATA MINING DALAM DETEKSI KECURANGAN LAPORAN KEUANGAN - A SYSTEMATIC LITERATURE REVIEW**

- Latar Belakang

Fraud adalah tindakan tidak jujur yang dilakukan untuk memperoleh keuntungan secara ilegal, termasuk kecurangan laporan keuangan yang menipu investor dan stakeholder. Praktik ini dapat merugikan entitas dan berdampak negatif pada ekonomi. Metode konvensional untuk mendeteksi kecurangan sering kali tidak efektif dan mahal, meningkatkan risiko kesalahan manusia. Dengan kemajuan teknologi, metode deteksi yang lebih canggih seperti machine learning, deep learning, dan data mining menjadi penting untuk menemukan pola tersembunyi dalam data dan meningkatkan akurasi deteksi. Penelitian ini bertujuan untuk meninjau penerapan teknologi tersebut dalam deteksi kecurangan laporan keuangan serta mengeksplorasi potensi, dampak positif, dan faktor pendorong dalam penerapannya.

- Tujuan Masalah

Penelitian ini bertujuan untuk melakukan tinjauan literatur sistematis mengenai penerapan teknologi tersebut dalam deteksi kecurangan laporan keuangan, serta mengeksplorasi potensi, dampak positif, dan faktor pendorong dalam penerapannya.

- Metodologi Penelitian

1. Systematic Literature Review (SLR)

Penelitian ini menggunakan metode tinjauan literatur sistematis untuk menganalisis berbagai artikel yang diterbitkan dalam jurnal akademik yang terindeks di Scopus dari tahun 2020 hingga 2024.

2. Proses Pencarian

Artikel dicari dengan menggabungkan kata kunci yang relevan menggunakan operator boolean (AND, OR), yang mencakup istilah seperti "machine learning," "deep learning," "data mining," dan "financial statement fraud."

3. Kriteria Inklusi dan Eksklusi

Kriteria literatur terdiri dari kriteria inklusi (artikel yang diterbitkan dalam rentang waktu 2020-2024, dalam bahasa Inggris, dan terindeks di Scopus) dan kriteria eksklusi (artikel yang tidak memenuhi kriteria tersebut).

#### 4. Metode PRISMA

Peneliti menerapkan metode PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) yang melibatkan beberapa tahapan penyaringan untuk mendapatkan jumlah artikel yang sesuai dengan rumusan masalah, yang akhirnya menghasilkan 21 artikel untuk dianalisis lebih lanjut.

#### 5. Analisis Data

Artikel yang terpilih dianalisis untuk mengeksplorasi topik, hasil, metodologi, rekomendasi, dan keterbatasan yang terdapat dalam penelitian sebelumnya.

- Teori Penelitian

1. Machine Learning
2. Deep Learning
3. Data Mining
4. Systematic Literature Review (SLR)

- Hasil Penelitian

1. Melalui penggabungan berbagai kekuatan machine learning, deep learning, dan data mining menjadi model yang komprehensif mampu memberikan solusi yang akurat dan efektif dalam mengidentifikasi anomali atau pola mencurigakan dalam data keuangan.
2. Teknik machine learning misalnya Natural Language Processing (NLP) dapat membantu mengubah informasi teks menjadi numerik yang nantinya dapat digunakan untuk menganalisis laporan keuangan dan mendeteksi tanda-tanda kecurangan atau manipulasi melalui analisis informasi.
3. Penggunaan teknik deep learning tidak hanya membantu meningkatkan efisiensi audit keuangan, tetapi juga mendukung prinsip keberlanjutan seperti dalam green finance, dimana investor dapat memastikan bahwa dana mereka dialokasikan dengan benar dan tidak disalahgunakan untuk tujuan yang tidak sesuai.