

Optimalisasi Analisis Sentimen Publik terhadap Marketplace di Indonesia melalui Aplikasi Decision Tree

Auralea Alvinia Syaikha - 13522148¹

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia

¹13522148@itb.ac.id

Abstract—Perkembangan pesat teknologi informasi di Indonesia telah mengubah pola konsumsi masyarakat, khususnya dalam konteks transaksi jual-beli melalui *marketplace*. Sentimen publik terhadap platform-platform ini menjadi kunci dalam membentuk citra dan keberlanjutan bisnis. Makalah ini bertujuan untuk mengulas optimalisasi analisis sentimen publik terhadap *marketplace* di Indonesia dengan memanfaatkan aplikasi *Decision Tree*. Metode ini dipilih sebagai pendekatan utama untuk menyederhanakan kompleksitas analisis sentimen dan memfasilitasi pengambilan keputusan. Melalui kerangka konseptual ini, diharapkan dapat memberikan wawasan yang bermanfaat bagi pemangku kepentingan dalam meningkatkan interaksi positif dengan pengguna pada pasar *online* mereka.

Keywords—*Decision Tree, Marketplace, Sentimen, Ulasan*

I. PENDAHULUAN

Dalam era globalisasi dan kemajuan teknologi informasi, pertumbuhan industri *e-commerce* di Indonesia telah mengalami perkembangan yang pesat. Fenomena ini tidak terlepas dari transformasi perilaku konsumen yang beralih dari pembelian konvensional ke penggunaan *platform online*, seperti *marketplace*. Kehadiran *marketplace* di Indonesia telah memberikan kemudahan bagi pelaku bisnis untuk menjangkau pasar lebih luas, sementara bagi konsumen, memberikan akses mudah untuk memenuhi kebutuhan belanja mereka.

Seiring dengan berkembangnya ekosistem *marketplace*, penting bagi pelaku bisnis dan pemerhati pasar untuk memahami sentimen publik terkait layanan dan produk yang ditawarkan. Analisis sentimen publik menjadi kunci untuk memahami respons masyarakat terhadap berbagai aspek dalam ekosistem *marketplace*. Dalam konteks ini, penggunaan teknologi analisis sentimen menjadi semakin penting sebagai alat untuk menggali informasi berharga dari data publik, khususnya di ranah *online*.

Makalah ini bertujuan untuk mengeksplorasi optimalisasi analisis sentimen publik terhadap *marketplace* di Indonesia, dengan menggunakan pendekatan yang canggih yaitu aplikasi *Decision Tree*. *Decision Tree* merupakan salah satu metode dalam *machine learning* yang dapat digunakan untuk mengklasifikasikan dan menganalisis data kompleks dengan cepat dan efisien. Melalui penerapan *Decision Tree* dalam konteks analisis sentimen, diharapkan dapat ditemukan pola-

pola unik yang dapat memberikan wawasan lebih dalam terkait preferensi dan persepsi masyarakat terhadap berbagai aspek dalam ekosistem *marketplace*.

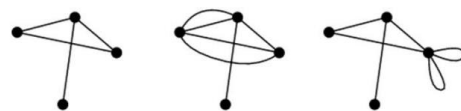
Penelitian ini tidak hanya akan memberikan kontribusi teoretis dalam pengembangan metode analisis sentimen, tetapi juga memberikan pandangan praktis bagi pelaku bisnis, peneliti pasar, dan pembuat kebijakan untuk memahami lebih baik bagaimana meningkatkan layanan dan produk mereka sesuai dengan kebutuhan dan harapan konsumen. Dengan demikian, makalah ini menjadi sebuah langkah strategis dalam meningkatkan kualitas layanan dan daya saing *marketplace* di Indonesia, melalui pemanfaatan teknologi analisis sentimen yang lebih canggih dan efektif.

II. TEORI DASAR

A. Teori Graf

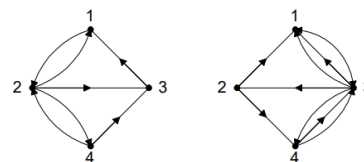
Graf digunakan untuk menggambarkan objek-objek yang bersifat diskrit dan interaksi di antara mereka. Interaksi ini tercermin dalam sisi-sisi graf, yang menghubungkan objek-objek tersebut. Definisi graf secara umum adalah $G = (V, E)$, dengan V sebagai kumpulan simpul yang tidak kosong, dan E sebagai kumpulan sisi yang menghubungkan simpul-simpul tersebut [1]. Bergantung pada orientasi arah pada sisi, graf dapat dibagi menjadi dua jenis utama:

1. Graf Tak-Berarah, graf yang sisinya tidak mempunyai orientasi arah.



Gambar 2.1 Graf Tak-Berarah [1]

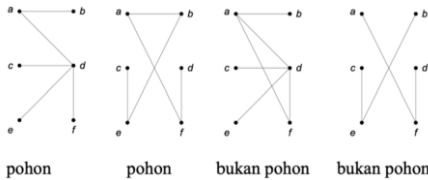
2. Graf Berarah, graf yang setiap sisinya diberikan orientasi arah.



Gambar 2.2 Graf Berarah [1]

B. Teori Pohon

Pohon merupakan sebuah graf, dengan syarat, dimana pada gambarannya digambarkan sama seperti pohon pada umumnya, namun memiliki orientasi terbalik. Sebuah pohon memiliki akar (root) sebagai simpul awal, cabang yang menghubungkan simpul-simpul tersebut, dan daun (leaf) sebagai simpul-simpul terakhir dalam hirarki. Setiap simpul dalam pohon memiliki tepat satu simpul yang disebut sebagai induk, kecuali akar yang tidak memiliki induk. Selain itu, setiap simpul dapat memiliki nol atau lebih simpul anak, dan simpul ini disebut sebagai daun jika tidak memiliki anak [2].



Gambar 2.2 Ilustrasi definisi pohon [2]

Agar dapat digolongkan sebagai pohon, suatu graf harus memenuhi beberapa sifat sebagai berikut:

1. Tiap pasang simpul dalam graf G terhubung dengan lintasan tunggal.

Lintasan tunggal berupa suatu busur yang menghubungkan dua buah simpul yang berbeda dan tidak ada lintasan lain yang menghubungkan kedua simpul tersebut.

2. Graf G merupakan graf terhubung dengan n sisi dan memiliki $(n-1)$ jumlah sisi.

Graf terhubung merupakan graf yang tiap pasang simpulnya memiliki lintasan yang menghubungkan. Graf tergolong terhubung kuat (*strongly connected graph*) jika untuk tiap pasang simpul terdapat lintasan berarah dari satu simpul dari simpul lainnya dan juga sebaliknya. Graf tergolong terhubung lemah (*weakly connected graph*) jika untuk tiap pasang simpul terdapat lintasan tak berarah yang menghubungkan.

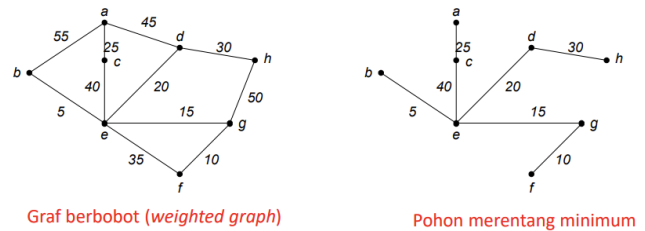
3. Graf G tidak mengandung sirkuit.

Sirkuit adalah lintasan yang menghubungkan seluruh simpul pada suatu graf yang berawal dan berakhir pada simpul yang sama.

4. Penambahan 1 sisi pada graf G akan membentuk tepat 1 buah sirkuit.

C. Pohon Merentang (Spanning Tree)

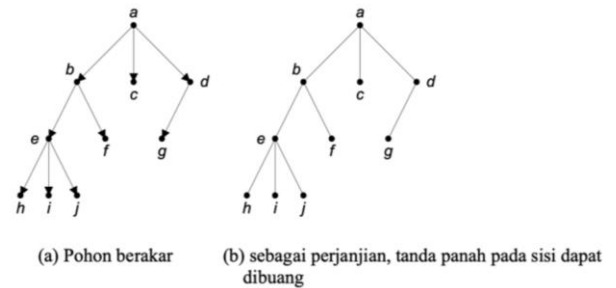
Pohon merentang, yang berasal dari graf yang terhubung, adalah subgraf merentang yang membentuk struktur pohon. Pohon merentang diperoleh dengan cara memotong sirkuit di dalam graf yang bersangkutan. Setiap kali terdapat graf yang terhubung, setidaknya terdapat satu pohon merentang yang dapat diidentifikasi. Untuk mencari minimum *spanning tree*, terdapat dua algoritma yang umum digunakan, yaitu algoritma Prim dan algoritma Kruskal [3].



Gambar 2.3 Ilustrasi Pohon Merentang [3]

D. Pohon Berakar

Pohon berakar (rooted tree) adalah tipe pohon di mana terdapat satu simpul yang memegang peran sebagai akar, dan sisinya diberi arah sebagai penunjuk. Namun, tanda panah pada sisi dapat diabaikan [3].



Gambar 2.4 Ilustrasi Pohon Berakar [3]

Pohon berakar memiliki terminologi sebagai berikut:

1. Anak (child) dan Orang Tua (parent)

Pada Gambar 2.3, simpul a adalah orang tua dari simpul b , c , dan d . Simpul b , c , dan d adalah anak dari simpul a .
2. Lintasan (path)

Lintasan adalah urutan sisi-sisi yang menghubungkan dua simpul. Sebagai contoh, pada Gambar 2.3, lintasan dari simpul a ke simpul g adalah a - d - g .
3. Saudara kandung (sibling)

Suatu simpul disebut bersaudara kandung dengan simpul lainnya jika simpul-simpul tersebut memiliki orang tua yang sama. Sebagai contoh, pada Gambar 2.3, simpul c dan d adalah saudara kandung simpul b .
4. Upapohon (subtree)

Suatu pohon bisa disebut upapohon dari pohon asal jika simpul-simpul dan sisi-sisi dari pohon tersebut merupakan bagian dari simpul-simpul dan sisi-sisi pohon asal.



Gambar 2.5 Ilustrasi Upapohon

Pohon di atas adalah salah satu contoh dari upapohon. Simpul b , e , dan f terdapat pada pohon asal dan sisi yang menghubungkan simpul-simpulnya juga terdapat pada pohon asal.

5. Derajat (degree)

Derajat suatu simpul menunjukkan jumlah anak yang dimiliki simpul tersebut. Sebagai contoh, pada Gambar 2.3, derajat simpul b adalah dua, dan derajat simpul c adalah nol.

6. Daun (leaf)

Suatu simpul disebut daun jika ia tidak memiliki anak (atau berderajat nol). Sebagai contoh, pada Gambar 2.3, simpul c, f, g, h, i, dan j adalah daun.

7. Simpul Dalam

Suatu simpul disebut simpul dalam jika ia memiliki orang tua dan anak. Sebagai contoh, pada Gambar 2.3, simpul b, d, e, dan g adalah simpul dalam.

8. Aras (level)/Tingkat

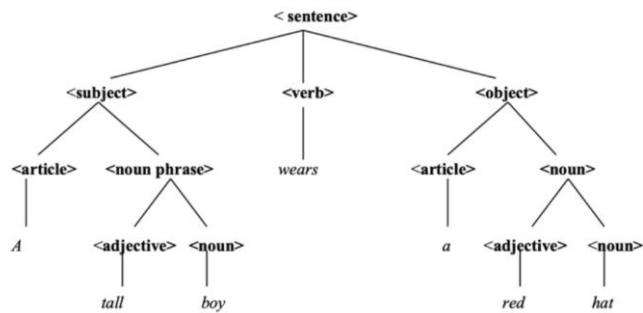
Tingkat suatu simpul menunjukkan jarak/banyak sisi suatu lintasan yang menghubungkan simpul tersebut dari akar pohon tersebut. Sebagai contoh, pada Gambar 2.3, Tingkat simpul f adalah dua, tingkat simpul h adalah 4, dan tingkat simpul a adalah nol.

9. Tinggi (height)/Kedalaman (depth)

Tinggi suatu pohon menunjukkan tingkat terbesar suatu simpul pada pohon tersebut. Sebagai contoh, pada Gambar 2.3, tinggi pohon adalah empat.

E. Pohon N-aire (N-ary)

Pohon n-aire adalah pohon berakar yang pada tiap simpulnya terdapat maksimal n anak. Pohon n-aire memiliki sifat teratur/penuh jika pada tiap simpulnya memiliki tepat sejumlah n anak [3].



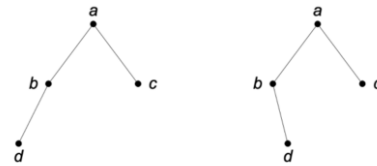
Gambar 2.6 Ilustrasi Pohon N-aire [3]

F. Pohon Biner

Pohon biner, juga dikenal sebagai binary tree, termasuk dalam kategori pohon n-aire dengan nilai n yang setara dengan dua. Pohon ini tergolong sebagai pohon teratur, yaitu pohon berakar yang menjaga urutan anak-anaknya sebagai informasi penting. Sehingga, jika urutan simpul-simpul diubah, maka akan membentuk pohon yang berbeda. Salah satu aspek unik dari pohon biner adalah bahwa setiap anak dari setiap simpulnya diberi penamaan spesifik, dengan salah satu anak disebut sebagai anak kiri (left child) dan yang lainnya sebagai anak kanan (right child) [3].

Penerapan pohon biner sangat beragam dalam kehidupan sehari-hari. Contohnya mencakup penggunaannya dalam pohon ekspresi, pohon keputusan, kode awalan, kode Huffman, dan

pohon pencarian biner. Keberagaman ini menunjukkan fleksibilitas pohon biner dalam mendukung berbagai aplikasi yang melibatkan struktur data hirarkis.

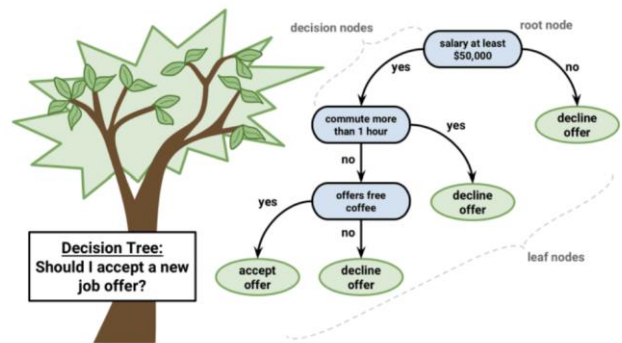


Gambar 2.7 Ilustrasi dua Pohon Biner yang berbeda [3]

G. Pohon Keputusan (Decision Tree)

Pohon keputusan merupakan implementasi dari konsep pohon berakar, di mana ujung akar menyajikan opsi yang dapat diikuti hingga mencapai daun atau solusi dari opsi tersebut. Proses ini menunjukkan bahwa untuk mencapai solusi, diperlukan rangkaian pilihan yang harus diambil [3].

Sebagai metode klasifikasi yang sering digunakan, Pohon Keputusan sangat mudah diaplikasikan dalam kehidupan sehari-hari. Model prediksi ini menggunakan struktur pohon atau hirarki untuk mengorganisir data. Ide dasar dari Pohon Keputusan adalah mentransformasikan data menjadi struktur pohon keputusan dengan menerapkan aturan-aturan yang melekat pada pohon keputusan itu sendiri.



Gambar 2.8 Ilustrasi contoh Pohon Keputusan [3]

H. Marketplace

Marketplace adalah suatu platform daring yang memberikan tempat bagi berbagai penjual dan pembeli untuk menjalankan transaksi. Prinsip dasar dari marketplace adalah menciptakan suatu ekosistem di mana penjual dapat menawarkan produk atau layanan kepada calon pembeli yang mencari kebutuhan atau keinginan tertentu. Model ini telah menjadi dasar pertumbuhan e-commerce dan digitalisasi bisnis. Dalam lingkungan marketplace, transaksi dilakukan secara elektronik melalui platform online, memberikan Marketplace pada umumnya memberikan keuntungan bagi penjual dengan menyediakan pangsa pasar yang lebih luas, sementara pembeli dapat dengan mudah mencari produk atau layanan, membandingkan harga, dan mengalami pengalaman berbelanja yang terintegrasi. Dalam ekosistem ini, pemahaman mengenai dinamika persaingan, sistem pembayaran, keamanan transaksi, dan mekanisme umpan

balik pelanggan menjadi penting karena memiliki dampak signifikan terhadap kesuksesan suatu *marketplace*. Salah satu contoh marketplace di Indonesia adalah Tokopedia, Bukalapak, Lazada, Shopee, Blibli, dan sebagainya [4].



Gambar 2.9 Beberapa Contoh *Marketplace* di Indonesia [4]

I. Analisis Sentimen

Analisis sentimen adalah suatu teknik dalam bidang komputasi dan pemrosesan bahasa alami yang bertujuan untuk menganalisis dan memahami perasaan atau sentimen yang terkandung dalam teks, seperti ulasan pelanggan, komentar, atau kiriman di media sosial. Dasar dari analisis sentimen mencakup interpretasi makna kata dan frasa, serta kemampuan untuk mengenali dan mengklasifikasikan emosi yang muncul dalam teks. Metode ini umumnya menggunakan teknik pembelajaran mesin (*machine learning*), seperti klasifikasi teks dan pemrosesan bahasa alami (*natural language processing / NLP*), untuk mengidentifikasi pola dan konteks yang dapat mengungkapkan sentimen positif, negatif, atau netral.

Penerapan analisis sentimen dapat dilakukan dalam berbagai konteks, termasuk dalam ranah bisnis, media sosial, dan survei pelanggan, guna membantu organisasi memahami pandangan dan respons masyarakat terhadap produk, layanan, atau isu tertentu. Selain itu, dasar-dasar teori analisis sentimen juga melibatkan aspek-aspek linguistik, psikologis, dan komputasional untuk memahami kompleksitas bahasa manusia serta variasi dalam ekspresi sentimen dalam berbagai konteks komunikasi [5].

III. PEMBAHASAN

A. Analisis Sentimen Publik

Ledakan data yang terdapat di internet, terutama di media sosial, merupakan harta karun yang berupa informasi penting apabila diolah dengan tepat. Media sosial tak luput menjadi media bagi konsumen untuk menyampaikan pendapatnya mengenai suatu produk ataupun jasa tertentu. Oleh karena itu, menganalisis sentimen dari opini-opini di media sosial tersebut penting untuk dilakukan dalam rangka penambangan opini.

Berdasarkan klasifikasi, analisis sentimen adalah mengklasifikasikan sentiment teks ke dalam positif atau negative. Hal ini adalah proses penting untuk menyimpulkan opini bernilai positif, negatif maupun netral. Oleh karena itu, terdapat dua kategori yang akan menjadi basis dalam pembuatan decision tree, yaitu polaritas sentimen dan subjektivitas/objektivitas sebagai berikut.

Analisis sentimen melibatkan penentuan polaritas dari teks atau data yang diamati. Sentimen dapat bersifat positif, negatif, atau netral. Dalam konteks marketplace, pemahaman terhadap polaritas ini memberikan wawasan terhadap bagaimana masyarakat menilai layanan, produk, dan pengalaman

berbelanja.

B. Penerapan Decision Tree dalam Analisis Sentimen

Berdasarkan sub-bab sebelumnya, telah disebutkan bahwa terdapat dua kategori yang akan menjadi basis pembuatan decision tree untuk menganalisis sentimen, yaitu analisis berdasarkan polaritasnya dan subjektivitas/objektivitasnya.



Gambar 3.1 Pohon Analisis Polaritas

Pohon di atas adalah representasi visual dari serangkaian pertanyaan dan keputusan yang diambil untuk mengklasifikasikan sentimen berdasarkan polaritasnya. Jika pada data review orang-orang terhadap marketplace terkait mengandung teks dengan kata kunci positif (“bagus”, “fantastis”, “luar biasa”, dan sejenisnya), maka review diklasifikasikan sebagai review positif. Sebaliknya, jika review mengandung kata kunci negatif (“buruk”, “kecewa”, dan sejenisnya), maka review akan diklasifikasikan sebagai review negatif.

C. Metode Penelitian

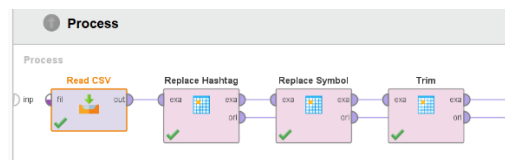
Dalam melakukan penelitian analisis sentiment, terdapat beberapa tahapan yang harus dilakukan, sebagai berikut. Dalam makalah ini, digunakan aplikasi RapidMiner, dimana RapidMiner adalah aplikasi yang digunakan dalam pembelajaran ilmu *data mining*.

1. Pengumpulan Dataset

Dataset dapat diperoleh melalui ulasan yang terdapat di berbagai sosial media, contohnya Twitter. Pada Twitter terdapat API yang membantu untuk mencari kata kunci tertentu yang relevan dengan penelitian.

2. Data Cleansing

Setelah dataset terkumpul, dapat dilakukan data cleansing.



Gambar 3.2 Data Cleansing Menggunakan RapidMiner

Proses data cleansing disesuaikan dengan dataset masing-masing, karena pada dataset yang diambil dalam makalah ini hanya terdapat hashtag dan symbol bebas, maka hanya perlu dilakukan replace untuk Hashtag dan juga Symbol. Replace hashtag

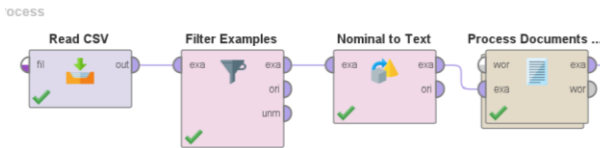
digunakan untuk menghapus hashtag yang terdapat pada dataset. Replace Symbol digunakan untuk menghapus simbol-simbol yang terdapat dalam dataset.

3. Pelabelan Data dan Ekstraksi Fitur

sentimen	text
negatif	maaf ulasan drone bagus sliipp mantap konek wifi drone tolong petunjuknya kecewa
postif	lumayan dronanya baterainya sebentar kejauhan drone lost kontrol hilang ngloyor jalan pesan
postif	barang bagus kurir ramah lucu lecut bagus pokok barang thank lazada ngga nyesel pokok belu lazada
negatif	kecewa banget pesen barang udah nunggu hasil zonk pokoknya uang
negatif	penipuan banget
negatif	kecewa paking bagus buka barang rusak setik berpungsi kembalikan kali kecewakan pengitman barang
postif	dapetnya susah setting coneck lumayan tinggal terbang
postif	semoga bermanfaat anak anak senang
postif	pengiriman barangnya berfungsi mantull mantul
postif	lazada best
postif	semogadrone pesanaman coba gejala
postif	makasih lazada
postif	mantul
negatif	barang rusak tolong ganti
negatif	barang rebang kamera nyalah sehid

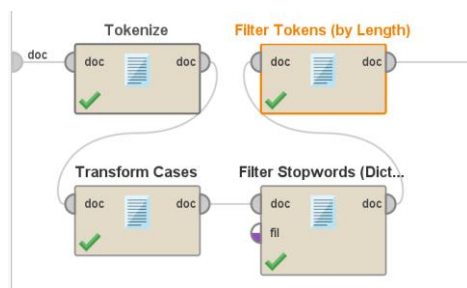
Gambar 3.3 Contoh Pelabelan Data

Pelabelan data secara manual hanya dilakukan untuk sebagian data saja, untuk sisanya akan digunakan pelabelan otomatis menggunakan algoritma decision tree.



Gambar 3.4 Ekstraksi Fitur

Pertama-tama data yang sudah dilabeli secara manual di filter terlebih dahulu. Lalu, karena tipe data tersebut masih dalam bentuk polynomial, maka harus diubah terlebih dahulu ke dalam bentuk teks. Selanjutnya, dilakukan tokenisasi dan beberapa proses filtering seperti *filter dictionary* dan *filter token by length* yang digunakan untuk menghapus teks yang hanya memiliki sedikit karakter. Proses tokenisasi dan filtering ini di mapping ke dalam satu bagian yang dinamakan *process documents*. Berikut adalah isi dari *process documents*, dimana terdapat tokenisasi dan berbagai proses filtering.

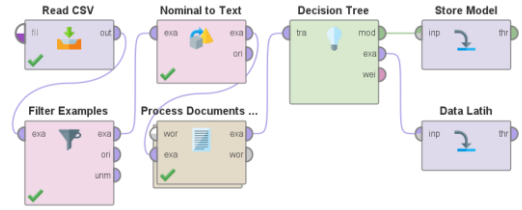


Gambar 3.5 Isi dari *Process Documents*

Pada tahap ekstraksi fitur, teks diubah menjadi representasi numerik yang dapat diproses oleh algoritma machine learning. Ini melibatkan ekstraksi

fitur, seperti TF-IDF (*Term Frequency-Inverse Document Frequency*) atau word embeddings.

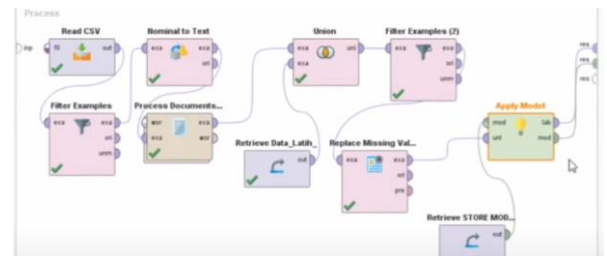
4. Pelatihan Model Decision Tree



Gambar 3.6 Pelatihan Model Decision Tree

Pada tahap ini, dilakukan pelatihan model decision tree, dimana algoritma decision tree diterapkan pada data pelatihan. Decision tree adalah struktur pohon yang dibentuk oleh aturan keputusan berdasarkan fitur-fitur data pelatihan. Algoritma decision tree secara iteratif memilih fitur terbaik untuk membagi data berdasarkan kriteria tertentu. Proses ini berlanjut hingga tercapai kondisi berhenti yang ditentukan (misalnya, kedalaman maksimum pohon atau jumlah sampel minimum di suatu cabang).

5. Pengaplikasian Model

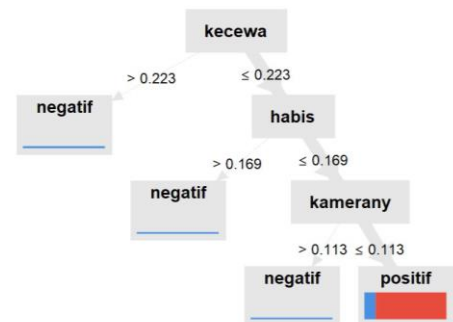


Gambar 3.7 Pengaplikasian Model

Tahap ini adalah tahap dimana model yang telah dilatih sebelumnya akan dihubungkan dengan data baru untuk melakukan prediksi atau klasifikasi. Operator "Apply Model" dikonfigurasi dengan memilih model yang ingin digunakan dari daftar model yang telah dilatih sebelumnya.

6. Analisis Hasil

Berikut adalah hasil yang didapatkan setelah melakukan pengaplikasian model tersebut.



Gambar 3.8 Hasil Model *Decision Tree*

Beberapa model memberikan probabilitas untuk setiap kelas sentimen. Misalnya, sebuah model

bisa memberikan prediksi "Positif" dengan probabilitas 0.8 dan "Negatif" dengan probabilitas 0.2. Ini memberikan pemahaman lebih mendalam tentang keyakinan model terhadap prediksi tertentu.

C. Optimalisasi Decision Tree

Aplikasi Decision Tree dalam analisis sentimen publik terhadap marketplace di Indonesia dapat dioptimalkan dengan sejumlah strategi untuk meningkatkan akurasi dan efisiensi hasil analisis sebagai berikut.

1. Pemilihan Fitur yang Sesuai

Dalam hal ini fitur yang dipilih harus relevan dengan sentimen yang dianalisis. Fitur-fitur yang dapat digunakan antara lain:

- a. Hashtag
- b. Emoji
- c. Frasa yang sering digunakan untuk menyatakan sentimen tertentu, dan lain-lain.

2. Pemilihan Metode Pelabelan Data

Metode pelabelan data dalam melabeli sentimen harus akurat dan konsisten, contohnya dalam melabeli sentiment perlu adanya kesepakatan antar pakar untuk menentukan kata-kata atau frasa yang termasuk kategori positif, negative, atau netral. Dalam hal ini, dapat digunakan Teknik seperti Crowdsourcing dapat membantu bisnis untuk mendapatkan hasil yang lebih baik karena dapat menjangkau kumpulan orang yang lebih luas dan beragam.

3. Validasi Silang (*Cross-Validation*)

Metode validasi silang dapat digunakan sebagai cara untuk lebih objektif menilai kinerja model dan mengurangi potensi *overfitting*. Melakukan pembagian dataset menjadi subset pelatihan dan pengujian secara bergantian dapat memberikan evaluasi yang lebih akurat terhadap sejauh mana model berperforma dengan baik. Untuk menghasilkan hasil yang lebih akurat dan efisien, dapat digunakan metode *K-fold Cross Validation*.

Metode ini membagi data menjadi k bagian sama besar, lalu menggunakan k-1 bagian untuk pelatihan dan 1 bagian untuk pengujian. Proses ini dilakukan sebanyak k kali, dengan setiap bagian data pelatihan dan data pengujian berbeda-beda. Hasil akurasi dari k kali proses tersebut kemudian dirata-rata untuk mendapatkan akurasi akhir model *machine learning*. Metode ini dinilai lebih akurat dan efisien karena menggunakan lebih banyak data untuk pelatihan dan pengujian, serta dapat dilakukan secara parallel sehingga menghemat waktu dan memberikan gambaran yang lebih lengkap terkait kinerja model *machine learning*. Gambar dibawah ini menunjukkan contoh algoritma *K-fold cross validation* menggunakan Python dan *library* scikit-learn.

```
#Importing required libraries
from sklearn.datasets import load_marketplace_sentiment
import pandas as pd
from sklearn.model_selection import KFold
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

#Loading the dataset
data = load_marketplace_sentiment(as_frame = True)
df = data.frame
X = df.iloc[:, :-1]
y = df.iloc[:, -1]

#Implementing cross validation
k = 5
kf = KFold(n_splits=k, random_state=None)
model = LogisticRegression(solver='liblinear')

acc_score = []

for train_index , test_index in kf.split(X):
    X_train , X_test = X.iloc[train_index, :], X.iloc[test_index, :]
    y_train , y_test = y[train_index] , y[test_index]

    model.fit(X_train, y_train)
    pred_values = model.predict(X_test)

    acc = accuracy_score(pred_values , y_test)
    acc_score.append(acc)

avg_acc_score = sum(acc_score)/k

print('accuracy of each fold - {}'.format(acc_score))
print('Avg accuracy : {}'.format(avg_acc_score))
```

Gambar 3.5 Contoh *K-fold Cross Validation* untuk k=5

4. Pemilihan Parameter Algoritma

Penyesuaian *hyperparameter* dengan optimal pada model *Decision Tree*, seperti mengatur kedalaman maksimum atau jumlah minimum sampel untuk split dapat meningkatkan akurasi model dan mengurangi risiko terjadinya *overfitting* atau *underfitting*.

IV. KESIMPULAN

Dari makalah ini, dapat disimpulkan bahwa penerapan ilmu matematika diskrit dalam kehidupan memberikan kontribusi yang signifikan. Berbagai konsep ilmu matematika diskrit ternyata dapat diterapkan secara praktis dalam memecahkan suatu masalah di kehidupan sehari-hari. Seperti yang dibahas dalam makalah ini, penerapan pohon keputusan (*decision tree*) telah mampu memberikan pemahaman yang lebih mendalam terhadap pola sentimen masyarakat terhadap *platform marketplace*. Hal ini dapat membantu dalam menangani permasalahan di bidang bisnis dan memberikan wawasan yang berharga kepada para pelaku bisnis untuk meningkatkan layanan mereka. Dengan menerapkan pendekatan ini, harapannya adalah memberikan pandangan yang lebih holistic bagi pelaku industry, regulator, dan pemangku kepentingan lainnya untuk mengoptimalkan layanan dan pengalaman pengguna di dunia *marketplace*.

V. UCAPAN TERIMA KASIH

Dalam kesempatan ini, penulis ingin menyampaikan ucapan terima kasih kepada semua pihak yang telah turut serta mendukung dan berkontribusi dalam penyelesaian makalah ini. Terima kasih kepada:

1. Tuhan Yang Maha Esa, berkat rahmat dan hidayah-Nya, makalah ini dapat terselesaikan dengan lancar
2. Keluarga dan teman yang telah memberikan dukungan kepada penulis selama pengerjaan makalah ini
3. Dr. Ir. Rinaldi, M.T., selaku dosen mata kuliah IF2120 Matematika Diskrit yang telah memberikan ilmu-ilmu dan materi dasar yang membantu dalam penulisan makalah ini

Ucapan terima kasih yang tulus kepada semua pihak yang telah turut serta dalam perjalanan penulisan makalah ini.

REFERENSI

- [1] R. Munir, 'Homepage Rinaldi Munir'. <file:///C:/Users/Auralea%20A%20S/Documents/SEMESTER%203/MATDIS/Graf-Bagian1-2023.pdf> [Diakses pada 8 Desember 2023]
- [2] R. Munir, 'Homepage Rinaldi Munir'. <file:///C:/Users/Auralea%20A%20S/Documents/SEMESTER%203/MATDIS/Pohon-Bag1-2023.pdf> [Diakses pada 7 Desember 2023]
- [3] R. Munir, 'Homepage Rinaldi Munir'. <file:///C:/Users/Auralea%20A%20S/Documents/SEMESTER%203/MATDIS/Pohon-Bag2-2023.pdf> [Diakses pada 7 Desember 2023]
- [4] Sekawanmedia.co.id, 'Kenali apa itu marketplace beserta jenis dan contoh penerapannya'. <https://www.sekawanmedia.co.id/blog/pengertian-marketplace/> [Diakses pada 8 Desember 2023]
- [5] Researchgate.net, 'Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014'. https://www.researchgate.net/publication/309543552_Analisis_Sentimen_Masyarakat_terhadap_Calon_Presiden_Indonesia_2014_berdasarkan_Opini_dari_Twitter_Menggunakan_Metode_Naive_Bayes_Classifier [Diakses pada 8 Desember 2023]

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 8 Desember 2023



Auralea Alvinia Syaikha
13522148

