



PART 8

TEXT MINING & FORECASTING

VERSI 2025

Pembahasan

- Studi Kasus: Text Mining & Forecasting
Prediksi Minat Konsumen terhadap Produk
Machine Learning for Business Intelligence



Tujuan Pembelajaran

- 1) Memahami proses text mining untuk analisis minat konsumen
- 2) Melakukan transformasi opini → metrik BI
- 3) Menganalisis sentimen dan topik utama
- 4) Melakukan forecasting minat konsumen
- 5) Menyusun strategi peningkatan minat berdasarkan hasil prediksi
- 6) Mengintegrasikan hasil analisis ke dashboard BI

Konteks Bisnis

- Konsumen mengeluarkan opini di platform digital
- Perusahaan perlu menangkap sinyal minat secara real-time
- Text Mining mengungkap persepsi
- Forecasting Machine Learning memprediksi perubahan minat



Data Collection

- Sumber data: Shopee, Tokopedia, Instagram, Twitter/X
- Struktur data: teks ulasan, rating, tanggal, lokasi, engagement
- Tantangan: spam, bot, teks campuran, noise besar



Arsitektur Data Flow

User → Platform → Scraping/API →
Preprocessing → Text Mining → Sentiment
Score → Time-series → Forecasting Model → BI
Dashboard



Preprocessing

- Cleaning: URL, hashtag, mention, tanda baca
- Case folding: lowercasing
- Normalization: slang → formal, typo correction
- Tokenization
- Stopword removal
- Lemmatization/Stemming



Contoh Before–After Preprocessing

Teks asli → “Produk ini bagusss bgt!!!
pengiriman cepatttt”

Setelah preprocessing → “produk bagus
kirim cepat”



Feature Extraction

- TF-IDF
- Bag-of-Words
- Word Embedding (Word2Vec, FastText)
- BERT Embedding



Perbandingan Representasi Teks

- Bag-of-Words: sederhana
- TF-IDF: menekankan kata penting
- Word2Vec: memahami konteks
- BERT: memahami makna kalimat



Sentiment Analysis

- Tujuan: mengukur persepsi positif/negatif
- Algoritma: Logistic Regression, SVM, Random Forest, BERT
- Output: skor sentimen per periode



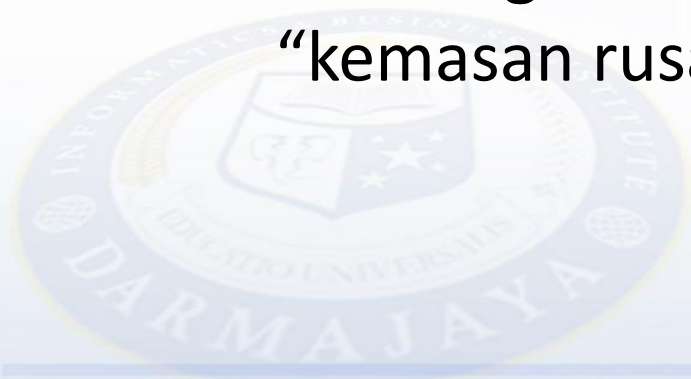
Contoh Output Sentimen

Positif: 73%

Negatif: 12%

Netral: 15%

Trending kata negatif: “bau”, “cepat habis”,
“kemasan rusak”



Evaluasi Model Sentimen

- Accuracy
- F1-score
- Confusion Matrix



Topic Modelling

- LDA (Latent Dirichlet Allocation)
- Output: tema ulasan → keluhan, pujian, preferensi



Contoh Topic Modeling Produk

Topic 1: Aroma → “bau”, “menyengat”, “kimia”

Topic 2: Pengemasan → “bocor”, “koyak”, “bubble wrap”

Topic 3: Harga → “murah”, “worth it”, “diskon”



Transformasi ke BI

Konversi ulasan → metrik BI

- Sentiment score mingguan
- Volume ulasan
- Keyword frequency
- Topic severity index

Dataset time-series

Date | Sentiment Score | Review Count | Keyword:
“murah” | Keyword: “rusak”

Forecasting Models

1. ARIMA (Autoregressive Integrated Moving Average)

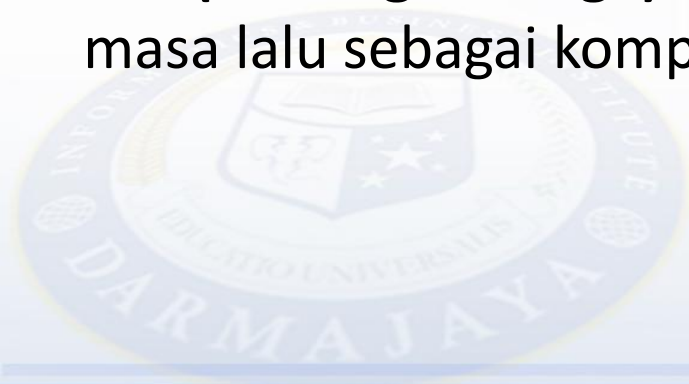
ARIMA adalah model statistik klasik untuk melakukan *time-series forecasting* yang bekerja berdasarkan pola masa lalu.

Komponen model:

AR (Autoregressive): menggunakan nilai masa lalu untuk memprediksi nilai saat ini.

I (Integrated): melakukan differencing untuk membuat data menjadi stasioner.

MA (Moving Average): menggunakan kesalahan prediksi masa lalu sebagai komponen koreksi.



ARIMA

Kelebihan:

Cocok untuk pola yang *linear*.

Baik untuk dataset yang tidak terlalu besar.

Mudah diinterpretasikan.

Kekurangan:

Tidak optimal untuk pola yang sangat fluktuatif atau *non-linear*.

Mebutuhkan data stasioner (pola stabil).

Relevansi untuk kasus minat konsumen:

Jika sentimen konsumen secara historis menunjukkan pola musiman (contoh: meningkat saat Ramadhan), ARIMA dapat mengcapture pola tersebut dengan baik.

Forecasting Models

2. Prophet (Facebook Prophet)

Model *forecasting* yang dirancang untuk data bisnis harian/bulanan dengan pola musiman kuat dan adanya event atau hari libur.

Karakteristik utama:

Menangani tren linear dan non-linear.

Menangani *outliers* dan data hilang.

Mendukung event khusus seperti promosi atau launching produk.

Hasil prediksi mudah dipahami melalui komponen:

Tren, Musim, Event

Relevansi untuk studi kasus:

Prediksi minat konsumen sering dipengaruhi event seperti diskon 11.11, Ramadhan, atau kampanye influencer. Prophet unggul untuk pola seperti ini.

Forecasting Models

3. LSTM (Long Short-Term Memory Neural Network)

Model deep learning yang dirancang menangkap **pola jangka panjang** dan relasi antar waktu (*long-term dependencies*).

Kekuatan LSTM:

Menangani data non-linear yang kompleks.

Dapat mempelajari pola yang berubah-ubah tanpa harus stasioner.

Sangat efektif untuk *sentiment score time series* yang berfluktuasi.

Kekurangan:

Membutuhkan dataset cukup besar.

Tidak se-transparan model statistik (black-box model).

Membutuhkan waktu pelatihan lebih lama.

Relevansi untuk BI:

Jika sentimen berasal dari teks dinamis (trending topic, isu viral, komentar harian), LSTM biasanya menghasilkan prediksi paling akurat.

Forecasting Pipeline

Sentiment Score Weekly → Train-Test Split
→ Forecasting → Evaluate (RMSE/MAPE)

Train-Test Split

Dataset dibagi menjadi:

- **Training set:** 80% periode awal waktu
- **Testing set:** 20% periode akhir untuk evaluasi

Tujuan: mengevaluasi apakah model bisa memprediksi masa depan dengan benar.

Evaluation

Evaluasi dilakukan menggunakan data asli pada testing set dan membandingkannya dengan prediksi.

Metrik:

RMSE (Root Mean Square Error) → semakin kecil semakin baik

MAE (Mean Absolute Error)

MAPE (Mean Absolute Percentage Error) → persentase error terhadap nilai asli

Trend Accuracy → apakah model menangkap arah tren

Contoh Grafik Forecasting:
Prediksi sentimen 3 bulan
ke depan

→ tren naik 10–15%

Interpretasi Hasil

- Konsumen semakin positif
- Kata “murah” naik 47%
- Kata “bau” turun 18%
- Risiko churn rendah

Strategi Bisnis (Produk)

- Reformulasi aroma
- Perbaiki pengemasan
- Penambahan varian produk

Strategi Bisnis (Marketing)

- Personalisasi kampanye
- Ads berbasis keyword
- Influencer endorsement

Strategi Bisnis (Operasional)

- Perencanaan stok berdasarkan forecasting
- Promosi musiman
- Distribusi inventori berdasarkan minat wilayah

Integrasi dalam BI

- Data warehouse
- ML Layer (sentiment + forecasting)
- Dashboard Layer

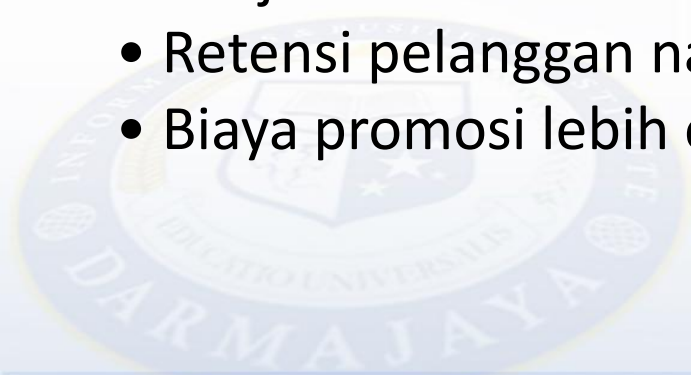


Produk: Serum X

- 15.000 ulasan
- Sentimen positif 73%
- Topic modeling: aroma & kemasan
- Prediksi minat naik 12%
- Strategi: perbaikan aroma + kampanye target

Dampak Bisnis

- Penjualan naik 15% dalam 4 bulan
- Retensi pelanggan naik
- Biaya promosi lebih efisien



Akhir bagian Part 8

