

Analisis Kritis Output Phase 3: Fitur Konten & Perilaku

Output dari Phase 3 adalah titik balik dalam proyek kita. Kita telah berhasil mengubah data interaksi dan teks mentah menjadi serangkaian fitur numerik yang kaya makna. Mari kita bedah hasilnya secara mendalam.

1. Evaluasi Output: Apakah Hasil Ini Bagus?

Jawaban Singkat: Ya, hasil ini sangat bagus dan sangat menjanjikan.

Penjelasan Rinci:

a. DataFrame `final_features.csv`

Ini adalah *output* utama dan paling berharga dari tahap ini. Keberhasilannya terletak pada **fusi fitur multi-dimensi**:

- **Fitur Jaringan (dari Phase 2):** pagerank, betweenness, in_degree, out_degree. Ini memberitahu kita **posisi dan peran** sebuah akun dalam jaringan.
- **Fitur Konten (Baru):** narrative_similarity. Ini memberitahu kita **seberapa seragam** konten yang diproduksi sebuah akun dibandingkan dengan komunitasnya.
- **Fitur Perilaku (Baru):** tweet_frequency, reply_ratio. Ini memberitahu kita **seberapa aktif dan bagaimana pola** sebuah akun berinteraksi.

Dengan menggabungkan ketiga jenis informasi ini, kita telah menciptakan sebuah "sidik jari" digital yang jauh lebih komprehensif untuk setiap akun. Model machine learning di Phase 4 akan jauh lebih kuat karena bisa belajar dari berbagai sinyal ini secara bersamaan.

b. Visualisasi Scatter Plot: Aktivitas vs. Keseragaman Narasi

Visualisasi ini adalah bukti keberhasilan dari rekayasa fitur kita. Grafik ini dengan jelas memetakan perilaku akun dan memungkinkan kita untuk mengidentifikasi kluster yang mencurigakan secara visual:

- **Sumbu Y (Skor Kesamaan Narasi):** Semakin tinggi, semakin "mirip-mirip" atau seragam tweet dari sebuah akun dengan grupnya.
- **Sumbu X (Aktivitas Menyebar / Out-Degree):** Semakin ke kanan, semakin aktif sebuah akun dalam menyebarkan pesan (me-reply).
- **Ukuran Titik (PageRank):** Semakin besar, semakin berpengaruh akun tersebut di dalam jaringan.

Interpretasi Kunci:

Area kanan atas pada grafik adalah "Zona Bahaya Buzzer". Akun-akun yang berada di sini menunjukkan dua karakteristik utama buzzer:

1. Mereka **sangat aktif** menyebarkan pesan (Out-Degree tinggi).
2. Narasi yang mereka sebarakan **sangat seragam** (Skor Similaritas tinggi).

Fakta bahwa kita bisa melihat titik-titik (akun) mengelompok di area ini menunjukkan bahwa fitur yang kita buat berhasil menangkap sinyal yang kita cari.

2. Kesesuaian & Keberhasilan Algoritma

a. Apakah Algoritma Ini Cocok?

Ya, sangat cocok.

- **TF-IDF & Cosine Similarity:** Algoritma ini dipilih secara sengaja **bukan** untuk memahami makna (semantik) dari sebuah teks, melainkan untuk mengukur **kesamaan leksikal (kata-kata yang digunakan)**. Untuk tujuan mendeteksi narasi seragam atau *copy-paste*, pendekatan ini jauh lebih efisien dan tepat sasaran daripada model bahasa yang kompleks. Tujuannya adalah mendeteksi pola, bukan pemahaman.
- **Rekayasa Fitur Perilaku:** *tweet_frequency* dan *reply_ratio* adalah metrik sederhana yang secara langsung mengkuantifikasi anomali perilaku. Akun organik cenderung memiliki frekuensi yang lebih wajar, sementara akun terkoordinasi seringkali menunjukkan aktivitas yang sangat tinggi dan terfokus pada topik tertentu.

b. Bagaimana Algoritma Ini Berhasil?

Keberhasilannya terletak pada kemampuannya untuk **mengubah konsep abstrak menjadi angka terukur**:

- Konsep "narasi seragam" berhasil diubah menjadi fitur numerik *narrative_similarity*.
- Konsep "aktivitas tidak wajar" berhasil diubah menjadi fitur *tweet_frequency* dan *reply_ratio*.

Dengan mengubah konsep-konsep kualitatif ini menjadi fitur kuantitatif, kita telah membuka jalan bagi algoritma *machine learning* untuk mengenali pola-pola ini dalam skala besar.

3. Permasalahan & Keterbatasan yang Ditemukan

Analisis yang baik juga harus mengakui adanya keterbatasan. Berikut adalah beberapa masalah potensial pada algoritma dan dataset kita:

a. Permasalahan pada Algoritma:

1. **Keterbatasan TF-IDF:** Seperti yang dijelaskan, TF-IDF tidak memahami sinonim atau konteks. Dua kalimat dengan makna yang sama persis tetapi menggunakan kata-kata yang berbeda akan dianggap tidak mirip. Ini berarti **buzzer cerdas yang mampu memparafrasekan narasinya** mungkin akan mendapat skor similaritas yang lebih rendah dan bisa lolos dari deteksi.
2. **Asumsi Centroid Komunitas:** Metode kita mengasumsikan setiap komunitas memiliki satu "narasi utama" (centroid). Jika sebuah komunitas besar memiliki beberapa sub-topik

yang valid, anggota yang fokus pada sub-topik tersebut bisa jadi keliru mendapat skor similaritas yang rendah.

b. Permasalahan pada Dataset:

1. **Kekurangan Fitur Krusial (Umur Akun):** Dataset yang kita kumpulkan dari *scraper* tidak memiliki informasi kapan sebuah akun dibuat (`user.created_at`). Ini adalah sebuah keterbatasan signifikan, karena **umur akun** adalah salah satu prediktor terkuat untuk mendeteksi "peternakan bot" (bot farm) yang seringkali dibuat secara massal dalam waktu yang berdekatan.
2. **Data yang Jarang (Sparsity):** Banyak pengguna dalam dataset mungkin hanya muncul dengan satu atau dua tweet. Untuk pengguna seperti ini, fitur `tweet_frequency` dan `narrative_similarity` menjadi kurang andal. Fitur kita paling kuat untuk akun-akun yang lebih aktif.
3. **Representasi Sesaat:** Dataset kita adalah potret dari beberapa isu dalam rentang waktu tertentu. Kita tidak bisa melihat perilaku jangka panjang sebuah akun. Aktor canggih bisa saja bertindak normal 99% waktunya dan hanya aktif sebagai buzzer saat ada "proyek". Model kita kemungkinan besar akan mengklasifikasikan akun seperti ini sebagai pengguna normal.