

SISTEM PENDETEKSIAN JARINGAN BOT PADA PLATFORM TIKTOK MENGUNAKAN *CANOPY CLUSTERING* DAN ALGORITMA KRUSKAL

Intan Febyola Putri Dwina Sidabutar

ABSTRAK

Maraknya aktivitas bot terkoordinasi di platform TikTok menimbulkan ancaman manipulasi opini publik yang tidak dapat dideteksi menggunakan metode konvensional berbasis analisis akun individual. Pendekatan *supervised learning* memiliki keterbatasan dalam konteks penelitian ini karena membutuhkan data berlabel yang sulit diperoleh pada TikTok, serta pada beberapa model modern dapat memerlukan sumber daya komputasi tinggi. Penelitian ini bertujuan merancang sistem deteksi jaringan bot pada TikTok berbasis analisis graf secara *unsupervised* menggunakan integrasi *Canopy Clustering* dan algoritma Kruskal (*Minimum Spanning Tree*). Penelitian menggunakan metodologi *Design Research Methodology* (DRM) dan *SDLC Prototyping*. Data komentar publik TikTok dikumpulkan menggunakan TikTokApi dengan empat variasi *dataset* (100–1.000 komentar). Sistem dibangun sebagai *pipeline* enam modul berbasis FastAPI dan Cytoscape.js. Komentar direpresentasikan menggunakan TF-IDF, lalu *Canopy Clustering* berbasis HNSW diterapkan sebagai *soft pre-filtering* sebelum graf berbobot dibangun dari relasi *mention*, *reply*, kemiripan konten, dan *co-thread*. Kruskal MST kemudian membentuk klaster komunitas dengan memotong *edge* lemah berdasarkan median bobot MST per komponen. Setiap klaster dinilai menggunakan skor komposit dari lima indikator: *cluster density*, *content repetition*, *temporal burst*, *high frequency*, dan *account age*. Evaluasi dilakukan melalui pengujian fungsional, integrasi, performa, serta metrik *Modularity*, *Silhouette Score*, dan *Conductance*.

Kata Kunci : Analisis Graf, *Canopy Clustering*, Deteksi Bot, Kruskal MST, TikTok

A Bot Network Detection System for the TikTok Platform Using Canopy Clustering and Kruskal's Algorithm

Intan Febyola Putri Dwina Sidabutar

ABSTRACT

The rise of coordinated bot activity on TikTok poses a serious threat of public opinion manipulation that cannot be detected using conventional individual-account analysis methods. Supervised learning approaches have limitations in the context of this study because they require labeled data, which is difficult to obtain on TikTok, and some modern models may also demand high computational resources. This study aims to design an unsupervised Graph-based bot network detection system for TikTok by integrating Canopy Clustering and Kruskal's algorithm (Minimum Spanning Tree) into a single modular pipeline. This study employs Design Research Methodology (DRM) combined with an SDLC Prototyping model. Data were collected from public TikTok video comments using TikTokApi across four dataset sizes (100–1,000 comments). The system was built as a six-module pipeline using FastAPI and Cytoscape.js. Comments are represented as TF-IDF vectors, then Canopy Clustering using HNSW is applied as soft pre-filtering before a weighted graph is constructed from mention, reply, content similarity, and co-thread relations. Kruskal MST forms community clusters by cutting weak edges based on the median MST edge weight per component. Each cluster is scored using five indicators: cluster density, content repetition, temporal burst, high frequency, and account age. Evaluation is conducted through functional, integration, and performance testing, as well as Modularity, Silhouette Score, and Conductance metrics.

Keywords: *Bot Detection, Canopy Clustering, Graph Analysis, Kruskal MST, TikTok.*