

RISET INFORMATIKA

PROBLEM FORMULATION

“Identifikasi Marka Linguistik Pembeda Teks Generatif AI dan Teks Manusia Menggunakan SHAP (SHapley Additive exPlanations)”



Dosen Pendamping:

Dr. Basuki Rahmat, S.Si. MT

Disusun Oleh:

Muhammad Hidayat Nurwahid (22081010300)

Universitas Pembangunan Nasional “Veteran” Jawa Timur

Surabaya

2025

PROBLEM FORMULATION

1.1.Latar Belakang

Kehadiran model bahasa generatif seperti model GPT dan Gemini telah mengubah pendekatan dalam pembuatan konten secara fundamental (Saha, 2025). Teknologi ini mampu menghasilkan teks dengan kualitas yang mirip dengan tulisan manusia (Eapen dkk., 2023). Sehingga, mengaburkan batas antara karya orisinal dan hasil rekayasa mesin. Kemampuan ini, meski inovatif, membawa implikasi serius yang menuntut perhatian.

Implikasinya terasa di berbagai sektor krusial. Dalam jurnalisme, risiko penyebaran misinformasi semakin tinggi. Di sektor komersial, rekayasa ulasan produk dapat merusak kepercayaan konsumen. Namun, tantangan terbesarnya mungkin terjadi di dunia akademik, di mana integritas karya ilmiah terancam oleh praktik plagiarisme yang semakin canggih dan sulit dilacak (Eke, 2023).

Sebagai respons, berbagai alat deteksi konten AI mulai bermunculan. Sayangnya, mayoritas alat ini beroperasi layaknya "kotak hitam" (*black box*). Alat-alat tersebut hanya memberikan vonis akhir "AI" atau "Manusia" tanpa disertai penjelasan yang dapat diverifikasi. Pendekatan ini melahirkan dilema baru: ketika seorang mahasiswa dituduh menggunakan AI, tidak ada bukti konkret yang bisa disajikan selain label dari sebuah sistem yang tidak transparan. Ketiadaan akuntabilitas ini membuat alat deteksi konvensional tidak dapat diandalkan dalam pengambilan keputusan yang adil.

Di sinilah letak urgensi penelitian ini. Masalah mendasarnya bukanlah ketiadaan detektor, melainkan ketiadaan penjelasan. Kita perlu beralih dari pertanyaan apakah sebuah teks ditulis oleh AI, ke pertanyaan mengapa sebuah teks dianggap demikian. Untuk itu, penelitian ini akan memanfaatkan *Explainable AI* (XAI) melalui metode SHAP (*SHapley Additive exPlanations*). Tujuannya adalah untuk "membongkar" kotak hitam tersebut, mengidentifikasi "sidik jari" atau marka linguistik yang secara konsisten menjadi pembeda antara teks AI dan tulisan manusia, dan pada akhirnya, menciptakan dasar untuk sistem deteksi yang transparan dan dapat dipertanggungjawabkan.

1.2.Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, penelitian ini berupaya menjawab pertanyaan-pertanyaan berikut:

1. Marka linguistik apa saja (meliputi aspek leksikal, sintaksis, dan stilistik) yang menjadi pembeda paling signifikan antara teks buatan *Generative AI* dan teks tulisan manusia?
2. Bagaimana metode SHAP dapat digunakan untuk mengidentifikasi dan memvisualisasikan kontribusi setiap marka linguistik terhadap keputusan model dalam mengklasifikasikan sebuah teks?
3. Sejauh mana konsistensi marka linguistik tersebut dapat diandalkan pada genre teks yang berbeda, seperti teks berita dan esai kreatif?