

RINGKASAN

PENGOLAHAN DATA DENGAN PYTHON UNTUK MACHINE LEARNING PADA APLIKASI SIMARGA INSIGHT (STUDI KASUS DPU BINA MARGA PROVINSI JAWA TIMUR) Ryan Pratama Putra, NIM E31232406, Tahun 2025, Teknologi Informasi, Politeknik Negeri Jember Taufiq Rizaldi, S.ST., M.T., (Pembimbing Magang), Ario Bagus Nugroho, S.Kom., M.M., (Pembimbing Lapangan).

Kebutuhan akan informasi yang cepat dan akurat mendorong instansi pemerintah untuk memanfaatkan data yang berasal dari media daring sebagai sumber evaluasi kerja. Berita *online* dan komentar masyarakat yang memuat opini dan pemberitahuan publik mengenai pembangunan infrastruktur memiliki nilai penting apabila dapat dianalisis secara sistematis. Oleh karena itu, kegiatan magang ini berfokus pada penerapan analisis sentimen pada aplikasi *Simarga Insight* di lingkungan DPU Bina Marga Provinsi Jawa Timur.

Berbagai portal berita *online* dan komentar youtube yang berkaitan dengan DPU Bina Marga Provinsi Jawa Timur. Data tersebut direkap dalam file excel dan csv yang berisi informasi judul, tautan sumber, serta kutipan. Selanjutnya, data diproses menggunakan bahasa pemrograman *python* melalui tahapan *pre-processing* teks, seperti pembersihan karakter, pengubahan huruf, untuk mempersiapkan data sebelum dianalisis.

Tahap berikutnya Data berita yang berkaitan dengan Dinas Pekerjaan Umum Provinsi Jawa Timur awalnya dikumpulkan secara manual dari beberapa situs berita online. File Excel dan CSV yang berisi judul, tautan sumber, dan cuplikan berita digunakan untuk mengorganisir data. Data tersebut kemudian diproses menggunakan bahasa pemrograman *python* melalui proses pra-pemrosesan teks, seperti mengubah huruf menjadi huruf kecil dan membersihkan karakter.

Langkah berikutnya adalah penggunaan analisis sentimen dengan model *Transformer* yang telah dilatih sebelumnya (Indonesia RoBERTa) tanpa *retraining*. Model ini digunakan untuk mengkategorikan opini berita ke dalam kelompok

positif, negatif, dan netral. Untuk meningkatkan efisiensi, proses analisis dilakukan secara bertahap, dan hasil prediksi disimpan kembali ke dalam dataset. Berdasarkan temuan magang, aplikasi SIMARGA *Insight* memiliki kemampuan untuk mengorganisir data berita dan menghasilkan hasil analisis sentimen yang informatif. Lembaga dapat menggunakan data ini sebagai bukti untuk mendukung pengambilan keputusan berbasis data dan memantau opini publik.