

**PENGUNAAN ALGORITMA RANDOM FOREST DALAM MENILAI
SENTIMEN PENGGUNA TWITTER TERKAIT ISU PERCERAIAN**

RISET INFORMATIKA



Oleh:

MUHAMMAD RAFLI AULIA ROJANI LUTFI

20081010061

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”

JAWA TIMUR

2023

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perceraian adalah isu yang kompleks dan sensitif dalam masyarakat Indonesia. Menurut [1] kasus perceraian di Indonesia dipicu oleh berbagai macam faktor, antara lain perselisihan yang terus berulang, kekerasan dalam rumah tangga, tidak adanya keharmonisan antara suami dan istri, permasalahan ekonomi, salah satu pihak pergi meninggalkan pihak lainnya, kecemburuan berlebih terhadap pasangan, dan adanya campur tangan pihak ketiga. Berdasarkan data dari laporan Statistik Indonesia 2023, jumlah kasus perceraian di Indonesia diperkirakan mencapai 516.334 pada tahun 2022. Angka ini mengalami peningkatan sebesar 15,31% dibandingkan dengan tahun 2021 yang mencatat 447.743 kasus. Angka tersebut merupakan jumlah kasus perceraian tertinggi dalam enam tahun terakhir di Indonesia.

Dalam era digital saat ini, media sosial seperti Twitter menjadi platform yang populer untuk berbagi pendapat dan pengalaman terkait perceraian. Hal ini dapat menimbulkan beragam sentiment. Sentimen masyarakat yang beragam dapat menjadi tolak ukur seseorang menilai suatu fenomena yang sedang berlangsung.

Analisis sentimen atau opinion mining adalah bidang studi yang menganalisis pendapat, sentimen, evaluasi, penilaian, sikap, dan emosi masyarakat terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atribut [2]. Analisis sentimen diperlukan guna mengetahui opini masyarakat terhadap kebijakan pemerintah. Dataset adalah hal yang sangat penting dalam analisis sentimen. Analisis sentimen dilakukan terhadap ratusan ribu data tweet Twitter yang didapatkan melalui TwitterAPI untuk mengetahui kecenderungan masyarakat Indonesia terhadap isu perceraian, apakah ke arah positif, negatif, atau netral. Analisis sentimen terhadap dataset yang besar atau big data memiliki karakteristik yaitu ukuran data, variasi fitur, dan frekuensi kemunculan [3]. Jumlah dan variasi dataset mempengaruhi hasil klasifikasi sentimen, semakin besar data yang digunakan pada proses training semakin bagus performa model klasifikasi yang dihasilkan [4]. Penggunaan dataset yang besar juga memastikan hasil klasifikasi sentimen valid karena telah mewakili suara setiap orang.

Random Forest juga tergolong memiliki akurasi yang tinggi jika di banding algoritma klasifikasi lainnya karena memiliki karakteristik ensemble learning. Dibuktikan dengan penelitian

yang telah dilakukan oleh [5] terhadap sentimen penilaian kustomer pada produk-produk media sosial, dari hasil perbandingan diketahui bahwa Random Forest paling unggul. Penelitian lain juga dilakukan oleh [6] guna meningkatkan performa hasil klasifikasi sentimen sebelumnya terhadap ulasan produk seluler di Kanada, diketahui bahwa Random Forest mampu meningkatkan akurasi. Melalui beberapa penelitian tersebut, maka Random Forest dipilih karena dapat menghasilkan akurasi yang tinggi dengan karakteristiknya yang ensemble learning.

1.2 Rumusan Masalah

Permasalahan yang dikaji dalam penelitian ini adalah sebagai berikut:

1. Bagaimana mengimplementasikan masing-masing algoritma Random Forest untuk mengklasifikasi sentimen masyarakat terhadap isu perceraian di Indonesia?
2. Bagaimana hasil visualisasi data sebelum dilakukan analisis menggunakan Random Forest?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Mengidentifikasi keadaan data dari hasil visualisasi sebelum dilakukan klasifikasi menggunakan algoritma Random Forest.
2. Mengidentifikasi model berdasarkan hasil tuning hyperparameter pada model klasifikasi Random Forest.

1.4 Manfaat Penelitian

Manfaat penelitian ini yaitu menjadi acuan informasi dan bahan evaluasi mengenai isu perceraian di Indonesia. Memanfaatkan bidang keilmuan data mining dengan menggunakan metode Random Forest untuk klasifikasi, sehingga dapat digunakan sebagai berikut:

1. Menambah wawasan mengenai analisis sentimen menggunakan metode klasifikasi Random Forest.
2. membantu pihak-pihak terkait, seperti pengacara atau psikolog, dalam memahami pandangan dan perasaan masyarakat terkait isu perceraian.
3. Dapat menjadi referensi dalam pengembangan ide penelitian selanjutnya.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah data sekunder yang didapat merupakan data tweet dari platform Twitter terkait topik isu perceraian di Indonesia per tanggal 29 November 2023 – 13 Desember 2023.
2. Menerapkan tuning hyperparameter yaitu minimum sample split dan minimum sample leaf pada pengujian model Random Forest