

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pengangguran masih menjadi fokus utama dalam permasalahan di setiap negara khususnya negara berkembang seperti Indonesia. Pengangguran merupakan masalah yang sangat kompleks karena saling mempengaruhi dan dipengaruhi oleh beberapa faktor yang berinteraksi mengikuti pola-pola yang tidak selalu mudah dipahami (Apriliani dkk., 2023). Rendah atau tingginya tingkat pengangguran dapat memberikan pengaruh terhadap kondisi perekonomian suatu daerah dan erat hubungannya dengan terciptanya kemakmuran serta kesejahteraan dalam pembangunan suatu daerah (Yuniarti & Imaningsih, 2022).



Gambar 1.1 Tingkat Pengangguran Terbuka (TPT) di Indonesia Berdasarkan Provinsi Tahun 2023

Indikator yang dapat digunakan dalam mengukur pengangguran merupakan Tingkat Pengangguran Terbuka (TPT). Menurut Badan Pusat Statistika (BPS) pada Survei Angkatan Kerja Nasional (SAKERNAS) Agustus 2023, Tingkat Pengangguran Terbuka Indonesia sebesar 5,32 persen atau masih melampaui asumsi yang ditetapkan sebelumnya. Dalam asumsi sebelumnya, tingkat pengangguran tersebut dipatok hanya 4,8-5 persen (Ismail dkk., 2024). Pada Gambar 1.1 memperlihatkan nilai TPT di beberapa daerah Indonesia yang menunjukkan nilai TPT masih terbilang tinggi bahkan nilai angkanya lebih besar daripada TPT Nasional.

Pada beberapa tahun terakhir, bidang *machine learning* telah berkembang pesat, dengan fokus tujuan pada pengembangan algoritma yang mampu belajar dari kumpulan data dan membuat prediksi atau keputusan yang lebih cerdas berdasarkan pembelajaran tersebut (Gabrikova dkk., 2023). Namun, literatur ilmiah mengenai prediksi pada bidang sosial sejauh ini masih minim daripada bidang industri dan teknologi. *Machine learning* adalah pendekatan yang masih relatif baru dan belum banyak dimanfaatkan dalam ilmu sosial dan berbasis kehidupan manusia, sehingga dibutuhkan lebih banyak eksplorasi dan implementasi dalam konteks permasalahan sosial (Kuikka, 2024).

Klasifikasi merupakan salah satu metode utama *machine learning* yang banyak digunakan. Pada konteks sosial yang pernah dilakukan sebelumnya, *machine learning* dapat membantu dalam klasifikasi dan prediksi masalah sosial berdasarkan faktor demografi ekonomi dan sosialnya (Yulistiani dkk., 2020). Klasifikasi adalah pendekatan *machine learning* yang memungkinkan kategori/kelas dapat diprediksi berdasarkan nilai atribut prediktornya dan dapat menemukan relasi dari variabel input (independen variabel) dan target variabel (dependen variabel) (Soofi & Awan, 2017). Demikian itu, mengingat masalah pengangguran yang masih tinggi, eksplorasi lebih mendalam terhadap potensi memanfaatkan klasifikasi dalam prediksi kategori TPT menjadi sangat relevan dan menarik untuk diteliti.

Salah satu metode yang banyak digunakan dalam klasifikasi adalah *ensemble learning*, yaitu teknik yang menggabungkan beberapa model untuk meningkatkan performa prediksi. Dua model yang populer dan sering digunakan dalam pendekatan ini adalah Random Forest dan Extreme Gradient Boosting (XGBoost) yang terbukti memiliki performa tinggi. Selain itu, kedua algoritma pohon keputusan tersebut dikenal cocok untuk menangani data yang umum dan kompleks pada bidang ilmu sosial (Verhagen, 2022). Namun demikian, pendekatan dengan memanfaatkan masing-masing antara algoritma Random Forest dan XGBoost secara spesifik belum pernah diaplikasikan sampai saat ini pada bidang sosial di sektor ketenagakerjaan, khususnya masalah pengangguran. Hal ini membuka peluang untuk menguji dan membandingkan kedua algoritma Random

Forest dan XGBoost dalam menangani masalah kompleks seperti Tingkat Pengangguran Terbuka (TPT).

Penelitian terdahulu telah menunjukkan bahwa metode *machine learning* memiliki potensi dalam melakukan klasifikasi Tingkat Pengangguran Terbuka (TPT) di Indonesia. Salah satunya penelitian yang dilakukan oleh Dewi dkk. (2024) dalam melakukan perbandingan algoritma *Naïve Bayes Classifier* dan *Support Vector Machine* pada data TPT Indonesia tahun 2020-2023. Hasil penelitian tersebut menunjukkan bahwa kedua algoritma mampu melakukan klasifikasi TPT dengan baik dan mencapai akurasi sebesar 75%. Temuan ini memberikan dasar bahwa pendekatan klasifikasi dapat digunakan dalam pemetaan status pengangguran. Namun demikian, penelitian tersebut menggunakan cakupan data pada tingkat yang belum menangkap variasi wilayah secara lebih terperinci, serta evaluasi performa model masih berfokus pada metrik akurasi. Hal ini membuka peluang untuk melakukan eksplorasi lanjutan dengan pendekatan algoritma lain, cakupan wilayah yang lebih granular, serta evaluasi model yang lebih menyeluruh.

Pada penelitian ini akan berfokus pada upaya mengidentifikasi algoritma *machine learning* dengan performa yang paling optimal dalam melakukan klasifikasi TPT di Indonesia. Perbandingan algoritma Random Forest dan XGBoost akan dilakukan dalam penelitian ini menggunakan data TPT tingkat kabupaten/kota sebanyak 514 data. Untuk memperoleh konfigurasi model terbaik, dilakukan proses *hyperparameter tuning* terhadap masing-masing algoritma guna mengoptimalkan parameter yang digunakan. Selanjutnya, dilakukan evaluasi menyeluruh menggunakan *confusion matrix* untuk analisis pola kesalahan klasifikasi, serta analisis kurva *Receiver Operating Characteristic* (ROC) dan nilai *Area Under Curve* (AUC) untuk menilai performa model terbaik secara holistik.

Berdasarkan uraian di atas, penelitian ini akan membahas penggunaan dan melakukan evaluasi performa masing-masing algoritma Random Forest dan XGBoost dalam klasifikasi Tingkat Pengangguran Terbuka di Indonesia. Harapan dari penelitian ini adalah dapat menghasilkan sebuah model yang optimal dalam mengklasifikasi TPT menurut Kabupaten/Kota di Indonesia dan juga untuk memperdalam pengetahuan tentang variabel-variabel yang mempunyai pengaruh terhadap TPT di Indonesia.

1.2 Identifikasi Masalah

Dari latar belakang diatas terdapat masalah yang timbul dan dapat diidentifikasi sebagai berikut:

1. Pengangguran masih menjadi permasalahan kompleks di Indonesia, disebabkan oleh berbagai faktor yang saling mempengaruhi. TPT sebagai indikator utama menunjukkan bahwa beberapa daerah memiliki nilai lebih tinggi dibanding rata-rata nasional.
2. Pendekatan berbasis *machine learning* telah terbukti efektif dalam banyak bidang, namun dalam konteks sosial masih tergolong terbatas.
3. Belum ada kajian yang secara khusus membandingkan Random Forest dan XGBoost dalam klasifikasi TPT dengan evaluasi metrik yang menyeluruh.

1.3 Pembatasan Masalah

Agar penyusunan tugas akhir ini tidak terlalu meluas dan keluar dari pokok permasalahan yang dirumuskan, maka berikut ini batasan permasalahan penelitian:

1. Penelitian ini menggunakan data sekunder yang diambil dari badan resmi Badan Pusat Statistika (BPS) yang telah tervalidasi dan banyak digunakan dalam penelitian sejenis.
2. Klasifikasi yang dilakukan menggunakan metode Random Forest dan XGBoost dengan mengevaluasi performa model menggunakan perhitungan *Accuracy*, *Precision*, *Recall*, *F1-Score* dan analisis kurva ROC & nilai AUC.
3. Penelitian ini menggunakan 5 variabel independen berupa Tingkat Partisipasi Angkatan Kerja, Indeks Pembangunan Manusia Penduduk, Presentase Penduduk Miskin, Produk Domestik Regional Bruto menurut Kapita, dan Kepadatan Penduduk.

1.4 Perumusan Masalah

Berdasarkan uraian latar belakang tersebut, maka dirumuskan masalah yaitu:

1. Bagaimana perbandingan Random Forest dan XGBoost dalam melakukan klasifikasi pada Tingkat Pengangguran Terbuka (TPT)?
2. Bagaimana performa yang dihasilkan dari perbandingan Random Forest dan XGBoost pada klasifikasi TPT di Indonesia?

1.5 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah di uraikan sebelumnya, penelitian ini mempunyai tujuan. Tujuan yang ingin dicapai melalui penelitian ini adalah sebagai berikut:

1. Menerapkan model Random Forest dan XGBoost untuk melakukan klasifikasi Tingkat Pengangguran Terbuka menurut Kabupaten/Kota di Indonesia dengan mempertimbangkan variabel fitur demografi.
2. Mengevaluasi performa model Random Forest dan XGBoost dalam mengklasifikasikan Tingkat Pengangguran Terbuka (TPT) menurut Kabupaten/Kota di Indonesia.

1.6 Manfaat Penelitian

Adapun manfaat dari penelitian ini yang dapat dipetik adalah sebagai berikut:

a. Bagi Penulis

Penelitian ini dapat memperluas wawasan penulis dalam melakukan analisis data dan mengimplementasi pengklasifikasian dengan pendekatan *machine learning* yaitu menggunakan Random Forest dan XGBoost pada bidang sosial khususnya Tingkat Pengangguran Terbuka (TPT) di Indonesia.

b. Bagi Universitas

Penelitian ini dapat menambah literatur mengenai *machine learning* pada analisis data khususnya model Random Forest dan XGBoost sebagai algoritma klasifikasi Tingkat Pengangguran Terbuka (TPT) di Indonesia dengan melihat evaluasi performa masing-masing model sebagai acuan, pembaharuan, dan pelengkap untuk penelitian yang akan hadir di masa depan.

c. Bagi Masyarakat dan Pemerintah

Penelitian ini dapat memberikan kontribusi terhadap pihak-pihak terkait seperti pemerintah, serikat kerja, dan peneliti sosial sebagai acuan dalam mengambil langkah konkrit untuk membuat kebijakan atau keputusan dalam masalah kasus pengangguran dengan lebih tepat.