

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pengangguran merupakan salah satu permasalahan yang sering dihadapi oleh negara berkembang termasuk Indonesia. Banyak faktor yang saling berhubungan mengikuti pola yang sulit dipahami sehingga membuat pengangguran termasuk masalah yang kompleks. Kompleksitas permasalahan ini tidak hanya melibatkan terbatasnya jumlah lapangan pekerjaan, tetapi juga ketidaksesuaian antara keterampilan tenaga kerja dan kebutuhan pasar. Kondisi ini menjadikan pengangguran sebagai isu multidimensional yang berdampak langsung terhadap pertumbuhan ekonomi, peningkatan kemiskinan, serta ketimpangan pembangunan antar wilayah. Topik pengangguran bahkan masih sering diperbincangkan oleh para politisi dalam perdebatan politik untuk dijadikan bahan membuat kebijakan dalam mengatasi masalah ini.

Salah satu indikator penting untuk mengukur tingkat pengangguran dan menilai kondisi ketenagakerjaan di suatu negara adalah Tingkat Pengangguran Terbuka (TPT). Tingkat pengangguran terbuka didefinisikan sebagai presentase penduduk yang termasuk angkatan kerja yang tidak bekerja dan sedang mencari pekerjaan terhadap jumlah angkatan kerja. Ukuran ini dapat digunakan untuk melihat seberapa besar tenaga kerja yang tidak terserap oleh pasar kerja di suatu wilayah. Masalah ketersediaan lapangan kerja di Indonesia yang relatif terbatas ini menyebabkan tingginya angka pengangguran karena tidak mampu menyerap para pencari kerja yang senantiasa bertambah setiap tahun seiring dengan

bertambahnya jumlah penduduk (Mulyadi, 2016).

Berdasarkan data Badan Pusat Statistika (BPS), tingkat pengangguran terbuka di Indonesia dalam 10 tahun terakhir mengalami fluktuasi yang signifikan dimana angka tertinggi terjadi pada Agustus tahun 2020 sebesar 7,07% dan menurun secara bertahap menjadi 4,91% pada Agustus 2024. Rasio angkatan kerja Indonesia yang menganggur pada tahun 2024 menurut *International Monetary Fund (IMF)* adalah 4,9% yang mana angka tersebut menempati posisi teratas tingkat pengangguran tertinggi di Asia Tenggara. Hal ini menunjukkan bahwa meskipun tingkat pengangguran terbuka di Indonesia mengalami penurunan, masalah pengangguran tetap menjadi masalah yang perlu perhatian serius karena kondisi ini menggambarkan adanya ketimpangan antara jumlah lapangan pekerjaan dengan jumlah tenaga kerja yang tersedia. Tingkat pengangguran terbuka yang tinggi di suatu negara menyebabkan rendahnya partisipasi masyarakat dalam pertumbuhan perekonomian yang akan berakibat juga pada penurunan tingkat pembangunan nasional (Fatika et al., 2022).

Penelitian mengenai tingkat pengangguran terbuka telah banyak dilakukan salah satunya menggunakan analisis *data mining*. Alat analisis yang digunakan dalam *data mining* adalah *machine learning* yang merupakan cabang ilmu dari *artificial intelligence* (kecerdasan buatan) yang memungkinkan komputer bertindak cerdas seperti manusia dan membuat keputusan berdasarkan data. Salah satu metode pengolahan data mining adalah klasifikasi yang bertujuan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan karakteristik yang dimilikinya. Teknik klasifikasi yang digunakan dalam penelitian ini adalah *random forest*.

Random Forest yang dikenalkan oleh Leo Breiman pada tahun 2001 merupakan salah satu metode klasifikasi yang dikembangkan dari metode *Classification and Regression Tree* (CART) dengan membangun sekumpulan pohon keputusan. Setiap pohon keputusan yang dibangun merupakan model yang akan digunakan untuk membuat prediksi berdasarkan data. Proses pembangunan pohon keputusan dalam *random forest* ini melibatkan elemen acak untuk menciptakan model yang kuat dan mampu menangani variasi dalam data sehingga dapat menghasilkan prediksi yang lebih akurat dan stabil. *Random forest* dapat mengukur seberapa penting dan besar kontribusi masing-masing variabel yang digunakan dalam model untuk membuat klasifikasi dan prediksi sehingga prediksi dapat lebih akurat dan fokus pada variabel yang benar-benar relevan. Salah satu metode yang digunakan adalah dengan mengacak nilai fitur tertentu dan mengamati perubahan pada akurasi model. Jika terjadi perubahan yang signifikan maka fitur tersebut dianggap penting dan memiliki pengaruh.

Penelitian tentang analisis Tingkat Pengangguran Terbuka dengan menggunakan *machine learning* diantaranya oleh Yulistiani et al (2020) yang menggunakan metode *Classification and Regression Tree* (CART) untuk mengklasifikasi dan prediksi tingkat pengangguran terbuka di Indonesia diperoleh *accuracy* sebesar 91,17%, *error ratio* sebesar 8,83% dan dari lima variabel yang diteliti, diperoleh tiga variabel yang paling berpengaruh terhadap TPT yaitu rata-rata lama sekolah, kemiskinan, dan angka partisipasi sekolah. Penelitian oleh Fatika et al (2022) dengan metode *Classification and Regression Tree* (CART) dan C4.5 untuk klasifikasi tingkat pengangguran terbuka di Indonesia diperoleh *accuracy* untuk kedua metode tersebut sama yaitu 100% dan dari delapan variabel

yang diteliti, diperoleh satu variabel yang berpengaruh terhadap klasifikasi TPT yaitu tingkat kesempatan kerja. Penelitian yang dilakukan oleh Dewi et al (2024) dengan membandingkan metode *Naive Bayes Classifier* dan *Support Vector Machine* untuk klasifikasi TPT di Indonesia dengan delapan variabel bebas diperoleh *accuracy* untuk setiap algoritma yang digunakan adalah sama yaitu 75%.

Algoritma klasifikasi telah banyak diterapkan, namun studi yang secara spesifik mengimplementasikan algoritma *Random Forest* untuk klasifikasi Tingkat Pengangguran Terbuka di Indonesia, khususnya dengan fokus pada penentuan variabel paling berpengaruh melalui analisis *feature importance* masih sangat terbatas atau belum dieksplorasi secara mendalam. *Random Forest* yang dibangun dari sekumpulan pohon keputusan dengan elemen acak, dikenal mampu menghasilkan model klasifikasi yang *robust*, stabil, dan akurat, bahkan dalam menghadapi variasi data yang kompleks. Keunggulan signifikan dari *Random Forest* terletak pada kemampuannya untuk mengukur dan mengidentifikasi seberapa penting kontribusi masing-masing variabel prediktor terhadap hasil klasifikasi. *Random Forest* diharapkan dapat memberikan pemahaman yang lebih komprehensif mengenai faktor-faktor kunci yang mempengaruhi Tingkat Pengangguran Terbuka, yang pada akhirnya dapat menjadi dasar perumusan kebijakan yang lebih tepat sasaran untuk mengatasi permasalahan pengangguran di Indonesia.

Berdasarkan uraian latar belakang tersebut, akan dilakukan penelitian dengan judul “Implementasi Algoritma *Random Forest* untuk Menentukan Variabel yang Berpengaruh pada Klasifikasi Tingkat Pengangguran Terbuka di

Indonesia”. Penelitian ini bertujuan untuk mengimplementasikan algoritma *random forest* untuk mengklasifikasikan tingkat pengangguran terbuka di setiap provinsi Indonesia ke dalam kelompok berstatus tingkat pengangguran terbuka rendah, sedang, atau tinggi serta mengetahui faktor yang memiliki pengaruh terhadap penentuan kategori tingkat pengangguran terbuka di Indonesia.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang diatas, rumusan masalah dalam penelitian adalah sebagai berikut :

1. Bagaimana proses implementasi algoritma *random forest* dalam membangun model untuk mengklasifikasikan Tingkat Pengangguran Terbuka di Indonesia?
2. Bagaimana kinerja algoritma *random forest* yang telah diimplementasikan dalam mengklasifikasikan Tingkat Pengangguran Terbuka berdasarkan metrik *accuracy*, *precision*, dan *recall*?
3. Variabel manakah yang paling berpengaruh terhadap klasifikasi Tingkat Pengangguran Terbuka di Indonesia dengan metode *random forest*?

1.3 Batasan Masalah

Batasan masalah yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistika (BPS) pada tahun 2015 – 2024. Variabel yang digunakan meliputi satu variabel dependen, yaitu Tingkat Pengangguran Terbuka (TPT), serta delapan variabel independen, yaitu presentase Indeks Pembangunan Manusia (IPM), presentase penduduk miskin, Upah Minimum Provinsi, presentase rata-rata lama sekolah (RLS), presentase laju pertumbuhan PDRB atas harga konstan, angka partisipasi kasar SMA, investasi

asing, dan kepadatan penduduk. Metode yang digunakan dalam penelitian ini adalah *random forest*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

1. Menerapkan algoritma *random forest* dalam membangun model untuk mengklasifikasikan Tingkat Pengangguran Terbuka di Indonesia.
2. Menganalisis kinerja algoritma *random forest* yang telah diimplementasikan dalam mengklasifikasikan Tingkat Pengangguran Terbuka berdasarkan metrik *accuracy*, *precision*, dan *recall*.
3. Menentukan faktor yang berpengaruh terhadap kategori Tingkat Pengangguran Terbuka di Indonesia.