

BAB I

PENDAHULUAN

1.1 Latar Belakang

Salah satu indikator yang sering digunakan untuk menggambarkan kondisi pasar tenaga kerja di suatu wilayah adalah Tingkat Pengangguran Terbuka (TPT). Indikator ini mencerminkan rasio antara jumlah penganggur dan total angkatan kerja pada titik waktu tertentu. Fluktuasi pada titik ini tidak hanya menandakan kemampuan suatu daerah untuk menyediakan kesempatan kerja, melainkan cenderung menunjukkan ketajaman pribadi secara sosial dan finansial. Berdasarkan hal tersebut, perkembangan TPT perlu dievaluasi karena akan menjadi pedoman dalam kebijakan ketenagakerjaan yang berorientasi kebijakan dan perencanaan regional yang berkelanjutan (Firmansyah & Huda, 2025).

Salah satu provinsi dengan populasi tertinggi adalah Jawa Timur dan struktur ekonomi yang sangat beragam di Indonesia. Hal ini mengakibatkan tenaga kerja pada daerah tadi terus berubah, baik karena perpindahan tenaga kerja antar sektor, keluarnya industri baru, atau melemahnya sektor-sektor yang sebelumnya menjadi andalan. Perkembangan pada sektor industri, perdagangan, investasi, serta pembenahan sistem keuangan dan ekonomi nasional, yang bertujuan untuk memperkuat perekonomian, tak jarang menyebabkan ketimpangan antara lapangan kerja yang tersedia serta jumlah tenaga kerja yang perlu diserap. Data taraf Pengangguran Terbuka pada tingkat kabupaten atau kota memberikan pola fluktuatif serta berbeda-beda antar daerah selama beberapa tahun. Oleh karena itu, perkembangan taraf Pengangguran Terbuka perlu terus dianalisis, sebab hasilnya dapat menjadi acuan dalam menyusun kebijakan ketenagakerjaan maupun perencanaan pembangunan regional yang lebih tepat targetnya (Afifah & Salma, 2026) (Halif et al., 2025).

Seperti yang dilaporkan dalam beberapa publikasi dari Badan Pusat Statistik (BPS) Provinsi Jawa Timur, nilai TPT pada beberapa tahun cenderung bervariasi dan dipengaruhi oleh aspek ekonomi dan sosial. Hal ini menunjukkan bahwa

kondisi pengangguran bukan merupakan peristiwa yang berdiri sendiri, melainkan perlu dikaitkan dengan variabel lain seperti Tingkat Partisipasi Angkatan Kerja (TPAK), inflasi, rata-rata lama sekolah, Produk Domestik Regional Bruto (PDRB) per kapita, dan Upah Minimum Regional (UMR). Memandangkan hubungan antarvariabel tersebut cukup kompleks, diperlukan pendekatan yang lebih analitis yang dapat menghasilkan prediksi dengan lebih presisi untuk menjadi dasar dalam mendukung proses pengambilan keputusan (Annastasya et al., 2025).

Memprediksi apa yang akan terjadi dengan menggunakan data yang cenderung mengikuti pola tertentu telah mendorong penerapan berbagai metode. Salah satu metode yang masih sering digunakan adalah *Regresi Linear*. Ini adalah metodologi statistik yang menelaah hubungan antara variabel keluaran (efek) dan satu atau lebih variabel paparan (penyebab). Salah satu manfaat pendekatan ini adalah relatif mudah untuk dihitung serta mudah diinterpretasikan secara logis. Namun, kekuatan *Regresi Linear* sangat bergantung pada asumsi bahwa hubungan antar variabel berjalan dengan lurus. Ketika pola hubungan menjadi lebih rumit, kemampuan model ini bisa menurun sehingga ketepatan prediksi menjadi kurang baik. *Regresi Linear* memang memiliki keterbatasan ketika pola hubungan antar variabel menjadi kompleks, sehingga perlu dipertimbangkan untuk menggunakan metode lain yang lebih sesuai dengan pola data yang ada. Dengan demikian, *Regresi Linear* masih bisa menjadi pilihan yang baik jika pola hubungan antar variabel bersifat linear (Halif et al., 2025)(Annastasya et al., 2025).

Dengan berkembangnya teknologi analisis data, algoritma berbasis *Machine Learning* kini semakin banyak dipakai untuk meningkatkan akurasi perkiraan. Salah satu algoritma yang bekerja dengan baik dalam menyelesaikan masalah *regresi* adalah *Random Forest Regression*. Cara kerja algoritma ini adalah dengan cara menggabungkan beberapa pohon keputusan yang dibuat, lalu menghasilkan satu prediksi akhir berdasarkan kombinasi dari semua pohon tersebut. Pendekatan ini memungkinkan *Random Forest* mengenali hubungan antar variabel yang lebih rumit dibandingkan metode statistik biasa, sehingga biasanya menghasilkan prediksi yang lebih konsisten pada data yang memiliki berbagai jenis karakteristik (Annastasya et al., 2025).

Meskipun kedua metode memiliki karakteristik yang berbeda, keduanya sama-sama memiliki potensi untuk diterapkan dalam prediksi Tingkat Pengangguran Terbuka. *Regresi Linear* menawarkan kemudahan interpretasi sehingga pengaruh setiap variabel independen terhadap nilai TPT dapat dijelaskan secara kuantitatif. Sebaliknya, *Random Forest* lebih unggul dalam mempelajari pola hubungan yang tidak selalu *linear* dan cenderung lebih adaptif terhadap variasi data. Perbedaan karakteristik tersebut menjadikan kedua algoritma menarik untuk dibandingkan pada data ketenagakerjaan di Provinsi Jawa Timur (Annastasya et al., 2025).

Sejumlah penelitian yang sebelumnya memperlihatkan bahwa memilih algoritma dapat memberikan hasil yang berbeda meskipun menggunakan dataset yang sama. Perbedaan ini disebabkan oleh beberapa faktor, seperti karakteristik data, jumlah variabel yang digunakan, proses awal sebelum memproses data, dan metode evaluasi yang dipakai. Karena itu, perlu dilakukan penelitian yang membandingkan langsung performa *Regresi Linear* dan *Random Forest Regression* menggunakan data yang sama agar hasil evaluasi bisa dilakukan dengan adil. Dengan membandingkan keduanya, diharapkan kita bisa mendapatkan informasi tentang algoritma yang paling tepat untuk memprediksi Tingkat Pengangguran di Provinsi Jawa Timur (Annastasya et al., 2025).

Penelitian ini menggunakan lima hal yang memengaruhi hasilnya, yaitu Produk Domestik Regional Bruto per kapita, Upah Minimum Regional, Tingkat Partisipasi Angkatan Kerja, inflasi, dan rata-rata lama sekolah. Kelima hal ini dipilih karena berkaitan dengan keadaan lapangan kerja dan ada datanya yang diterbitkan secara teratur oleh Badan Pusat Statistik. Dengan menggunakan kelima hal ini, diharapkan bisa memahami perubahan Tingkat Pengangguran Terbuka, sehingga bisa digunakan untuk membuat model yang mencerminkan keadaan sebenarnya.

Data yang digunakan dalam penelitian ini mencakup 38 kabupaten/kota di Provinsi Jawa Timur dengan periode pengamatan 2021–2025. Dengan menggunakan data dari seluruh wilayah administratif tersebut, diharapkan kita dapat memahami variasi karakteristik ketenagakerjaan di Jawa Timur dengan lebih

baik. Sebelum melakukan analisis, data harus diproses terlebih dahulu untuk memastikan kualitasnya baik dan dapat digunakan untuk mendapatkan hasil prediksi yang akurat.

Dalam menilai kinerja dua algoritma *regresi linier* dan *random forest* kita menggunakan beberapa ukuran seperti kesalahan rata-rata absolut, kesalahan kuadrat rata-rata, akar dari kesalahan kuadrat rata-rata, serta koefisien determinasi. Dengan menggunakan lebih dari satu metrik, kita dapat memahami tidak hanya seberapa besar kesalahan prediksi, tetapi juga seberapa baik model dapat menjelaskan variasi data nyata. Dengan demikian, *Regresi linier* dan *random forest regresi* dapat dibandingkan dengan lebih baik.

Dalam memprediksi Tingkat Pengangguran di Jawa Timur, penelitian ini membandingkan kinerja *Regresi Linear* dan *Regresi Random Forest*. Hasil penelitian ini diharapkan dapat membantu menentukan algoritma mana yang memiliki akurasi prediksi tertinggi. Selain itu, Studi ini dapat digunakan sebagai referensi untuk studi lain mendatang atau sebagai dasar untuk pengambilan kebijakan di bidang ketenagakerjaan yang didukung oleh analisis.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana hasil perkiraan Metode Algoritma *Regresi Linier* untuk Menentukan Tingkat Pengangguran Terbuka di Jawa Timur?
2. Bagaimana hasil perkiraan Metode Algoritma algoritma *Random Forest Regression Linier* untuk Menentukan Tingkat Pengangguran Terbuka di Jawa Timur?
3. Algoritma mana yang memberikan hasil prediksi terbaik berdasarkan nilai MAE, MSE, RMSE, dan R² dalam memperkirakan Taraf Pengangguran Terbuka di Jawa Timur?

1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan terfokus, maka batasan penelitian ditetapkan sebagai berikut:

1. Penelitian menggunakan data Tingkat Pengangguran Terbuka (TPT) pada 38 kabupaten dan kota di Jawa Timur selama periode 2021–2025.
2. Variabel independen yang digunakan meliputi PDRB per kapita, Tingkat Pengangguran Terbuka (TPT), inflasi, Tingkat Partisipasi Angkatan Kerja (TPAK), dan rata-rata lama pendidikan adalah variabel terikatnya.
3. Algoritma yang dibandingkan di riset ini adalah *Regresi Linear* dan *Random Forest Regression*.
4. Performa model dievaluasi menggunakan metrik MAE, MSE, RMSE, dan R2.
5. Penelitian ini menerapkan bahasa pemrograman *Python* melalui lingkungan pengembangan *Google Colaboratory*.

1.4 Tujuan dan Manfaat

Tujuan dan manfaat berdasarkan rumusan masalah yang telah ditetapkan, penelitian ini bertujuan untuk:

a. Tujuan

1. Mengembangkan model prediksi Tingkat Pengangguran Terbuka (TPT) Jawa Timur yang dihitung dengan algoritma *regresi linier*.
2. Mengembangkan model prediksi Tingkat Pengangguran Terbuka (TPT) Jawa Timur yang dihitung dengan algoritma *Random Forest Regression*.
3. Membandingkan performa kedua algoritma tersebut berdasarkan nilai MAE, MSE, RMSE, dan R2, sehingga terdapat metode yang memiliki tingkat akurasi prediksi terbaik.

b. Manfaat

1. Manfaat Teoritis

Hasil penelitian diharapkan dapat memperkaya referensi mengenai penerapan algoritma *Regresi Linear* dan *Random Forest Regression* dalam bidang prediksi, khususnya pada penelitian yang berkaitan dengan Tingkat Pengangguran Terbuka (TPT). Selain itu, penelitian ini dapat menjadi salah satu sumber rujukan bagi peneliti berikutnya yang mengembangkan kajian serupa.

1. Manfaat Praktis

Secara praktis, penelitian ini diharapkan mampu memberikan informasi yang dapat dimanfaatkan sebagai bahan pertimbangan dalam proses pengambilan keputusan terkait kebijakan ketenagakerjaan. Selain itu, hasil penelitian juga dapat menjadi referensi dalam memilih algoritma prediksi yang sesuai untuk menganalisis data sosial ekonomi, khususnya data ketenagakerjaan di Provinsi Jawa Timur.

1.5 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain *predictive-comparative* untuk menganalisis faktor-faktor yang memengaruhi Tingkat Pengangguran Terbuka (TPT) di Jawa Timur. Data numerik diambil dari Badan Pusat Statistik (BPS) dan mencakup variabel inflasi, PDRB per kapita, Upah Minimum Kabupaten, Lama Pendidikan Rata-rata, dan Tingkat Partisipasi Angkatan Kerja. Penelitian ini membangun model untuk memprediksi TPT dan Tingkat Partisipasi Angkatan Kerja, Lama Pendidikan Rata-rata, dan Upah Minimum Kabupaten dibandingkan dengan kinerja algoritma *regresi linier* dan *regresi random forest*.

Proses penelitian dimulai dengan pemrosesan data, termasuk pengecekan *missing value* dan data duplikat, analisis deskriptif, serta standardisasi data. Kumpulan data tersebut dibagi menjadi 20% untuk pengujian dan 80% untuk pelatihan. Model yang dibangun dievaluasi menggunakan metrik seperti MAE, MSE, dan RMSE untuk menentukan algoritma terbaik dalam memprediksi TPT Jawa Timur.

1.6 Prosedur Penelitian

Pendekatan studi ini dirancang secara cermat untuk memastikan validitas dan akurasi hasil pemodelan. Langkah pertama melibatkan mengumpulkan data, khususnya data sekunder dari kantor Badan Pusat Statistik (BPS) Jawa Timur. Studi ini menggunakan data dari 38 kabupaten dan kota di Jawa Timur untuk periode tahun 2021 hingga 2025. Variabel yang digunakan mencakup Tingkat Pengangguran Terbuka (TPT) sebagai variabel dependen serta lima variabel

independen inflasi, Angka Partisipasi Angkatan Kerja (TPAK), Produk Domestik Regional Bruto (PDRB), Upah Minim Regional (UMR), dan lama sekolah rata-rata.

Setelah pengumpulan data, dilakukan tahap pra-pemrosesan untuk menjamin kualitas data. Saat ini, data diperiksa untuk mengidentifikasi nilai yang hilang dan data ganda, dianalisis secara deskriptif maupun antarvariabel, serta distandardisasi menggunakan metode *StandardScaler*. Selanjutnya, dataset dibagi menjadi data pelatihan (80%) dan data pengujian (20%) untuk mendukung proses pelatihan dan pengujian model.

Langkah berikutnya adalah membuat model prediktif dengan *Regresi Linier* dan *Random Forest*. *Regresi Linier* digunakan sebagai model acuan untuk menganalisis hubungan *linier* antara variabel independen dan variabel dependen, dan *Random Forest* dipakai untuk menemukan pola yang terhubung dengan kompleks melalui pendekatan pembelajaran *ensemble* yang terdiri dari banyak pohon keputusan.

Pada tahap akhir, kinerja kedua model dievaluasi menggunakan data pengujian. Hasil prediksi dari masing-masing model dibandingkan dengan angka Tingkat Pengangguran Terbuka (TPT) aktual menggunakan metrik seperti *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), serta *koefisien determinasi* (R²). Berdasarkan hasil evaluasi tersebut, model dengan kinerja terbaik dipilih dengan memprioritaskan nilai MAE, MSE, dan RMSE yang lebih rendah serta nilai R² yang lebih tinggi. Model yang menunjukkan kinerja terbaik kemudian dijadikan dasar dalam merumuskan kesimpulan penelitian.

1.7 Sistematika Penulisan

Sistematika penulisan Tugas Akhir ini disusun berdasarkan urutan bab yang terstruktur agar pembahasan dan analisis dapat disajikan secara logis dan terperinci, sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang penelitian, rumusan masalah, tujuan penelitian,

manfaat penelitian, batasan penelitian, metode penelitian, prosedur penelitian, serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Membahas teori dan penelitian terdahulu yang berkaitan dengan Tingkat Pengangguran Terbuka, Regresi Linear, Random Forest Regression, serta evaluasi model regresi.

BAB III PERANCANGAN

Di bab ini akan dijelaskan perancangan penelitian yang meliputi arsitektur sistem, alur penelitian, perancangan dataset, tahapan pra-pemrosesan data, serta perancangan model prediksi yang digunakan dalam penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini mencakup implementasi dan hasil pengujian model *Regresi Linear* dan *Regresi Random Forest*, analisis kinerja model menggunakan metrik evaluasi, serta pembahasan temuan penelitian dalam kaitannya dengan teori dan penelitian terdahulu.

BAB V PENUTUP

Bagian ini mencakup hasil penelitian, kesimpulan, dan saran untuk penelitian lanjutan.