

BAB I

PENDAHULUAN

1.1 Latar Belakang

Transformasi digital dan peningkatan aktivitas daring masyarakat telah menyebabkan jumlah data yang dihasilkan menjadi sangat besar dan belum pernah terjadi sebelumnya. Setiap detik, interaksi manusia dengan teknologi, mulai dari transaksi *e-commerce*, kegiatan media sosial, hingga pencatatan sensor otomatis, menciptakan jejak digital yang terakumulasi secara eksponensial. Akumulasi data yang tidak hanya masif dari segi volume, tetapi tumbuh dengan kecepatan tinggi (*velocity*) dan memiliki format yang beragam (*variety*), melahirkan fenomena kompleks yang dikenal sebagai *Big Data*. Pengelolaan entitas data dengan karakteristik tersebut memerlukan infrastruktur teknologi analitik yang tidak hanya kuat dalam fungsinya, tetapi efisien dalam penggunaan sumber daya.

Dalam ekosistem analitik ini, *R* (melalui *RStudio*) dan *Python* menjadi dua platform utama yang digunakan oleh para praktisi data. *R* dikenal memiliki kelebihan dalam hal pengolahan statistik dan visualisasi data, sedangkan *Python* memberikan fleksibilitas dengan ekosistem yang luas, terutama dalam hal integrasi pembelajaran mesin dan otomatisasi (Ahmed, dkk., 2023). Namun, dalam proses *Big Data Analytics*, kebutuhan terhadap sumber daya komputasi seperti energi listrik, siklus *CPU*, dan kapasitas memori menjadi semakin besar. Jika tidak diatur, penggunaan sumber daya yang berlebihan ini dapat menyebabkan biaya operasional yang tidak efisien dan dampak negatif bagi lingkungan.

Seiring meningkatnya perhatian terhadap keberlanjutan teknologi informasi, prinsip *Green Computing* menjadi perspektif penting dalam evaluasi platform analitik. Menurut penelitian terbaru, komputasi berkelanjutan telah muncul sebagai bidang studi yang kritis, yang membahas dampak lingkungan dari sistem komputasi melalui inovasi dalam perangkat lunak, perangkat keras, dan algoritma (Kautish, dkk., 2025). Performa platform tidak hanya dinilai dari kecepatan atau keluaran analitik, tetapi dari efisiensi pemanfaatan *CPU*, memori, dan energi untuk mendukung proses komputasi yang lebih berkelanjutan. Pada penelitian ini, istilah *Big Data Analytics* digunakan untuk merujuk pada pendekatan *pipeline* analitik dan kebutuhan efisiensi komputasi, sedangkan pengujian dilakukan pada analitik data skala menengah pada perangkat kelas menengah sesuai skenario eksperimen yang ditetapkan.

Beberapa penelitian terdahulu telah berusaha untuk membandingkan kedua platform tersebut. Penelitian oleh (Batko dan Ślęzak, 2022) serta (Maheshwari, 2024) menunjukkan adanya perbedaan dalam karakteristik performa antara *R* (melalui *RStudio*) dan *Python* namun umumnya masih berfokus pada aspek fungsional atau metrik kinerja komputasi konvensional. Studi lainnya, seperti yang dilakukan oleh (Nahrstedt, dkk., 2024), mulai membahas aspek energi namun masih terbatas pada perbandingan pustaka di dalam *Python* (seperti *pandas* dengan *polars*) dan belum mencakup perbandingan antar platform secara komprehensif.

Berdasarkan pemetaan literatur awal, Kajian yang secara sistematis membandingkan performa *R* (melalui *RStudio*) dan *Python* dalam satu skenario eksperimen *pipeline* analitik yang terkontrol dengan beban kerja yang simetris

masih sangat terbatas. Dari 15 penelitian terdahulu yang ditelaah, tercatat hanya 4 yang memenuhi kriteria kesetaraan metodologis tersebut.

Selain itu, integrasi evaluasi performa dengan perspektif *Green Computing* yang menekankan efisiensi energi dan sumber daya perangkat lunak juga masih jarang dibahas secara eksplisit. Kondisi ini menunjukkan bahwa kajian komparatif yang mengukur waktu eksekusi, utilisasi *CPU*, penggunaan memori, dan konsumsi energi secara terpadu dalam satu rancangan uji yang setara masih belum banyak dilakukan. Oleh karena itu, penelitian ini bertujuan mengevaluasi keempat parameter tersebut pada *R* (melalui *RStudio*) dan *Python* untuk menghasilkan rekomendasi berbasis data.

1.2 Rumusan Masalah

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

1. Bagaimana perbandingan performa pemanfaatan sumber daya komputasi antara *R* (melalui *RStudio*) dan *Python*, yang diukur berdasarkan waktu eksekusi, utilisasi *CPU*, penggunaan memori, dan konsumsi energi dalam *Big Data Analytics* berbasis *Green Computing*?
2. Bagaimana karakteristik performa masing-masing platform dalam mendukung pengolahan data skala menengah berdasarkan parameter waktu eksekusi, utilisasi *CPU*, penggunaan memori, dan konsumsi energi?

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Mengevaluasi dan membandingkan performa *R* (melalui *RStudio*) dan *Python* dalam *Big Data Analytics* melalui pengukuran waktu eksekusi, utilisasi *CPU*, penggunaan memori, dan konsumsi energi sebagai parameter utama *Green Computing*.
2. Mengidentifikasi karakteristik performa dan *trade-off* performa pada masing-masing platform sebagai dasar rekomendasi pemilihan teknologi dalam komputasi berbasis *Green Computing*.

1.4 Manfaat Penelitian

Penelitian ini memiliki beragam manfaat yang dapat diharapkan, antara lain:

1. Manfaat Teoritis :
 - a. Menambah literatur ilmiah mengenai komparasi performa *R* (melalui *RStudio*) dan *Python* dalam konteks *Big Data Analytics*.
 - b. Memperkuat pemahaman penerapan prinsip *Green Computing* dalam evaluasi efisiensi pemanfaatan sumber daya komputasi.
2. Manfaat Praktis :
 - a. Memberikan rujukan berbasis bukti bagi praktisi dalam memilih platform analitik sesuai ketersediaan sumber daya perangkat keras.
 - b. Menjadi pertimbangan bagi organisasi dalam merancang sistem pengolahan data yang lebih efisien energi dan ramah lingkungan.

1.5 Batasan Masalah

Untuk menjaga penelitian tetap terarah dan fokus pada tujuan yang ingin dicapai, ditetapkan batasan masalah sebagai berikut:

1. Penelitian membandingkan performa platform *R* (melalui *RStudio*) dan *Python* sebagai alat analitik utama, tanpa melibatkan platform *cloud* atau *distributed computing* lainnya.
2. Penilaian performa dibatasi pada empat parameter teknis yang selaras dengan prinsip *Green Software Foundation* (GSF), yaitu waktu eksekusi, utilisasi *CPU*, penggunaan memori, dan konsumsi energi.
3. Dataset yang digunakan adalah data publik “*AI4I 2020 Predictive Maintenance*” yang bersifat sintesis untuk memastikan keterulangan eksperimen.
4. Skenario analisis difokuskan pada tahapan *data ingestion*, *data cleaning & preprocessing*, dan *transformation*. Sebagai bagian dari *benchmarking*, digunakan model klasifikasi dasar (*Logistic Regression*) serta teknik *oversampling* (*SVM-SMOTE*) semata-mata sebagai *workload* pembanding untuk memastikan *pipeline* berjalan *end-to-end* dan pengukuran parameter *Green Computing* secara konsisten.
5. Pengukuran dilakukan pada perangkat keras kelas menengah dan menggunakan alat pemantauan (*Intel Power Gadget*) sesuai rancangan pengujian.