

BAB I

PENDAHULUAN

1.1 Latar Belakang Penelitian

Dalam era transformasi digital yang sedang berkembang pesat, penggunaan teknologi *cloud computing* menjadi suatu pertimbangan bagi para pengguna, baik suatu perusahaan maupun pengembang aplikasi karena efisiensi biaya, fleksibilitas yang tinggi dalam pengelolaan dan pemantauan data secara terpusat serta kemampuan menangani permasalahan data dengan cepat (Cole & Moore, 2018). Selain itu, teknologi *cloud* menyediakan layanan server yang bervariasi untuk mendukung penerapan aplikasi, salah satunya adalah layanan komputasi. Google Cloud Platform (GCP) merupakan salah satu penyedia layanan *cloud* yang menyediakan berbagai layanan komputasi yang terdiri dari layanan *server-based* dan *serverless*. Kedua layanan komputasi tersebut digunakan untuk *deployment* sebuah aplikasi, salah satunya model *machine learning*.

Machine learning adalah bagian dari kecerdasan buatan (AI) yang dapat mempelajari pola dari data tanpa perlu mendefinisikannya secara langsung (Lee & Shin, 2020). Penerapan model *machine learning* saat ini menjadi perhatian karena pemanfaatan kecerdasan buatan (AI) pada aplikasi sedang menjadi tren. Survey mengenai tren teknologi di masa depan menunjukkan bahwa kecerdasan buatan unggul berada di posisi pertama dengan persentase sebesar 26% dibandingkan teknologi-teknologi lainnya, salah satunya platform digital dan aplikasi yang memiliki persentase sebesar 12,7% (Statista Research Department, 2023). Berdasarkan peringkat indeks kesiapan kecerdasan buatan pada tahun 2022, Indonesia berada pada peringkat ke-43 dari 181 negara dengan skor indeks sebesar 60,89 dari 100. Berdasarkan data tersebut, Indonesia mendapatkan skor terendah pada sektor teknologi. Hal tersebut mengindikasikan perlunya penelitian mengenai teknologi kecerdasan buatan sebagai sumber pengetahuan agar pengembang di Indonesia dapat memanfaatkan teknologi kecerdasan buatan secara masif (Statista Research Department, 2024), termasuk *machine learning*. Seperti aplikasi pada umumnya *machine learning* juga dapat diterapkan melalui sever lokal maupun

cloud. Namun, terdapat kekurangan yang disebabkan melalui penggunaan server lokal untuk mengimplementasikan *machine learning*, salah satunya adalah akses dan pemrosesan data yang terbatas (Arbain dkk., 2022). Teknologi *cloud* memiliki potensi untuk mengatasi kekurangan tersebut dengan layanan-layanan yang tersedia sehingga akses terhadap aplikasi mudah dan waktu pemrosesan data serta penerapan *machine learning* dapat dilakukan lebih cepat (Xu, 2020; Zhu dkk., 2022).

Wu dkk. (2022) mengatakan bahwa *serverless computing* berpotensi menyediakan layanan yang *high performance*, *low cost*, dan pengolaan yang mudah. Meskipun begitu, layanan *server-based* juga menyediakan layanan *high performance* dan fleksibilitas yang tinggi melalui pengelolaan kombinasi perangkat komputasi oleh pengguna seperti CPU dan GPU, Cloud Storage, dan fitur lainnya sesuai dengan kebutuhan untuk mengimplementasikan *data science* atau *machine learning* yang berkinerja tinggi (Abraham & Yang, 2023a; Barua, 2021). Dengan keunggulan performa yang diberikan oleh masing-masing layanan, para pengguna pasti memiliki kesulitan dan memiliki pertimbangan untuk memilih layanan yang sesuai dengan kebutuhan untuk *deploy* model *machine learning*. Eisa dkk. (2020) menyatakan bahwa proses pemilihan layanan *cloud* tidak dapat dianggap sepele dan faktor konsumen sulit dalam memilih layanan *cloud* di antaranya adalah tingkat pengetahuan dan pemahaman tentang *requirement* layanan *cloud* yang sangat bervariasi dan penyedia layanan yang menyediakan bermacam-macam layanan berdasarkan variasi kinerja, harga layanan, dan lainnya. Hal tersebut juga disampaikan oleh de Moraes dkk. (2022) bahwa pemilihan layanan *cloud* menjadi tantangan bagi para *customer* untuk mengetahui layanan yang sesuai dengan kebutuhan.

Oleh karena itu, diperlukan penelitian yang dapat menjadi acuan pertimbangan para pengguna dalam memilih layanan komputasi *cloud* yang sesuai untuk *deployment* aplikasi khususnya dalam konteks model *machine learning* agar pengguna dapat mengetahui dan memaksimalkan layanan *cloud* untuk digunakan sesuai dengan kebutuhan. Model yang digunakan adalah model *smart agriculture*

berbasis *machine learning* yang dapat digunakan para petani untuk optimalisasi lahan dengan memberikan rekomendasi bibit tanaman (Pratama dkk., 2024).

1.2 Rumusan Masalah Penelitian

Berdasarkan latar belakang yang telah dipaparkan, terdapat rumusan masalah yang diuraikan sebagai berikut:

1. Bagaimana analisis performa layanan *server-based* yang dihasilkan melalui *deployment* model *machine learning*?
2. Bagaimana analisis performa layanan *serverless* yang dihasilkan melalui *deployment* model *machine learning*?
3. Bagaimana analisis perbandingan layanan terbaik untuk menerapkan model *machine learning* berdasarkan hasil pengukuran performa layanan?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan, tujuan dari penelitian ini yaitu sebagai berikut:

1. Melakukan analisis performa layanan *server-based* yang dihasilkan melalui *deployment* model *machine learning*.
2. Melakukan analisis performa layanan *serverless* yang dihasilkan melalui *deployment* model *machine learning*.
3. Melakukan analisis perbandingan layanan terbaik untuk menerapkan model *machine learning* berdasarkan hasil pengukuran performa layanan.

1.4 Batasan Masalah Penelitian

Adapun batasan dari penelitian ini dapat diuraikan sebagai berikut:

1. Platform *cloud* yang digunakan hanya Google Cloud Platform.
2. Layanan yang digunakan untuk melakukan *deployment* model *machine learning* pada penelitian ini adalah layanan *server-based* yaitu Compute Engine dan layanan *serverless* yang terdiri dari Kubernetes Engine, App Engine, Cloud Run, Cloud Functions, dan Vertex AI.

3. Model *machine learning* yang diterapkan pada layanan *cloud* sudah dilatih secara eksternal menggunakan metode pendekatan *supervised learning* pada *framework* scikit-learn dengan algoritma Decision Tree.
4. Parameter *endpoint* hasil *deployment* model *machine learning* diuji melalui *request* HTTP GET pada Apache JMeter.
5. Parameter yang digunakan sebagai alat ukur performa dalam penelitian ini: CPU dan *memory utilization*, *latency*, serta *pricing*. Untuk parameter komparasi tambahan: *developer experiences* yang meliputi kompatibilitas *framework machine learning*, tingkat kemudahan penerapan, dan tingkat ketersediaan dokumentasi.
6. Tidak membahas pengembangan model *machine learning* melalui *cloud*.
7. Biaya *cloud* yang disajikan pada penelitian ini diasumsikan saat nominal kurs *dollar* ke mata uang rupiah sebesar Rp 16.209,99.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini diuraikan sebagai berikut:

1. Manfaat teoritis

Penelitian ini dapat memberikan pengetahuan mengenai teknologi *cloud computing*, layanan *server-based* dan *serverless computing* termasuk hasil performa layanan.

2. Manfaat praktis

- a. Bagi Peneliti

Penelitian ini dapat menambah pengetahuan peneliti terhadap layanan *server-based* dan *serverless* di GCP. Lalu, menambah kemampuan penulis dalam memanfaatkan layanan-layanan *cloud* di GCP yang digunakan pada penelitian ini.

- b. Bagi Perusahaan dan Pengembang Aplikasi

Penelitian ini dapat menjadi bahan pertimbangan perusahaan dan pengembang aplikasi dalam memilih layanan *cloud* yang sesuai untuk *deploy* model *machine learning*.

- c. Bagi Universitas

Penelitian ini dapat menjadi referensi untuk menyusun penelitian yang serupa dalam bidang *cloud*.

1.6 Sistematika Penulisan

Untuk memberikan dan mengetahui garis besar pembahasan penelitian ini secara menyeluruh, diperlukan sebuah sistematika penulisan sebagai kerangka penulisan skripsi. Ada pun sistematika penulisannya adalah sebagai berikut:

1. BAB I Pendahuluan

BAB I terdiri dari latar belakang penelitian, rumusan masalah penelitian, tujuan penelitian, batasan masalah penelitian, manfaat penelitian, dan sistematika penulisan penelitian.

2. BAB II Kajian Pustaka

BAB II terdiri dari landasan teori yang berkaitan dengan penelitian yang dilakukan dan beberapa penelitian yang relevan.

3. BAB III Metodologi Penelitian

BAB III terdiri dari jenis, metode, dan penjelasan setiap alur penelitian sebagai acuan dalam melakukan penelitian.

4. BAB IV Hasil dan Pembahasan

BAB IV terdiri dari hasil penelitian yang sudah dilakukan serta pembahasan terkait analisis dari hasil penelitian tersebut.

5. BAB V Simpulan, Implikasi, dan Rekomendasi

BAB ini terdiri dari kesimpulan, implikasi, dan rekomendasi dari hasil penelitian yang sudah dilakukan.