

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam dunia komputasi modern, *benchmark* merupakan proses menjalankan program atau beban kerja tertentu untuk mengukur kinerja sistem komputer secara akurat. Pendekatan ini digunakan untuk mengevaluasi performa perangkat keras dalam menjalankan tugas spesifik, membandingkan konfigurasi antar PC, serta memverifikasi kesesuaian performa dengan spesifikasi yang diiklankan [1][2][3]. Contoh aplikasinya seperti 3DMark yang menjadi standar untuk mengukur kemampuan *rendering* 3D [4], PCMark untuk mengukur produktivitas perkantoran [5][6], serta Standard Performance Evaluation Corporation (SPEC) CPU, yang secara luas digunakan oleh vendor prosesor, penyedia layanan cloud, perancang server, dan operator pusat data untuk mengevaluasi kinerja komputasi intensif secara objektif dan terstandarisasi [7]. Dalam praktik industri, benchmark SPEC CPU digunakan sebagai acuan dalam pemilihan prosesor untuk kebutuhan enterprise, perencanaan kapasitas data center, serta evaluasi performa infrastruktur cloud. Pengukuran SPEC CPU terbagi menjadi empat metrik utama yang merepresentasikan beban kerja dunia nyata. Metrik CINT2017 (CINT) digunakan untuk mengevaluasi performa aplikasi berbasis integer yang bersifat latensi-sensitif, seperti kompilasi perangkat lunak, query database, dan transaksi sistem. CFP2017 (CFP) merepresentasikan beban kerja komputasi floating-point yang umum pada simulasi ilmiah, analisis teknik, dan pemodelan numerik [8][9]. Sebaliknya, metrik CINT2017rate dan CFP2017rate digunakan untuk mengukur kapasitas throughput sistem secara paralel, yang menjadi acuan penting bagi operator data center, penyedia layanan cloud, serta sistem high-performance computing (HPC) dalam menangani beban kerja multi-user dan multi-thread secara simultan [10][11].

Pemilihan empat metrik ini krusial karena merepresentasikan berbagai karakteristik beban kerja komputasi dunia nyata. Pentingnya pemisahan ini karena penyesuaian operasi fundamental prosesor yang berbeda, di mana CINT menguji pemrosesan bilangan bulat (*logic* dan *control-flow*) yang dominan pada aplikasi *database* SQL, kompilasi kode perangkat lunak, dan *Artificial Intelligence* (AI) level dasar yang krusial bagi *Software House* dan *Data Center*. Sementara itu, CFP menguji pemrosesan bilangan pecahan (*mathematical computation intensif*) yang sangat penting untuk simulasi ilmiah, pemodelan finansial, dan *rendering* grafis bagi insinyur dan peneliti [7][12][13]. Lu et al. [1] mengadopsi pemisahan ini sebagai dasar metodologi dengan membagi *dataset* menjadi empat sub-kelompok agar model dapat secara spesifik mempelajari karakteristik beban kerja yang unik dari setiap aspek, yang pada akhirnya berkontribusi pada akurasi prediksi yang tinggi.

Dalam konteks evaluasi performa prosesor, pihak yang terdampak langsung dari keterbatasan pengujian *benchmark* adalah vendor prosesor, penyedia layanan cloud, operator pusat data, serta peneliti dan praktisi yang membutuhkan estimasi performa CPU secara cepat dan objektif. Permasalahan utama yang dihadapi adalah keterbatasan pendekatan prediksi performa berbasis *machine learning* dalam mempertahankan akurasi ketika diaplikasikan pada spesifikasi perangkat keras yang tidak terdapat dalam data pelatihan. Secara khusus, metrik yang menjadi fokus dalam penelitian ini adalah selisih antara skor *benchmark* SPEC CPU2017 hasil prediksi model dan nilai *benchmark* aktual pada CPU serta compiler yang tidak terdapat dalam dataset. Deviasi prediksi diukur menggunakan selisih absolut skor *benchmark* sebagai indikator kuantitatif kemampuan generalisasi model. Pada kondisi saat ini, karakteristik deviasi prediksi tersebut belum terdokumentasi secara jelas, sehingga diperlukan analisis lebih lanjut untuk menilai kelayakan pendekatan prediksi performa pada perangkat keras generasi baru.

Namun, pelaksanaan *benchmark* spesifik industri Standard Performance Evaluation Corporation (SPEC) secara manual menghadapi hambatan teknis dan logistik. Berbeda dengan pengujian sintesis sederhana, standar SPEC

mewajibkan kepatuhan terhadap aturan eksekusi yang ketat, dimana penguji harus melakukan kompilasi kode sumber secara manual dan menjalankan pengujian tersebut minimal tiga kali iterasi untuk memastikan validitas statistik [14]. Proses ini sangat memakan waktu; satu siklus pengujian lengkap untuk metrik CINT dan CFP dari kombinasi perangkat keras tertentu, penguji wajib memiliki akses ke *hardware* seperti prosesor, memori, dan motherboard tersebut [12]. Masalah ini menjadikan pengujian manual untuk ribuan kombinasi konfigurasi *hardware* menjadi sulit secara ekonomi dan logistik bagi organisasi.

Menanggapi tantangan ini, berbagai penelitian mulai mengkaji pendekatan yang lebih efisien menggunakan kecerdasan buatan (Artificial Intelligence, AI), khususnya Machine Learning (ML), untuk memprediksi skor *benchmark* berdasarkan spesifikasi perangkat keras tanpa harus menjalankan pengujian secara langsung [1][2][15]. Salah satu penelitian yang menjadi pijakan penting dalam bidang ini adalah karya Meiliu Lu, Leonardo Lopez, dan Michael Guynn yang berjudul “*Predicting Computer Performance Based on Hardware Configuration Using Multiple Neural Networks*” [1].

Untuk metode yang digunakan oleh Meiliu Lu, penelitian tersebut menggabungkan berbagai teknik analisis dan model ML untuk memprediksi performa perangkat keras berdasarkan spesifikasi sistem. Kumpulan teknologi dan metode yang digunakan di penelitian dapat dilihat di tabel 1.1 :

Tabel 1.1 Teknologi atau Metode pada Penelitian Meiliu Lu

Teknologi/Metode	Deskripsi
Dataset	SPEC CPU2006 = 48,129 data, SPEC CPU2017 = 1,941 data (Total ~ 50, 000 data).
Jenis <i>Benchmark</i>	CINT2017, CFP2017, CINT2017rate, CFP2017rate.
Fitur Utama (Input)	System, Number of Cores, Number of Chips, Number of Cores per Chip, Number of Threads per Core, Processor, Processor MHz, Processor Characteristics, 1st Level Cache, 2nd Level Cache, 3rd Level Cache, Memory, Compiler.
Fitur Utama (Output)	Result.
Preprocessing	Konversi string ke numerik (pengelompokkan prosesor berdasarkan <i>family-model type-iteration</i> dan diberikan <i>class number</i> , pengelompokan memori berdasarkan total memori (GB) dan jumlah modul, dan compiler diberikan <i>class number</i> serta diurut berdasarkan versinya).
Analisis fitur	Principal Component Analysis (PCA) menggunakan FactoMineR untuk kombinasi fitur numerik dan kategorikal.
Model & Struktur	Artificial Neural Networks (ANN) dengan 13 neuron input, 3 hidden layer (50-100-50), dan 1 neuron output.
Algoritma Pelatihan	<i>Backpropagation</i> dengan <i>train-test-reverse approach</i>
Jumlah Epoch	50 epoch untuk mencapai konvergensi
Pengujian Model	Membandingkan Result dari dataset dengan hasil prediksi Result.

Pada tabel 1.1, secara keseluruhan, pendekatan ANN serta teknologi dan metode yang digunakan di penelitian terbukti efektif dalam memperkirakan performa perangkat keras dengan akurasi yang tinggi. Namun, peneliti juga

menyadari adanya keterbatasan yang perlu diselesaikan dalam penelitian lanjutan. Salah satu tantangan utama adalah proses konversi *string* ke numerik yang masih terbatas pada CPU dan *compiler* yang ada di *dataset*. Artinya, model belum sepenuhnya mampu mengenali atau memprediksi performa perangkat keras baru yang belum pernah dilatih sebelumnya. Keterbatasan utama yang harus diatasi adalah ketidakmampuan model beradaptasi terhadap perangkat keras terbaru, dan kebutuhan untuk meningkatkan akurasi model, melampaui capaian sebelumnya.

Berdasarkan keterbatasan tersebut, penelitian ini berfokus pada adaptasi metodologi Lu et al. agar model mampu memprediksi performa CPU dan compiler generasi terbaru yang tidak terdapat dalam dataset pelatihan. Target utama penelitian ini adalah mempertahankan tingkat akurasi yang sebanding dengan studi Lu et al., sekaligus memastikan model tetap stabil ketika diaplikasikan pada perangkat keras yang belum pernah digunakan selama proses pembelajaran.

Sebagai bagian dari evaluasi kemampuan generalisasi model, penelitian ini menerapkan skema pemisahan data yang lazim digunakan dalam machine learning, yaitu dengan membagi dataset SPEC CPU2017 ke dalam data pelatihan, validasi, dan pengujian. Evaluasi performa model dilakukan menggunakan data uji (test set) yang dipisahkan sejak awal dari dataset utama dan tidak pernah digunakan pada proses pelatihan maupun validasi. Pendekatan ini memastikan bahwa pengukuran kinerja model mencerminkan kemampuan prediksi terhadap data baru yang belum pernah dilihat sebelumnya, namun masih berasal dari distribusi dataset yang sama.

Untuk setiap metrik *benchmark* SPEC CPU2017, yaitu CINT2017, CFP2017, CINT2017rate, dan CFP2017rate, model dilatih secara terpisah dan dievaluasi menggunakan data uji yang sesuai. Selisih antara nilai aktual dan hasil prediksi pada data uji digunakan sebagai indikator kemampuan generalisasi model. Dengan demikian, hasil evaluasi yang diperoleh tidak hanya merepresentasikan performa model pada data pelatihan, tetapi juga

menggambarkan sejauh mana model mampu mempertahankan akurasi prediksi ketika diterapkan pada data yang tidak terlibat dalam proses pembelajaran.

1.2 Pertanyaan Penelitian

Berdasarkan latar belakang, masalah yang terjadi, identifikasi masalahnya adalah “Bagaimana kemampuan model prediksi berbasis machine learning dalam mempertahankan akurasi prediksi skor *benchmark* SPEC CPU2017 ketika diuji menggunakan data uji?”

1.3 Batasan Penelitian

Beberapa batasan yang digunakan pada penelitian ini adalah sebagai berikut:

1. Penelitian berfokus pada prediksi skor benchmark berbasis SPEC CPU2017, yaitu CINT2017, CFP2017, CINT2017rate, dan CFP2017rate, berdasarkan spesifikasi perangkat keras.
2. Skor ground truth untuk pengujian generalisasi diperoleh dari hasil benchmark SPEC CPU2017 pada beberapa CPU yang masing-masing mewakili satu metrik benchmark.
3. Penelitian ini mengadaptasi dan mengevaluasi pendekatan model prediksi performa dari penelitian terdahulu, dengan fokus pada kemampuan generalisasi model terhadap data uji.
4. Prediksi skor benchmark dibatasi pada parameter spesifikasi perangkat keras yang tersedia dalam dataset, sehingga tidak mencakup seluruh aspek performa sistem secara menyeluruh.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk mengevaluasi kemampuan generalisasi model prediksi performa berbasis machine learning dalam memprediksi skor benchmark SPEC CPU2017. Selain itu, penelitian ini bertujuan untuk menganalisis sejauh mana tingkat akurasi prediksi yang diperoleh pada data pelatihan dan validasi dapat dipertahankan ketika model diuji menggunakan data uji.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan studi prediksi performa perangkat keras dengan menerapkan evaluasi generalisasi model yang sesuai dengan praktik ML. Kemudian, penelitian ini memberikan gambaran mengenai kemampuan model prediksi dalam mempertahankan akurasi pada data uji, sehingga dapat menjadi referensi bagi penelitian selanjutnya yang memanfaatkan dataset *benchmark* standar seperti SPEC CPU2017.

1.6 Sistematika Penulisan

1. BAB I PENDAHULUAN

Bab I membahas masalah yang diteliti, identifikasi masalah, batasan masalah, tujuan penelitian, manfaat dari penelitian, dan sistematika penulisan.

2. BAB II TINJAUAN PUSTAKA

Bab II membahas landasan penelitian terdahulu yang mendukung penelitian ini, dan membahas landasan teori yang digunakan.

3. BAB III ANALISIS DAN PERANCANGAN SISTEM

Bab III membahas metode penelitian, studi literatur, pengembangan model, dan pengembangan skrip.

4. BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM

Bab IV membahas spesifikasi sistem, implementasi model, hasil pengujian model, dan hasil pengujian skrip.

5. BAB V KESIMPULAN DAN SARAN

Bab V membahas analisis dan hasil dari penelitian yang telah dilakukan.

