

BAB I PENDAHULUAN

1.1 Latar Belakang

Pentingnya mengidentifikasi dan mengelola penyakit daun kentang merupakan aspek krusial dalam pertanian. Kentang, sebagai salah satu tanaman pangan utama, menjadi sumber makanan yang signifikan di banyak wilayah dunia. Namun, tantangan utama yang dihadapi petani adalah masalah yang terkait dengan kesehatan tanaman, khususnya penyakit pada daun kentang. Penyakit ini dapat mengakibatkan kerugian produksi yang substansial jika tidak diidentifikasi dan ditangani dengan tepat.

Teknologi dalam bidang kecerdasan buatan telah membuka peluang baru untuk mendeteksi dan mengatasi masalah pertanian, termasuk dalam identifikasi penyakit pada tanaman. Salah satu pendekatan yang menonjol adalah penggunaan *Convolutional Neural Network* (CNN), sebuah jenis algoritma *deep learning* yang sangat efektif dalam pengenalan pola visual pada gambar.

Penelitian mengenai klasifikasi penyakit pada daun kentang telah dilakukan secara luas dengan melibatkan berbagai jenis metode, antara lain *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM) [1], dan *Convolutional Neural Network* (CNN) [2] [3]. Dari penelitian yang memanfaatkan *Convolutional Neural Network* (CNN) menunjukkan bahwa metode tersebut mampu mengenali pola dan ciri khas penyakit daun kentang dengan tingkat akurasi yang baik dibandingkan metode lain. Namun, keunggulan ini diimbangi dengan biaya waktu komputasi yang lebih tinggi karena kemampuan CNN dalam mengekstrak fitur-fitur kompleks dari citra. Penggunaan CNN membutuhkan proses yang intensif dalam pelatihan karena memiliki banyak lapisan (*layer*) dan parameter yang harus disesuaikan dan dioptimalkan selama proses tersebut. Terutama saat menangani banyak citra, intensitas komputasi CNN semakin meningkat, memperlihatkan bahwa pembangunan model memerlukan sumber daya yang signifikan.

Selain kompleksitas model, faktor lain yang mempengaruhi adalah cara pelatihan model dilakukan secara sekuensial. Program pelatihan dibagi menjadi sejumlah instruksi atau tugas yang dieksekusi secara berurutan, satu per satu, dalam

satu alur eksekusi tunggal. Pendekatan ini dikenal sebagai komputasi sekuensial atau *single-threaded*. Dalam lingkungan *single-threaded*, satu inti pemrosesan menyelesaikan satu tugas pada satu waktu dan proses komputasi dilakukan langkah demi langkah.

Seiring dengan perkembangan model CNN, tidak hanya memungkinkan untuk meningkatkan komponen perangkat keras komputer, seperti peningkatan memori, CPU yang lebih cepat, GPU yang lebih canggih, dan sebagainya, tetapi juga dapat meningkatkan jumlah komputer yang terlibat secara bersamaan dalam pelatihan satu model [4]. Perlu diakui bahwa utilisasi rendah dan keterbatasan paralelisme dapat memperlambat waktu pelatihan, sementara *overhead* komunikasi antara CPU dan GPU serta skalabilitas terbatas juga menjadi kendala. Hal ini mencerminkan tren menuju komputasi yang lebih terdistribusi dan *parallel* untuk meningkatkan efisiensi.

Mendistribusikan *mini-batch* secara paralel menjadi alternatif menarik dalam upaya meningkatkan efisiensi pelatihan model, terutama pada CNN dan proses klasifikasi. Penelitian ini mengimplementasikan algoritma *Mini-Batch Gradient Descent*, suatu variasi dari algoritma *gradient descent*. Algoritma ini memperkenalkan konsep pengoptimalan berbasis *batch* yaitu parameter model diperbarui berdasarkan subset kecil dari data pelatihan pada setiap iterasi yang disebut *mini-batch* [5]. Penerapan *Mini-Batch Gradient Descent* membawa sejumlah keuntungan, terutama ketika dihadapkan dengan set pelatihan yang sangat besar [6].

Dengan memanfaatkan paralelisasi data yang membagi *mini-batch* menjadi batch yang lebih kecil dan mendistribusikannya ke beberapa pekerja (*Worker*), setiap pekerja dapat menjalankan model yang sama dan menghitung gradien secara independen berdasarkan *mini-batch* yang diterima, sebelum disinkronkan untuk memperbarui bobot model secara konsisten. Teknik ini sejalan dengan penelitian terkait, seperti yang dilakukan oleh Abadi M., et al. pada tahun 2016, yang menggarap sistem *parameter server*, dan menawarkan model pemrograman seragam yang memungkinkan pengguna memanfaatkan sistem heterogen dalam skala besar, baik untuk tugas produksi maupun eksperimen dengan pendekatan

baru. Pemanfaatan beberapa pekerja ini bertujuan untuk menjalankan bagian-bagian dari graf komputasi. Setiap pekerja akan bekerja pada satu atau lebih perangkat (seperti CPU) dan dapat berkomunikasi dengan pekerja lain untuk sinkronisasi dan pertukaran data selama proses pelatihan. Dalam skenario distribusi, pekerja beroperasi secara paralel untuk meningkatkan efisiensi dan kecepatan pelatihan model [7].

Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan *Mini-Batch Gradient Descent* menggunakan sistem komputasi terdistribusi yang mendukung multi-proses dalam pelatihan, dengan pendekatan data parallelism dan replikasi model. Implementasi ini akan memanfaatkan *TensorFlow* terdistribusi dengan strategi pelatihan *ParameterServerStrategy*, yang dijalankan pada lingkungan lokal dengan *TensorFlow* versi CPU-only. Harapannya, metode ini memiliki potensi untuk mempercepat proses pelatihan model *Convolutional Neural Network* (CNN), mengurangi biaya waktu komputasi, dan pada akhirnya meningkatkan produktivitas dalam konteks pertanian.

1.2 Rumusan Masalah

Berdasarkan latar belakang, maka rumusan masalah pada penelitian ini adalah, sebagai berikut:

1.2.1 Permasalahan

Dalam penelitian ini, permasalahan yang perlu dipecahkan adalah:

1. Bagaimana mengimplementasikan *Mini-Batch Gradient Descent* secara efisien pada lingkungan multi-proses untuk mencapai konvergensi yang lebih baik atau meningkatkan kecepatan pelatihan.
2. Bagaimana memilih pendekatan multi-proses yang optimal dalam penerapan *Mini-Batch Gradient Descent*.

1.2.2 Metode Usulan

Penelitian ini menggunakan teknik distribusi data dan komputasi untuk menjalankan *Mini-Batch Gradient Descent* secara paralel. Dilakukan perbandingan performa dari pendekatan: multi-proses yaitu beberapa *worker* menggunakan

pelatihan *parameter server* pada CPU dan sekuensial, dalam konteks klasifikasi penyakit daun kentang menggunakan *Convolutional Neural Network* (CNN). Analisis akan difokuskan pada faktor-faktor seperti kecepatan, konvergensi model, dan penggunaan sumber daya, dengan tujuan menentukan metode paralelisasi yang paling optimal.

1.2.3 Pertanyaan Penelitian

Dalam mengarahkan penelitian ini, pertanyaan penelitian yang akan dijawab meliputi:

1. Bagaimana kinerja *Mini-Batch Gradient Descent* dalam klasifikasi penyakit daun kentang menggunakan CNN, dan apa dampak dari implementasi multi-proses pada metode tersebut terhadap waktu pelatihan dan kualitas model?

1.3 Tujuan dan Manfaat

Tujuan dan manfaat yang dihasilkan dari penelitian ini adalah sebagai berikut :

1.3.1 Tujuan

Tujuan penelitian ini adalah untuk mengevaluasi dan memahami sejauh mana pengaruh penerapan integrasi *Mini-Batch Gradient Descent* terhadap optimisasi model *Convolutional Neural Network* (CNN) dalam klasifikasi penyakit daun kentang, terutama pada pengurangan waktu pelatihan model dan mempertahankan kualitas model melalui pendekatan paralelisasi dengan mengimplementasikan *multi-worker*.

1.3.2 Manfaat

Hasil dari penelitian ini diharapkan dapat membawa manfaat signifikan dengan mengurangi biaya waktu komputasi dalam pelatihan model *Convolutional Neural Network* (CNN) dan meningkatkan efisiensi klasifikasi penyakit daun kentang.

1.4 Batasan Masalah

Batasan yang diterapkan pada penelitian ini adalah:

1. Penelitian ini akan memanfaatkan dataset penyakit daun kentang [8] dari Kaggle (<https://www.kaggle.com/datasets/rizwan123456789/potato-disease-leaf-datasetp1d>) yang diambil pada saat pelaksanaan *capstone project* program

Bangkit 2023 pada tanggal 9 Mei – 12 Juni 2023. Dataset ini telah terlabel menjadi 3 kelas; *Early blight*, *late blight* dan *healthy*, dengan 4062 gambar dan terbagi menjadi *subset training* sebanyak 3251 gambar, *testing* dengan total 405 gambar, dan *validation* sebanyak 416 gambar.

2. Penelitian ini akan menggunakan model *Convolutional Neural Network* (CNN) yang telah dikembangkan sebelumnya. Dengan kata lain, penelitian ini tidak mencakup pengembangan model CNN dari awal; sebaliknya, fokus utamanya adalah pada upaya untuk mengoptimalkan model yang sudah ada melalui penerapan *Mini-Batch Gradient Descent* yang diparalelisasikan.

1.5 Sistematika Penulisan

Sistematika penulisan yang digunakan pada penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN, Bab ini mencakup pendahuluan yang memberikan ikhtisar mengenai strukturnya. Ini diuraikan dan dibagi menjadi latar belakang, perumusan masalah, batasan masalah, tujuan serta manfaat penelitian, dan sistematika penyusunan penelitian.

BAB II KAJIAN PUSTAKA, Bab ini memuat tinjauan pustaka yang melibatkan penelitian terdahulu yang relevan untuk mendukung penyusunan penelitian. Bab ini juga menguraikan landasan teori yang digunakan sebagai sumber dalam memahami permasalahan dalam melakukan penelitian ini.

BAB III METODE USULAN, Bab ini menguraikan metodologi yang diterapkan dalam penelitian ini. Bagian ini akan membahas tentang arsitektur sistem dan struktur penelitian, mencakup aspek seperti objek penelitian dan sumber data, serta prosedur yang terlibat dalam penelitian, termasuk pemahaman sistem, studi literatur, dan perancangan sistem.

BAB IV HASIL DAN PEMBAHASAN, Bab ini membahas penerapan pendekatan untuk menyelesaikan masalah yang telah dijelaskan sebelumnya, termasuk setiap langkah yang diperlukan untuk menyiapkan lingkungan dan alat-alat yang dibutuhkan.

BAB V PENUTUP, Bab ini memuat kesimpulan mengenai performa dan efisiensi dari pelatihan yang telah dilakukan serta kinerja yang dicapai. Selain itu, terdapat saran yang dapat dijadikan pertimbangan bagi berbagai pihak untuk penelitian selanjutnya.