

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di Indonesia pertanian sering kali menjadi mata pencaharian utama bagi sebagian besar penduduk di daerah pedesaan. Petani dan pekerja di sektor pertanian menggantungkan hidup mereka dari hasil produksi pertanian. Sehingga pertanian memiliki peran penting dalam pengurangan kemiskinan, penciptaan lapangan kerja, dan kesejahteraan ekonomi bagi komunitas lokal. Dari Badan Pusat Statistik (BPS) indeks produksi pertanian tahun 2020 meningkat sebesar 5,12 poin dibanding tahun 2019, yaitu dari 162,43 (angka tetap) pada tahun 2019 menjadi 167,55 (angka sementara) pada tahun 2020. Hal tersebut disebabkan oleh peningkatan indeks produksi peternakan hortikultura, perkebunan, dan peternakan[1].

Kacang merupakan salah satu sumber pangan penting di seluruh dunia. Kacang dapat digunakan sebagai bahan baku untuk berbagai jenis makanan, seperti bubur, pasta, roti, dan kudapan. Selain itu, kacang juga mengandung banyak nutrisi yang baik untuk kesehatan, seperti protein, serat, vitamin, dan mineral, jenis kacang-kacangan yang akan di bahas memiliki berbagai jenis, seperti *Seker*, *Barbunya*, *Bombay*, *Cali*, *Horoz*, *Sira*, dan *Dermason*. Kacang kering (*Phaseolus vulgaris L.*) merupakan tumbuhan berjenis pulse atau tanaman menghasilkan kulit (*Fabaceae - Leguminosae*), kacang juga menjadi tanaman terpenting dan memiliki produksi terbanyak di dunia[2].

Kacang-kacangan adalah sumber dari senyawa fenolik yang berperan dalam berbagai proses fisiologi dan metabolisme pada manusia[3]. Kacang kering juga berperan penting dalam pertanian di Turki. Setiap jenis kacang memiliki karakteristik yang unik, seperti bentuk, warna, dan ukuran. Selain itu, kacang juga dapat diolah menjadi berbagai produk makanan, seperti tepung, pasta, dan kudapan. Selain itu dari segi pertanian, cukup banyak petani yang belum mengetahui jenis kacang yang mereka tanam sehingga dengan penelitian ini dapat mempermudah dalam mendeskripsikan jenis kacang baru yang mereka tanam. karena terdapat banyak jenis kacang yang tersedia dengan karakteristik dan sifat yang berbeda, dan tujuan dasar penelitian ini untuk menentukan klasifikasi jenis kacang yang paling sesuai dari data yang tersedia.

Klasifikasi merupakan pengelompokan data atau objek baru ke dalam kelas atau label berdasarkan atribut-atribut tertentu[4], dengan klasifikasi ini dapat pula digunakan untuk mengklasifikasikan tanaman kacang dengan berbagai metode yang ada di dalamnya, klasifikasi

biasanya dilakukan dengan menggunakan algoritma pembelajaran mesin. Algoritma ini akan mempelajari pola atau hubungan antara fitur-fitur yang dimiliki oleh objek-objek dalam data latih atau training data, dan kemudian membuat model atau aturan-aturan untuk memprediksi kelas atau label dari objek-objek yang belum diketahui kelas nya dalam data uji atau testing data.

SMOTE adalah teknik yang digunakan untuk menangani ketidakseimbangan kelas dalam dataset, terutama ketika jumlah sampel dari kelas minoritas jauh lebih sedikit dibandingkan dengan kelas mayoritas. Ketidakseimbangan ini dapat menyebabkan algoritma machine learning lebih condong untuk memprediksi kelas mayoritas, sehingga kinerja model terhadap kelas minoritas menjadi buruk. Keunggulan dari smote dapat meningkatkan kinerja model, dengan menggunakan smote hal ini dapat mengurangi bias terhadap kelas mayoritas, dikarenakan data yang digunakan sangat tidak seimbang sehingga agar model menjadi seimbang, mencegah overfitting: Karena data sintetis dibuat dengan interpolasi, SMOTE juga menghindari masalah duplikasi data. Menggunakan Informasi Struktur Data: SMOTE menggunakan hubungan antara data minoritas untuk membuat sampel baru yang realistis.

Dalam penelitian ini data yang diperoleh memiliki kelas atau label yang cukup banyak lebih tepatnya 7 label sehingga dibutuhkan metode yang dapat menghitung multiclass data Salah satu metode yang cukup umum digunakan dalam klasifikasi yaitu *decision tree*, *Decision tree* adalah sebuah struktur pohon, dimana setiap simpul (*node*) pohon merepresentasikan atribut yang telah diuji. Setiap cabang merupakan pembagian hasil uji dan *node* daun (*leaf*) merepresentasikan kelompok kelas tertentu. Level *node* teratas dari sebuah *decision tree* adalah akar (*root*) yang biasanya berupa atribut yang paling memiliki pengaruh terbesar pada suatu kelas tertentu. Pada umumnya *Decision tree* melakukan strategi pencarian secara top-down untuk solusinya. Pada proses klasifikasi nilai atribut akan diuji dengan cara mempelajari jalur dari *node* akar (*root*) sampai ke *node* akhir (*leaf*) baru kemudian kelas baru akan ditentukan[5].

Dalam pemilihan menggunakan metode ini dikarenakan Mendukung pengambilan keputusan, hal ini sangat penting karena dengan dapat menghasilkan sebuah keputusan, maka proses klasifikasi akan jelas. Namun tidak hanya itu, metode ini juga menghasilkan model yang dapat diuji dan diverifikasi, hal ini juga dapat mempengaruhi kepercayaan bagi pengguna, lalu beberapa alasan lain seperti mudah dipahami, selain itu juga mampu mengatasi atribut yang tidak penting, klasifikasi kacang ini terdiri dari 16 atribut dan 13611 data, dikarenakan atribut terlalu banyak maka akan dilakukan seleksi fitur yang paling mempengaruhi korelasi dengan atribut lainnya.

Dan juga dapat mengatasi data yang hilang atau yang kosong sehingga perhitungan dapat selesai dengan akurat dan tepat, dari penelitian yang dikemukakan pada tahun 2022 yang ditulis oleh Aulia Rahmayanti, Lili Rusdiana, dan Suratno, penelitian ini menggunakan data nilai mahasiswa dengan data berjumlah 7 atribut. Penelitian ini mengulik tentang perbandingan antara metode naïve bayes dan decision tree, penelitian ini mengukur tingkat kelulusan tepat waktu mahasiswa STMIK Palangkaraya yang memperoleh hasil akurasi dari kedua algoritma tersebut menunjukkan tingkat akurasi algoritma C4.5 sebesar 90% sedangkan algoritma Naïve Bayes sebesar 85%. Algoritma C4.5 juga menghasilkan nilai recall dan precision sebesar 92% dan 94%, nilai ini lebih baik dibandingkan algoritma Naïve Bayes dengan nilai 86% dan 93%.[6]

Beberapa penelitian sebelumnya yaitu pada tahun 2020 Hananda Hafizan, Anggita Nadia Putri melakukan penelitian tentang gizi bayi dengan klasifikasi, data yang digunakan berisi data 3 atribut dan memiliki 2 kelas, penelitian ini digunakan untuk dapat melakukan pemeriksaan dan pengukuran antropometri agar dapat mengetahui status gizi balita dapat mendatangi tempat pelayanan kesehatan masyarakat dan menghasilkan perhitungan software RapidMiner menunjukkan akurasi menggunakan tools performance adalah 100% [7]. dan juga penelitian dari Asmaul Husnah Nasrullah tahun 2021 yang bertema tentang penerapan *decision tree* pada produk terlaris, penelitian ini menggunakan data dengan atribut dan 2 kelas. Dari hasil penelitian yang telah dilakukan untuk data penjualan dan stok barang di UD. Cipta Karya Gorontalo maka diperoleh akurasi sebesar 90% dan nilai AUC 0.709 [5].

1.2 Rumusan masalah

1.2.1 Permasalahan

Klasifikasi kacang terdiri dari 16 atribut dan 13611. Sistem membutuhkan rule/aturan agar dapat mengklasifikasikan dengan baik. Proses klasifikasi membutuhkan data yang telah dinormalisasi dan seimbangkan melakukan seleksi atribut.

1.2.2 Metode Usulan

Dari permasalahan yang telah jelaskan, penelitian ini menggunakan metode *Decision tree* dikarenakan kemampuan untuk menangani data yang kompleks dan dapat memberikan interpretasi yang mudah dipahami. Metode ini menggunakan sistem pohon keputusan yang dibangun dari sebuah *node* akar yang merepresentasikan seluruh dataset,

kemudian dilakukan *splitting* pada *node* tersebut berdasarkan informasi atau atribut yang paling penting untuk membagi dataset menjadi dua atau lebih subset yang lebih homogen.

1.2.3 Pertanyaan Penelitian

1. Bagaimana pengaruh penyeimbangan data Smote terhadap hasil performa (akurasi, presisi, *recall*) sistem?
2. Bagaimana hasil dari evaluasi dengan menggunakan Confusion Matrix terhadap hasil performa (akurasi, presisi, *recall*) sistem?

1.3 Tujuan dan manfaat

1.3.1 Tujuan Penelitian

Penelitian ini ditujukan untuk mengetahui bagaimana proses klasifikasi serta bagaimana hasil mengklasifikasi data dengan menggunakan metode *decision tree* untuk data kacang. Dan juga untuk mengetahui bagaimana pengaruh penyeimbangan, seleksi fitur untuk mengetahui hasil tingkat akurasi, presisi, dan *recall*, dari metode klasifikasi *Decision tree* menggunakan dataset kacang.

1.3.2 Manfaat Penelitian

Penelitian ini diharapkan untuk bermanfaat bagi peneliti yang memiliki data kacang belum diidentifikasi, dan dapat dijadikan sebagai referensi untuk penelitian selanjutnya tentang penggunaan metode *decision tree* dalam proses klasifikasi, serta dapat meningkatkan wawasan pembaca dalam penerapan sistem algoritma C4.5.

1.4 Batasan-Batasan

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan data berasal dari peneliti bernama Murat Koklu dan berikut adalah link download.
(<https://www.kaggle.com/datasets/muratkokludataset/dry-bean-dataset?resource=download>)
2. Data yang digunakan berjumlah 13611 data dengan 7 *class* dan 16 atribut data.

3. Data tiap *class* yaitu *Seker* 2027 data, *Barbunya* 1322 data, *Bombay* 522 data, *Cali* 1630 data, *Dermason* 3546 data, *Horoz* 1928 data dan *Sira* 2636 data

1.5 Sistematika Skripsi

Sistematika penulisan Skripsi ini terdiri dari beberapa bab, yaitu:

BAB I PENDAHULUAN

Pada poin ini membahas tentang latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, serta sistematika penulisan laporan.

BAB II KAJIAN PUSTAKA

Pada bab ini membahas tentang landasan teori dalam penyusunan penelitian ini. Landasan teori tersebut diantaranya adalah bagaimana proses klasifikasi untuk mengelompokkan jenis kacang berdasarkan atribut yang dimiliki.

BAB III METODE USULAN

Pada bab ini membahas mengenai metode yang direncanakan, arsitektur (Input, Proses, Output), data penelitian, tahapan penelitian serta rencana pengujian dan juga mengandung jadwal penelitian untuk penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tentang bagaimana proses yang dilakukan serta dibahas tentang bagaimana perhitungan yang dijalankan mulai dari data sebelum perhitungan dan juga langkah-langkah yang dilakukan sehingga sesuai kriteria dari metode yang dilakukan, sehingga menghasilkan nilai yang akurat dan benar.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi tentang bagaimana hasil yang diperoleh dari perhitungan serta poin-poin yang dapat di garis bawahi untuk mengembangkan penelitian berikutnya, serta menambahkan sedikit saran untuk pembaca yang akan menggunakan penelitian ini sebagai referensi.

DAFTAR PUSTAKA

Berisi tentang daftar jurnal-jurnal atau referensi yang dijadikan acuan dalam penelitian ini