# BAB I
# PENDAHULUAN

## 1.1 Latar Belakang

Indonesia menjadi salah satu negara yang terkenal dengan keanekaragaman hayati yang kaya dan melimpah. Dari keanekaragaman tersebut dibagi kembali menjadi beberapa jenis salah satunya adalah jamur. Secara garis besar, jamur dibagi menjadi 2 jenis yaitu jamur yang dapat dimakan dan jamur yang beracun, bahkan beberapa diantaranya sangat mematikan[1]. Namun, kenyataannya masih terdapat banyak kasus dimana orang tidak bisa membedakan jenis jamur sehingga dapat menyebabkan efek samping berupa keracunan.

Tampilan jamur sebagai salah satu faktor utama dalam proses menentukan jenis jamur, diantaranya: bentuk tudung jamur, permukaan tudung jamur dan lampiran insang. Selain itu, tipe cincin yang ada pada jamur dapat memberikan fitur diagnostik kritis [2]. Hal ini menunjukan bahwa dalam proses menentukan suatu jenis jamur diperlukan pertimbangkan dari beberapa faktor yang dapat mempengaruhi interaksi antar fitur[3].

Dalam beberapa penelitian sebelumnya seperti [4][5][6][7][8], dilakukan eksperimen untuk bisa menyelesaikan permasalahan dibidang klasifikasi jenis jamur. Namun, dalam penelitian tersebut belum menemukan proses yang memuaskan. Beberapa diantaranya memiliki masalah seperti ketidakseimbangan data dan skala fitur yang tidak menentu. Menurut penelitian [9], kumpulan data yang tidak seimbang dapat menyebabkan prediksi bias, di mana model mendukung kelas mayoritas, mengakibatkan akurasi yang buruk untuk kelas minoritas. Selain itu menurut penelitian [10], skala fitur yang tidak menentu dapat menghambat kemampuan model untuk menggeneralisasi terutama dalam kumpulan data yang tidak seimbang di mana besarnya fitur dapat mempengaruhi prediksi secara tidak proporsional.

*Synthetic Minority Over-sampling Technique (SMOTE)* menjadi salah satu Teknik untuk mengatasi ketidakseimbangan data. *SMOTE* sangat baik dalam

menghasilkan sampel sintetis untuk kelas minoritas, meningkatkan representasi kelas yang kurang terwakili dalam kumpulan data. Selain itu, dengan menyediakan lebih banyak data pelatihan untuk kelas minoritas, *SMOTE* dapat menyebabkan kinerja prediktif yang lebih baik dan mengurangi bias dalam model klasifikasi [11].

Dalam proses analisis machine learning terdapat beberapa metode yang sering digunakan diantaranya *Naïve Bayes*, *Logistic Regression* dan *KNN*. *Naïve Bayes* efisien secara komputasi, sehingga cocok untuk kumpulan data besar dan data dimensi tinggi, karena memiliki waktu pelatihan yang singkat. Namun, *Naïve Bayes* kurang optimal dalam menangani data yang hilang, karena secara inheren tidak menangani kasus seperti itu dengan baik[12][13]. *Logistic Regression* memiliki kinerja baik, terutama ketika banyak variabel kategoris, karena dapat menangani hasil biner secara efektif[14]. *KNN* dapat menangani tugas klasifikasi dan regresi, menangkap hubungan non-linier secara efektif. Namun, *KNN* memerlukan pemindaian seluruh kumpulan data pelatihan untuk setiap klasifikasi, yang mengarah ke peningkatan kompleksitas waktu, terutama dengan kumpulan data besar[15].

*Support Vector Machine (SVM)* merupakan salah satu metode yang lebih tahan terhadap dimensi tinggi dibandingkan *KNN* karena mencari *hyperplane* optimal untuk memisahkan kelas dalam ruang fitur. Namun, meskipun *SVM* lebih *robust* terhadap data berdimensi tinggi, metode ini tetap sensitif terhadap skala fitur, sehingga normalisasi atau standarisasi data seringkali diperlukan untuk hasil yang optimal[16]. *SVM* juga dikenal karena akurasinya yang tinggi dalam tugas klasifikasi, mencapai akurasi hingga 95% dalam aplikasi medis seperti deteksi tumor otak[17]. Namun, dalam tugas klasifikasi, dataset berdimensi tinggi seringkali menyebabkan *overfitting*, yang dapat memperumit proses klasifikasi. Hal ini terjadi karena semakin banyak fitur dalam dataset, semakin besar kemungkinan model terlalu menyesuaikan diri dengan data latih, sehingga kurang mampu menggeneralisasi data baru dengan baik[18].

Untuk membedakan jenis jamur antara yang beracun dan yang tidak beracun, terdapat berbagai tantangan dalam proses klasifikasi. Ketidakseimbangan

data serta skala fitur yang tidak konsisten dapat menurunkan akurasi model dalam mengklasifikasikan jenis jamur. Meskipun berbagai metode machine learning telah diterapkan, performa model dalam menangani ketidakseimbangan data masih belum optimal. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan metode yang lebih efektif dalam mengklasifikasikan jenis jamur, dengan mempertimbangkan solusi untuk mengatasi ketidakseimbangan data, skala fitur yang tidak seragam serta mengurangi risiko *overfitting* pada dataset berdimensi tinggi.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah diatas, terdapat beberapa rumusan masalah, antara lain:

1. Bagaimana pengaruh penerapan teknik *SMOTE* terhadap hasil keseluruhan dalam klasifikasi jamur beracun?
2. Bagaimana perbandingan algoritma *machine learning* yang dilakukan dalam proses evaluasi menggunakan *accuracy, precision, recall, F1-score dan confusion matrix*?
3. Algoritma mana yang memberikan hasil terbaik dalam klasifikasi jamur?

## 1.3 Batasan Masalah

Dalam penelitian ini terdapat batasan masalah yang ditentukan oleh peneliti, antara lain:

1. Algoritma *machine learning* yang digunakan antara lain *Naïve Bayes, Logistic Regression dan Support Vector Machine*.
2. Teknik seleksi fitur yang digunakan adalah *Mutual Information*.
3. Teknik mengatasi tidak keseimbangan data adalah *SMOTE*.
4. Teknik normalisasi data yang digunakan adalah *MinMaxScaler*.
5. *Hyperparameter tuning* yang digunakan dalam penelitian ini ditentukan secara *default*.
6. Teknik evaluasi kinerja akhir menggunakan *accuracy*, *precision*, *recall*, *F1-score* dan *confusion matrix*.

## 1.4 Tujuan Penelitian

Penelitian ini dilakukan dengan tujuan untuk memperoleh pemahaman yang lebih mendalam mengenai efektivitas algoritma *machine learning* dalam klasifikasi jamur beracun dan tidak beracun. Adapun tujuan spesifik dari penelitian ini adalah sebagai berikut:

1. Untuk menganalisis dampak penggunaan teknik *SMOTE* terhadap hasil keseluruhan dalam klasifikasi jamur.
2. Untuk membandingkan performa tiga algoritma *machine learning (Naïve Bayes, Logistic Regression,* dan *Support Vector Machine)* dalam proses klasifikasi jamur.
3. Untuk menentukan algoritma terbaik di antara *Naïve Bayes, Logistic Regression,* dan *Support Vector Machine* dalam klasifikasi jamur beracun dan tidak beracun.

## 1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

a. Teoritis: Penelitian ini bertujuan untuk meningkatkan pemahaman kita tentang klasifikasi jamur menggunakan metode *machine learning*, terutama dengan melihat seberapa efektif algoritma *Naïve Bayes, Logistic Regression,* dan *Support Vector Machine*. Hasil penelitian ini diharapkan dapat digunakan sebagai referensi untuk studi selanjutnya di bidang *data mining* dan klasifikasi data tidak seimbang.

b. Praktis: Penelitian ini diharapkan dapat membantu peneliti, praktisi, serta pihak yang berkepentingan dalam bidang pertanian dan keamanan pangan dalam mengembangkan sistem klasifikasi jamur yang lebih akurat. Dengan demikian, penelitian ini dapat berkontribusi dalam mengurangi risiko konsumsi jamur beracun serta meningkatkan keselamatan bagi masyarakat yang mengkonsumsi jamur liar.

## 1.6 Sistematika Penulisan

Skripsi ini disusun dalam lima bab utama yang masing-masing menjelaskan tahap-tahap penting dalam proses penelitian, sebagai berikut:

**BAB I PENDAHULUAN**, Bab ini membahas latar belakang permasalahan mengenai pentingnya klasifikasi jamur beracun, rumusan masalah, batasan masalah, tujuan, dan manfaat penelitian, serta sistematika penulisan skripsi secara umum.

**BAB II TINJAUAN PUSTAKA**, Bab ini menyajikan studi literatur dari penelitian-penelitian sebelumnya yang relevan serta dasar teori yang mendukung penelitian. Teori yang dijelaskan meliputi algoritma *machine learning* (*Naïve Bayes, Logistic Regression,* dan *SVM*), *preprocessing data*, *SMOTE*, *Mutual Information*, dan evaluasi model klasifikasi seperti *confusion matrix*.

**BAB III METODE PENELITIAN**, Bab ini menjelaskan secara sistematis tahapan-tahapan penelitian yang dilakukan, mulai dari pengumpulan data, *exploratory data analysis (EDA)*, *preprocessing*, *modeling*, hingga evaluasi model. Penelitian menggunakan dataset dari *Kaggle/UCI*, dengan *tools* seperti *Python* di *Google Colab* dan berbagai *library* pendukung seperti *Scikit-learn* dan *Imbalanced-learn*.

**BAB IV HASIL DAN PEMBAHASAN**, Bab ini menyajikan hasil dari setiap tahapan yang telah dilakukan dalam penelitian, mulai dari analisis data, proses *preprocessing*, hasil *modeling* menggunakan tiga algoritma, serta evaluasi menggunakan *classification report* dan *confusion matrix*. Disajikan pula hasil perbandingan sebelum dan sesudah penggunaan teknik *SMOTE*.

**BAB V PENUTUP**, Bab ini berisi kesimpulan dari hasil penelitian serta saran untuk penelitian selanjutnya, seperti penggunaan teknik *tuning hyperparameter* atau penerapan metode *ensemble* untuk meningkatkan performa model.