

BAB II KAJIAN PUSTAKA

2.1 Penelitian Relevan

Dalam melakukan penelitian yang berjudul Klasifikasi Jamur Layak Konsumsi Berdasarkan Ciri Morfologi Menggunakan Algoritma Klasifikasi *K-Nearest Neighbor* ini, terdapat beberapa penelitian terkait yang telah dilakukan sebelumnya dan dapat dijadikan sebagai acuan untuk melakukan penelitian ini. Hal ini bertujuan sebagai bentuk perbandingan dan pengukuran tingkat keberhasilan penelitian ini.

Penelitian yang dilakukan oleh (Prayoga , Nawangsih, & Wiyatno, 2019) dengan judul “Implementasi Metode Naïve Bayes Classifier Untuk Identifikasi Jenis Jamur”. Indonesia merupakan Negara dengan kekayaan alam dan keanekaragaman hayati yang berlimpah. Dengan iklim yang sangat cocok untuk tumbuhnya berbagai jenis tanaman, maka salah satu komoditas yang memiliki prospek cerah adalah pertanian dalam bidang *hortikultura* sebagai swasembada pangan, selain sebagai salah satu cara untuk meningkatkan perekonomian masyarakat, hal ini juga diharapkan dapat meningkatkan gizi masyarakat. Salah satu komoditas *hortikultura* yang banyak dijumpai dalam hal pangan adalah budidaya jamur. Terdapat ribuan jenis jamur yang tumbuh dalam wilayah dengan suhu yang berbeda-beda. Salah satu jenis jamur yang paling banyak dijumpai di alam terbuka dan dibudidayakan untuk dikonsumsi adalah jenis jamur yang tergolong dalam genus *agaricus* dan *lepiota*. Namun dari banyaknya jamur genus *agaricus* dan *lepiota* yang tumbuh tidak semua jamur dapat dikonsumsi, terdapat pula beberapa jamur yang beracun atau tidak aman untuk dikonsumsi. Dengan ciri-ciri jamur yang hampir mirip antara satu dengan yang lainnya tentunya menjadi kesulitan tersendiri untuk membedakan antara jamur yang aman dan tidak aman untuk dikonsumsi. Sehingga tidak sedikit kasus kematian yang diakibatkan oleh keracunan jamur. Untuk membedakan jenis jamur yang dapat dikonsumsi dan beracun perlu adanya pengklasifikasian terhadap dua jenis jamur tersebut. Metode klasifikasi yang digunakan adalah metode *Naïve Bayes Classifier*,

metode ini dapat memprediksi peluang di masa depan berdasarkan pengalaman sebelumnya. Sehingga dengan adanya penelitian ini dapat mengklasifikasikan jenis jamur yang layak dikonsumsi dan beracun. Hasil yang diperoleh dari penelitian ini adalah metode *Naïve Bayes Classifier* dapat diterapkan dalam mengidentifikasi jamur *agaricus* dan *lepiota* masuk dalam kategori layak konsumsi dan beracun dengan akurasi yang tergolong *good classification* sebesar 86,06%.

Penelitian dengan judul “Identifikasi Jamur Menggunakan Metode *K-Nearest Neighbor* Dengan Ekstraksi Ciri Morfologi” yang diteliti oleh (Zubair & Muslihk, 2017). Dalam penelitiannya peneliti membahas mengenai jamur yang menjadi salah satu bahan pangan yang banyak dikonsumsi oleh masyarakat umum. Namun dari sekian banyak jamur yang tumbuh, tidak semua jenis jamur dapat dikonsumsi oleh masyarakat, dikarenakan terdapat beberapa jamur yang beracun. Untuk membedakan jenis-jenis jamur yang beracun dan jamur yang layak dikonsumsi, terdapat beberapa cara yang dapat dilakukan, salah satu cara yang dapat dijadikan sebagai tolak ukur pembedanya adalah berdasarkan ciri morfologi jamur tersebut. Adapun ciri morfologi yang dimaksud adalah ciri-ciri yang dapat dilihat dari sebuah jamur, seperti bentuk payung, warna, bau, dan lain sebagainya. Untuk mengklasifikasikan jenis jamur layak konsumsi dan beracun, peneliti memanfaatkan ekstraksi ciri morfologi dengan metode klasifikasi KNN. Data yang digunakan dalam penelitian ini di dapat dari situs *UCI Machine Learning Repository*. Selanjutnya data yang didapat diklasifikasikan menggunakan metode KNN dengan memanfaatkan *third party Add-ins* dari Ms. Excel yaitu XLSTAT. Setelah melakukan percobaan sebanyak 10 kali dengan nilai k yang berbeda-beda, nilai akurasi paling tinggi didapatkan sebesar 99% pada percobaan ke-10 dengan nilai $k=60$. Dengan demikian dapat disimpulkan bahwa metode klasifikasi KNN dapat mengklasifikasikan jamur layak konsumsi dan beracun dengan sangat baik.

Penelitian relevan ketiga dilakukan oleh (Nasution, Nababan, Syaliman, Novelan, & Jannah, 2019) dengan judul “Penerapan *Principal Component Analysis* (PCA) dalam Penentuan Faktor Dominan yang Mempengaruhi Pengidap Kanker Serviks”. Kanker serviks merupakan salah satu jenis kanker yang paling

banyak diderita oleh masyarakat. Penderita awal penyakit kanker umumnya tidak terlihat seperti orang yang sedang sakit. Skrining merupakan bentuk upaya deteksi dini yang dapat dilakukan untuk mengidentifikasi penyakit atau kelainan yang secara klinis belum jelas dengan menggunakan tes, pemeriksaan atau prosedur tertentu. Di mana skrining ini dapat digunakan untuk membedakan seseorang yang menderita suatu kelainan namun terlihat sehat secara cepat. Data yang digunakan dalam penelitian ini adalah dataset *Cervical Cancer* sebanyak 858 data dengan 32 variabel serta 4 label kelas meliputi : *Hinselmann*, *Schiller*, *Cytology* dan *Biopsy*, dataset didapat dari situs *UCI Machine Learning Repository*. Dengan banyaknya variabel skrining yang digunakan yakni sebanyak 32 variabel, peneliti bermaksud untuk menyederhanakan variabel skrining dengan menghilangkan beberapa variabel yang kurang dominan tanpa mengurangi maksud dan tujuan dari data aslinya menggunakan PCA. Setelah dilakukan pengujian, hasil yang didapat adalah algoritma PCA dapat mereduksi atau menyederhanakan variabel skrining untuk pendeteksian dini kanker serviks yang awalnya sebanyak 32 variabel menjadi 9 variabel dengan total varians sebesar 99%. Sehingga dapat disimpulkan bahwa algoritma PCA dapat mereduksi variabel skrining kanker serviks dengan sangat baik.

Dari ketiga penelitian terkait di atas, terdapat beberapa persamaan sekaligus perbedaan dengan penelitian ini. Persamaan antara penelitian ini dengan penelitian pertama terletak pada topik yang dibahas yakni mengklasifikasi jamur layak konsumsi dan beracun, kemudian perbedaannya terletak pada metode yang digunakan, di mana penelitian ini menggunakan metode KNN dengan dataset jamur yang didapat dari situs *UCI Machine Learning Repository*, sedangkan pada penelitian pertama metode yang digunakan adalah *Naïve Bayes Classifier* dengan menggunakan data yang sama dengan penelitian ini. Penelitian terkait kedua memiliki topik sekaligus metode yang sama dengan penelitian ini yakni klasifikasi jamur menggunakan algoritma KNN, yang membedakan adalah pada penelitian ini ditambahkan metode PCA untuk menyeleksi atribut atau variabel dari dataset yang digunakan. Kemudian pada penelitian terkait terakhir perbedaannya terletak pada topik yang diteliti, namun memiliki kesamaan dalam metode yang diterapkan yakni metode PCA untuk menyeleksi atribut.

2.2 Landasan Teori

2.2.1 Data Mining

Menurut (Cahyanti, Rahmayani, & Husniar, 2020) data mining diartikan sebagai proses penemuan pola pada suatu data. Adapun tugas dari data mining dibagi menjadi enam kelompok yaitu, deskripsi, estimasi, asosiasi, klasifikasi, prediksi, dan *clustering*.

Data mining merupakan suatu proses untuk menemukan informasi-informasi terbaru yang bermanfaat dan belum diketahui sebelumnya di dalam suatu *database* yang besar, di mana proses yang digunakan adalah teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin (Mulyati, Husein, & Ramdhan, 2020).

Menurut (Mustakim, Giantika O, 2016) terdapat tiga karakteristik dari data mining yaitu:

1. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
2. Data mining biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
3. Data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategi

2.2.2 Klasifikasi

Klasifikasi merupakan proses penemuan fungsi atau model yang membedakan kelas data dengan tujuan bisa digunakan untuk memprediksi kelas dari objek yang belum diketahui labelnya. Berikut macam-macam algoritma klasifikasi yang telah banyak digunakan diantaranya, *Decision/classification tree*, *Bayesian classifiers*/*Naïve Bayes classifiers*, *Neural networks*, Analisa Statistik, Algoritma Genetika, *Rough sets*, *k-nearest neighbor*, Metode *Rule Based*, *Memory based reasoning*, dan *Support vector machines* (SVM) (Annur, 2018).

Menurut (Nasution, Khotimah, & Chamidah, 2019) klasifikasi adalah pengelompokan data atau objek baru kedalam suatu kelas atau label

berdasarkan variable-variabel yang telah ada. Klasifikasi dilakukan dengan tujuan memprediksi kelas dari suatu objek yang belum diketahui kelasnya.

Tahapan dari klasifikasi ada tiga:

1. Pembangunan model

Tahap pertama adalah proses pembangunan model berdasarkan data latih yang telah memiliki atribut dan kelas.

2. Penerapan model

Pada tahap ini data-data diterapkan untuk menentukan kelas dari data atau objek yang baru.

3. Evaluasi

Tahap terakhir adalah proses evaluasi, proses evaluasi ini bertujuan untuk melihat hasil akurasi dari kedua tahap sebelumnya yakni pembangunan dan penerapan model.

2.2.3 Jamur

Jamur merupakan *kingdom fungi*, yang mana jamur tidak memiliki daun dan akar yang sejati seperti halnya tumbuhan. Umumnya jamur memiliki sel banyak (multiseluler) seperti jamur tempe dan jamur merang, namun ada pula jamur yang hanya memiliki sel tunggal (uniseluler) salah satunya adalah ragi. Jamur multiseluler tersusun dari benang-benang yang di sebut hifa. Yang mana jika dilihat menggunakan mikroskop, hifa memiliki bentuk yang bersekat-sekat (bersepta) namun ada pula yang tidak bersekat (Zubair & Muslikh, 2017).

Menurut (Setiorini P, Astian, & E, 2018) jamur bersifat pathogen dan dapat tumbuh dan hidup pada pH kurang dari 7, tanah yang lembab, memiliki tajuk yang rapat (70% - 95%), suhu rendah (20°C - 26°C), kelembapan yang tinggi, dan pohon yang telah lapuk, namun ada pula jamur yang tumbuh pada batang kayu yang masih hidup.

Berdasarkan kedua jurnal diatas dapat disimpulkan bahwa jamur tidak sama dengan tumbuhan yang mendapatkan makanannya dengan proses fotosintesis karena memiliki daun yang dapat menghasilkan klorofil, lain halnya dengan jamur yang mendapatkan makanan dari lingkungan sekitarnya

dengan bantuan hifa. Jamur hanya bisa tumbuh dan berkembang pada daerah-daerah yang lembab.

2.2.4 *Principal Component Analysis*

Principal Component Analysis (PCA) merupakan kombinasi linear dari variabel awal yang secara geometris kombinasi linear ini merupakan sistem koordinat baru yang diperoleh dari rotasi sistem semula. Metode *Principal Component Analysis* sangat berguna jika data yang ada memiliki jumlah variabel dalam jumlah besar dan memiliki korelasi antar variabelnya. Perhitungan dari atau *Principal Component Analysis* didasari pada perhitungan nilai *eigen* dan vektor *eigen* yang menyatakan penyebaran data dari suatu dataset (Fauzi, Supriyadi, & Maulidah, 2020).

Dalam jurnal (Nasution, Nababan, Syaliman, Novelan, & Jannah, 2019) Tujuan dari prosedur pengerjaan *Principal Component Analysis* adalah untuk menyederhanakan atribut atau fitur dengan hanya memilih atribut yang paling dominan dari dataset tanpa mengurangi maksud dan tujuan dari data asli. Adapun tahapan dalam penerapan algoritma PCA adalah sebagai berikut:

1. Menghitung matrik varians kovarian dari data observasi.

$$Var(x) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n - 1)} \quad (2.1)$$

$$Cov(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n - 1)} \quad (2.2)$$

Keterangan:

x = atribut acak x

x_i = nilai dari atribut x ke- i

\bar{x} = nilai rata-rata dari atribut x

y = atribut acak y

y_i = nilai dari atribut y ke- i

\bar{y} = nilai rata-rata dari atribut y

n = jumlah data observasi

2. Mencari *eigenvalues* dari matrik kovarian yang telah diperoleh.

$$Det(M - \lambda I) = 0 \quad (2.3)$$

Keterangan:

M = matrik $n \times n$

λ = nilai *eigenvalue*

I = matrik identitas

3. Menentukan nilai *proporsi Principal Component*

$$PC(\%) = \frac{\text{NilaiEigen}}{\text{VarianceCovarian}} \times 100\% \quad (2.4)$$

4. Mencari *eigenvector* dari matrik kovarian yang telah diperoleh.

$$Mv = \lambda v \quad (2.5)$$

Keterangan:

M = matrik $n \times n$

λ = nilai *eigenvalue*

v = *eigenvector*

Sehingga dihasilkan kombinasi linier yaitu:

$\lambda_1, \lambda_2, \lambda_3 \dots \lambda_n$ *eigenvalue* dari matrik M

$v_1, v_2, v_3 \dots v_n$ *eigenvector* berdasarkan *eigenvalue*-nya (λ_n)

2.2.5 *K-Nearest Neighbor*

K-Nearest Neighbor merupakan suatu metode klasifikasi terhadap sekumpulan data berdasarkan data yang telah terklasifikasikan sebelumnya. Adapun prinsip kerja algoritma KNN adalah bekerja berdasarkan jarak terdekat dari data uji ke data latih (Cahyanti, Rahmayani, & Husniar, 2020). Metode KNN dibagi dalam dua fase yaitu pembelajaran (*training*) dan pengujian atau klasifikasi (*testing*). Pada fase *training* proses yang dilakukan hanya menyimpan vektor-vektor fitur dan klasifikasi dari data pembelajaran, sedangkan pada fase *testing* dilakukan perhitungan untuk data yang akan diuji (yang belum diketahui klasifikasinya) (Baharuddin, Hasanuddin, & Azis, 2019).

Langkah-langkah metode klasifikasi *K-Nearest Neighbor* :

1. Menentukan nilai K

Tahap pertama yang harus dilakukan adalah menentukan nilai K , nilai K dapat ditentukan dengan bebas.

2. Menghitung jarak

Tahap yang kedua adalah menghitung jarak antara data yang akan diuji dengan data latih. Dalam penelitian ini metode yang digunakan untuk menghitung jarak ada 2, yaitu *Euclidean Distance* dan *Manhattan Distance*.

Berikut persamaan dari kedua metode perhitungan jarak tersebut:

$$\text{Euclidean Distance} = d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.6)$$

$$\text{Manhattan Distance} = d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (2.7)$$

Keterangan:

d = jarak antara data *training* (x) dan data *testing* (y)

x_i = nilai dari data *training* (x) ke- i

y_i = nilai dari data *testing* (y) ke- i

3. Menentukan nilai tetangga terdekat

Tahap ketiga adalah menentukan nilai tetangga terdekat dari masing-masing jarak. Di mana hasilnya berupa kategori kelas yang akan dijadikan sebagai acuan ketika klasifikasi dilakukan berdasarkan nilai tetangga terdekat.

4. Menentukan klasifikasi

Tahap terakhir adalah menentukan klasifikasi berdasarkan nilai tertinggi tetangga terdekat. Contoh nilai tetangga terdekat adalah jamur masuk kedalam kategori kelas beracun maka pada data uji yang belum diketahui kelasnya jamur tersebut tergolong ke dalam kelas beracun.

Berikut adalah kelebihan serta kekurangan penerapan metode KNN:

a. Kelebihan

- Mudah dimengerti dan diimplementasikan
- Efektif untuk data besar
- *Nonlinier*

b. Kekurangan

- Perlu menentukan parameter K
- Sensitif terhadap data pencilan
- Rentan pada dimensionalitas tinggi
- Nilai komputasi tinggi

- Tidak bisa menangani nilai yang hilang secara implisit
- Rentan terhadap variabel non-informatif
- Rentan terhadap perbedaan rentang variabel

2.2.6 Confusion Matrix

Confusion matrix adalah suatu metode evaluasi yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Akurasi dalam klasifikasi adalah persentase ketepatan rekam data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi (Sari, Astuti, Saluza, Faradillah, & Yunita, 2021).

Tabel 2. 1 Model *Confusion Matrix*

		<i>Predicted Class</i>	
		<i>Class Yes</i>	<i>Class No</i>
<i>Actual Class</i>	<i>Class Yes</i>	<i>TP</i>	<i>FN</i>
	<i>Class No</i>	<i>FP</i>	<i>TN</i>

Keterangan variabel dari table 2.1 adalah sebagai berikut:

TP (True Positive) : Jumlah data dengan nilai aktual bernilai positif dan nilai prediksinya positif

FP (False Positive) : Jumlah data dengan nilai aktual bernilai negatif dan nilai prediksinya positif

FN (False Negative) : Jumlah data dengan nilai aktual bernilai positif dan nilai prediksinya negatif

TN (True Negative) : Jumlah data dengan nilai aktual bernilai negatif dan nilai prediksinya negatif

Untuk mengetahui hasil keberhasilan dari klasifikasi yang dilakukan dapat diketahui dengan menghitung hasil akurasi. Akurasi dapat didefinisikan sebagai tingkat kedekatan antara nilai prediksi yang didapatkan dengan nilai aktual. Berikut rumus untuk menghitung nilai akurasi:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.8)$$