

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Gambaran Umum Objek Penelitian

Pada penelitian ini, objek penelitian utamanya adalah *dataset* tentang *Livestock Symptoms and Diseases*. Pemilihan *Livestock Symptoms and Diseases* ini diambil dari *website* kaggle.com pada saat penelitian dilakukan (23 Februari 2024). *Dataset* ini dibuat secara sintetis untuk tujuan penelitian dan dapat digunakan untuk berbagai visualisasi dan analisis. Pemilihan ini didasarkan atas kelengkapan informasi pada dataset yang mencakup objek-objek penelitian terkait hubungan penyakit pada hewan ternak kambing dengan suhu tubuh.

	Animal	Age	Temperature	Symptom 1	Symptom 2	Symptom 3	Disease
0	cow	3	103.1	depression	painless lumps	loss of appetite	pneumonia
1	buffalo	13	104.5	painless lumps	loss of appetite	depression	lumpy virus
2	sheep	1	100.5	depression	painless lumps	loss of appetite	lumpy virus
3	cow	14	100.3	loss of appetite	swelling in limb	crackling sound	blackleg
4	sheep	2	103.6	painless lumps	loss of appetite	depression	pneumonia
...	...	...	...	...	...	...	...
43773	goat	9	102.2	swelling in muscle	lameness	crackling sound	blackleg
43774	buffalo	3	101.8	loss of appetite	sores on mouth	difficulty walking	foot and mouth
43775	buffalo	15	104.1	swelling in extremities	chills	shortness of breath	anthrax
43776	cow	9	104.9	crackling sound	lameness	swelling in muscle	blackleg
43777	buffalo	4	103.5	difficulty walking	sores on gums	loss of appetite	foot and mouth

43778 rows × 7 columns

Gambar 3. 1 Dataset Livestock Symptoms and Diseases

Gambar 3.1 merupakan dataset *Livestock Symptoms and Diseases* yang digunakan sebagai objek penelitian utama dalam penelitian ini. Pada penelitian untuk pembuatan model pendeteksian penyakit hewan ternak kambing berdasarkan suhu badan menggunakan metode *knowledge discovery in database* (KDD). Tahapan yang ada pada metode ini terdapat *pre-KDD*, *selection*, *pre-processing*, *transformation*, *data mining*, *evaluation*, dan *post-*

*KDD*. Tahapan-tahapan tersebut diterapkan pada proses pembuatan model deteksi penyakit hewan ternak kambing berdasarkan suhu badan.

### 3.2 Metode Penelitian

#### 3.2.1 Perbandingan Metode KDD & CRISP-DM

Terdapat beberapa metodologi yang biasa digunakan untuk proses data mining, berikut adalah perbandingan antara metode KDD dan CRISP-DM yang biasa digunakan dalam data mining.

Tabel 3. 1 Perbandingan Metodologi Data Mining; CRISP-DM dan KDD [13]

Indikator	CRISP-DM	KDD
Definisi	Metodologi proses yang terstruktur untuk data mining yang terdiri dari enam tahap: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penyebaran.	Proses yang lebih luas mencakup langkah-langkah untuk menemukan pengetahuan yang berharga dari data, termasuk pemahaman domain, pemrosesan data, penemuan pola, evaluasi pola, dan penyebaran pengetahuan.
Kelebihan	Struktur yang jelas dan tahapan yang terdefinisi dengan baik membuatnya mudah diikuti dan diterapkan dalam berbagai proyek data mining.	Pendekatan yang lebih holistic memungkinkan untuk penemuan pengetahuan yang lebih mendalam dari data, mencakup aspek-aspek seperti pemahaman domain dan evaluasi yang lebih luas.
Kekurangan	Terkadang dianggap terlalu rigid dan tidak cukup fleksibel untuk menangani situasi yang kompleks atau tidak terstruktur dengan baik.	Proses yang lebih kompleks dan luas dapat menjadi sulit untuk diimplementasikan dalam skala besar atau dalam

		proyek dengan batasan waktu yang ketat.
--	--	---

Berdasarkan kelebihan dan kekurangan yang disajikan pada tabel 3.1, pendekatan KDD (*Knowledge Discovery in Database Process*) memiliki beberapa keunggulan dibandingkan dengan CRISP-DM (*Cross-Industry Standard Process for Data Mining*) karena KDD lebih luas dan menyeluruh. KDD mencakup tahap pemahaman domain yang lebih luas, yang memungkinkan peneliti untuk memahami konteks penyakit hewan ternak kambing secara menyeluruh. Hal ini tentunya dapat membantu dalam menentukan faktor-faktor suhu badan yang paling relevan dalam mendeteksi penyakit. KDD juga melibatkan langkah-langkah yang lebih luas dalam pemrosesan data, termasuk integrasi data dari berbagai sumber dan transformasi data yang diperlukan. Dalam konteks ini, KDD dapat membantu dalam menggabungkan suhu badan dengan data kesehatan hewan lainnya untuk pemahaman yang lebih komprehensif. KDD juga mencakup tahap penemuan pola yang lebih mendalam, yang dapat membantu dalam mengidentifikasi pola suhu badan yang terkait dengan penyakit tertentu pada hewan ternak kambing. Hal ini dapat memungkinkan pengembangan model deteksi penyakit yang lebih akurat. KDD juga mencakup tahap penemuan pola yang mendalam, yang dapat membantu dalam mengidentifikasi pola suhu badan yang terkait dengan penyakit tertentu pada hewan ternak kambing. Hal ini dapat memungkinkan pengembangan model deteksi penyakit yang lebih akurat. KDD juga mencakup evaluasi pola yang lebih luas, yang dapat membantu dalam mengukur kualitas model deteksi penyakit yang dikembangkan. Evaluasi yang komprehensi ini penting untuk memastikan keakuratan dan keandalan model.

Dalam penelitian berbasis Knowledge Discovery in Databases (KDD), pemilihan algoritma yang tepat sangat penting untuk mendapatkan hasil yang akurat dan bermakna. Berikut adalah penjelasan tentang empat

algoritma yang sering digunakan: Decision Tree, K-Nearest Neighbors (KNN), Random Forest, dan Naive Bayes, serta alasan mengapa memilih algoritma-algoritma ini dalam penelitian.

Decision Tree adalah algoritma yang sangat mudah dipahami dan diinterpretasikan karena strukturnya yang mirip dengan diagram alir. Ini membuatnya sangat berguna bagi peneliti dan pemangku kepentingan yang mungkin tidak memiliki latar belakang teknis yang mendalam. Selain itu, Decision Tree mampu menangani data kategori dan numerik tanpa memerlukan normalisasi atau skala fitur, serta dapat menangkap interaksi yang kompleks antara fitur dalam data.

Algoritma K-Nearest Neighbors (KNN) dipilih karena kesederhanaannya dalam implementasi dan pemahaman. KNN efektif dalam mendeteksi pola di mana data serupa memiliki label yang sama, dan karena merupakan algoritma berbasis instance, KNN tidak memerlukan fase pelatihan eksplisit. Fleksibilitasnya dalam digunakan untuk klasifikasi dan regresi menjadikannya pilihan yang baik dalam banyak konteks penelitian.

Random Forest adalah metode ensemble yang menggunakan banyak pohon keputusan untuk menghasilkan prediksi yang lebih akurat dan stabil. Dengan menggabungkan hasil dari beberapa pohon, Random Forest mampu mengurangi overfitting dan meningkatkan generalisasi, menjadikannya sangat cocok untuk dataset dengan banyak fitur. Selain itu, Random Forest dapat memberikan estimasi tentang pentingnya fitur, yang sangat membantu dalam pemahaman data dan pengambilan keputusan.

Naive Bayes adalah algoritma klasifikasi berbasis teorema Bayes dengan asumsi independensi yang kuat antara fitur. Meskipun asumsi independensi jarang berlaku sepenuhnya, Naive Bayes sering kali memberikan hasil yang baik pada data yang bervariasi, terutama data teks. Kecepatan dan efisiensinya dalam pelatihan dan prediksi membuatnya cocok untuk dataset yang besar. Selain itu, performa yang baik pada data

teks seperti analisis sentimen dan spam filtering menambah nilai lebih dari algoritma ini.

Secara keseluruhan, kombinasi algoritma Decision Tree, KNN, Random Forest, dan Naive Bayes dipilih dalam penelitian KDD ini karena masing-masing menawarkan keunggulan unik yang dapat membantu menghasilkan model yang robust, akurat, dan mudah diinterpretasikan. Dengan memilih algoritma ini, diharapkan penelitian dapat mencapai hasil yang optimal dan sesuai dengan tujuan yang diinginkan. Dengan demikian, menggunakan pendekatan KDD dalam penelitian ini dapat memberikan pemahaman yang lebih komprehensi tentang hubungan antara suhu badan dan penyakit pada hewan ternak kambing, serta membantu dalam pengembangan model deteksi penyakit yang lebih efektif.

### **3.2.2 Tahapan Penelitian**

Berikut ini adalah tahapan kerja yang dilakukan dalam penelitian ini dengan menggunakan metode KDD:

#### **1. *Pre-KDD***

Tahapan awal penelitian dilakukan dengan pemahaman dari tujuan penelitian dan domain penelitian. Hal ini dilakukan agar identifikasi tujuan penelitian jelas, pemahaman domain penelitian juga dilakukan, termasuk penyakit-penyakit yang umumnya terjadi atau mempengaruhi hewan ternak kambing dan kaitannya dengan suhu badan.

#### **2. *Selection***

Pada tahap ini dilakukan pemilihan dan integrasi data yang diperlukan, seperti data suhu badan dan data penyakit hewan ternak kambing.

#### **3. *Pre-processing***

Pada tahap ini dilakukan pembersihan data dari *noise*, *missing values* dan outlier. Selain itu, dilakukan juga penggabungan data

suhu badan dengan data penyakit hewan ternak kambing jika diperlukan.

#### 4. *Transformation*

Pada tahap ini dilakukan normalisasi data suhu badan dan data penyakit untuk memperoleh skala yang seragam, kemudian juga mengubah data kategorikal menjadi representasi numerik jika diperlukan.

#### 5. *Data Mining*

Pada tahap ini menentukan secara jelas masalah yang ingin diselesaikan, yaitu bagaimana menggunakan data suhu badan untuk mendeteksi penyakit. Kemudian, menggunakan teknik data mining seperti pemodelan prediktif atau *clustering* untuk mengembangkan model deteksi penyakit.

#### 6. *Evaluation*

Pada tahap ini dilakukan evaluasi kinerja model deteksi penyakit menggunakan metrik seperti akurasi, presisi, *recall* dan *F1-score*, dan menarik kesimpulan dari evaluasi model untuk memahami sejauh mana model efektif dalam mendeteksi penyakit.

#### 7. *Post-KDD*

Pada tahap ini dilakukan penyajian hasil penelitian dan interpretasi temuan-temuan penting kepada pemangku kepentingan terkait. Kemudian, menyebarluaskan pengetahuan yang ditemukan melalui publikasi ilmiah atau presentasi.

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A