

**LAPORAN**

**RENCANA TUGAS MAHASISWA (RTM) Ke-5**

**MATA KULIAH ANALISIS DATA EKSPLORATIF**

**“PENGEMBANGAN MODEL DAN EVALUASI”**



**DISUSUN OLEH:**

Muhammad Aryasatya Nugroho ( 22083010085 )

**DOSEN PENGAMPU:**

Tresna Maulana Fahrudin S.ST., M.T. (NIP. 199305012022031007)

**PROGRAM STUDI SAINS DATA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAWA TIMUR**

**2023**

Menghitung manual supervised learning menggunakan metode Naïve Bayes (kategorikal) pada animal dataset.

### a) Dataset Overview

No	Indeks	Jumlah Kaki	Ukuran Tubuh	Cara Melahirkan	Sayap	Tanduk	Gigi Taring	Lengan Pencapit
0	1	Gajah	4	besar	beranak	tidak	tidak	ya
1	2	Kanguru	4	sedang	beranak	tidak	tidak	tidak
2	3	Sapi	4	besar	beranak	tidak	ya	tidak
3	4	Kelelawar	4	kecil	beranak	ya	tidak	ya
4	5	Kelinci	4	kecil	beranak	tidak	tidak	tidak
5	6	Tikus	4	kecil	beranak	tidak	tidak	tidak
6	7	Musang	4	kecil	beranak	tidak	tidak	ya
7	8	Kuda Nil	4	besar	beranak	tidak	tidak	ya
8	9	Babon	4	kecil	beranak	tidak	tidak	ya
9	10	Paus Sperma	0	besar	beranak	tidak	tidak	ya
10	11	Babi Hutan	4	sedang	beranak	tidak	tidak	ya

Dataset ini berisi informasi morfologi beberapa hewan, termasuk jumlah kaki, ukuran tubuh, cara melahirkan, keberadaan sayap, tanduk, gigi taring, dan lengan pencapit, serta kelas hewan (Mamalia atau bukan). Data ini dapat digunakan untuk analisis perbandingan karakteristik morfologis antara hewan-hewan tersebut.

### b) Kode Script Penerapan Naïve Bayes Pada Dataset Animal Morphology

```

1 import pandas as pd
2
3 class AnimalClass:
4     def __init__(self, file_path):
5         self.data = pd.read_csv(file_path, sep=";")
6         self.training_data = self.data.drop(index=[15, 49])
7
8         self.mammal_count = self.training_data[self.training_data["Kelas"] == "Mamalia"].shape[0]
9         self.insect_count = self.training_data[self.training_data["Kelas"] == "Insekta"].shape[0]
10        self.total_training_data = self.training_data.shape[0]
11
12        self.prob_mammal = self.mammal_count / self.total_training_data
13        self.prob_insect = self.insect_count / self.total_training_data
14
15    def calc_feat_prob(self, feature, feature_value, target_class):
16        feat_class_count = self.training_data[(self.training_data["Kelas"] == target_class) & (self.training_data[feature]
17        if feat_class_count > 0:
18            target_class_count = self.training_data[self.training_data["Kelas"] == target_class].shape[0]
19            feat_class_prob = feat_class_count / target_class_count
20        else:
21            target_class_count = self.training_data[self.training_data["Kelas"] == target_class].shape[0]
22            feat_class_prob = target_class_count / self.total_training_data
23        return feat_class_prob
24
25    def calc_test_prob(self, test_data, class_prob, target_class):
26        data_prob = class_prob
27        for feature in test_data.columns[2:-1]:
28            feature_value = test_data[feature].values[0]
29            feature_prob = self.calc_feat_prob(feature, feature_value, target_class)
30            data_prob *= feature_prob
31        print(f"Probability of '{feature}' being '{feature_value}' given '{target_class}': {feature_prob:.4f}")
32        return data_prob
33
34    def predict_class(self, test_data, prob_mammal, prob_insect):
35        prob_mammal_test = self.calc_test_prob(test_data, prob_mammal, "Mamalia")
36        prob_insect_test = self.calc_test_prob(test_data, prob_insect, "Insekta")
37
38        if prob_mammal_test > prob_insect_test:
39            return "Mamalia"
40        else:
41            return "Insekta"
42
43    if __name__ == "__main__":
44        classifier = AnimalClass("animal_morphology_classify.csv")
45
46        test_data_rhinoceros = classifier.data.iloc[[15]]
47        test_data_millipede = classifier.data.iloc[[49]]
48
49        print("\nCalculations for Data Uji Ke-16 'Badak':")
50        prediction_rhinoceros = classifier.predict_class(test_data_rhinoceros, classifier.prob_mammal, classifier.prob_insect)
51        print("\nCalculations for Data Uji Ke-50 'Kaki Seribu':")
52        prediction_millipede = classifier.predict_class(test_data_millipede, classifier.prob_mammal, classifier.prob_insect)
53
54        print("Prediction for Data Uji Ke-16 'Badak' :", prediction_rhinoceros)
55        print("Prediction for Data Uji Ke-50 'Kaki Seribu' :", prediction_millipede)

```

Kode diatas adalah implementasi klasifikasi sederhana menggunakan Naïve Bayes untuk memprediksi kelas hewan berdasarkan morfologis yang terdapat dalam dataset. Langkah-langkah yang dilakukan:

**1) Inisialisasi Kelas**

Pertama-tama kita membaca dataset dari file CSV yang diberikan dan pra-pemrosesan data dengan menghilangkan dua baris (indeks 15 dan 49) untuk dihitung.

**2) Perhitungan Probabilitas Kelas**

Menghitung kemunculan kelas mamalia dan insekta dalam data yang telah dihilangkan, lalu Menghitung probabilitas masing-masing kelas berdasarkan jumlah data.

**3) Perhitungan Probabilitas Fitur**

Menggunakan Naive Bayes untuk menghitung probabilitas fitur-fitur seperti jumlah kaki, ukuran tubuh, cara melahirkan, dan atribut lainnya berdasarkan data latih. Ini memungkinkan pembuatan model sederhana untuk memprediksi kelas ("Mamalia" atau "Insekta") berdasarkan nilai fitur pada data uji.

**4) Perhitungan Probabilitas Data Uji**

Memproses data uji dengan menghitung probabilitas terhadap kedua kelas menggunakan informasi fitur yang dimiliki oleh data uji.

**5) Prediksi Kelas**

Membandingkan probabilitas hasil perhitungan terhadap kedua kelas ("Mamalia" dan "Insekta"), lalu menentukan kelas mana yang memiliki probabilitas lebih tinggi sebagai prediksi akhir.

**6) Evaluasi pada Data Uji**

Menggunakan model yang telah dibuat untuk memprediksi kelas dari dua data uji, yaitu "Badak" dan "Kaki Seribu" dan menampilkan hasil prediksi kelas untuk kedua data uji.

Jadi, kode diatas bertujuan untuk melakukan klasifikasi hewan berdasarkan ciri morfologisnya yang diberikan dalam dataset. Dengan Menghitung probabilitas fitur-fitur untuk setiap kelas, metode Naïve Bayes digunakan untuk memprediksi kelas dari data uji yang diberikan. Dari kode diatas, dapat dilihat output seperti berikut:

```
Calculations for Data Uji Ke-16 'Badak':
Probability of 'Jumlah Kaki' being '4' given 'Mamalia': 0.9583
Probability of 'Ukuran Tubuh' being 'besar' given 'Mamalia': 0.3333
Probability of 'Cara Melahirkan' being 'beranak' given 'Mamalia': 1.0000
Probability of 'Sayap' being 'tidak' given 'Mamalia': 0.9583
Probability of 'Tanduk' being 'ya' given 'Mamalia': 0.2500
Probability of 'Gigi Taring' being 'tidak' given 'Mamalia': 0.5000
Probability of 'Lengan Pencapit' being 'tidak' given 'Mamalia': 1.0000
Probability of 'Jumlah Kaki' being '4' given 'Insekta': 0.5000
Probability of 'Ukuran Tubuh' being 'besar' given 'Insekta': 0.5000
Probability of 'Cara Melahirkan' being 'beranak' given 'Insekta': 0.5000
Probability of 'Sayap' being 'tidak' given 'Insekta': 0.2083
Probability of 'Tanduk' being 'ya' given 'Insekta': 0.0417
Probability of 'Gigi Taring' being 'tidak' given 'Insekta': 1.0000
Probability of 'Lengan Pencapit' being 'tidak' given 'Insekta': 0.9167

Calculations for Data Uji Ke-50 'Kaki Seribu':
Probability of 'Jumlah Kaki' being '400' given 'Mamalia': 0.5000
Probability of 'Ukuran Tubuh' being 'sangat kecil' given 'Mamalia': 0.5000
Probability of 'Cara Melahirkan' being 'bertelur' given 'Mamalia': 0.5000
Probability of 'Sayap' being 'tidak' given 'Mamalia': 0.9583
Probability of 'Tanduk' being 'tidak' given 'Mamalia': 0.7500
Probability of 'Gigi Taring' being 'tidak' given 'Mamalia': 0.5000
Probability of 'Lengan Pencapit' being 'tidak' given 'Mamalia': 1.0000
Probability of 'Jumlah Kaki' being '400' given 'Insekta': 0.5000
Probability of 'Ukuran Tubuh' being 'sangat kecil' given 'Insekta': 1.0000
Probability of 'Cara Melahirkan' being 'bertelur' given 'Insekta': 1.0000
Probability of 'Sayap' being 'tidak' given 'Insekta': 0.2083
Probability of 'Tanduk' being 'tidak' given 'Insekta': 0.9583
Probability of 'Gigi Taring' being 'tidak' given 'Insekta': 1.0000
Probability of 'Lengan Pencapit' being 'tidak' given 'Insekta': 0.9167
Prediction for Data Uji Ke-16 'Badak' : Mamalia
Prediction for Data Uji Ke-50 'Kaki Seribu' : Insekta
```

Output kode diatas menunjukkan perhitungan probabilitas kedua data uji, yaitu data uji ke-16 'badak' dan data uji ke-50 'kaki seribu', terhadap kelas 'Mamalia' dan 'Insekta'. Dari hasil perhitungan, prediksi untuk uji ke-16 adalah "Mamalia" karena probabilitasnya sendiri lebih tinggi dibandingkan dengan kelas "Insekta". Sementara itu, untuk data uji-50, prediksi kelasnya adalah 'insekta'. Hasil ini diperoleh dari perhitungan probabilitas fitur-fitur menggunakan model Naïve Bayes berdasarkan data yang telah diberikan.