

LAPORAN
MATA KULIAH ANALISIS DATA EKSPLORATIF (A)
“PENGEMBANGAN MODEL DAN EVALUASI”



DISUSUN OLEH:

Reza Putri Angga (22083010006)

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

STUDI KASUS DAN PEMBAHASAN

Untuk melakukan perhitungan manual *supervised learning* menggunakan metode *Naive Bayes* langkah demi langkah untuk menyelesaikan permasalahan pada animal dataset yang memiliki fitur 8 fitur di antaranya yakni, jumlah kaki (kontinyu), ukuran tubuh (nominal), cara melahirkan (nominal), sayap (nominal), tanduk (nominal), gigi taring (nominal), lengan pencapit (nominal), dan label kelas (mamalia dan insekta)

Dengan menggunakan simulasi perhitungan untuk data uji ke-16 (Badak) dan data uji ke-50 (Kaki Seribu), pisahkan data uji tersebut dari data latih lainnya. Dan melakukan perhitungan manual *supervised learning* metode *Naive Bayes* secara manual ke dalam dataset di perlukan beberapa langkah-langkah, di antaranya yakni :

A. Melakukan Proses *Load* (Pembacaan) Dataset Yang Akan Di Analisis Lebih Lanjut

Di Tampilkan Dataset Animal Yang Akan Di Analisis Lebih Lanjut

Out[2]:

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

Pada kode script di atas, di lakukan proses *load* atau pembacaan dataset *animal* dengan nama file “*animal_morphology_classift.csv*” dan dapat di perhatikan bahwa dataset *animal* yang akan di analisis lebih lanjut tersebut memiliki 8 fitur, 2 kelas, dan terdapat sebanyak 50 *record* data. Setelah melakukan dan mengetahui dataset yang akan di analisis, selanjutnya bisa di lanjutkan dengan pembuatan *kode script* penerapan *Naive Bayes*.

B. Kode Script Penerapan Naive Bayes Pada Dataset Animal

Kode Script Penerapan Naive Bayes Pada Dataset Animal

```
In [70]: import pandas as pd

##di lakukan penerapan naive bayes classifier
#dengan langkah-langkah, yakni :
print("'Di Lakukan Proses Perhitungan Penerapan Naive Bayes Classifier Untuk Data Animal'")

#inisiasi naive bayes class untuk perhitungan probabilitas
class NaiveBayesClass:

    #untuk konstruktor kelas
    def __init__(self, DataAnimal):
        self.Data = pd.read_csv(DataAnimal, sep = ";")
        self.DataLatih = self.Data.drop(index = [15, 49]) #inisiasi data latih

        self.JumlahMamalia = self.DataLatih[self.DataLatih["Kelas"] == "Mamalia"].shape[0] #jumlah data kelas
        self.JumlahInsekta = self.DataLatih[self.DataLatih["Kelas"] == "Insekta"].shape[0] #jumlah data kelas
        self.JumlahDataLatih = self.DataLatih.shape[0]

        self.ProbabilitasMamalia = self.JumlahMamalia / self.JumlahDataLatih #perhitungan probabilitas mamalia
        self.ProbabilitasInsekta = self.JumlahInsekta / self.JumlahDataLatih #perhitungan probabilitas insekta

    #untuk perhitungan probabilitas tiap fitur
    def PerhitunganProbabilitasFitur(self, Fitur, NilaiFitur, Kelas):
        #perhitungan muncul fitur dalam kelas
        JumlahFiturKelas = self.DataLatih[(self.DataLatih["Kelas"] == Kelas) & (self.DataLatih[Fitur] ==
                                                                                               NilaiFitur)].shape[0]

        if JumlahFiturKelas > 0: #kondisi apakah fitur ditemukan/tidak
            #jika iya, di lakukan perhitungan probabilitas fitur
            JumlahDataKelas = self.DataLatih[self.DataLatih["Kelas"] == Kelas].shape[0]
            ProbabilitasFiturKelas = JumlahFiturKelas / JumlahDataKelas
        else:
            JumlahDataKelas = self.DataLatih[self.DataLatih["Kelas"] == Kelas].shape[0]
            ProbabilitasFiturKelas = 0.0 #jika tidak, di lakukan inisiasi nilai probabilitas 0,0
        return ProbabilitasFiturKelas
```

```
#untuk perhitungan probabilitas data uji
def HitungProbabilitasDataUji(self, DataUji, ProbabilitasKelas, Kelas):
    #inisiasi awal, bahwa probabilitas data adalah probabilitas kelas
    ProbabilitasData = ProbabilitasKelas
    for Fitur in DataUji.columns[2:-1]:
        NilaiFitur = DataUji[Fitur].values[0] #mengambil nilai fitur dari data uji
        #perhitungan probabilitas fitur kelas
        ProbabilitasFiturKelas = self.PerhitunganProbabilitasFitur(Fitur, NilaiFitur, Kelas)
        ProbabilitasData *= ProbabilitasFiturKelas
    print(f"Probabilitas {Fitur} : {NilaiFitur} Di kelas {Kelas} : {ProbabilitasFiturKelas}")
    return ProbabilitasData

#untuk perhitungan prediksi kelas
def PrediksiKelas(self, DataUji, ProbabilitasMamalia, ProbabilitasInsekta):
    #perhitungan probabilitas kelas mamalia
    ProbabilitasUjiMamalia = self.HitungProbabilitasDataUji(DataUji, ProbabilitasMamalia, "Mamalia")
    #perhitungan probabilitas kelas insekta
    ProbabilitasUjiInsekta = self.HitungProbabilitasDataUji(DataUji, ProbabilitasInsekta, "Insekta")

    #penampilan hasil nilai data uji
    print("'Di Peroleh Nilai Dan Perbandingan Untuk Data Uji'")
    print(f"Di Peroleh Nilai Probabilitas Data Uji Terhadap Kelas Mamalia : {ProbabilitasUjiMamalia}")
    print(f"Di Peroleh Nilai Probabilitas Data Uji Terhadap Kelas Insekta : {ProbabilitasUjiInsekta}")

    if ProbabilitasUjiMamalia > ProbabilitasUjiInsekta: #pembuatan kondisi
        return "Mamalia"
    else:
        return "Insekta"
```

```
#untuk menjalankan kode script
if __name__ == "__main__":
    DatasetAnimal = NaiveBayesClass("animal_morphology_classify.csv") #inisiasi objek
    DataUjiBadak = DatasetAnimal.Data.iloc[[15]]
    DataUjiKakiSeribu = DatasetAnimal.Data.iloc[[49]]

    #penampilan hasil probabilitas di tiap kelas
    print("\n'Perhitungan Probabilitas Di Masing-Masing Kelas'")
    print(f"Di Peroleh Jumlah Total Kelas Mamalia Sebesar : {DatasetAnimal.JumlahMamalia}")
    print(f"Di Peroleh Jumlah Total Kelas Insekta Sebesar : {DatasetAnimal.JumlahInsekta}")
    print(f"Di Peroleh Jumlah Total Data Latih Sebesar : {DatasetAnimal.JumlahDataLatih}")
    print(f"Di Peroleh Nilai Probabilitas Kelas Mamalia Sebesar : {DatasetAnimal.ProbabilitasMamalia}")
    print(f"Di Peroleh Nilai Probabilitas Kelas Insekta Sebesar : {DatasetAnimal.ProbabilitasInsekta}")

    #penampilan hasil probabilitas fitur di tiap kelas
    print("\n'Perhitungan Probabilitas Fitur Di Masing-Masing Kelas'")
    print("\n'Nilai Probabilitas Masing-Masing Fitur Pada Setiap Kelas Untuk Data Uji Ke-16 Badak'")

    #prediksi data uji ke-16 "badak"
    HasilPrediksiBadak = DatasetAnimal.PrediksiKelas(DataUjiBadak, DatasetAnimal.ProbabilitasMamalia,
                                                         DatasetAnimal.ProbabilitasInsekta)
    print(f"Hasil Prediksi Data Uji Ke-16 'Badak' : {HasilPrediksiBadak}")

    #prediksi data uji ke-50 "kaki seribu"
    print("\n'Nilai Probabilitas Masing-Masing Fitur Pada Setiap Kelas Untuk Data Uji Ke-50 Kaki Seribu'")
    HasilPrediksiKakiSeribu = DatasetAnimal.PrediksiKelas(DataUjiKakiSeribu, DatasetAnimal.ProbabilitasMamalia,
                                                            DatasetAnimal.ProbabilitasInsekta)
    print(f"Hasil Prediksi Data Uji Ke-50 'Kaki Seribu' : {HasilPrediksiKakiSeribu}")
```

Pada kode script di atas, di lakukan proses penerapan dan perhitungan *Naive Bayes* secara manual untuk menentukan nilai probabilitas baik di setiap kelas dan di setiap fitur pada semua kelas untuk melakukan prediksi yang di berikan untuk data uji ke-16, yakni “Badak” dan data uji ke-50, yakni “Kaki Seribu”. Tentunya dalam menjalankan kode ini di perlukan penggunaan library *pandas* yang di permissalkan sebagai pd. Untuk penerapan perhitungan *Naive Bayes*, di perlukan pembuatan *ClassNaiveBayes* yang memiliki konstruktor yang di pergunakan menghitung jumlah data dan nilai probabilitas, baik untuk kelas “Mamalia” maupun kelas “Insekta”.

Di dalam *class* tersebut, juga terdapat beberapa *define function* yang di pergunakan sebagai langkah-langkah perhitungan dalam *Naive Bayes*. Untuk *define function* PerhitunganProbabilitasFitur di pergunakan untuk melakukan perhitungan probabilitas kemunculan nilai fitur pada kelas tertentu. Terdapat variabel JumlahFiturKelas yang di pergunakan untuk menghitung jumlah baris dalam data latih sesuai kelas. Dan di buat dua kondisi, jika JumlahFitur dalam kelas tersebut muncul lebih besar dari 0 kali, maka akan di lakukan perhitungan probabilitas yang di simpan dalam variabel ProbabilitasFiturKelas. Sedangkan, jika JumlahFitur dalam kelas tersebut muncul sebesar 0, maka nilai ProbabilitasFiturKelas tersebut akan di insiasi sebesar 0,0.

Untuk *define function* HitungProbabilitasDataUji di pergunakan untuk melakukan perhitungan probabilitas data uji terhadap suatu kelas dengan memperhitungkan probabilitas di masing-masing fitur yang telah di tuliskan sebelumnya. Terdapat proses *looping for* untuk melakukan proses iterasi pengambilan nilai fitur untuk data uji. Lalu,

terdapat variabel `ProbabilitasFiturKelas` yang di pergunakan untuk melakukan perhitungan probabilitas fitur dalam kelas tertentu.

Untuk *define function* `PrediksiKelas` di pergunakan untuk menghitung probabilitas data uji kedua kelas dan memberikan keputusan hasil prediksi. Terdapat variabel `ProbabilitasUjiMamalia` yang di pergunakan untuk menghitung probabilitas data uji terhadap kelas “Mamalia” dan `ProbabilitasUjiInsekta` yang di pergunakan untuk menghitung probabilitas data uji terhadap kelas “Insekta”. Dan di buat kondisi jika hasil `ProbabilitasUjiMamalia > ProbabilitasUjiInsekta` maka hasil prediksi data uji tersebut, masuk ke dalam kelas “Mamalia”, begitupula sebaliknya.

Langkah yang terakhir, tentunya setelah melakukan proses pembuatan *class* dan juga *define function* di lakukan proses menjalankan kode script menggunakan “`if __name__ == “__main__”`” dengan menggunakan dataset *animal* akan di tampilkan hasil probabilitas di masing-masing kelas, hasil probabilitas pada tiap-tiap fitur, dan hasil prediksi dari data uji. Untuk penjelasan dari kode script ini akan di jelaskan lebih lanjut dalam poin pembahasan selanjutnya.

C. Output Kode Script Penerapan *Naive Bayes* Pada Dataset *Animal*

```
'Di Lakukan Proses Perhitungan Penerapan Naive Bayes Classifier Untuk Data Animal'

'Perhitungan Probabilitas Di Masing-Masing Kelas'
Di Peroleh Jumlah Total Kelas Mamalia Sebesar : 24
Di Peroleh Jumlah Total Kelas Insekta Sebesar : 24
Di Peroleh Jumlah Total Data Latih Sebesar : 48
Di Peroleh Nilai Probabilitas Kelas Mamalia Sebesar : 0.5
Di Peroleh Nilai Probabilitas Kelas Insekta Sebesar : 0.5

'Perhitungan Probabilitas Fitur Di Masing-Masing Kelas'

'Nilai Probabilitas Masing-Masing Fitur Pada Setiap Kelas Untuk Data Uji Ke-16 Badak'
Probabilitas Jumlah Kaki : 4 Di kelas Mamalia : 0.9583333333333334
Probabilitas Ukuran Tubuh : besar Di kelas Mamalia : 0.3333333333333333
Probabilitas Cara Melahirkan : beranak Di kelas Mamalia : 1.0
Probabilitas Sayap : tidak Di kelas Mamalia : 0.9583333333333334
Probabilitas Tanduk : ya Di kelas Mamalia : 0.25
Probabilitas Gigi Taring : tidak Di kelas Mamalia : 0.5
Probabilitas Lengan Pencapit : tidak Di kelas Mamalia : 1.0
Probabilitas Jumlah Kaki : 4 Di kelas Insekta : 0.0
Probabilitas Ukuran Tubuh : besar Di kelas Insekta : 0.0
Probabilitas Cara Melahirkan : beranak Di kelas Insekta : 0.0
Probabilitas Sayap : tidak Di kelas Insekta : 0.20833333333333334
Probabilitas Tanduk : ya Di kelas Insekta : 0.041666666666666664
Probabilitas Gigi Taring : tidak Di kelas Insekta : 1.0
Probabilitas Lengan Pencapit : tidak Di kelas Insekta : 0.9166666666666666

'Di Peroleh Nilai Dan Perbandingan Untuk Data Uji'
Di Peroleh Nilai Probabilitas Data Uji Terhadap Kelas Mamalia : 0.019133391203703703
Di Peroleh Nilai Probabilitas Data Uji Terhadap Kelas Insekta : 0.0
Hasil Prediksi Data Uji Ke-16 'Badak' : Mamalia
```

```
'Nilai Probabilitas Masing-Masing Fitur Pada Setiap Kelas Untuk Data Uji Ke-50 Kaki Seribu'
Probabilitas Jumlah Kaki : 400 Di kelas Mamalia : 0.0
Probabilitas Ukuran Tubuh : sangat kecil Di kelas Mamalia : 0.0
Probabilitas Cara Melahirkan : bertelur Di kelas Mamalia : 0.0
Probabilitas Sayap : tidak Di kelas Mamalia : 0.9583333333333334
Probabilitas Tanduk : tidak Di kelas Mamalia : 0.75
Probabilitas Gigi Taring : tidak Di kelas Mamalia : 0.5
Probabilitas Lengan Pencapit : tidak Di kelas Mamalia : 1.0
Probabilitas Jumlah Kaki : 400 Di kelas Insekta : 0.0
Probabilitas Ukuran Tubuh : sangat kecil Di kelas Insekta : 1.0
Probabilitas Cara Melahirkan : bertelur Di kelas Insekta : 1.0
Probabilitas Sayap : tidak Di kelas Insekta : 0.2083333333333334
Probabilitas Tanduk : tidak Di kelas Insekta : 0.9583333333333334
Probabilitas Gigi Taring : tidak Di kelas Insekta : 1.0
Probabilitas Lengan Pencapit : tidak Di kelas Insekta : 0.9166666666666666
'Di Peroleh Nilai Dan Perbandingan Untuk Data Uji'
Di Peroleh Nilai Probabilitas Data Uji Terhadap Kelas Mamalia : 0.0
Di Peroleh Nilai Probabilitas Data Uji Terhadap Kelas Insekta : 0.0
Hasil Prediksi Data Uji Ke-50 'Kaki Seribu' : Insekta
```

```
'Di Lakukan Proses Perhitungan Penerapan Naive Bayes Classifier Untuk Data Animal'

'Perhitungan Probabilitas Di Masing-Masing Kelas'
Di Peroleh Jumlah Total Kelas Mamalia Sebesar : 24
Di Peroleh Jumlah Total Kelas Insekta Sebesar : 24
Di Peroleh Jumlah Total Data Latih Sebesar : 48
Di Peroleh Nilai Probabilitas Kelas Mamalia Sebesar : 0.5
Di Peroleh Nilai Probabilitas Kelas Insekta Sebesar : 0.5

'Perhitungan Probabilitas Fitur Di Masing-Masing Kelas'

'Nilai Probabilitas Masing-Masing Fitur Pada Setiap Kelas Untuk Data Uji Ke-16 Badak'
Probabilitas Jumlah Kaki : 4 Di kelas Mamalia : 0.9583333333333334
Probabilitas Ukuran Tubuh : besar Di kelas Mamalia : 0.3333333333333333
Probabilitas Cara Melahirkan : beranak Di kelas Mamalia : 1.0
Probabilitas Sayap : tidak Di kelas Mamalia : 0.9583333333333334
Probabilitas Tanduk : ya Di kelas Mamalia : 0.25
Probabilitas Gigi Taring : tidak Di kelas Mamalia : 0.5
Probabilitas Lengan Pencapit : tidak Di kelas Mamalia : 1.0
Probabilitas Jumlah Kaki : 4 Di kelas Insekta : 0.0
Probabilitas Ukuran Tubuh : besar Di kelas Insekta : 0.0
Probabilitas Cara Melahirkan : beranak Di kelas Insekta : 0.0
Probabilitas Sayap : tidak Di kelas Insekta : 0.2083333333333334
Probabilitas Tanduk : ya Di kelas Insekta : 0.0416666666666666
Probabilitas Gigi Taring : tidak Di kelas Insekta : 1.0
Probabilitas Lengan Pencapit : tidak Di kelas Insekta : 0.9166666666666666
'Di Peroleh Nilai Dan Perbandingan Untuk Data Uji'
Di Peroleh Nilai Probabilitas Data Uji Terhadap Kelas Mamalia : 0.019133391203703703
Di Peroleh Nilai Probabilitas Data Uji Terhadap Kelas Insekta : 0.0
Hasil Prediksi Data Uji Ke-16 'Badak' : Mamalia
```

Untuk melakukan pembahasan *output* dari masing-masing perhitungan dapat di jabarkan sebagai berikut.

➤ Probabilitas Masing-Masing Kelas

Untuk menghitung probabilitas di masing-masing kelas, yakni kelas “Mamalia” dan “Insekta”. Untuk probabilitas di kelas “Mamalia” di lakukan perhitungan jumlah total kelas sebesar 28 di bagi dengan jumlah total data latih sebesar 48. Dan di peroleh nilai probabilitas kelas “Mamalia” sebesar 0,5. Untuk probabilitas di kelas “Insekta” di lakukan perhitungan jumlah total kelas sebesar 28 di bagi dengan jumlah total data latih sebesar 48. Dan di peroleh nilai probabilitas kelas “Insekta” sebesar 0,5.

➤ ***Naive Bayes Untuk Data Uji Ke-16 “Badak”***

Untuk menghitung probabilitas di masing-masing nilai fitur di lakukan proses perhitungan nilai masing-masing fitur kemudian nilai-nilai dari masing-masing fitur tersebut di kalikan dengan nilai probabilitas kelas. Untuk kelas “Mamalia” maka semua fitur “Mamalia” di kali dengan nilai probabilitas kelas “Mamalia”. Begitupula untuk kelas “Insekta”.

Di peroleh nilai probabilitas untuk data uji ke-16 “Badak” terhadap kelas “Mamalia” sebesar 0,0193 dan terhadap kelas “Insekta” sebesar 0,0. Maka, dapat di peroleh hasil prediksi data uji ke-16 “Badak” termasuk ke dalam kelas “Mamalia”.

➤ ***Naive Bayes Untuk Data Uji Ke-50 “Kaki Seribu”***

Untuk menghitung probabilitas di masing-masing nilai fitur di lakukan proses perhitungan nilai masing-masing fitur kemudian nilai-nilai dari masing-masing fitur tersebut di kalikan dengan nilai probabilitas kelas. Untuk kelas “Mamalia” maka semua fitur “Mamalia” di kali dengan nilai probabilitas kelas “Mamalia”. Begitupula untuk kelas “Insekta”.

Di peroleh nilai probabilitas untuk data uji ke-50 “Kaki Seribu” terhadap kelas “Mamalia” sebesar 0,0 dan terhadap kelas “Insekta” sebesar 0,0. Di karenakan, ProbabilitasMamalia tidak lebih besar dari ProbabilitasInsekta. Maka, dapat di peroleh hasil prediksi data uji ke-50 “Kaki Seribu” termasuk ke dalam kelas “Insekta”.