**SISTEM CERDAS DAN PENDUKUNG KEPUTUSAN**

**LAPORAN PROYEK AKHIR**



**DISUSUN OLEH:**

|  |  |  |  |
| --- | --- | --- | --- |
| **ANGGOTA** | **:** | **FAZRI RIDWAN** | **123190045** |
|  |  | **DHIMAS WAHYU ASSHIDIQ** | **123190071** |
| **KELAS** | **:** | **B** | |

**PROGRAM STUDI INFORMATIKA**

**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK INDUSTRI**

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”**

**YOGYAKARTA**

**2021**

**KATA PENGANTAR**

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa yang telah melimpahkan rahmat dan karunia-Nya yang sangat besar sehingga kami dapat menyelesaikan mata kuliah Sistem Cerdas dan Pendukung Keputusan serta laporan proyek akhir yang berjudul Klasifikasi Penggunaan Metode Kontrasepsi Menggunakan Metode Naïve Bayes. Adapun laporan ini berisi tentang proyek akhir yang kami pilih dari hasil pembelajaran selama mata kuliah berlangsung.

Rasa terima kasih juga kami ucapkan kepada Dosen mata kuliah yang telah membimbing kami dan mengajari kami sehingga proyek ini dapat disusun dengan baik.

Selayaknya kalimat yang menyatakan bahwa tidak ada sesuatu yang sempurna, kami juga menyadari bahwa proyek ini juga masih memiliki banyak kekurangan. Maka dari itu kami mengharapkan saran serta masukan dari para pembaca sekalian demi penyusunan laporan proyek akhir ini dengan lebih baik lagi.

|  |  |
| --- | --- |
|  | Yogyakarta, 15 Juni 2021  Penyusun |

**DAFTAR ISI**

DAFTAR ISI

[KATA PENGANTAR ii](#_Toc75332807)

[I. DAFTAR ISI iii](#_Toc75332808)

[I. JUDUL 1](#_Toc75332809)

[II. PENDAHULUAN 1](#_Toc75332810)

**[2.1](#_Toc75332811)****[Latar Belakang](#_Toc75332811)** [1](#_Toc75332811)

[III. ISI DAN PEMBAHASAN 2](#_Toc75332812)

[IV. Kesimpulan dan Saran 13](#_Toc75332813)

# 

# JUDUL

**KLASIFIKASI PENGGUNAAN METODE KONTRASEPSI MENGGUNAKAN METODE BAYES**

# PENDAHULUAN

## **Latar Belakang dan Tujuan**

Kontrasepsi adalah alat untuk mencegah kehamilan,kurangnya pengetahun tentang metode kontrasepsi menyebabkan pasangan usia subur menggunakan kontrasepsi yang tidak sesuai dengan kondisi tubuhnya atau bahkan tidak menggunakan kontrasepsi. Klasifikasi dengan algoritma Naive Bayes adalah sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan, dinilai berpotensi baik dalam mengklasifikasikan dokumen dibandingkan dengan metode pengklasifikasian lain dalam hal akurasi dan efisiensi untuk menentukan metode kontrasepsi yang akan digunakan, maka perlu data yang harus di input yang terdiri dari faktor, usia istri, edukasi, jumlah anak, agama, dan faktor lainya. Pada akhirnya diharapkan sistem ini akan memberikan solusi bagi pasangan yang ingin menggunakan alat kontrasepsi.

Pada penelitian ini akan diterapkan algoritma Naive Bayes dalam penentuan Metode Kontrasepsi. Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma Naive Bayes adalah salah satu metode *data mining* yang termasuk kedalam sepuluh klasifikasi data mining yang paling popular diantara algoritma-algoritma lainnya.

Memilih metode kontrasepsi bukanlah suatu hal yang mudah karena efek yang berdampak terhadap tubuh tidak akan diketahui selama belum menggunakannya, sehingga perlu pengetahuan yang luas dan tepat mengenai kekurangan dan kelebihan dari masing masing metode kontrasepsi yang kemudian disesuaikan dengan kondisi tubuh pengguna. Bagi setiap pasangan harus mempertimbangkan penggunaan metode kontrasepsi secara rasional, efisien dan efektif. Berdasarkan masalah tersebut, perlu dicarikan upaya untuk memecahkan masalah tersebut. Salah satu cara yang ditawarkan adalah dengan mengklasifikasi tiap data.

# ISI DAN PEMBAHASAN

* 1. **Metode Naïve Bayes**

**Algoritma Naive Bayes** merupakan sebuah metode klasifikasi menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes. **Algoritma Naive Bayes** memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dr Naïve Bayes Classifier ini adalah asumsi yg sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian.

Naive Bayes Classifier bekerja sangat baik dibanding dengan model classifier lainnya. Hal ini dibuktikan pada jurnal Xhemali, Daniela, Chris J. Hinde, and Roger G. Stone. “Naive Bayes vs. decision trees vs. neural networks in the classification of training web pages.” (2009), mengatakan bahwa “Naïve Bayes Classifier memiliki tingkat akurasi yg lebih baik dibanding model classifier lainnya”.

Keuntungan penggunan adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter yg diperlukan dalam proses pengklasifikasian. Karena yg diasumsikan sebagai variabel independent, maka hanya varians dari suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians.

Tahapan dari proses algoritma Naive Bayes adalah:

* 1. Menghitung jumlah kelas / label.
  2. Menghitung Jumlah Kasus Per Kelas
  3. Kalikan Semua Variable Kelas
  4. Bandingkan Hasil Per Kelas

#### **Kelebihan & Kekurangan Naive Bayes**

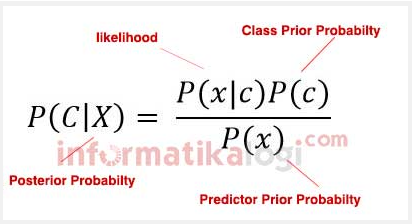
**Kelebihan**

* Bisa dipakai untuk data kuantitatif maupun kualitatif
* Tidak memerlukan jumlah data yang banyak
* Tidak perlu melakukan data training yang banyak
* Jika ada nilai yang hilang, maka bisa diabaikan dalam perhitungan.
* Perhitungannya cepat dan efisien
* Pengklasifikasian dokumen bisa dipersonalisasi, disesuaikan dengan kebutuhan setiap orang

**Kekurangan**

* Asumsi independence antar atribut membuat akurasi berkurang (karena biasanya ada keterkaitan)
* Apabila probabilitas kondisionalnya bernilai nol, maka probabilitas prediksi juga akan bernilai nol
* Asumsi bahwa masing-masing variabel independen membuat berkurangnya akurasi, karena biasanya ada korelasi antara variabel yang satu dengan variabel yang lain
* Keakuratannya tidak bisa diukur menggunakan satu probabilitas saja. Butuh bukti-bukti lain untuk membuktikannya.
* Untuk membuat keputusan, diperlukan pengetahuan awal atau pengetahuan mengenai masa sebelumnya.
* Banyak celah yang bisa mengurangi efektivitasnya seperti hanya bisa mendeteksi kata-kata saja, tidak bisa berupa gambar

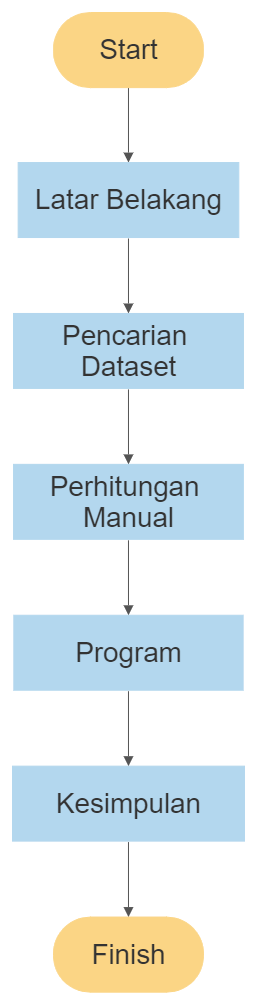
Rumus Naïve Bayes



***x***  : Data dengan class yang belum diketahui  
***c***  : Hipotesis data merupakan suatu class spesifik  
***P(c|x)*** : Probabilitas hipotesis berdasar kondisi (posteriori probability)  
***P(c)*** : Probabilitas hipotesis (prior probability)  
***P(x|c)*** : Probabilitas berdasarkan kondisi pada hipotesis  
***P(x)*** : Probabilitas c

Rumus diatas menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik karakteristik sampel pada kelas C (disebut juga likelihood), dibagi dengan peluang kemunculan karakteristik  sampel secara global ( disebut juga evidence).

* 1. **Tahapan**



* 1. **Pemahaman Dataset**

Tahapan pemahaman data dimulai dengan pengumpulan dari dataset yang didapatkan melalui internet, dari data tersebut dapat di identifikasi beberapa hal baik dalam masalah data, mendeteksi subyek menarik dari data, dan memperoleh informasi yang tersembunyi. Dari dataset pasien yang menggunakan alat kontrasepsi diperoleh beberapa atribut antara lain:

1. Usia Istri Berisi data usia istri, berkisar antara umur 16 tahun sampai dengan 49 tahun. Dimana untuk umur dibawah 26 tahun termasuk kategori 1, umur dibawah 41 tahun termasuk kategori 2 dan untuk umur 41 tahun atau lebih merupakan kategori 3.
2. Pendidikan Istri Berisi pendidikan formal istri, mulai dari pendidikan dasar, SMP, SMA, dan Sarjana.
3. Pendidikan Suami Berisi pendidikan formal istri, mulai dari pendidikan dasar, SMP, SMA, dan Sarjana.
4. Jumlah Anak Berisi jumlah anak yang dilahirkan oleh istri.(kategori 1 apabila jumlah anak < 4, kategori 2 apabila jumlah anak 2..10, kategori 3 apabila jumlah anak lebih dari 10)
5. Agama yang dianut jika beragara islam maka dijawab ya dan selain agama islam dijawan tidak.
6. Istri Bekerja Berisi jawaban istri, jika bekerja dijawab ya, jika tidak bekerja dijawab tidak.
7. Kesibukan Suami Berisi aktifitas pekerjaan suami yaitu kesibukan sangat tinggi, tinggi, sedang, dan rendah.
8. Standar Hidup Berisi pola hidup dari keluarga pasien, yaitu pola hidup sangat tinggi, tinggi, sedang, dan rendah. (1 = *low* … 4 = *high*)
9. Media Exposure berisi apakah pasangan mendapatkan paparan media yang bagus atau tidak. (0 = *good,* 1 = not good)
10. Metode Kontrasepsi Berisi label yaitu no-use, short-term dan long-term.(1 = *no use,* 2 = *long-term*, 3 = *short-term*)

kemudian atribut tersebut disesuaikan dengan metode data mining Naïve Bayes yaitu untuk klasifikasi dalam bentuk label polynominal yaitu no-use, short-term dan long-term. Apabila atribut belum sesuai maka perlu di transformasi data agar dapat dimodelkan sesuai dengan metode algoritma Naïve Bayes

* 1. **Perhitungan manual**

Pada bagian ini akan dijelaskan bagaimana perhitungan pada metode Bayes berjalan. Dataset yang digunakan adalah dataset yang telah dijelaskan diatas, untuk datasetnya sendiri bisa di akses melalui link berikut. <https://www.kaggle.com/joelzcharia/contraceptive-prevalence-survey?select=1987+Indonesia+Contraception+Prevalence+Study.csv>

Tahap pertama adalah menentukan data test yang nantinya akan digunakan untuk mengklasifikasi metode kontrasepsi yang digunakan. Adapun data test yang akan digunakan adalah sebagai berikut:

* Untuk kategori yang pertama, yaitu umur, akan di *set* dengan nilai 3 (*old*)
* Kategori pendidikan istriakan di *set* dengan nilai 3 (SMA)
* Pendidikan pasangan akan di set dengan nilai 4 (Sarjana)
* Jumlah anak tiga, maka kategorinya 1
* Agama di *set* dengan nilai 1 (Islam)
* Pekerjaan di *set* dengan nilai 1 (Tidak Bekerja)
* Kesibukan suami di *set* nilai 2
* Standar hidup di *set* 3
* *Media Exposure* di *set* 0 (*good*)

Sehingga didapatkan kumpulan data test (3, 3, 4, 1, 1, 1, 2, 3, 0).

Tahap selanjutnya adalah menghitung probabilitas kemunculan setiap kategori dan setiap keputusan(Metode Kontrasepsi yang digunakan).

|  |
| --- |
| Perhitungan probabilitas untuk kategori umur:  P(age->3|cmu->1) = freq(age->3|cmu->1)/freq(cmu->1)  = 175/629  = 0.2782194  P(age->3|cmu->2) = freq(age->3|cmu->2)/freq(cmu->2)  = 84/333  = 0.25225225  P(age->3|cmu->3) = freq(age->3|cmu->3)/freq(cmu->3)  = 47/511  = 0.09197652 |

|  |
| --- |
| Perhitungan probabilitas untuk kategori pendidikan:  P(edu->3|cmu->1) = freq(edu->3|cmu->1)/freq(cmu->1)  = 175/629  = 0.278219396  P(edu->3|cmu->2) = freq(edu->3|cmu->2)/freq(cmu->2)  = 80/333  = 0.24024024  P(edu->3|cmu->3) = freq(edu->3|cmu->3)/freq(cmu->3)  = 155/511  = 0.30332681 |

|  |
| --- |
| Perhitungan probabilitas untuk kategori pndidikan pasangan:  P(partner\_edu->4|cmu->1) = freq(partner\_edu ->4|cmu->1)/freq(cmu->1)  = 338/629  = 0.53736089  P(partner\_edu ->4|cmu->2) = freq(partner\_edu ->34cmu->2)/freq(cmu->2)  = 257/333  = 0.77177177  P(partner\_edu ->4|cmu->3) = freq(partner\_edu ->4|cmu->3)/freq(cmu->3)  = 304/511  = 0.59491194 |

|  |
| --- |
| Perhitungan probabilitas untuk kategori jumlah anak:  P(num->1|cmu->1) = freq(num ->1|cmu->1)/freq(cmu->1)  = 422/629  = 0.6709062  P(num ->1|cmu->2) = freq(num ->1|cmu->2)/freq(cmu->2)  = 172/333  = 0.516516517  P(num ->1|cmu->3) = freq(num ->1|cmu->3)/freq(cmu->3)  = 314/511  = 0.614481409 |

|  |
| --- |
| Perhitungan probabilitas untuk kategori agama:  P(rel->1|cmu->1) = freq(rel ->1|cmu->1)/freq(cmu->1)  = 554 /629  = 0.880763116  P(rel ->1|cmu->2) = freq(rel ->1|cmu->2)/freq(cmu->2)  = 257 /333  = 0.771771772  P(rel ->1|cmu->3) = freq(rel ->1|cmu->3)/freq(cmu->3)  = 442 /511  = 0.864970646 |

|  |
| --- |
| Perhitungan probabilitas untuk kategori bekerja:  P(work->1|cmu->1) = freq(work ->1|cmu->1)/freq(cmu->1)  = 459 /629  = 0.72972973  P(work ->1|cmu->2) = freq(work ->1|cmu->2)/freq(cmu->2)  = 244 /333  = 0.732732733  P(work ->1|cmu->3) = freq(work ->1|cmu->3)/freq(cmu->3)  = 401 /511  = 0.784735812 |

|  |
| --- |
| Perhitungan probabilitas untuk kategori kesibukan suami:  P(occu->2|cmu->1) = freq(occu ->2|cmu->1)/freq(cmu->1)  = 200 /629  = 0.317965024  P(occu ->2|cmu->2) = freq(occu ->2|cmu->2)/freq(cmu->2)  = 79 /333  = 0.237237237  P(occu ->2|cmu->3) = freq(occu ->1|cmu->3)/freq(cmu->3)  = 146 /511  = 0.285714286 |

|  |
| --- |
| Perhitungan probabilitas untuk kategori standar hidup:  P(liv->3|cmu->1) = freq(liv->3|cmu->1)/freq(cmu->1)  = 184 /629  = 0.292527822  P(liv->3|cmu->2) = freq(liv->3|cmu->2)/freq(cmu->2)  = 90/333  = 0.27027027  P(liv->3|cmu->3) = freq(liv->3|cmu->3)/freq(cmu->3)  = 157/511  = 0.307240705 |

|  |
| --- |
| Perhitungan probabilitas untuk kategori *media exposure*:  P(med->0|cmu->1) = freq(med->0|cmu->1)/freq(cmu->1)  = 555/629  = 0.882352941  P(med->0|cmu->2) = freq(med->0|cmu->2)/freq(cmu->2)  = 323/333  = 0.96996997  P(med->0|cmu->3) = freq(med->0|cmu->3)/freq(cmu->3)  = 486/511  = 0.951076321 |

|  |
| --- |
| Probabilitas kemunculan cmu = 1  P(cmu->1) = freq(cmu->1)/total data  = 629/1473  = 0.427019688  Probabilitas kemunculan cmu = 2  P(cmu->2) = freq(cmu->2)/total data  = 511/1473  = 0.226069246  Probabilitas kemunculan cmu = 3  P(cmu->3) = freq(cmu->3)/total data  = 333/1473  = 0.346911066 |

Setelah mendapatkan probabilitas untuk setiap kategori dan perobabilitas untuk setiap cmu, langkah selanjutnya adalah mencari probabilitas untuk setiap kasifikasi berdasarkan data tes yang dimiliki.

|  |
| --- |
| Probabilitas tidak menggunakan alat kontrasepsi:  P(test|p) \* P(cmu->1) = 0.2782194 \* 0.278219396 \* 0.53736089 \* 0.6709062 \* 0.880763116 \* 0.72972973 \* 0.317965024 \* 0.292527822 \* 0.882352941 \* 0.427019688 = **0.00062858**  Probabilitas penggunaan jangka Panjang:  P(test|p) \* P(cmu->2) = 0.25225225 \* 0.24024024 \* 0.77177177 \* 0.516516517 \* 0.771771772 \* 0.732732733 \* 0.237237237 \* 0.27027027 \* 0.96996997 \* 0.226069246 = **0.00019207**  Probabilitas penggunaan jangka pendek:  P(test|p) \* P(cmu->3) = 0.09197652 \* 0.30332681 \* 0.59491194 \* 0.614481409 \* 0.864970646 \* 0.784735812 \* 0.285714286 \* 0.307240705 \* 0.951076321 \* 0.346911066 = **0.0002005** |

Jadi klasifikasi penggunaan metode kontrasepsi untuk data tes diatas kemungkinan adalah tidak menggunakan alat kontrasepsi.

* 1. **Analisis program(penjelasan program)**

Bahasa pemrogramman yang digunakan adalah bahasa Python:

|  |
| --- |
| **# 1. Age - age of the woman**  **# if age < 26 then 'young', category = 1**  **# if age < 41 then 'adult', category = 2**  **# else 'old', category = 3**  **# 2. Education - level of education woman has received (1=low, 4=high)**  **# 3. Partner Education - level of education partner has received (1=low, 4=high)**  **# 4. Number of Children - number of kids mothered by woman**  **# if num < 4 then 'low', category = 1**  **# if num < 10 then 'medium', category = 2**  **# else then 'high', category = 3**  **# 5. Religion=Islam - woman that identify as Muslim (0=No, 1=Yes)**  **# 6. Currently Working - woman is currently employed (0=Yes, 1=No)**  **# 7. Husbands Occupation - Not specified (categorical 1-4)**  **# 8. Standard of Living - based on the standard of living index (1=low, 4=high)**  **# 9. Media exposure - quality of media exposure (0=Good, 1=Not good)**  **# 10. Contraceptive Method Used - 1=No-use, 2=Long-term, 3=Short-term**  **import pandas as pd**  **import numpy as np**  **from sklearn.model\_selection import train\_test\_split**  **from sklearn.metrics import accuracy\_score**  **df = pd.read\_csv('1987\_Indonesia\_Contraception\_Prevalence\_Study.csv')**  **# menentukan kategori decision yang ada pada data frame**  **class\_dec = np.array(df['Contraceptive Method Used'].drop\_duplicates())**  **# menentukan jumlah klasifikasi decision dan jumlah kategori/kolom**  **sizeof\_class = class\_dec.size**  **sizeof\_columns = df.shape[1]-1**  **sizeof\_line = df.shape[0]**  **# labeling data**  **age = df['Age'].values**  **num\_child = df['Number of Children'].values**  **i = 0**  **while i < sizeof\_line:**  **# labeling pada kolom age**  **if age[i] < 26:**  **age[i] = 1**  **elif age[i] < 41:**  **age[i] = 2**  **else:**  **age[i] = 3**    **#labeling pada kolom number of children**  **if num\_child[i] < 4:**  **num\_child[i] = 1**  **elif num\_child[i] < 10:**  **num\_child[i] = 2**  **else:**  **num\_child[i] = 3**  **i = i + 1**  **df['Age'] = age**  **df['Number of Children'] = num\_child**  **# ---------------------------membagi dataset---------------------------------------------**  **x = df.drop(['Contraceptive Method Used'], axis=1)**  **x = x.values**  **y = df['Contraceptive Method Used'].values**  **x\_train, x\_test, y\_train, y\_test = train\_test\_split(x,y, test\_size=0.2, random\_state=1234)**  **# --------------------------------------------------------------------------------------**  **#data\_test = x\_test[15]**  **#print(data\_test)**  **# fungsi perkalian**  **def perkalian(p\_c\_d, i, j):**  **global sizeof\_columns**  **if i == sizeof\_columns:**  **return 1**  **else:**  **return p\_c\_d[i][j] \* perkalian(p\_c\_d, i+1, j)**  **def train\_bayes(df, data\_test):**  **# ambil data**  **count\_age = df[df['Age']==data\_test[0]]**  **count\_edu = df[df['Education']==data\_test[1]]**  **count\_partner\_edu = df[df['Partner Education ']==data\_test[2]]**  **count\_num\_child = df[df['Number of Children']==data\_test[3]]**  **count\_relig = df[df['Religion = Islam']==data\_test[4]]**  **count\_current\_work = df[df['Currently working']==data\_test[5]]**  **count\_husb\_occu = df[df['Husband Occupation ']==data\_test[6]]**  **count\_standart\_liv = df[df['Standard of Living']==data\_test[7]]**  **count\_media\_expo = df[df['Media Exposure']==data\_test[8]]**    **# menghitung banyaknya kemunculan untuk setiap decision/class**  **i = 0**  **freq\_d = np.zeros((sizeof\_class))**  **while i < sizeof\_class:**  **freq\_d[i] = len(df[df['Contraceptive Method Used']==class\_dec[i]])**  **i = i + 1**    **# menghitung banyaknya kemunculan untuk setiap kombinasi kategori dan decision**  **j = 0**  **freq\_c\_d = np.zeros((sizeof\_columns, sizeof\_class))**  **while j < sizeof\_class:**  **freq\_c\_d[0][j] = len(count\_age[count\_age['Contraceptive Method Used']==class\_dec[j]])**  **freq\_c\_d[1][j] = len(count\_edu[count\_edu['Contraceptive Method Used']==class\_dec[j]])**  **freq\_c\_d[2][j] = len(count\_partner\_edu[count\_partner\_edu['Contraceptive Method Used']==class\_dec[j]])**  **freq\_c\_d[3][j] = len(count\_num\_child[count\_num\_child['Contraceptive Method Used']==class\_dec[j]])**  **freq\_c\_d[4][j] = len(count\_relig[count\_relig['Contraceptive Method Used']==class\_dec[j]])**  **freq\_c\_d[5][j] = len(count\_current\_work[count\_current\_work['Contraceptive Method Used']==class\_dec[j]])**  **freq\_c\_d[6][j] = len(count\_husb\_occu[count\_husb\_occu['Contraceptive Method Used']==class\_dec[j]])**  **freq\_c\_d[7][j] = len(count\_standart\_liv[count\_standart\_liv['Contraceptive Method Used']==class\_dec[j]])**  **freq\_c\_d[8][j] = len(count\_media\_expo[count\_media\_expo['Contraceptive Method Used']==class\_dec[j]])**  **j = j + 1**    **# menghitung probabilitas kemunculan untuk setiap kombinasi kategori dan decision/class**  **p\_c\_d = np.zeros((sizeof\_columns, sizeof\_class))**  **i = 0**  **while i < sizeof\_columns:**  **j = 0**  **while j < sizeof\_class:**  **p\_c\_d[i][j] = freq\_c\_d[i][j]/freq\_d[j]**  **j = j + 1**  **i = i + 1**    **# mnghitung probabilitas kemunculan setiap decision yang ada pada data frame**  **p\_d = np.zeros((sizeof\_class))**  **i = 0**  **while i < sizeof\_class:**  **p\_d[i] = freq\_d[i]/sizeof\_line**  **i = i + 1**    **# menghitung probabilitas dari decision saat diketahui kondisi data\_test**  **p\_test\_d = np.zeros((sizeof\_class))**  **i = 0**  **while i < sizeof\_class:**  **p\_test\_d[i] = perkalian(p\_c\_d, 0, i) \* p\_d[i]**  **i = i + 1**    **print(f'freq-c-d : \n{freq\_c\_d}\n')**  **print(f'freq-d:\n{freq\_d}\n')**  **print(f'prob(c|d):\n{p\_c\_d}\n')**  **print(f'prob(d):\n{p\_d}\n')**  **print(f'prob(test|d):\n{p\_test\_d}\n')**  **if max(p\_test\_d) == p\_test\_d[0]:**  **return 1**  **elif max(p\_test\_d) == p\_test\_d[1]:**  **return 2**  **else:**  **return 3**  **'''**  **test di 20% data asli**  **y\_pred = np.zeros((x\_test.shape[0]))**  **for i in range(x\_test.shape[0]):**  **y\_pred[i] = train\_bayes(df, x\_test[i])**    **acc\_score = accuracy\_score(y\_test, y\_pred)**  **print(f'accuracy score:{acc\_score}')**  **'''**  **'''**  **test di 20 data asli**  **test = pd.read\_excel('train.xlsx')**  **x\_test2 = test.drop(['Contraceptive Method Used'], axis=1)**  **x\_test2 = x\_test2.values**  **y\_test2 = test['Contraceptive Method Used'].values**  **y\_pred = np.zeros((x\_test2.shape[0]))**  **for i in range(x\_test2.shape[0]):**  **y\_pred[i] = train\_bayes(df, x\_test2[i])**  **'''**  **pred = train\_bayes(df, [3, 3, 4, 1, 1, 1, 2, 3, 0])**  **if pred == 1:**  **print('cmu : no-use')**  **elif pred == 2:**  **print('cmu : long-term')**  **else:**  **print('cmu : short-term')** |

Adapun evaluasi terhadap sistem dengan menghitung akurasi dari model yang telah dibuat. Untuk perhitungan akurasi kami menggunakan dua pendekatan yaitu dengan menguji pada 20 data acak dari total keseluruhan data dan menguji pada 20% data dari total keseluruhan data. Data yang digunakan adalah sebagai berikut:

* Pengujian terhadap 20 data acak

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age | Education | Partner Education | Number of Children | Religion = Islam | Currently working | Husband Occupation | Standard of Living | Media Exposure | Contraceptive Method Used |
| 1 | 3 | 3 | 1 | 1 | 0 | 3 | 2 | 0 | 1 |
| 2 | 2 | 3 | 1 | 1 | 1 | 3 | 4 | 0 | 1 |
| 3 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 |
| 2 | 1 | 3 | 1 | 1 | 0 | 3 | 3 | 1 | 1 |
| 2 | 4 | 4 | 1 | 1 | 0 | 1 | 4 | 0 | 3 |
| 1 | 2 | 2 | 1 | 1 | 1 | 3 | 2 | 0 | 3 |
| 3 | 4 | 4 | 2 | 0 | 0 | 2 | 4 | 0 | 2 |
| 2 | 4 | 4 | 2 | 0 | 0 | 2 | 4 | 0 | 2 |
| 3 | 4 | 4 | 2 | 0 | 0 | 2 | 4 | 0 | 2 |
| 3 | 4 | 4 | 1 | 0 | 1 | 1 | 4 | 0 | 2 |
| 2 | 4 | 4 | 1 | 1 | 1 | 1 | 4 | 0 | 3 |
| 2 | 3 | 3 | 2 | 1 | 0 | 1 | 4 | 0 | 3 |
| 2 | 3 | 3 | 2 | 1 | 0 | 2 | 2 | 0 | 3 |
| 1 | 3 | 3 | 1 | 1 | 1 | 2 | 4 | 0 | 3 |
| 1 | 3 | 1 | 1 | 1 | 0 | 3 | 4 | 0 | 2 |
| 2 | 4 | 4 | 2 | 1 | 1 | 2 | 4 | 0 | 2 |
| 3 | 4 | 4 | 2 | 1 | 0 | 1 | 4 | 0 | 2 |
| 2 | 4 | 4 | 2 | 1 | 0 | 3 | 3 | 0 | 2 |
| 1 | 4 | 4 | 1 | 1 | 0 | 2 | 2 | 0 | 2 |
| 2 | 3 | 4 | 1 | 0 | 1 | 1 | 4 | 0 | 2 |

Dengan menggunakan program yang telah dibuat, kami mendapatkan hasil dari kolom *Contraceptive Method Used* adalah sebagai berikut:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age | Education | Partner Education | Number of Children | Religion = Islam | Currently working | Husband Occupation | Standard of Living | Media Exposure | Contraceptive Method Used | Hasil dari program |
| 1 | 3 | 3 | 1 | 1 | 0 | 3 | 2 | 0 | 1 | 1 |
| 2 | 2 | 3 | 1 | 1 | 1 | 3 | 4 | 0 | 1 | 3 |
| 3 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 |
| 2 | 1 | 3 | 1 | 1 | 0 | 3 | 3 | 1 | 1 | 1 |
| 2 | 4 | 4 | 1 | 1 | 0 | 1 | 4 | 0 | 3 | 2 |
| 1 | 2 | 2 | 1 | 1 | 1 | 3 | 2 | 0 | 3 | 1 |
| 3 | 4 | 4 | 2 | 0 | 0 | 2 | 4 | 0 | 2 | 2 |
| 2 | 4 | 4 | 2 | 0 | 0 | 2 | 4 | 0 | 2 | 2 |
| 3 | 4 | 4 | 2 | 0 | 0 | 2 | 4 | 0 | 2 | 2 |
| 3 | 4 | 4 | 1 | 0 | 1 | 1 | 4 | 0 | 2 | 2 |
| 2 | 4 | 4 | 1 | 1 | 1 | 1 | 4 | 0 | 3 | 2 |
| 2 | 3 | 3 | 2 | 1 | 0 | 1 | 4 | 0 | 3 | 2 |
| 2 | 3 | 3 | 2 | 1 | 0 | 2 | 2 | 0 | 3 | 3 |
| 1 | 3 | 3 | 1 | 1 | 1 | 2 | 4 | 0 | 3 | 3 |
| 1 | 3 | 1 | 1 | 1 | 0 | 3 | 4 | 0 | 2 | 1 |
| 2 | 4 | 4 | 2 | 1 | 1 | 2 | 4 | 0 | 2 | 2 |
| 3 | 4 | 4 | 2 | 1 | 0 | 1 | 4 | 0 | 2 | 2 |
| 2 | 4 | 4 | 2 | 1 | 0 | 3 | 3 | 0 | 2 | 2 |
| 1 | 4 | 4 | 1 | 1 | 0 | 2 | 2 | 0 | 2 | 1 |
| 2 | 3 | 4 | 1 | 0 | 1 | 1 | 4 | 0 | 2 | 2 |

Terlihat dari data diatas terdapat tujuh data yang tidak sesuai dari data aslinya atau hanya 13 dari 20 data yang tepat sasaran. Disini bisa dicari akurasi dari model tersebut.

akurasi = (13/20)\*100

akurasi = **65%**

* Pengujian terhadap 20% data acak

Untuk pengujian ini kami menggunakan *function* yang disediakan python yaitu *train\_test\_split()*, fungsi ini akan membagi dataset yang dimiliki, dalam kasus ini menjadi 20% dari data asli. Unutk *syntax* nya adalah sebagai berikut:

|  |
| --- |
| 1  x = merupakan data yang akan diuji  y = merupakan data target  test\_size=0.2, membagi data menjadi 20% dari data asli  random\_state, mengambil data\_test secara acak |

Data yang akan di uji adalah x\_test yang jumlahnya 20% dari data asli, untuk *syntax*nya adalah sebagai berikut:

|  |
| --- |
| 2 |

Dengan menggunakan fungsi *accuracy\_score()* akan dibandingkan kecocokan antara y\_test(data target) dan y\_pred(data prediksi). Setelah dihitung, fungsi tersebut akan mengembalikan nilai dari 0 sampai 1, semakin suatu nilai dekat dengan angka satu maka semakin baik. Dari hasil tersebut didapat akurasi skor sebesar **0.535593220338983.**

Dari perbandingan tersebut bisa dilihat bahwa akurasi untuk data 20% lebih kecil dari 20 data.

# Kesimpulan dan Saran

**Kesimpulan**

Berdasarkan permasalahan yang telah dibahas maka dapat diambil keputusan bahwa, model tersebut dapat dijadikan acuan untuk memilih metode kontrasepsi yang akan digunakan berdasarkan probabilitas untuk setiap metode kontrasepsinya.

Bisa disimpulkan pula metode ini kurang cocok untuk digunakan pada data yang sangat besar, tetapi bisa digunakan untuk data yang relatif sedikit. Hal ini karena semakin banyak data akan semakin banyak pula ketidakkonsistenan dari suatu data tersebut, hal ini akan menyebabkan kecilnya tingkat akurasi dari hasil prediksi. Untuk sebuah model yang memiliki akurasi diatas 50% sebenarnya sudah cukup baik apalagi data yang digunakan melebihi 1400 data.

**Saran**

Setelah perancangan metode klasifikasi ini dilakukan, terdapat beberapa saran yang dapat diterapkan untuk pengembangan yang lebih lanjut, diantaranya:

# Diharapkan adanya pengembangan ruang lingkup sistem lebih lanjut mengenai data anamnesa akseptor KB dengan menambahkan kriteria kondisi tubuh yang lebih terperinci sesuai dengan jenis kontrasepsi yang akan digunakan agar menambah keakuratan dan menjadikan sistem ini lebih dikompleks.

# Untuk kedepannya diharapkan sistem klasifikasi ini dibuat lebih dinamis dan tidak hanya dikembangkan secara offline saja, tetapi juga bisa digunakan secara online dalam bentuk aplikasi mobile yang dapat diinstal di perangkat mobile seperti Os Android maupun yang berbasis web sehingga lebih mudah digunakan.

# Diperlukan penggunaan algoritma yang berbeda pada klasifikasi ini dan dilakukan riset yang lebih mendalam

Referensi :

<https://informatikalogi.com/algoritma-naive-bayes/>

<https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes/>

<https://www.kaggle.com/joelzcharia/contraceptive-prevalence-survey?select=1987+Indonesia+Contraception+Prevalence+Study.csv>