**Analisa Data Mining Untuk Memprediksi Skill Mahasiswa Menggunakan Metode Naïve Bayes di Program Studi Sistem Informasi UIN Imam Bonjol Padang**

Yaslinda Lizar1,\*, Alya Sahira Firrizqi2, Asriwan Guci3, Joko Sunadi4

12Fakultas Sains dan Teknologi UIN Imam Bonjol, Padang, Indonesia

3STIKes MERCUBAKTIJAYA Padang, Indonesia

4Universitas Putra Indonesia YPTK Padang, Indonesia

|  |  |  |
| --- | --- | --- |
| **Article Information** |  | **ABSTRACT** |
| *Article History:*  Accepted by the Editor:  Final Revision:  Published Online: | Mahasiswa yang memiliki suatu *skill* tertentu merupakan sesuatu yang dapat membawa dampak yang baik bukan hanya kepada diri mahasiswa itu sendiri namun juga memiliki dampak yang baik terhadap Program Studi di suatu Fakultas serta Universitas. Tidak hanya IPK yang tinggi, mahasiswa yang memiliki *skill* tertentu akan memiliki nilai plus serta akan mampu bersaing ketika terjun ke dunia kerja nantinya. Oleh karena itu untuk mengetahui mahasiswa yang memiliki *skill* maka dapat dilakukan dengan menggunakan *data mining* dengan menggunakan metode *Naïve Bayes*. Penelitian ini dilakukan untuk memprediksi *skill* yang dimilikimahasiswa program studi Sistem Informasi di UIN Imam Bonjol Padang. Dan batasan dari penelitian ini ialah hanya memprediksi *skill* mahasiswa dalam bidang pengolahan data dan programer atau *coding.* Data yang akan diambil yaitu dari nilai matakuliah mahasiswa yang berkaitan dengan pengolahan data dan pemrograman. Metode penelitian yang digunakan ialah metode kuantitatif, dan tools yang digunakan ialah menggunakan aplikasi data mining Orange serta untuk model pengujian yaitu menggunakan *Confusion Matrix*. Hasil penelitian prediksi *skill* yang dimiliki mahasiswa program studi sistem informasi di UIN Imam Bonjol Padang yang diambil 14 sampel data ialah memiliki kemampuan atau *skill* dalam bidang pemrograman atau coding dan memiliki confusion matriks atau tingkat ke akuratan sebesar 50 %.  [https://licensebuttons.net/l/by-sa/3.0/88x31.png](http://creativecommons.org/licenses/by-sa/4.0/)This is an open access article under the CC–BY-SA license |
| **Keywords** |
| Data Mining  Metode Naïve Bayes  Klasifikasi Data  Mahasiswa  *Skill* |
| **Correspondence** |
| E-mail: yaslinda@uinib.ac.id\* |

# Introduction

Mahasiswa yang memiliki nilai potensi diatas rata-rata merupakan mahasiswa yang tidak hanya unggul dibidang akademik saja, tetapi juga selalu berkreasi, berinovasi dan berkreatif serta juga memiliki skill atau kemampuan khusus [2]. Saat ini dunia globalisasi sedang masuk kedalam sendi kehidupan manusia, hampir mempengaruhi seluruh aspek kehidupan manusia saat ini. Bahkan, [3] menggambarkan bahwa dunia saat ini seperti ketika seseorang sedang tertidur lalu bangun pagi dan melihat semuanya telah berubah. Dengan perubahan jaman yang semakin besar, hal ini juga berdampak kepada tuntutan – tuntutan di dalam dunia kerja semakin berat.

Generasi yang hidup di jaman atau di era digital ini tidak hanya melakukan kerja dalam setting kantoran, namun dengan begitu majunya ilmu teknologi sekarang, pekerjaan akan dapat dilakukan dimana saja [2]. Artinya ialah dengan adanya perubahan gaya dalam bekerja yang biasanya hanya dilakukan di dalam kantor namun sekarang sudah dapat dilakukan dimana saja dengan adanya teknologi bahkan orang bisa bekerja dari rumah dan melakukan meeting dengan klien luar negeri tanpa perlu keluar rumah dengan *teleconference*, maka siapapun sekarang dituntut atau dipaksa untuk dapat berimprovisasi dan memiliki skill tertentu yang dapat ditonjolkan untuk masuk kedalam dunia kerja di era sekarang ini. Tidak hanya mengandalkan nilai akademis, namun keterampilan suatu *skill* juga sangat diperlukan.

Oleh sebab itu, untuk mengetahui *skill* yang dimiliki oleh mahasiswa diperlukannya suatu cara yang dapat menjawab pertanyaan diatas. Pada masa saat ini, *data mining* digunakan untuk dapat membantu menyelesaikan masalah serta dapat membantu dalam pengambilan suatu keputusan. Beberapa penelitian yang berkaitan dengan *data mining* yaitu dipergunakan untuk meningkatkan akreditas suatu perguruan tinggi Universitas Dian Nuswantoro, 2012 yang diambil dari banyak nya data mahasiswa dan data jumlah kelulusan maka untuk dapat mengambil pelajaran digunakanlah data mining untuk proses penambangan informasi sehingga berguna untuk pihak Universitas [4].

*Data mining* adalah suatu teknik yang digunakan untuk penambangan data untuk membuat suatu model, lalu dari model tersebut dilakukan pengenalan pola data dan diambilnya sebuah informasi [5]. Dari banyaknya teknik dalam *data mining*, salah satu tekniknya ialah klasifikasi. Teknik klasifikasi ini merupakan teknik pembelajaran untuk melakukan sebuah prediksi nilai dari suatu target nilai [6]. Karena dalam penelitian ini memiliki dua variable yaitu mahasiswa dengan *skill* pengolahan data dan mahasiswa yang memiliki *skill coding* maka penelitian ini akan menggunakan teknik klasifikasi dalam *data mining*.

# Literature Review

## Data Mining

*Data Mining* merupakan sebuah proses yang memiliki banyak teknik pembelajaran komputer atau yang disebut sebagai *machine learning* yang digunakan untuk menganalisis dan juga memasukkan atau meng-ekstraksi pengetahuan yang dilakukan secara otomatis. *Data mining* juga sering dikaitkan dengan sesuatu cara yang digunakan untuk menguraikan suatu penemuan serta mencari informasi dan pengetahuan yang ada di dalam sebuah *database, data warehouse* dan juga *big data*.

Dari arti *data mining* sendiri ialah menambang, dimana data mining adalah suatu proses menggali atau menambang pengetahuan dari sekumpulan data yang banyak atau dalam jumlah yang sangat besar [2]. Tujuan dari data mining sendiri yaitu untuk dapat menemukan suatu hubungan yang terkait atau suatu pola yang dapat memberikan hasil indikasi yang bermanfaat untuk penggunanya.

## Teknik Klasifikasi

Klasifikasi adalah salah satu teknik dari banyak nya teknik yang ada di dalam *data mining*. Teknik klasifikasi ini dapat digunakan untuk memprediksi atau meramalkan kecenderungan yang akan terjadi di masa yang akan datang [4].

Dalam klasifikasi ini proses yang akan dilakukan ialah nantinya akan menggolongkan data kedalam beberapa variable dengan tujuan untuk membentuk suatu model dan memperhatikan atribut yang berpengaruh [7].

Komponen – komponen dari sebuah Teknik klasifikasi ini ialah sebagai berikut:

1. Kelas, ialah merupakan label yang ditentukan dari hasil klasifikasi
2. *Predicator*, ialah merupakan variable bebas yang ditentukan berdasarkan hasil dari karakteristik atribut yang telah diklasifikasikan
3. *Data Training*, ialah merupakan banyak data yang dikumpulkan yang di dalamnya berisi kelas dan juga predicator yang bertujuan untuk dilatih agar dapat dikelompokkan ke dalam kelas benar dan tepat.
4. *Data Testing*, ialah merupakan sekumpulan data baru yang nantinya akan dikelompokkan yang bertujuan untuk mengetahui tingkat akurasi dari model yang telah dibuat sebelumnya. Biasanya *data testing* jumlahnya akan lebih sedikit dari pada *data training.*

## Algoritma Naïve Bayes

Algoritma *Naïve Bayes* ialah klasifikasi digunakan untuk memprediksi probabilitas. *Naïve bayes* ini didasarkan pada teorema Bayes yang memiliki kemampuan mengklasifikasikan yang hampir mirip dengan *decision tree*. Bayes ini memiliki akurasi yang tinggi saat diterapkan kedalam database [7]. Algoritma ini memiliki tingkat akurasi yang tinggi dalam mengklasifikasikan data [8]. Berikut adalah rumus dari teorema *Naïve Bayes*:

Text

Description automatically generated

Keterangan:

X : Data dengan class yang tidak diketahui

H : Hipotesis data X merupakan class yang spesifik

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X

P(H) : Probabilitas hipotesis H

P(X|H) : Probabilitas X berdasarkan kondisi H

P(X) : Probabilitas hipotesis X

## Confusion Matrix

*Confusion matrix* merupakan sebuah *tool* yang biasa digunakan di dalam *data mining* yang digunakan untuk mengevaluasi model klasifikasi untuk memprediksi sebuah objek yang benar atau objek yang salah [9]. *Confusion matrix* dapat juga diartikan sebagai tabel yang memberikan keseluruhan data yang telah diuji untuk mengetahui tingkat keakuratan nya [10]. Matrix dari prediksi akan dibandingkan dengan kelas *data testing* dan *training* yang berisi nilai *actual* dan prediksi pada klasifikasi.

Tabel 3. 1 *Confusion Matrix*

Table

Description automatically generated

Dan berikut adalah rumus untuk menghitung tingkat akurasi:

Text

Description automatically generated with low confidence

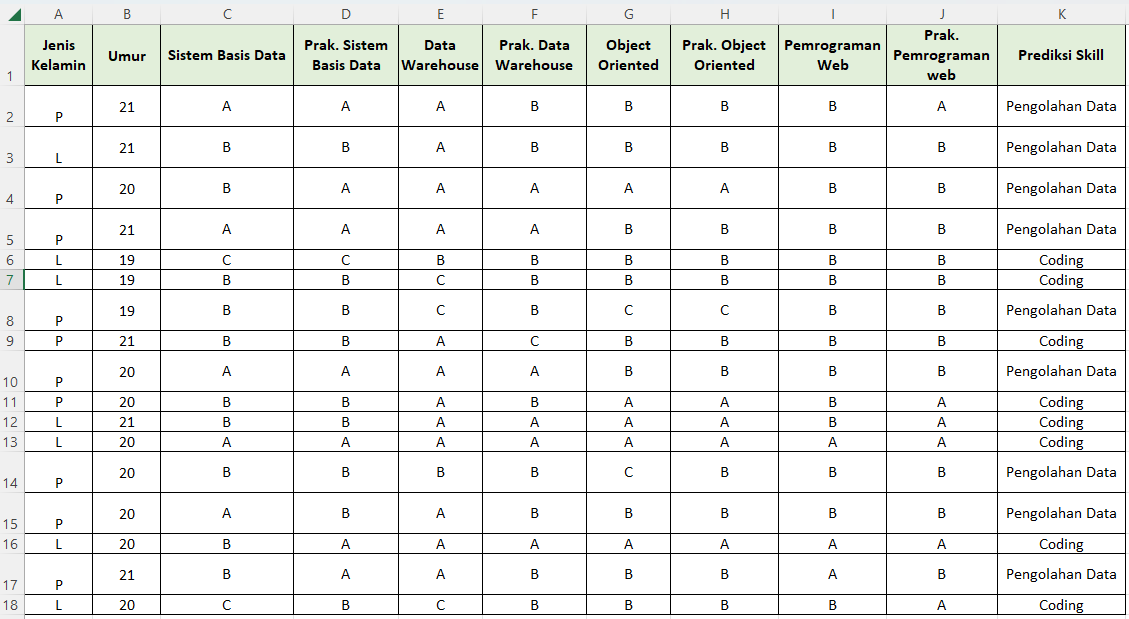
# Method

Metode penelitian yang digunakan di dalam penelitian ini ialah menggunakan metode penelitian kuantitatif, dengan teknik menganalisa data. *Tools* yang digunakan ialah menggunakan aplikasi *data mining* Orange serta untuk model pengujian yaitu menggunakan *Confusion Matrix.*

# Results and Discussion

## Pengumpulan Data Awal

Data yang diambil dalam penelitian ini ialah diambil dari mahasiswa program studi Sistem Informasi pada Fakultas Sains dan Teknologi di UIN Imam Bonjol Padang. Berikut adalah *dataset* yang terdiri dari 17 data.



Gambar 5. 1 *Dataset*

## Mendeskripsikan Data

*Dataset* dalam penelitian ini terdiri dari beberapa atribut yaitu jenis kelamin, umur, nilai dari matakuliah; sistem basis data, prak. sistem *basis data, data warehouse*, prak. *data warehouse, object oriented,* prak. *object orientd*, pemrograman web, dan prak. pemrograman web. Jumlah dari *dataset* dalam penelitian ini ialah 17 data.

## Evaluasi Kualitas

Evaluasi pada data yang digunakan dalam penelitian tidak menemukan adanya nilai null atau biasa yang disebut dengan *missing value* dalam dataset mahasiswa yang digunakan.

## Seleksi Data

Dataset yang berjumlah 17 data akan diambil *data testing* nya untuk diproses. Yang dimana dalam penelitian ini digunakan *data testing* yang dimana 20% dari *dataset* atau *data training*. Maka *data testing* yang diperoleh adalah berjumlah 4 *data testing* dari 17 *dataset.*

## Pengolahan Data Mentah

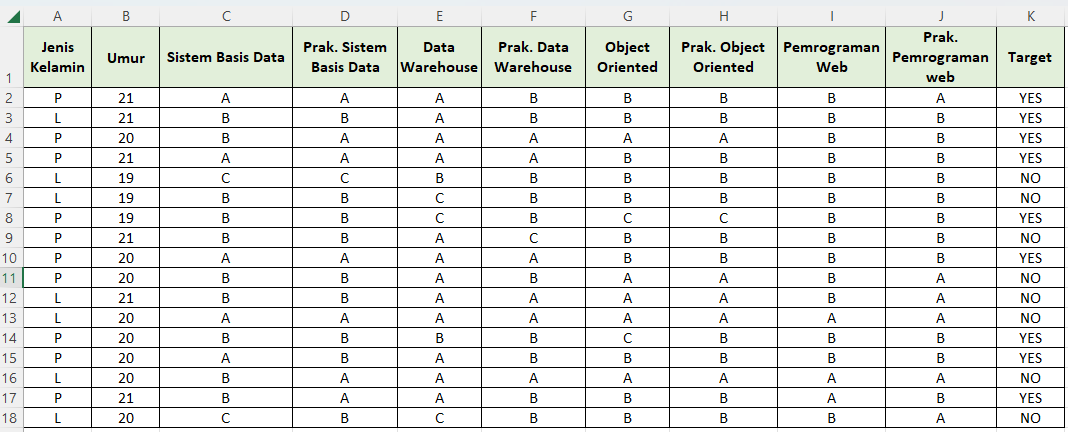
Pada pengolahan data adalah tahapan dimana untuk memastikan semua *dataset* yang dipilih telah benar untuk dilakukannya pengolahan atau yang disebut dengan *Preposessing Data*.

## Transformasi Data

Tahap ini adalah tahap melakukan inisialisasi pada data yang dijadikan target. Yang dimana pada penelitian kali ini target data yang dipilih ialah pada atribut prediksi *skill*.

1. Mahasiswa yang memiliki prediksi *skill* yang berinisial “YES” merupakan mahasiswa yang memiliki *skill* dibidang pengolahan data
2. Mahasiswa yang memiliki prediksi *skill* yang “NO” merupakan mahasiswa yang mahasiswa yang memiliki *skill* dibidang coding

Berikut adalah tabel data set yang telah dilakukan inisialisasi :



Gambar 5. 2 *Dataset After Initialization*

## Teknik Pemodelan

Teknik yang diambil dalam penelitian ini ialah menggunkan algoritma *Neive Bayes* dan *tool* yang dipakai adalah menggunakan aplikasi *data mining* Orange.

## Perhitungan Data Mining

Perhitungan *data mining* atau menghitung jumlah atau label atau target kelas “P(H)”. Perhitungan jumlah seluruh kelas berdasarkan klasifikasi yang terbentuk atau disebut sebagai *prior probability.*

*Class* 1 (C1) = *Skill* Pengolahan Data = YES = 9

*Class* 2 (C2) = *Skill* Coding = NO = 8

Total = 17 data

C1 (*Class* *Skill* Pengolahan Data = “YES”) = YES/Total

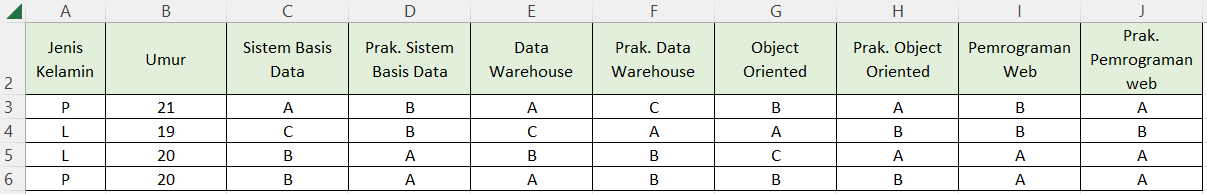
= 9/17 = 0,530

*C1 (Class Skill Coding* = “NO”) = NO/Total

= 8/17 = 0,470

## Perhitungan Kasus Perkelas

Pada tahap ini akan menghitung jumlah kasus yang sama “P(X|H)” dengan menggunakan *data testing* berikut.



Gambar 5. 3 Data Testing

1. X (Jenis kelamin=P, Umur=21, SBD=A, Prak SBD= B, DataWarehouse=A, Prak Dw= A, Obj Oriented=B, Prak Obj Oriented=A, Pemrog Web=B, Prak Pemrog Web = A)

Tabel 5. 1 Perhitungan pada Tabel Data Testing Pertama

|  |  |
| --- | --- |
| P(Jenis kelamin=P|YES)= 8/9 = 0,888  P(Jenis kelamin=P|NO) = 2/8 = 0,25 | P(Obj Oriented=B|YES) = 6/9 = 0,666  P(Obj Oriented=B|NO) = 4/8 = 0,5 |
| P(Umur=21|YES) = 4/9 = 0,444  P(Umur=21 |NO) = 2/8 = 0,25 | P(PrakObjOriented=A|YES)=1/9=0,111  P(Prak Obj Oriented=A|NO) =4/8 = 0,5 |
| P(SBD=A |YES) = 4/9 = 0,444  P(SBD=A |NO) = 1/8 = 0,125 | P(Pemrog Web=B |YES) = 8/9 = 0,888  P(Pemrog Web=B |NO) = 6/8 = 0,75 |
| P(Prak SBD= B |YES) = 4/9 = 0,444  P(Prak SBD= B |NO) = 5/8 = 0,625 | P(PrakPemrogWeb=A|YES)=1/9=0,111  P(PrakPemrogWeb=A|NO)=5/8=0,625 |
| P(DataWarehose=A|YES)=7/9=0,777  P(DataWarehose|NO) = 5/8 = 0,625 | P(Prak Dw= A |YES) = 3/9 = 0,333  P(Prak Dw= A|NO) = 3/8 = 0,375 |

1. Pengkalian Semua Variabel

Pada tahapan ini ialah mengalikan semua variable dengan rumus “P(X|H)”

1. P(X|YES) = 0,888 x 0,444 x 0,444 x 0,444 x 0,777 x 0,666 x 0,111 x 0,888 x 0,111 x 0,333

= 0,0001465

0,0001465 x 10.000 = 1,465

1. P(X|NO) = 0,25 x 0,25 x 0,125 x 0,625 x 0,625 x 0,5 x 0,5 x 0,75 x 0,625 x 0,375

= 0,0001341

0,00013411 x 10.000 = 1,341

1. Bandingkan Hasil Perkelas

Pada tahapan ini ialah mengalikan semua variable dengan rumus “P(X|H) x P(H)”

1. P(X|H) x P(H) = 0,530 x 1,465 = 0,776
2. P(X|H) x P(H) = 0,470 x 1,341= 0,630

**Perbandingan**

**Kesimpulan: C1 > C2 [YES]**

1. X (Jenis kelamin=L, Umur=19, SBD=A, Prak SBD= B, DataWarehouse=C, Prak Dw= A, Obj Oriented=A, Prak Obj Oriented=B, Pemrog Web=B, Prak Pemrog Web = B

Tabel 5. 2 Perhitungan pada Tabel Data Testing Kedua

|  |  |
| --- | --- |
| P(Jenis kelamin=L |YES) =1/9= 0,111  P(Jenis kelamin=L |NO) = 6/8 = 0,75 | P(Obj Oriented=A|YES) = 1/9 = 0,111  P(Obj Oriented=A |NO) = 4/8 = 0,5 |
| P(Umur=19|YES) = 1/9 = 0,111  P(Umur=19 |NO) = 2/8 = 0,25 | P(PrakObjOriented=B|YES)=1/9=0,111  P(PrakObjOriented=B|NO)=4/8= 0,5 |
| P(SBD=A |YES) = 4/9 = 0,444  P(SBD=A |NO) = 1/8 = 0,125 | P(Pemrog Web=B |YES) = 7/9 = 0,777  P(Pemrog Web=B |NO) = 4/8 = 0,5 |
| P(Prak SBD= B |YES) = 4/9 = 0,444  P(Prak SBD= B |NO) = 5/8 = 0,625 | P(PrakPemrogWeb=B|YES)=8/9=0,888  P(PrakPemrogWeb=B|NO)=3/8= 0,375 |
| P(DataWarehose=C|YES)=1/9=0,111  P(DataWarehose=C|NO)=2/8 = 0,25 | P(Prak Dw= A |YES) = 3/9 = 0,333  P(Prak Dw= A|NO) = 3/8 = 0,375 |

1. Pengkalian Semua Variabel

Pada tahapan ini ialah mengalikan semua variable dengan rumus “P(X|H)”

1. P(X|YES) = 0,111 x 0,111x 0,444 x 0,444 x 0,111 x 0,111 x 0,111 x 0,777 x 0,888 X 0,333

= 0,00000076

0,00000076 x 10.000 = 0,007

1. P(X|NO) = 0,75 x 0,25 x 0,125 x 0,625 x 0,25 x 0,5 x 0,5 x 0,5 x 0,375 x 0,375

= 0,0000643

0,0000402 x 10.000 = 0,643

1. Bandingkan Hasil Perkelas

Pada tahapan ini ialah mengalikan semua variable dengan rumus “P(X|H) x P(H)”

1. P(X|H) x P(H) = 0,530 x 0,007 = 0,003
2. P(X|H) x P(H) = 0,470 x 0,643 = 0,302

**Perbandingan**

**Kesimpulan: C1 < C2 [NO]**

1. X (Jenis kelamin=L, Umur=20, SBD=B, Prak SBD= A, DataWarehouse=B, Prak Dw= B, Obj Oriented=C, Prak Obj Oriented=A, Pemrog Web=A, Prak Pemrog Web = A

Tabel 5. 3 Perhitungan pada Tabel Data Testing Ketiga

|  |  |
| --- | --- |
| P(Jenis kelamin=L |YES) =1/9= 0,111  P(Jenis kelamin=L |NO) = 6/8 = 0,75 | P(Obj Oriented=C|YES) = 1/9 = 0,111  P(Obj Oriented=C |NO) = 4/8 = 0,5 |
| P(Umur=20 |YES) = 4/9 = 0,444  P(Umur=20 |NO) = 4/8 = 0,5 | P(PrakObjOriented=A|YES)=1/9=0,111  P(PrakObjOriented=A|NO)=4/8 = 0,5 |
| P(SBD=B |YES) = 4/9 = 0,444  P(SBD=B |NO) = 5/8 = 0,625 | P(Pemrog Web=A |YES) = 1/9 = 0,111  P(Pemrog Web= |NO) = 2/8 = 0,25 |
| P(Prak SBD= A |YES) = 4/9 = 0,444  P(Prak SBD= A |NO) = 5/8 = 0,625 | P(PrakPemrogWeb=A|YES)=1/9=0,111  P(PrakPemrogWeb =A|NO)=5/8= 0,625 |
| P(DataWarehose=B|YES)=1/9=0,111  P(DataWarehose=B|NO)=1/8 = 0,125 | P(Prak Dw= B |YES) = 6/9 = 0,666  P(Prak Dw= B |NO) =4/8 = 0,5 |

1. Pengkalian Semua Variabel

Pada tahapan ini ialah mengalikan semua variable dengan rumus “P(X|H)”

1. P(X|YES) = 0,111 x 0,444 x 0,444 x 0,444 x 0,111 x 0,111 x 0,111 x 0,111 x 0,111 x 0,666

= 0,00000011

0,000000 x 10.000 = 0,001

1. P(X|NO) = 0,75 x 0,5 x 0,625 x 0,625 x 0,125 x 0,5 x 0,5 x 0,25 x 0,625 x 0,5

= 0,0003576

0,0000402 x 10.000 = 3,576

1. **Bandingkan Hasil Perkelas**

Pada tahapan ini ialah mengalikan semua variable dengan rumus “P(X|H) x P(H)”

1. P(X|H) x P(H) = 0,530 x 0,001 = 0,005
2. P(X|H) x P(H) = 0,470 x 3,576 = 1,680

**Perbandingan**

**Kesimpulan: C1 < C2 [NO]**

1. X (Jenis kelamin=P, Umur=20, SBD=B, Prak SBD= A, DataWarehouse=A, Prak Dw= B, Obj Oriented=B, Prak Obj Oriented=B, Pemrog Web=A, Prak Pemrog Web = A

Tabel 5. 4 Perhitungan pada Tabel Data Testing Keempat

|  |  |
| --- | --- |
| P(Jenis kelamin=P |YES)=8/9 = 0,888  P(Jenis kelamin=P |NO) = 2/8 = 0,25 | P(ObjOriented=B |YES) = 6/9 = 0,666  P(Obj Oriented=B |NO) = 4/8 = 0,5 |
| P(Umur=20 |YES) = 4/9 = 0,444  P(Umur=20 |NO) = 4/8 = 0,5 | P(PrakObjOriented=B|YES)=1/9=0,111  P(PrakObjOriented=B|NO)=4/8= 0,5 |
| P(SBD=B |YES) = 4/9 = 0,444  P(SBD=B |NO) = 5/8 = 0,625 | P(Pemrog Web=A |YES) = 1/9 = 0,111  P(Pemrog Web= |NO) = 2/8 = 0,25 |
| P(Prak SBD= A|YES) = 4/9 = 0,444  P(Prak SBD= A|NO) = 5/8 = 0,625 | P(PrakPemrogWeb=A|YES)=1/9=0,111  P(PrakPemrogWeb=A|NO)=5/8= 0,625 |
| P(DataWarehose=A|YES)=7/9=0,777  P(DataWarehose=A|NO)=5/8 = 0,625 | P(Prak Dw=B|YES) = 6/9 = 0,666  P(Prak Dw=B|NO) =4/8 = 0,5 |

1. Pengkalian Semua Variabel

Pada tahapan ini ialah mengalikan semua variable dengan rumus “P(X|H)”

1. P(X|YES) = 0,888 x 0,444 x 0,444 x 0,444 x 0,777 x 0,666 x 0,111 x 0,111 x 0,111 x 0,666

= 0,0000366

0,0000366 x 10.000 = 0,366

1. P(X|NO) = 0,25 x 0,5 x 0,625 x 0,625 x 0,625 x 0,5 x 0,5 x 0,25 x 0,625 x 0,5 x

= 0,00005960

0,00005960 x 10.000 = 5,960

1. **Bandingkan Hasil Perkelas**

Pada tahapan ini ialah mengalikan semua variable dengan rumus “P(X|H) x P(H)”

1. P(X|H) x P(H) = 0,530 x 0,366 = 0,193
2. P(X|H) x P(H) = 0,470 x 5,960 = 2,801

**Perbandingan**

**Kesimpulan: C1 < C2 [NO]**

## Confusion Matrix

Tabel 5. 5 Confusion Matrix

|  |  |  |
| --- | --- | --- |
| Classification | Predicted class | |
| Class=Yes | Class=No |
| Class=Yes | TP = 1 | FN = 1 |
| Class=No | FP = 1 | TN = 3 |

Accuracy = (TP + TN / TP + FP + FN + TN) \*100%

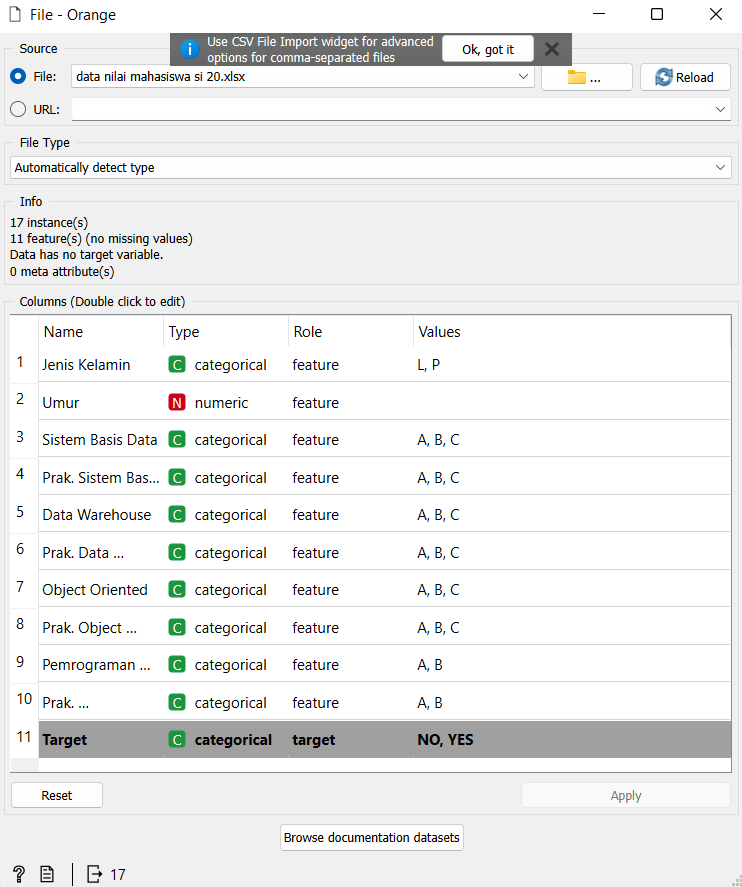
= (1 + 3 / 1 + 1 + 1 + 3) 100%

= 3/6 \* 100%

= 50%

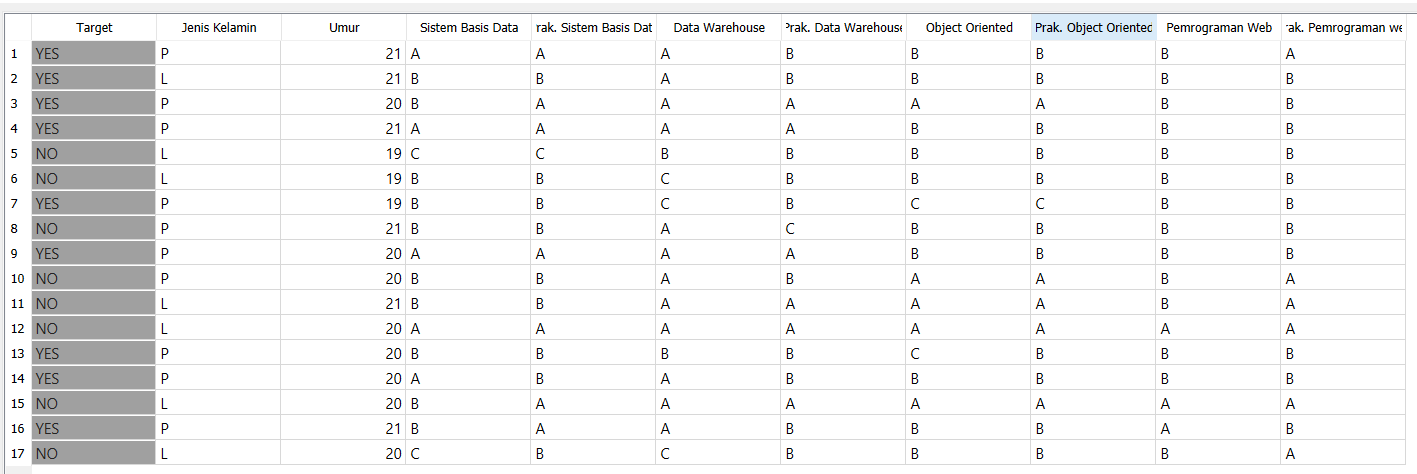
## Implementasi Aplikasi Orange

Berikut adalah gambar pengolahan data yang dilakukan di aplikasi data mining orange menggunakan algoritma *Naïve Bayes*:



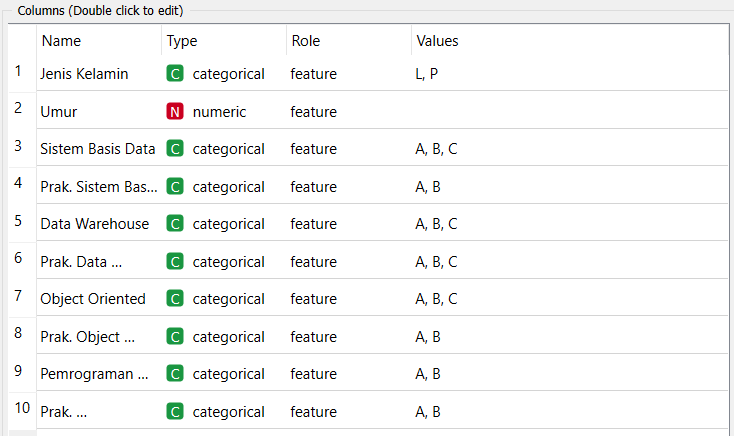
Gambar 5. 4 Columns Data Training

**Gambar 5.4** merupakan columns dari data training. Yang dimana pada tahap ini adalah menentukan target dari tabel data training. Pada penelitian ini yang menjadi targetnya ialah tabel Depresion.



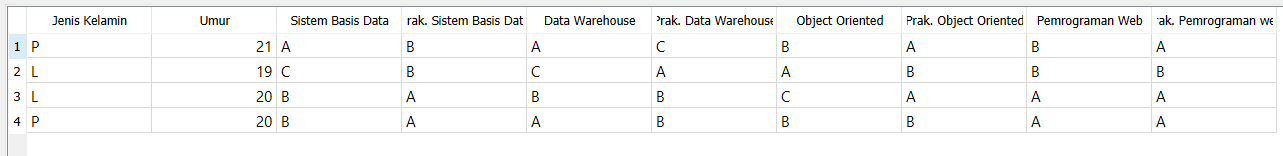
Gambar 5. 5 Isi Data Training

**Gambar 5.5** merupakan isi dari data training. Yang dimana data inilah nantinya yang akan diolah atau diproses dengan data testing.



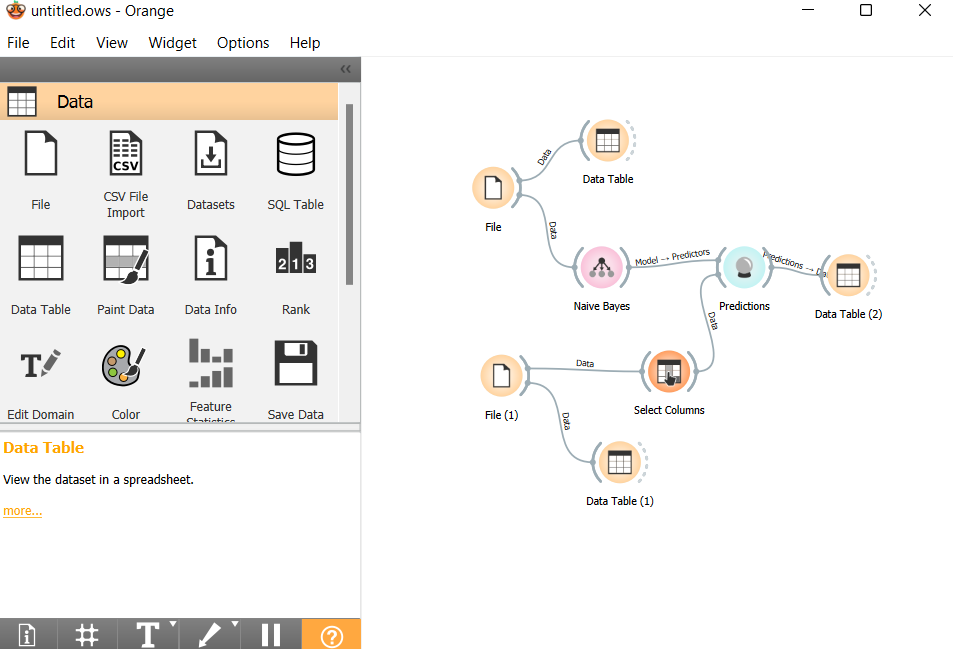
Gambar 5. 6 Data Testing

**Gambar 5.6** merupakan columns dari data testing.



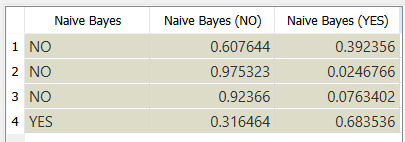
Gambar 5. 7 Isi Data Testing

**Gambar 5.7** merupakan isi dari data testing. Yang dimana data inilah nantinya yang akan diolah atau diproses dengan data training.



Gambar 5. 8 Proses Training dan Testing

**Gambar 5.8** merupakan proses yang dilakukan di dalam aplikasi orang menggunakan algoritma Naïve Bayes



Gambar 5. 9 *Result*

**Gambar 5.9** merupakan *result* atau hasil dari proses yang telah dilakukan. Dimana hasil dapat dilihat dibagian sebelah kiri yaitu pada tulisan *Naïve Bayes*. Dari hasil yang telah didapatkan menggunakan aplikasi orange yaitu pada proses pertama proses data testing mendapatkan kesimpulan “NO”, pada proses kedua mendapatkan hasil “NO”, pada proses ketiga mendapatkan hasil “NO” dan pada proses keempat mendapatkan hasil “YES”.

# Conclusion

Maka dari hasil penelitian yang telah dilakukan menggunakan 2 metode perhitungan, yaitu menggunakan rumus algoritma *Naïve Bayes* yang dihitung secara manual dan menggunakan *tool* aplikasi Orange yang dimana dengan jumlah 17 dataset dan 20 % data testing dari dataset maka jumlah dari *data testing* berjumlah 4 maka hasil yang didapat adalah :

1. Dari perhitungan *data mining* menggunakan rumus algoritma *Naïve Bayes* yang dihitung secara manual maka didapat lah hasil proses pertama C1 > C2 [YES], pada proses kedua C1 < C2 [NO], pada proses ketiga C1 < C2 [NO], dan pada proses keempat C1 < C2 [NO].
2. Dari perhitungan data mining menggunakan tool orange hasil yang didapatkan pada proses pertama C1 < C2 [NO], pada proses kedua C1 < C2 [NO], pada proses ketiga C1 < C2 [NO], dan pada proses keempat C1 > C2 [YES].
3. Pada perhitungan yang telah dilakukan baik menggunakan rumus secara manual maupun menggunakan *tool* aplikasi Orange hasil kesimpulan yang didapat kan ialah prediksi *skill* yang dimiliki mahasiswa program studi sistem informasi di UIN Imam Bonjol Padang yang diambil 14 sampel data ialah memiliki kemampuan atau *skill* dalam bidang *coding*.
4. Dapat kita lihat bahwa ada perbedaan ketika kita menggunakan rumus algoritma *Naïve Bayes* yang dihitung secara manual dan menggunakan *tool* aplikasi O range pada proses pertama dan terakhir. Jadi untuk menghitung tingkat akurasi ketepatan menggunakan aplikasi orange ini dapat telah dihitung menggunakan confusion matriks yang dimana tingkat akurasi atau ketepatan nya dengan jumlah dataset atau *data training* 17 data dan *data testing* nya 4 data maka memiliki *confusion matriksnya* adalah 50 %.

References

[1] A. R. Jannah and D. Arifianto, “PENERAPAN METODE CLUSTERING DENGAN ALGORITMA K-MEANS UNTUK PREDIKSI KELULUSAN MAHASISWA JURUSAN TEKNIK INFORMATIKA DI,” 2017.

[2] R. Ningsih, “PERUBAHAN DUNIA KERJA The Identity Status Profiles of Late Adolescents in Yogyakarta (Indonesia) and Its Microsystem Context Influence View project”, doi: 10.13140/RG.2.2.11172.01922.

[3] H. A. R. Tilaar, *Perubahan Sosial dan Pendidikan*. PT. Gramedia Widiasarana Indonesia, 2002.

[4] J. O. Ong, “Implementasi Algotritma K-means clustering untuk menentukan strategi marketing president university,” *J. Ilm. Tek. Ind.*, vol. vol.12, no, no. juni, pp. 10–20, 2013.

[5] O. : Yuda and S. Nugroho, “DATA MINING MENGGUNAKAN ALGORITMA NAÏVE BAYES UNTUK KLASIFIKASI KELULUSAN MAHASISWA UNIVERSITAS DIAN NUSWANTORO.”

[6] Kusrini and E. T. Luthfi, *Algoritma Data Mining*. Yogyakarta: Penerbit Andi, 2009.

[7] R. Kumara and C. Supriyanto, “KLASIFIKASI DATA MINING UNTUK PENERIMAAN SELEKSI CALON PEGAWAI NEGERI SIPIL 2014 MENGGUNAKAN ALGORITMA DECISION TREE C4.5,” 2013.

[8] F. Nurhuda, S. W. Sihwi, and A. Doewes, “Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier,” *ITSmart J. Teknol. dan Inf.*, vol. 2, no. 2, pp. 35–42, 2013, doi: doi:10.20961/its.v2i2.630.

[9] F. Gorunescu, *Data Mining Concept, Models and Techniques*, 1st ed. Springer Berlin, Heidelberg, 2011. [Online]. Available: https://link.springer.com/book/10.1007/978-3-642-19721-5

[10] D. Putra and A. Wibowo, “Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes,” *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 2, pp. 84–92, 2020, doi: DOI:10.30645/SENARIS.V2I0.147.