

# Klasifikasi Data Guru Sekolah Menengah Pertama (SMP) PNS & NON PNS Kabupaten Karawang Menggunakan Algoritma *Naïve Bayes*

Ishal Saepulloh  
Universitas Buana Perjuangan  
Karawang, Indonesia  
if18.ishalsaepulloh@mhs.ubpkarawang.ac.id

Tatang Rohana  
Universitas Buana Perjuangan  
Karawang, Indonesia  
tatang.rohana@ubpkarawang.ac.id

Cici Emilia Sukmawati  
Universitas Buana Perjuangan  
Karawang, Indonesia  
cici.emilia@ubpkarawang.ac.id

**Abstract** —Dinas Pendidikan Pemuda dan Olahraga (Disdikpora) merupakan salah satu lembaga yang melaksanakan urusan pemerintahan daerah berdasarkan asas otonomi dan tugas pembantuan di bidang pendidikan. Salah satu tugasnya yaitu melakukan pengelolaan kesekretariatan meliputi perencanaan umum, kepegawaian, keuangan, evaluasi dan pelaporan. Permasalahan yang terjadi di Dinas Pendidikan Pemuda dan Olahraga Kabupaten Karawang yaitu banyaknya jumlah Data Guru SMP PNS dan NON PNS seluruh Sekolah Negeri maupun Swasta yang ada di Kabupaten Karawang harus didata secara akurat dikarenakan agar tidak terjadi kesalahan data untuk *recruitment* guru baru untuk memenuhi kurangnya guru yg ada di sekolah, dan adanya pengangkatan guru NON PNS menjadi PNS yang diatur oleh Pemerintahan Dinas Pendidikan Pemuda dan Olahraga. Berdasarkan permasalahan tersebut, maka peneliti akan menggunakan teknik data mining dalam klasifikasi data Guru SMP PNS dan NON PNS Kabupaten Karawang yang menggunakan algoritma *Naïve Bayes* untuk mengetahui tingkat akurasi yang dihasilkannya. Dalam klasifikasi Data Guru SMP PNS dan NON PNS menghasilkan *confusion matrix* yang akan menjadi penunjang dalam proses perhitungan tingkat keakurasian data. Dataset dibagi menjadi dua yakni data training 174 dan data testing 35. Pengujian ini menggunakan *microsoft excel* untuk perhitungan manual dan pemrograman python dengan model data 174 yaitu data training dan 35 data testing. Didalam pembagian data untuk data training memiliki nilai akurasi sebesar 100%, presisi sebesar 100%, recall sebesar 100% dengan itu data tersebut sangat baik digunakan dalam klasifikasi menggunakan algoritma *Naïve Bayes*.

**Kata kunci** — *Data Mining, Guru, klasifikasi, Naïve Bayes, Python*

## I. PENDAHULUAN

Dinas Pendidikan Pemuda dan Olahraga (Disdikpora) merupakan salah satu lembaga yang melaksanakan urusan pemerintahan daerah berdasarkan asas otonomi dan tugas pembantuan di bidang pendidikan[1]. Salah satu tugasnya yaitu melakukan pengelolaan kesekretariatan meliputi perencanaan umum, kepegawaian, keuangan, evaluasi dan pelaporan[2]. Permasalahan yang terjadi di Dinas Pendidikan Pemuda dan Olahraga Kabupaten Karawang yaitu banyaknya jumlah Data Guru SMP PNS dan NON PNS seluruh Sekolah SMP Negeri maupun Swasta yang ada di Kabupaten Karawang. Data Guru SMP PNS dan NON PNS yang ditangani oleh bagian *Departement* Program dan Pelaporan, karena berwenang mengetahui total kelengkapan guru seluruh sekolah SMP Negeri maupun Swasta di Kabupaten Karawang. Guru yang PNS dan NON PNS haruslah didata secara akurat dikarenakan, agar tidak terjadi kesalahan data untuk *recruitment* guru baru agar memenuhi kurangnya guru yang ada di setiap sekolah SMP Negeri dan Swasta, atau adanya pengangkatan guru NON PNS menjadi PNS yang diatur oleh Pemerintah Dinas Pendidikan Pemuda dan Olahraga terutama pada sekolah menengah pertama (SMP) yang ada di Kabupaten Karawang.

Salah satu faktor yang mendukung adalah pengolahan data yang akurat dengan klasifikasi menggunakan algoritma *Naïve Bayes* yang nantinya dapat memberikan sebuah informasi dan data dapat dibutuhkan oleh sebuah badan atau lembaga dalam mengolah sebuah perusahaan, maupun dari pihak internal dan eksternal yang membutuhkan[3]. Dalam fungsi dan kinerja perusahaan maupun organisasi tersebut dapat berjalan dengan baik, terutama pada profesi Guru PNS dan NON PNS saat ini masih banyak dibicarakan baik dikalangan pendidikan [4]. Berdasarkan permasalahan tersebut, maka peneliti akan menggunakan teknik data mining untuk klasifikasi data Guru SMP PNS dan NON PNS Kabupaten Karawang yang menggunakan algoritma *Naïve Bayes* untuk mengetahui tingkat akurasi yang dihasilkannya. Penelitian mengenai klasifikasi dengan menggunakan algoritma *Naïve Bayes* telah ada sebelumnya yakni terdapat pada jurnal tentang penerapan algoritma *Naïve Bayes* guna memutuskan tingkatan ketertiban siswa, yang menghasilkan tingkat akurasi yang sudah baik untuk diterapkan dalam penelitian ini[5].

## II. METODE PENELITIAN

### A. Studi Literatur

Peneliti melakukan studi literatur dengan mempelajari jurnal mengenai karya ilmiah yang bersangkutan dengan teknik Data Mining dengan konsep klasifikasi menggunakan algoritma *Naïve Bayes* yang bersumberkan melalui internet.

**B. Pengumpulan Data**

Dataset yang didapatkan merupakan berupa data Guru SMP PNS & NON PNS yang diperoleh dari Dinas Pendidikan Pemuda dan Olahraga. Data yang didapatkan berupa *softfile* dengan jumlah 174 data yang diambil dari seluruh sekolah SMP Negeri dan Swasta di Kabupaten Karawang pada tahun 2020/2021. Pada tahapan pengumpulan data ini yang didapatkan berupa NSS (Nomer Statistik Sekolah), NPSN (Nomor Pokok Sekolah Nasional), Nama Sekolah, Alamat Sekolah, Guru PNS golongan II, Guru PNS golongan III, Guru PNS golongan IV, Guru Tidak Tetap, Total, dan Keterangan.

**C. Seleksi Data**

Tahapan proses seleksi data merupakan pemilihan data atribut yang diperlukan peneliti yang akan di proses untuk klasifikasi menggunakan algoritma Naïve Bayes. Pada tahapan Seleksi data pada penelitian ini yang diperlukan berupa No, Nama Sekolah, Guru PNS golongan II, Guru PNS golongan III, Guru PNS golongan IV, Total, dan Keterangan untuk variabel lainnya tidak digunakan dalam proses klasifikasi.

**D. Perhitungan Manual**

Peneliti melakukan proses klasifikasi dengan menggunakan perhitungan manual dengan aplikasi microsoft excel 2010 dan mengimplementasikan pada algoritma *Naïve Bayes* dalam melakukan klasifikasi data Guru SMP PNS dan NON PNS dengan kategori sebagai berikut :

1. No.
2. Nama Sekolah.
3. Guru PNS golongan II
4. Guru PNS golongan III.
5. Guru PNS golongan IV.
6. Total.
7. Keterangan.

Alur pada tahapan peroses perhitungan manual menggunakan *Microsoft Excel* sebagai berikut :

1. Baca data training.
2. Cari nilai mean dan standar deviasi dari masing-masing parameter yang merupakan data numerik.
3. Cari nilai probabilistik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
4. Mendapatkan nilai dalam tabel mean, standard deviasi dan probabilitas.
5. Solusi kemudian dihasilkan.

**E. Perhitungan Python**

Tahapan yang telah dilakukan melalui perhitungan manual menggunakan *Microsoft Excel* peneliti akan melakukan perbandingan hasil akurasi, persisi, dan recall dengan perhitungan menggunakan pemograman *python* dengan algoritma *Naïve Bayes*, sehingga dapat diamati perbedaanya.

**III. HASIL DAN PEMBAHASAN**

**A. Hasil Studi Literatur**

Hasil dari studi literatur peneliti mudah memahami alur tentang teknik data mining menggunakan metode klasifikasi dengan algoritma *Naïve Bayes* untuk diterapkan pada penelitian.

**B. Hasil Pengumpulan Data**

Data di dapatkan dari Dinas Pendidikan Pemuda dan Olahraga Kabupaten Karawang yaitu Data Guru SMP PNS dan NON PNS pada tahun 2020/2021 yang terdiri dari SMP Negeri, dan SMP Swasta total 174 data.

Tabel 1. Data Guru SMP PNS & NON PNS

No	NSS	NPS	Nama Sekolah	Alamat Sekolah	Guru PNS golongan II	Guru PNS golongan III
----	-----	-----	--------------	----------------	----------------------	-----------------------

1	20102210 0001	20217932	SMPN 1 BANYUSARI	BANYUSARI	1	13
2	20102212 4001	20255752	SMPN 2 BANYUSARI	BANYUSARI	0	5
3	20102210 4008	20217914	SMPN 1 BATUJAYA	BATUJAYA	0	1
4	20102210 8002	20265262	SMPN SATU ATAP 1 BATUJAYA	BATUJAYA	0	0
5	20102210 8003	20265265	SMPN SATU ATAP 2 BATUJAYA	BATUJAYA	1	6
.....	.....	.....	.....	.....	.....	.....
174		70004086	SMP IT IBNU LUTFI SYARIF	PAKISJAYA	1	3

Guru PNS golongan IV	Guru Tidak Tetap	Total	Keterangan
17	19	>20	Lengkap
0	8	<20	Tidak Lengkap
10	22	>20	Lengkap
0	9	<20	Tidak Lengkap
.....	.....	.....	.....
0	6	<20	Tidak Lengkap

C. Seleksi Data

Tahapan ini merupakan proses penyeleksian atribut yang akan digunakan. terdapat 8 variabel yang akan digunakan seperti No, Nama Sekolah, Guru PNS golongan II, Guru PNS golongan III, Guru PNS golongan IV, Guru Tidak Tetap, Total, dan Keterangan. Sehingga didapatkan sebuah dataset yang akan digunakan pada proses klasifikasi menggunakan algoritma *Naïve Bayes*.

Tabel 2. Dataset

No	Nama Sekolah	Guru PNS golongan II	Guru PNS golongan III	Guru PNS golongan IV	Guru Tidak Tetap	Total	Keterangan
1	SMPN 1 BANYUSARI	1	13	17	19	>20	Lengkap
2	SMPN 2 BANYUSARI	0	5	0	8	<20	Tidak Lengkap
3	SMPN 1 BATUJAYA	0	1	10	22	>20	Lengkap
4	SMPN SATU ATAP 1 BATUJAYA	0	0	0	9	<20	Tidak Lengkap
.....	.....	.....	.....	.....	.....	.....	.....
174	SMP IT IBNU LUTFI	1	3	0	6	< 20	Tidak Lengkap

SYARIF

D. Perhitungan Manual

Pada perhitungan manual ini menggunakan *microsoft excel* untuk menunjukan tingkat nilai akurasi, presisi, dan *recall* dengan menggunakan data testing sebanyak 35 yang di ambil dari data training.

Tabel 3. Perhitungan Confusion Matrix

Aktual	Lengkap	Tidak Lengkap
Lengkap	18	0
Tidak Lengkap	0	0

setelah itu mencari nilai *True Positive*, *True Negative*, *False Positive*, dan *False Negative*. Lalu menghitung tingkat akurasi, presisi, dan *recall*. Berikut merupakan hasil nilai akurasi, presisi, dan *recall* :

1) Menghitung nilai akurasi.

$$\text{Akurasi} = \frac{18+0}{18+0+0+0} = 100\%$$

2) Menghitung nilai presisi.

$$\text{Presisi} = \frac{18}{18+0} = 100\%$$

3) Menghitung nilai *recall*

$$\text{Recall} = \frac{18}{18+0} = 100\%$$

Pada tahapan perhitungan manual dengan menggunakan *microsoft excel* pada klasifikasi data Guru SMP PNS dan NON PNS mendapatkan nilai akurasi 100% , nilai presisi 100%, dan nilai *recall* 100%. Yang di mana perhitungan tersebut mendapatkan nilai yang sangat baik.

E. Program Python

Hasil prediksi yang di buat dari 35 data testing dengan kelas keterangan 1 “Lengkap”, dan 0 “Tidak Lengkap menggunakan program *python* mendapatkan nilai 1.0 yang berarti mendapatkan nilai tingkat akurasi sebesar 100%. Berikut ini adalah gambaran hasil data testing dibuat menggunakan perhitungan program *python* :

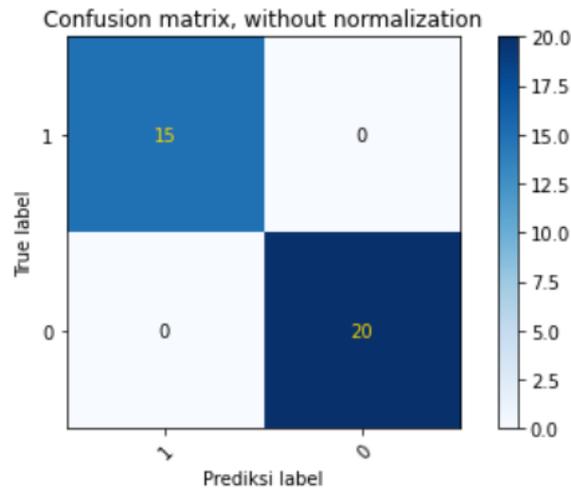
```
In [11]: #Prediksi terhadap model training yang sudah dibuat
prediction = NaiveBayes.predict(X_test)
accuracies = accuracy_score(y_test, prediction)

print(prediction)
print(accuracies)

[1 0 1 1 1 0 1 1 0 0 0 0 1 0 1 0 0 0 0 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 0]
1.0
```

Gambar 2. Hasil Data Testing Pada Perhitungan Python

Hasil perhitungan confusion matrix dari 35 data testing menggunakan program *python* mendapatkan nilai *True positive* 15, *True Negative* 0, *False Positive* 20, dan *Fales Negative* 0. Berikut ini gambaran hasil confusion matrix menggunakan perhitungan program *python* :



Gambar 3. Hasil Perhitungan Confusion Matrix

F. Hasil

Hasil dari perhitungan manual menggunakan *microsoft excel* dengan perhitungan menggunakan pemrograman *python* menunjukan hasil yang berbeda. Hasil perhitungan manual mendapatkan *True positive* 18, *True Negative* 0, *False Positive* 0, dan *False Negative* 0 kemudian hasil perhitungan pemrograman *python* mendapatkan *True positive* 15, *True Negative* 0, *False Positive* 20, dan *Fales Negative* 0. Tingkat akurasi, presisi, dan *recall* memiliki nilai yang sama dari perhitungan manual dan perhitungan pemrograman *python* mendapatkan hasil 100% dari 35 data testing. Maka dari itu Perhitungan manual dan perhitungan pemrograman *python* mendapatkan hasil yang sangat baik dalam klasifikasi algoritma *Naïve Bayes* untuk Data Guru SMP PNS & NON PNS Kabupaten Karawang. Berikut hasil penelitian yang dapat disajikan :

Tabel 4. Hasil Penelitian Confuison Matrix

Confusion Matrix	Manual	Python
<i>True Positive</i>	18	15
<i>True Negative</i>	0	0
<i>False Positive</i>	0	20
<i>False Negative</i>	0	0

Tabel 5. Hasil Perhitungan Manual dan Perhitungan Python

Nilai	Perhitungan Manual	Perhitungan Python
Akurasi	100%	100%
Presisi	100%	100%
<i>Recall</i>	100%	100%

IV. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilaksanakan dapat disimpulkan bahwa pengujian data taraining 174 dan data testing 35 dalam klasifikasi data guru SMP PNS & NON PNS Kabupaten Karawang yang telah dilakukan dengan perhitungan manual menggunakan *microsoft excel* dan perhitungan pemrograman *python* memiliki hasil nilai yang sama yaitu tingkat akurasi sebesar 100%, presis 100%, dan *recall* 100%.

Saran untuk penelitian selanjutnya diharapkan dapat menggunakan algoritma klasifikasi lainnya, agar dapat diketahui perbedaan perhitungan antara algoritma *Naïve Bayes* dan algoritma klasifikasi lainnya dan sekaligus sebagai tambahan referensi bagi penelitian selanjutnya dalam meneliti permasalahan yang sama.

PENGAKUAN

Naskah ilmiah ini adalah sebagian dari penelitian Tugas Akhir milik Ishal Saepulloh, dengan judul Implementasi Algoritma *Naïve Bayes* Untuk Klasifikasi Data Guru SMP PNS & NON PNS Kabupaten Karawang yang dibimbing oleh Tatang Rohana dan Cici Emilia Sukmawati.

DAFTAR PUSTAKA

- [1] Rezki, D. A. (2020). Analisa Kinerja Guru Sekolah di Dinas Pendidikan Dengan Menggunakan Metode Rought Set ( Studi Kasus : Dinas Pendidikan Provinsi Sumatera Utara ). *J. Comput. Systt. and Informatics*, 1(3), 172–177.
- [2] Sidik Rahmatullah, Supriyanto, Rustam, Merri Parida, I. P. (2021). Algoritma Naive Bayes Untuk Menentukan Tingkat Kedisiplinan Siswa. *Informasi Dan Komputer*, 9(1), 32–44.
- [3] Suhendro, D. (2017). Perancangan dan Implementasi Realisasi Anggaran Pendapatan ( Studi Kasus : Pengadilan Negeri Klas IB Pematangsiantar ). *Seminar Nasional Teknologi Informatika*, 30–36.
- [4] Haryati, S., Sudarsono, A., & Suryana, E. (2015). Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama*, 11(2), 130–138.
- [5] Meiza, C. (2016). Perbedaan Kebahagiaan Pada Guru Berstatus Pns Dan Honorer. *Jurnal Ilmiah Psikologi Gunadarma*, 9(2), 97461. <https://doi.org/10.35760/psi>