Implementasi Algoritma *Naive Bayes* Untuk Klasifikasi Menentukan Prestasi Siswa Berdasarkan Nilai Rata-Rata

Muh Nurhariza
Universitas Buana Perjuangan
Karawang, Indonesia
if18.muhnurhariza@mhs.ubpkarawang.ac.id

Ayu Ratna Juwita
Universitas Buana Perjuangan
Karawang, Indonesia
ayurj@ubpkarawang.ac.id

Dwi Sulistya Kusumaningrum
Universitas Buana Perjuangan
Karawang, Indonesia
dwi.sulistya@ubpkarawang.ac.id

***Abstract*—** Penentukan prestasi siswa umumnya menjadi permasalahan disetiap sekolah salah satunya pada SMK Tri Mitra Karawang. Karena di setiap sekolah terdapat siswa yang rajin belajar dan kurang rajin sehingga siswa memiliki prestasi yang berbeda beda. Pada penelitian ini ada beberapa tahapan yang digunakan, Tahap yang pertama adalah Analisis dimana pada tahap ini tujuannya untuk mendapatkan informasi dan data. Tahap yang kedua adalah Implementasi yang dimana tahap ini melakukan proses perhitungan dan tahap yang ketiga adalah Evaluasi dimana tahap ini untuk mengetahui hasil dari perhitungan. Pada penelitian ini peneliti menggunakan algoritma *Naïve Bayes* untuk mengklasifikasi menentukan prestasi siswa dengan menggunakan 400 data siswa atau data set yang dijadikan sebagai data latih dan 100 data siswa yang dijadikan sebagai data testing yang diperoleh dari data set. Hasil dari algoritma *Naïve Bayes* menggunakan *Confusion Matrix* mendapatkan hasil akurasi sebesar 100%, presisi 100% dan *recall* 100%.

***Kata kunci* —** Sekolah, Prestasi, *Naïve Bayes, Python*

I. PENDAHULUAN

Pendidikan merupakan salah satu upaya yang dilakukan untuk menentukan hasil akhir suatu negara. Instruksi yang bermanfaat akan terus memberikan usia yang hebat. Di samping kesulitan dunia yang cepat, kesulitan juga terdapat pada dunia pendidikan yang semakin luas, langkah ini sangat mendorong siswa untuk mencapai prestasi terbaik. Prestasi belajar ialah didapat oleh siswa dari belajar, khususnya mengingat kenyataan dan menyampaikan pengetahuannya baik dengan lisan ataupun dalam bentuk *hard copy*, apalagi dalam keadaan ujian. Dengan demikian, secara umum prestasi belajar didapat dari latihan-latihan disekolah yang bersifat mental maupun tidak baku dengan penilaian dan evaluasi. Prestasi belajar siswa dapat terlihat dengan nilai yang diberikan oleh guru dari banyaknya mata pelajaran yang telah dikonsentrasikan oleh siswa. Setiap pembelajaran, tentu saja, secara konsisten berharap untuk menciptakan pembelajaran yang paling ekstrem. Selama pencapaian waktu yang dihabiskan, prestasi belajar siswa sangat dipengaruhi elemen yang berbeda. Salah satu faktor utama yang sangat mempengaruhi hasil belajar adalah kehadiran pendidik. Mengingat kehadiran pendidik selama waktu mengajar dan latihan pembelajaran sangat menarik, maka, pada saat itu, sifat instruktur harus dipikirkan.

Permasalahan yang sering terjadi khususnya di Sekolah Profesi Tri Mitra Karawang, adalah banyak siswa yang malas untuk mengkaji, siswa yang malas belajar, tidak menonjolkan apa yang disadari, tugas sekolah yang dibebani, hasil pembelajaran. Siswa hanya pergi ke kelas atau lulus dari sekolah. tak henti-hentinya sekolah cukup bertemu sahabat dan membawa uang jajan. Itu saja gambaran latihan belajar siswa rendah. Oleh karena itu, mereka mengalami kesulitan belajar. Begitu banyak mahasiswa yang memiliki berbagai prestasi. Di setiap sekolah, khususnya di Sekolah Profesi Tri Mitra Karawang, terdapat siswa yang berprestasi atau berprestasi dan tidak berprestasi. Dengan tujuan agar pesta sekolah hendaknya mengharapkan dalam menggarap sifat pelatihan dalam mengelola siswa yang kurang berprestasi sehingga dapat menciptakan usia yang layak dan bekerja pada sifat sekolah. Oleh karena itu, untuk mengantisipasi prestasi siswa yang berbeda, pelatihan Sekolah Profesi Tri Mitra Karawang memerlukan informasi yang tepat tentang prestasi siswa dengan tujuan agar dapat menjadi acuan bagi daerah pengajaran untuk lebih mudah mengetahui prestasi siswa yang berhasil dan membutuhkan prestasi. . Maka dari permasalahan tersebut pencipta bermaksud untuk mengatasi permasalahan tersebut dengan memanfaatkan strategi penambangan informasi.

Naïve Bayes adalah salah satu algoritma klasifikasi yang sangat efektif dan efisien. Algoritma Naïve Bayes Classifier berguna untuk mengklasifikasikan data terhadap atribut tertentu. Dalam laporan sebelumnya mengenai perkiraan tingkat kelulusan mahasiswa, dikatakan bahwa ketepatan hasil review adalah 80,7% dari informasi persiapan yang menambahkan hingga 1162 informasi dan informasi pengujian menambahkan hingga 587 informasi. Dalam penelitian ini ia menggunakan perhitungan Credulous Bayes. Dalam bidang pendidikan, perhitungan semacam ini dapat digunakan untuk mengantisipasi tingkat prestasi siswa atau siswi sebagaimana penelitian sebelumnya yang dipimpin oleh Heri Susanto (2014) mengatakan bahwa prestasi siswa tergantung pada status keuangan wali, inspirasi, disiplin siswa dan prestasi. masa lalu. Variabel inspirasi adalah variabel yang menentukan kemampuan seorang siswa untuk berhasil atau tidak dalam prestasi belajarnya di kemudian hari.[1].

Berdasarkan latar belakang dalam penelitian ini Algoritma Naive Bayes sangat efektif digunakan dibandingkan dengan Algoritma data mining yang lainnya. Penulis berharap dengan menggunakan Algoritma Naive Bayes dalam Klasifikasi data sekolah dapat membantu memecahkan masalah dalam menentukan prestasi siswa di SMK Tri Mitra Karawang.

II. TINJAUAN PUSTAKA

A. Prestasi Belajar

Prestasi belajar adalah hasil belajar yang dicapai setelah melalui cara mendidik dan latihan belajar yang paling umum. Prestasi belajar dapat terlihat melalui nilai yang diberikan oleh seorang pengajar dari banyaknya bidang studi yang telah dikonsentrasikan oleh siswa. Setiap gerakan belajar tentunya secara konsisten berharap dapat memberikan pembelajaran yang terbaik. Selama pencapaian waktu yang dihabiskan, prestasi belajar secara tegas dipengaruhi oleh variabel yang berbeda. Salah satu faktor utama yang sangat persuasif dalam kemajuan pembelajaran adalah kehadiran tenaga pendidik. Mengingat kehadiran pendidik selama waktu mengajar dan latihan pembelajaran sangat persuasif, maka sifat instruktur harus dipikirkan[2].

B. Klasifikasi

Klasifikasi adalah prosedur penanganan informasi yang mempartisi objek menjadi beberapa kelas yang ditunjukkan oleh jumlah kelas yang ideal. Urutan adalah metode menemukan contoh yang dapat mengisolasi kelas informasi satu sama lain untuk memutuskan objek yang termasuk dalam klasifikasi tertentu dengan melihat cara berperilaku dan karakteristik kumpulan yang telah dicirikan. Prosedur ini dapat menjelaskan informasi baru dengan memanfaatkan hasil untuk memberikan berbagai prinsip. Saat memeriksa ketertiban, kami membahas kategorisasi ilmiah. Kategorisasi ilmiah (Tassein = tatanan + nomos = ilmu, peraturan) muncul pertama kali sebagai studi tentang pengelompokan makhluk hidup, tetapi kemudian dibuat sebagai studi tentang karakterisasi sebagai aturan, termasuk di sini standar pengelompokan. Selanjutnya, urutan (klasifikasi ilmiah) adalah cara yang paling umum untuk menempatkan item tertentu dalam sekelompok klasifikasi, mengingat setiap pengaturan artikel. Ada beberapa perhitungan yang dapat digunakan untuk menangani masalah urutan informasi, a.l. Pohon Pilihan, KNN (K-Closest Neighbor), ANN (Organisasi Otak Palsu), Guileless Bayes, SVM (Backing Vector Machine). Pada review ini, perhitungan yang digunakan untuk mengkarakterisasi adalah Choice Tree dengan deret C4.5[3].

C. Algoritma *Naive Bayes*

Algoritma Naïve Bayes Classifier merupakan salah satu metode yang bisa menyelesaikan prediksi pada masa yang akan datang dan secara langsung. Manfaat dari perhitungan ini adalah bahwa ia memiliki kapasitas yang sangat baik untuk mengukur informasi yang memiliki contoh pola dan harapan untuk bulan berikutnya. Dengan demikian, perhitungan Algoritma Naïve Bayes merupakan perhitungan yang masuk akal dalam menilai jumlah penawaran di bulan berikutnya untuk jenis nutrisi di toko obat yang penting bagi klien.[4]

Rumus *Naive Bayes* :

$$P(X|H) = P(H|X)/P(X) \quad (1)$$

X : Data dengan class yang belum diketahui

H : Hipotesis data X, merupakan suatu class spesifik

$P(H)|X)$: Probabilitas hipotesis H berdasar kondisi (posteriori probability)

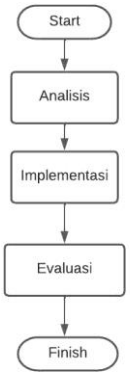
$P(H)$: Probabilitas hipotesis H (prior probability)

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

$P(X)$: Probabilitas dari X

D. Metode Penelitian

Penelitian ini dilaksanakan di SMK Tri Mitra Karawang pada bulan Desember 2021 sampai Juni 2022. Pada penelitian ini data yang digunakan adalah data siswa kelas XII tahun ajaran 2015-2017 sebanyak 400 data siswa. Dalam penelitian ini ada beberapa tahapan yang akan dilakukan seperti gambar berikut :



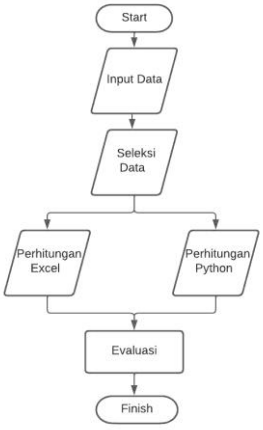
Gambar 1 Prosedur Penelitian

- Analisis

Pada tahap ini, penyelidikan masalah selesai, persyaratan untuk perbaikan dilakukan. Tahap investigasi berisi pemeriksaan kebutuhan. Pemeriksaan kebutuhan harus dimungkinkan dengan membedah apa yang dibutuhkan. Pada tahap ini akan dipahami apa yang harus diciptakan untuk menentukan prestasi siswa. Untuk tahap selanjutnya akan dilakukan persepsi untuk mendapatkan apa yang dibutuhkan oleh analis.

1. Observasi
   Observasi ini dilakukan peneliti dengan pihak sekolah terutama pada Kurikulum. Dilakukannya *Oberservasi* pada Kurikulum untuk mendapatkan apa yang dibutuhkan oleh peneliti yaitu data siswa. Pada tahap ini peneliti membutuhkan sebuah data siswa dari pihak sekolah untuk bahan uji pada penelitian ini.

- Implementasi



Gambar 2 Alur Sistem

1. Input Data
   Dalam tahap ini, peneliti menginput data siswa yang didapat dari pihak sekolah dan mendapatkan data nilai rapot siswa kelas XII SMK Tri Mitra Karawang tahun ajaran 2015-2017 yang dimana terdapat 402 data siswa untuk dijadikan bahan pada penelitian ini.
2. Seleksi Data
   Tahapan seleksi data merupakan proses untuk memilih data dari atribut yang dibutuhkan dan yang tidak dibutuhkan dalam penelitian ini.
3. Perhitungan Excel dan Perhitungan *Python*
   Pada tahapan ini, peneliti menggunakan aplikasi *Mircrosoft Excel* dan *Python* untuk menerapkan algoritma *Naïve Bayes*. Disini data akan dibagi menjadi 2, yaitu berprestasi dan tidak berprestasi. Setelah perhitungan menggunakan *Microsoft Excel* selesai selanjutnya akan melakukan perhitungan menggunakan *Python*, selanjutnya peneliti akan membandingkan antara hasil perhitungan menggunakan *Microsoft Excel* dengan hasil perhitungan menggunakan *Python* yang kemudian akan didapatkan hasil perbandingannya.
4. Evaluasi
   Setelah dilakukan implementasi *Naïve Bayes*, maka akan dihasilkan pengujian dan evaluasi *Confusion Matrix* untuk mengetahui performa akurasi, presisi dan *recall*. *Confusion matrix* adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah [5]. Rumusnya sebagai berikut :

   Akurasi $\text{Akurasi} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$

   Presisi $\text{Presisi} = \frac{\text{TP}}{\text{TP+FP}}$

   Recall $Recall = \frac{\text{TP}}{\text{TP+FN}}$

   - TP adalah *True* Positif, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.
   - TN adalah *True* Negatif, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
   - FP adalah *False* Positif, yaitu jumlah data positif namun terklasifikasi salah oleh sistem.
   - FN adalah *False* Negatif, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.

III. HASIL DAN PEMBAHASAN

1. Input Data

Data yang digunakan dalam penelitian ini yaitu dari data siswa kelas XII tahun ajaran 2015-2017 sebanyak 402 data siswa. Berikut adalah data dari nilai siswa yang diperoleh dari SMK Tri Mitra Karawang.

Tabel 1 Data Nilai Siswa

No	**Nama**	**No Induk**	**Nilai Rapot Semester 1**	**Nilai Rapot Semester 2**	**…**	**…**	**Nilai Rapot Semester 6**
1	Hilda Nathaniela	1520001501	80	90	…	…	77
2	Maulana Alif Anugrah	1520001502	87	86	…	…	88
3	Ckasinta Winda S	1520001503	79	86	…	…	89
4	Nadya Saphira Esfandiari	1520001504	89	78	…	…	79
5	Feggy Rizkiana Herman	1520001505	76	88	…	…	76
6	Peter Sulaeman	1520001506	85	81	…	…	76
7	Armelia nur asyiffa	1520001507	84	86	…	…	88
8	Arya Mahardika	1520001508	77	77	…	…	88
9	ahmad fauzan naufal	1520001509	79	89	…	…	75
10	Renaya Sarasti	1520001510	82	77	…	…	85
11	Lucky Wiratama Suganda	1520001511	81	81	…	…	76
12	Gracia vini	1520001512	76	77	…	…	84
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
402	Rizkiyah Nurbaity	1720001832	85	85	…	…	81

2. Seleksi Data

Tahap ini merupakan proses penyeleksian atribut, terdapat 6 variabel yang digunakan untuk parameter mendapatkan hasil siswa berprestasi dan tidak berprestasi, variabelnya adalah Nilai Rapot Semester 1, Nilai Rapot Semester 2, Nilai Rapot Semester 3, Nilai Rapot Semester 4, Nilai Rapot Semester 5 dan Nilai Rapot Semester 6. Berikut kriterianya :

Tabel 2 Seleksi Variabel Data

Variabel	Digunakan	Keterangan
Nama	√	Digunakan
Nilai Rapot Semester 1	√	Digunakan
Nilai Rapot Semester 2	√	Digunakan
Nilai Rapot Semester 3	√	Digunakan
Nilai Rapot Semester 4	√	Digunakan
Nilai Rapot Semester 5	√	Digunakan
Nilai Rapot Semester 6	√	Digunakan

Sehingga didapatkan sebuah data yang digunakan sebagai data set yang akan digunakan untuk melakukan perhitungan dalam penelitian ini. Data set dapat dilihat pada tabel berikut :

Tabel 3 Data Set

Nilai Rapot Smt 1	Nilai Rapot Smt 2	Nilai Rapot Smt 3	Nilai Rapot Smt 4	Nilai Rapot Smt 5	Nilai Rapot Smt 6	Status
80	90	89	90	76	77	Berprestasi
87	86	87	78	82	88	Berprestasi
79	86	82	90	81	89	Berprestasi
						Tidak
89	78	75	80	82	79	Berprestasi
						Tidak
76	88	85	78	79	76	Berprestasi
						Tidak
85	81	79	77	85	76	Berprestasi
84	86	76	86	87	88	Berprestasi
77	77	90	82	85	88	Berprestasi
79	89	84	86	90	75	Berprestasi
82	77	85	89	83	85	Berprestasi
						Tidak
81	81	82	88	78	76	Berprestasi
						Tidak
76	77	79	77	89	84	Berprestasi
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
85	85	79	86	87	81	Berprestasi

3. Perhitungan Manual

Hasil dari perhitungan manual menggunakan algoritma *Naïve Bayes* menggunakan *Microsoft Excel* menunjukan akurasi sebesar 100%. Data *training* sebanyak 402 siswa dan data *testing* sebanyak 100 data diambil data training. Langkah-langkah untuk perhitungan manual dengan menggunakan algoritma *Naive Bayes* adalah sebagai berikut :

1. Siapkan data set yang akan digunakan dalam perhitungan manual ini, Data set akan dijadikan sebagai data latih pada tabel dibawah :

Tabel 4 Data Latih

Nilai Rapot Smt 1	Nilai Rapot Smt 2	Nilai Rapot Smt 3	Nilai Rapot Smt 4	Nilai Rapot Smt 5	Nilai Rapot Smt 6	Status
80	90	89	90	76	77	Berprestasi
87	86	87	78	82	88	Berprestasi
79	86	82	90	81	89	Berprestasi
						Tidak
89	78	75	80	82	79	Berprestasi
76	88	85	78	79	76	Tidak

						Berprestasi
						Tidak
85	81	79	77	85	76	Berprestasi
84	86	76	86	87	88	Berprestasi
77	77	90	82	85	88	Berprestasi
79	89	84	86	90	75	Berprestasi
82	77	85	89	83	85	Berprestasi
						Tidak
81	81	82	88	78	76	Berprestasi
						Tidak
76	77	79	77	89	84	Berprestasi
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
85	85	79	86	87	81	Berprestasi

2. Menghitung probabilitas kelas dan atribut dengan kelas sebagai berikut :
   a) Menghitung probabilitas nilai atribut label

Tabel 5 Probabilitas Kelas

Status	**Nilai**
Berprestasi	0.6069652
Tidak Berprestasi	0.3930348
Hasil Probabilitas Kelas	100.00%

   b) Menghitung probabilitas atribut kelas yang diketahui

Tabel 6 Probabilitas Atribut Nilai Rapot Semester 1 Terhadap Kelas

Nilai Rapot Smt 1	**Berprestasi**	**Tidak Berprestasi**
75	0.032786885	0.082278481
76	0.008196721	0.139240506
77	0.036885246	0.113924051
78	0.036885246	0.107594937
79	0.06147541	0.088607595
80	0.073770492	0.101265823
81	0.06557377	0.075949367
82	0.077868852	0.101265823
83	0.073770492	0.044303797
84	0.073770492	0.037974684
85	0.057377049	0.044303797
86	0.053278689	0.012658228
87	0.102459016	0.012658228
88	0.081967213	0.006329114
89	0.086065574	0.025316456
90	0.077868852	0.006329114

Tabel 7 Probabilitas Atribut Nilai Rapot Semester 2 Terhadap Kelas

Nilai Rapot Smt 2	Berprestasi	Tidak Berprestasi
75	0	0.113924051
76	0.016393443	0.120253165
77	0.024590164	0.088607595
78	0.049180328	0.088607595
79	0.028688525	0.075949367
80	0.06147541	0.088607595
81	0.040983607	0.101265823
82	0.098360656	0.050632911
83	0.081967213	0.063291139
84	0.102459016	0.044303797
85	0.090163934	0.025316456
86	0.094262295	0.025316456
87	0.06147541	0.03164557
88	0.081967213	0.025316456
89	0.086065574	0.03164557
90	0.081967213	0.025316456

Tabel 8 Probabilitas Atribut Nilai Rapot Semester 3 Terhadap Kelas

Nilai Rapot Smt 3	Berprestasi	Tidak Berprestasi
75	0.045081967	0.139240506
76	0.049180328	0.056962025
77	0.028688525	0.120253165
78	0.040983607	0.082278481
79	0.040983607	0.094936709
80	0.045081967	0.094936709
81	0.057377049	0.03164557
82	0.053278689	0.126582278
83	0.045081967	0.037974684
84	0.098360656	0.037974684
85	0.077868852	0.044303797
86	0.090163934	0.050632911
87	0.06147541	0.044303797
88	0.102459016	0.006329114
89	0.077868852	0.03164557
90	0.086065574	0

Tabel 9 Probabilitas Atribut Nilai Rapot Semester 4 Terhadap Kelas

Nilai Rapot Smt 4	Berprestasi	Tidak Berprestasi
75	0.012295082	0.132911392
76	0.024590164	0.082278481
77	0.024590164	0.151898734
78	0.028688525	0.063291139
79	0.053278689	0.094936709
80	0.06147541	0.063291139
81	0.077868852	0.113924051
82	0.086065574	0.050632911
83	0.032786885	0.050632911
84	0.090163934	0.050632911
85	0.057377049	0.056962025
86	0.086065574	0.025316456
87	0.081967213	0.03164557
88	0.073770492	0.025316456
89	0.110655738	0
90	0.098360656	0.006329114

Tabel 10 Probabilitas Atribut Nilai Rapot Semester 5 Terhadap Kelas

Nilai Rapot Smt 5	Berprestasi	Tidak Berprestasi
75	0.020491803	0.063291139
76	0.028688525	0.113924051
77	0.024590164	0.069620253
78	0.028688525	0.139240506
79	0.045081967	0.088607595
80	0.06557377	0.132911392
81	0.040983607	0.075949367
82	0.081967213	0.063291139
83	0.057377049	0.044303797
84	0.106557377	0.037974684
85	0.081967213	0.025316456
86	0.086065574	0.050632911
87	0.094262295	0.03164557
88	0.073770492	0
89	0.069672131	0.037974684
90	0.094262295	0.025316456

Tabel 11 Probabilitas Atribut Nilai Rapot Semester 6 Terhadap Kelas

Nilai Rapot Smt 6	Berprestasi	Tidak Berprestasi
75	0.020491803	0.075949367
76	0.032786885	0.139240506
77	0.016393443	0.075949367
78	0.049180328	0.113924051
79	0.057377049	0.050632911
80	0.040983607	0.107594937
81	0.057377049	0.101265823
82	0.06147541	0.037974684
83	0.040983607	0.088607595
84	0.073770492	0.03164557
85	0.073770492	0.101265823
86	0.057377049	0.018987342
87	0.12295082	0.037974684
88	0.118852459	0
89	0.106557377	0.006329114
90	0.069672131	0.012658228

3. Mengkalikan semua atribut pada data testing dengan probabilitas dari semua hasil probabilitas atribu dan kelas. Data testing disini digunakan sebanyak 100 data dapat dilihat pada tabel dibawah :

Tabel 12 Data Testing

Nama	Nilai Rapot Smt 1	Nilai Rapot Smt 2	Nilai Rapot Smt 3	Nilai Rapot Smt 4	Nilai Rapot Smt 5	Nilai Rapot Smt 6	Status
Hilda Nathaniela	80	90	89	90	76	77	-
Maulana Alif Anugrah	87	86	87	78	82	88	-
Ckasinta Winda S	79	86	82	90	81	89	-
Nadya Saphira Esfandiari	89	78	75	80	82	79	-
Feggy Rizkiana Herman	76	88	85	78	79	76	-
Peter Sulaeman	85	81	79	77	85	76	-
Armelia Nur Asyiffa	84	86	76	86	87	88	-
Arya Mahardika	77	77	90	82	85	88	-
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
Dwi Laksana Bhakti	77	78	77	85	89	80	-
Deri Fauzi	83	85	84	82	85	82	-

Setelah mengkalikan semua atribut pada data testing diatas dengan semua hasil probabilitas atribut dan kelas, Dari 100 data diperoleh akurasi sebesar 98%, tabel dibawah menunjukan hasil prediksi sebagai berikut :

Tabel 13 Hasil Data Klasifikasi

Prediksi	**Berprestasi**	**Tidak Berprestasi**
Berprestasi	1.322	1.746
Berprestasi	1.007	0
Berprestasi	8.049	3.395
Berprestasi	3.348	2.489
Tidak Berprestasi	1.346	4.793
Tidak Berprestasi	3.865	8.963
Berprestasi	2.001	0
Berprestasi	3.972	0
…	…	…
…	…	…
Tidak Berprestasi	5.175	1.110
Berprestasi	1.722	8.148

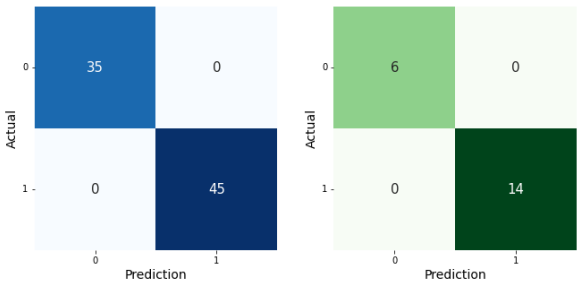
4. Perhitungan *Python*

```
: pipeline.predict(X_pred)

: array([1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1,
       1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 1,
       1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 1,
       1, 1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1,
       0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1], dtype=int64)
```

Gambar 3 Hasil Klasifikasi Data Testing

Hasil prediksi dari data testing yang telah dihitung sebanyak 100 data siswa dengan keterangan 0 untuk "Tidak Berprestasi" dan 1 untuk "Berprestasi". Dengan akurasi data testing terdiri dari 59 *True Positive* yaitu Berprestasi dan 41 *False Negative* yaitu Tidak Berprestasi dengan akurasi 100% pada data testing.



Gambar 4 Visual Data Testing

5. Evaluasi

Setelah melakukan perhitungan data menggunakan manual tahap selanjutnya adalah melakukan evaluasi menggunakan *Confusion Matrix*. Dalam tahap ini akan mencari nilai akurasi, presisi dan *recall*. Berdasarkan hasil dari data testing pada tabel berikut :

Tabel 14 Confusion Matrix

	Actual	
Prediction	Berprestasi	Tidak Berprestasi
Berprestasi	59	0

Tidak Berprestasi	0	41

1. Menghitung nilai akurasi
   Akurasi = $\frac{59+41}{59+41+0+0}$ = 100%

2. Menghitung nilai presisi
   Presisi = $\frac{59}{59+0}$ = 100%

3. Menghitung nilai *recall*
   Recall = $\frac{59}{59+0}$ = 100%

6. Hasil Pembahasan
   Berdasarkan perhitungan manual dan perhitungan python menggunakan algoritma *Naïve Bayes* memiliki akurasi yang baik. Maka penggunaan algoritma *Naïve Bayes* dapat digunakan untuk menentukan prestasi siswa. Hasil dari perhitungan manual dan perhitungan *python* mendapatkan hasil sebanyak 59 data *True Positive* dan 41 data *False Negative*. Berikut hasil perhitungan dari manual dan python :

Tabel 15 Hasil Perhitungan Manual dan *Python*

Perhitungan	
Manual	*Python*
Berprestasi	Berprestasi
Berprestasi	Berprestasi
Berprestasi	Berprestasi
Berprestasi	Tidak Berprestasi
Tidak Berprestasi	Tidak Berprestasi
Tidak Berprestasi	Tidak Berprestasi
Berprestasi	Berprestasi
Berprestasi	Berprestasi
Berprestasi	Berprestasi
Berprestasi	Berprestasi
…	…
…	…
Tidak Berprestasi	Tidak Berprestasi
Berprestasi	Berprestasi

Tabel 4.16 Hasil Evaluasi

Confusin Matrix	**Manual**	***Python***
True Positive	60	59
True Negative	0	0
False Positive	38	41
False Negative	2	0

IV. KESIMPULAN DAN SARAN

Berdasarkan hasil perhitugan dari klasifikasi menentukan prestasi siswa berdasarkan data siswa dengan 6 kriteria parameter menggunakan algoritma *Naïve Bayes* maka dapat disimpulkan sebagai berikut :

1. Hasil perhitungan algoritma *Naïve Bayes* dengan dengan data *testing* sebanyak 100 data siswa diambil dari data set 400 data siswa, mendapatkan hasil 41 siswa dinyatakan tidak berprestasi dan 59 siswa dinyatakan berprestasi.
2. Hasil klasifikasi algoritma *Naïve Bayes* dalam perhitungan manual dan *python* dengan *Confusion Matrix*. Hasil dari perhitungan manual menunjukan hasil Akurasi 98%, Presisi 100% dan *Recall* 98%. Dan untuk perhitungan *python* menunjukan hasil Akurasi 100%, Presisi 100% dan *Recall* 100%.

Untuk penelitian yang akan datang, peneliti mengharapkan agar menggunakan algoritma yang berbeda untuk melihat hasil

yang lebih akurat dan tepat. Serta menggunakan data yang lebih banyak dalam melakukan klasifikasi untuk mendapatkan hasil yang lebih baik.

PENGAKUAN

Naskah ilmiah ini adalah sebagian dari penelitian Tugas Akhir milik Muh Nurhariza dengan judul Implementasi Algoritma *Naïve Bayes* Untuk Klasifikasi Menentukan Prestasi Siswa Berdasarkan Nilai Rata-Rata yang dibimbing oleh Ibu Ayu Ratna Juwita, M. Kom dan Ibu Dwi Sulistya Kusumaningrum,M.Pd.

DAFTAR PUSTAKA

[1] Yusuf, B., Qalbi, M., Basrul, B., Dwitawati, I., Malahayati, M., & Ellyadi, M. (2020). Implementasi Algoritma Naive Bayes Dan Random Forest Dalam Memprediksi Prestasi Akademik Mahasiswa Universitas Islam Negeri Ar-Raniry Banda Aceh. *Cyberspace: Jurnal Pendidikan Teknologi Informasi*, *4*(1), 50. https://doi.org/10.22373/cj.v4i1.7247.

[2] Dan, A., & Yang, F. (2018). *Study About Student Learning Achievement Aspect And*. *2*(2), 115–123.

[3] Romli, I., & Zy, A. T. (2020). Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5. Jurnal Sains Komputer & Informatika (J-SAKTI, 4(2), 694–702.

[4] Wijaya, H. D., & Dwiasnati, S. (2020). *Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat*. *7*(1), 1–7.

[5] Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI*, *5*(2), 697–711.