

## BAB 2

### TINJAUAN PUSTAKA

#### 2.1 Indeks Prestasi Mahasiswa

Prestasi mahasiswa merupakan salah satu tolak ukur dalam menentukan apakah mahasiswa ini berhasil atau gagal dalam perkuliahan. Proses untuk menentukannya bisa dilihat dari IP mahasiswa, jika mahasiswa mendapatkan IP yang rendah maka penting bagi manajemen universitas untuk terus memonitoring proses pembelajaran dan melakukan kebijakan yang diperlukan untuk meningkatkan prestasi mahasiswa. Melakukan prediksi prestasi mahasiswa merupakan hal penting agar perkuliahan dapat berjalan dengan lancar, dengan Indeks Prestasi Semester (IPS) bisa di prediksi bagaimana prestasi nilai akademik mahasiswa kedepan. Prediksi ini sendiri bisa membantu Dosen untuk mengevaluasi teknik, kualitas, dan pendekatan dalam pembelajaran di dalam perkuliahan serta membantu mahasiswa itu sendiri dalam mengukur kemampuannya dalam mengikuti perkuliahan. Dosen dapat memantau Indeks Prestasi Semester (IPS) mahasiswa dan melakukan perbandingan dengan prediksi yang sudah dibuat (Hasudungan & Pranoto, 2021). Berikut merupakan tabel penelitian yang terkait membahas evaluasi prestasi mahasiswa ditunjukkan di tabel 2.1.

**Tabel 2.1 Penelitian Terkait**

Author	Metode	Hasil	Keterangan
Rofilde Hasudungan, Wawan Joko Pranoto, Rudiman (2020)	Naive Bayes dan Maximum Dependency of Attributes	Eksperimen dilakukan dengan data 40 pelajar dengan 28 fitur menunjukkan model memiliki akurasi 79%, hasil akurasi lebih tinggi dibandingkan Naive Bayes tanpa MDA yang memiliki akurasi 68%.	Naive bayes klasifier merupakan algoritma yang berdasar pada teorema naive. Naive bayes tidak memiliki kemampuan untuk memilih fitur terbaik untuk itu digunakanlah MDA sebagai metode untuk memilih fitur terbaik dalam prediksi.

**Tabel 2.1 Penelitian Terkait**

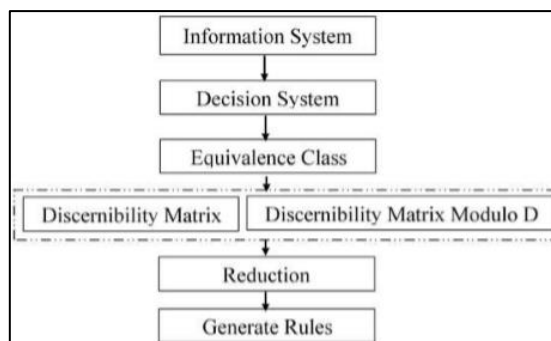
Author	Metode	Hasil	Keterangan
Rofilde Hasudungan ( 2021 )	Naive Bayes dan <i>Rough Set</i>	<i>Rough Set</i> dan Naive Bayes menunjukkan dengan melakukan seleksi fitur dan menggunakan Naive Bayes sebagai metode klasifikasi meningkatkan akurasi secara signifikan dari 68,09% ke 80,85% dan presisi kelas gagal dari 84,38% ke 86,64%.	Metode <i>Rough Set</i> dan Naive Bayes ditujukan untuk menganalisa data pelajar yang berjumlah 47 data dengan menggunakan <i>Rough Set</i> atribut dikurangi dari 28 ke 6 atribut dan Naive Bayes digunakan untuk metode klasifikasi dan analisis data.
Riswan Efendi , dkk ( 2018 )	<i>Rough Set</i> dan Jaccard	Berdasarkan hasil pengolahan data asli diperoleh 3 atribut yang mempengaruhi prestasi akademik mahasiswa jurusan matematika yaitu lamanya bermain handphone per hari, jumlah unduhan artikel ilmiah per hari dan jumlah halaman Al-Qur'an yang dibaca per hari oleh mahasiswa tersebut.	Mengaplikasikan pendekatan non-statistik untuk menganalisa hubungan antara atribut bersyarat dan atribut keputusan IPK mahasiswa dengan metode <i>Rough Set</i> dan Jaccard.
Mucahid Mustafa Saritas, Ali Yasar ( 2019 )	Naive Bayes	Naive Bayes berhasil mendapatkan akurasi sebesar 83,54% dan berguna sebagai pencegahan penyakit kanker.	Klasifikasi Naive Bayes diaplikasikan kedalam data dengan 9 input.
Yuhua Qian, dkk ( 2018 )	<i>Rough Set</i>	Teori analisis dan eksperimen menunjukkan hasil bahwa masing-masing algoritma dalam himpunan <i>Rough Set</i> secara signifikan mengungguli rekan	Sebagai metode pembelajaran mesin, teori <i>Rough Set</i> klasik menghadapi tiga tantangan untuk analisis <i>Rough Set</i> dalam big data, inefisiensi komputasi dan kecocokan yang

**Tabel 2.1 Penelitian Terkait**

Author	Metode	Hasil	Keterangan
		aslinya dalam teori <i>Rough Set</i> klasik.	berlebihan dalam pengurangan atribut.
Rofilde Hasudungan , Wawan Joko Pranoto ( 2021 )	Naive Bayes	Hasil penelitian menunjukkan bahwa algoritma naive bayes dapat memprediksi prestasi akademik mahasiswa dengan tingkat akurasi sebesar 77,5%	Penelitian ini mengungkap faktor keluarga terhadap prestasi menggunakan algoritma naive bayes dan enam belas parameter yang terlibat.

## 2.2 *Rough Set*

*Rough Set* merupakan salah satu metode matematika untuk menangani data yang tidak konsisten dan samar serta tidak lengkap. *Rough Set* dikembangkan oleh Zdzislaw Pawlak yang dapat digunakan sebagai metode matematika untuk menangani ketidakjelasan dan ketidakpastian dan telah berhasil diterapkan dalam berbagai tugas seperti fitur seleksi / ekstraksi , sintesis aturan dan klasifikasi (Yulianti, 2016). Berikut skema penyelesaian menggunakan algoritma *Rough Set* menurut (Hengki Juliansaa, Sarjon Defitb, 2018).



**Gambar 2.1 Skema Algoritma *Rough Set***

Berikut adalah keterangan dari skema penyelesaian menggunakan algoritma *Rough Set* :

1. **Information System** adalah tabel yang terdiri dari baris yang mempresentasikan data dan kolom yang merepresentasikan atribut atau

variabel dari data. Information System dapat dipresentasikan sebagai fungsi  
:  $S = \{U, A\}$ .

2. **Decision System** adalah information system dengan atribut tambahan yang dinamakan dengan decision atribut, dalam mining dikenal dengan nama kelas atau target. Decision System merupakan fungsi yang mendeskripsikan information system.
3. **Equivalence Class** adalah mengelompokkan objek – objek yang sama untuk atribut A ( $U, A$ ). Class BC5 adalah sebuah indeterminacy yang memberikan 2 keputusan yang berbeda. Situasi ini dapat ditangani dengan teknik data cleaning.
4. **Discrenibility Matrix** diberikan sebuah IS  $A = (U, A)$  dan B gabung A. Discrenibility Matrix dari A adalah MB, dimana tiap – tiap entri MB(Lj) terdiri dari sekumpulan atribut yang berbeda antara objek  $X_i$  dan  $X_j$ .
5. **Discrenibility Matrix Modulo D** didefinisikan seperti berikut dimana Modulo (ij) adalah sekumpulan atribut yang berbeda antara objek  $X_i$  dan  $X_j$  dan juga berbeda atribut keputusan. Diberikan sebuah DS  $A = (U, A\{d\})$  dan subset dari atribut B gabung A. Discernibility Matrix Modulo D dari A, MBd, didefinisikan seperti berikut di mana MB(i,j) adalah sekumpulan atribut yang berbeda antara objek  $x_i$  dan  $x_j$  dan juga berbeda atribut keputusan.
6. **Reduction** untuk data yang jumlah variabelnya tidak besar sangat tidak mungkin mencari seluruh kombinasi variabel yang ada, karena jumlah indiscernibility yang dicari  $= (2^n - 1) - 1$ . Reduction sendiri ialah suatu subset minimal atribut yang membentuk relasi *indiscernibility*. Misalkan  $S = (U, A, V, f)$  merupakan tabel informasi dan misalkan B merupakan sebarang subset dari A. Misalkan b anggota B, maka b dikatakan *dispensable* (berlebihan) dalam B jika  $U/(B - \{a\}) = U/B$ , sebaliknya b dikatakan *indispensable* dalam B.
7. **Generate Rules** merupakan proses utama dalam menemukan pengetahuan dalam database dengan ekstrasi aturan dari sistem pengambilan keputusan.

### 2.3 Maximum Dependency of Attributes

Seleksi fitur digunakan untuk memilih subset fitur dari semua fitur yang relevan dan ketergantungan tinggi terhadap data. Maximum Dependency of Attributes (*MDA*) adalah reduksi fitur yang berdasarkan kepada *Rough Set* yang menggunakan ketergantungan atribut dalam sistem informasi untuk memilih atribut pengelompokkan. Teknik ini memiliki kelebihan dalam menemukan atribut yang memiliki ketergantungan maksimum dan menghilangkan redundansi. Algoritma *MDA* terdiri dari beberapa langkah utama menurut (Hasudungan & Pranoto, 2020) :

1. Hitung kelas ekivalensi menggunakan relasi indiscernibility pada setiap atribut.
2. Tentukan derajat ketergantungan atribut  $\alpha$  , sehubungan dengan  $\alpha$  , dimana  $i \neq j$ .
3. Pilih tingkat maximum of dependency dari setiap atribut.

### 2.4 Naïve Bayes

*Naïve Bayes* merupakan salah satu pengklasifikasi statistik yang klasifikasinya dapat memprediksi probabilitas keanggotaan kelas dari suatu data yang akan masuk ke dalam kelas tertentu sesuai dengan perhitungan probabilitas. Klasifikasi *Naïve Bayes* diasumsikan bahwa ada atau tidak ciri dari tertentu dari kelas tidak ada hubungannya dengan kelas lainnya , jumlah data yang dibutuhkan oleh metode *Naïve Bayes* sedikit dan ini merupakan keunggulan dari metode ini (Hasudungan & Pranoto, 2021). Persamaan (1) dari metode Naive Bayes.

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \quad (1)$$

Keterangan :

$P(H|E)$  : Probabilitas akhir bersyarat suatu hipotesis H terjadi jika diberikan bukti E terjadi.

$P(E|H)$  : Probabilitas sebuah bukti E akan memengaruhi hipotesis H.

$P(H)$  : Probabilitas awal hipotesis H terjadi tanpa memandang bukti apapun.

$P(E)$  : Probabilitas awal bukti E terjadi tanpa memandang hipotesis/bukti lain.

## 2.5 Python

Python merupakan bahasa pemrograman tingkat tinggi interpreted, berorientasi objek dan tidak menggunakan compiler. Struktur data tingkat tinggi bawaan dikombinasikan dengan pengetikan dinamis dan pengikatan dinamis, tidak hanya menarik untuk pengembangan aplikasi tetapi juga digunakan sebagai bahasa script atau untuk menghubungkan komponen yang ada. Sintak python yang sederhana dan mudah dipelajari menekankan readability dan karena hal tersebut dapat mengurangi biaya pemeliharaan program. Python mendukung modul dan paket, memfasilitasi modularitas program dan penggunaan kembali kode.

## 2.6 Google Colab

Google Colab merupakan produk dari Google Research. Google Colab pada dasarnya berjalan dengan standar notebook jupyter gratis yang memungkinkan siapa saja untuk menulis kode dan mengeksekusi kode python melalui browser dan sepenuhnya berjalan di cloud. Google Colab tidak memerlukan penyiapan apapun dan notebook yang anda buat dapat diedit secara bersamaan dengan anggota tim. Google Colab sangat cocok untuk pembelajaran mesin, analisis data, dan pendidikan. Keuntungan terbesar Google Colab adalah mendukung library machine learning populer yang dapat dimuat dengan mudah di notebook.

Fitur yang diberikan Google Colab menurut (Dr. Michael J. Garbade, 2021) :

1. Menulis dan mengeksekusi kode python.
2. Membuat, mengunggah dan membagikan notebook.
3. Import atau simpan notebook dari atau ke Google Drive.
4. Import atau publikasi notebook dari GitHub.
5. Import dataset dari luar cloud.
6. Gratis cloud servis dengan GPU gratis.

Adapun kekurangan Google Colab yang dibahas oleh (Adam, 2021) adalah :

1. Tidak dapat mengetahui jenis GPU yang digunakan dan kapasitas memory yang bisa digunakan secara maksimal.
2. Google Colab secara default hanya menyediakan format notebook (.ipynb).

3. Waktu runtime yang dibatasi 12 jam setelah itu akan ada jeda dan harus menunggu beberapa waktu untuk bisa digunakan lagi.

## **2.7 Evaluasi**

Evaluasi merupakan kegiatan yang dilakukan untuk mengukur sesuatu dan membandingkan sesuatu dengan dasar ukuran tertentu. Kemudian akan dilakukan pengambilan keputusan terhadap sesuatu dengan mendasarkan diri pada ukuran yang telah ditentukan. Evaluasi mencakup dua kegiatan yaitu pengukuran dan penilaian (Hidayat, Asyafah, Indonesia, & Barat, 2019).

Tujuan dilakukannya kegiatan evaluasi adalah :

1. Untuk mengetahui kesulitan yang dialami seseorang dalam melakukan kegiatannya sehingga bisa dilakukan diagnosis.
2. Untuk mengetahui tingkat efektivitas dari metode – metode , media dan sumber daya lainnya dalam melakukan kegiatan.
3. Untuk mengetahui tingkat penguasaan seseorang terhadap kompetensi yang telah ditetapkan.
4. Sebagai umpan balik dan informasi penting bagi para pelaksana evaluasi untuk memperbaiki kekurangan yang ada yang dimana hal tersebut dapat menjadi patokan untuk mengambil keputusan di masa mendatang.

## **2.8 Akurasi**

Akurasi merupakan ketepatan suatu aspek yang tingkat nilai pengukurannya mendekati nilai benar. Akurasi menentukan seberapa dekat suatu pengukuran terhadap nilai yang diterima (Fitrya et al., 2017). Akurasi mengukur tingkat ketepatan dan kemiripan pada hasil untuk membandingkannya dengan nilai absolut , semakin mendekati ukurannya , semakin tinggi juga level akurasinya.

## **2.9 Presisi**

Presisi adalah penyempurnaan dalam pengukuran , perhitungan atau spesifikasi terutama yang berada dijumlah digit yang diberikan. Sebuah alat ukur dapat dikatakan presisi ketika digunakan dalam pengukuran besaran tertentu yang ketika diulang maka alat ukur tersebut mampu menghasilkan hasil ukur yang sama seperti sebelumnya (Paken Pandiangan, S.Si., M.Si. Artoto Arkundato, S.Si.,

n.d.). Kepresisian dari sebuah sistem pengukuran dapat disebut sebagai reproduktivitas atau pengulangan yang ketika diulang pengukurannya kondisinya tidak berubah dan tetap mendapatkan hasil yang sama. Semakin tinggi level dari presisi maka semakin kecil variasi antar pengukuran (Rabbani, 2020).