

## BAB 2

### TINJAUAN PUSTAKA

#### 2.1 Prediksi Prestasi Mahasiswa

Prediksi ialah proses yang mempertimbangkan dari berbagai sudut pandang untuk meramalkan peristiwa atau kejadian di masa depan. Tahapan peramalan mampu dilaksanakan apabila termuat dataset latihan yang mampu dikelola dengan baik alur ataupun polanya. Tidak sedikit algoritma yang dipergunakan teruntuk melakukan peramalan suatu data ataupun peristiwa yang akan terjadi di masa depan. Meramal mampu dilaksanakan dengan sebuah data musiman, yang dimaknai dengan data yang tertera mampu berubah tiap waktunya. (Rolansa et al., 2020). Melakukan prediksi terhadap nilai prestasi yang diperoleh mahasiswa ialah peranan utama yang perlu dipertimbangkan dan dipersiapkan supaya pembelajaran perkuliahan mampu berjalan optimal serta efektif. Oleh karenanya perlu mempergunakan IPS atau dikenal dengan Indeks Prestasi Semester serta jalur masuk perguruan tinggi, dengan keduanya mampu dilakukan peramalan mengenai prestasi yang nantinya diraih oleh mahasiswa. Tahapan prediksi mampu mempermudah pengajar dalam meningkatkan serta melakukan kemajuan pada prestasi anak didiknya. (Hasudungan & Pranoto, 2021). Berikut adalah tabel penelitian terkait yang membahas tentang prediksi menggunakan algoritma Rough set dan Naïve Bayes sebagai mana ditunjukkan pada tabel 2.1.

Tabel 2.1 Penelitian Terkait

Author	Metode	Hasil	Keterangan
Silvana Samara (2021)	Rough set	Penelitian mebghasilkan 14 buah rules berupa pola aturan sebagai rujukan untuk	Cara pengolahan data mengaplikasikan algoritma Rough Set, diawali dari pemilihan

Tabel 2.1 Penelitian Terkait (*Lanjutan*)

Auhtor	Metode	Hasil	Keterangan
		<p>memprediksi hasil belajar lulus, cukup dan tidak lulus. Berdasarkan rules yang dibuat, disimpulkan bahwa atribut kondisi yang paling berpengaruh dalam penentuan hasil belajar yaitu nilai UAS dilanjutkan dengan nilai tugas dan jumlah kehadiran</p>	<p>atribut kondisi dan atribut keputusan dilanjutkan dengan progres menghilangkan data ganda, hingga mendapatkan reduct dan menciptakan rules. Pengolahan data menggunakan software Rosetta.</p>
<p>Aswanda Putra, Zekson A Matondang, Noferianto Sitompul (2018)</p>	<p>Rough set</p>	<p>hasil prediksi menyatakan bahwa faktor yang paling mempengaruhi kecerdasan anak adalah berpikir abstrak, memecahkan masalah, memahami gagasan</p>	<p>untuk mengetahui anak yang berpotensi, sehingga anak tersebut dapat di arahkan atau didik</p>

Tabel 2.1 Penelitian Terkait (*Lanjutan*)

Author	Metode	Hasil	Keterangan
Sinta Novianti , Paska Marto Hasugian (2021)	Rough set	hasil dari analisa perhitungan ialah 152 siswa untuk prediksi jumlah siswa ditahun yang akan datang.	dalam penelitian ini menggunakan data pendaftaran siswa tahun 2018, 2019, 2020.
Mokhamad Ramdhani Raharjo, Agus Perdana Windart (2021)	Rough set	menghasilkan pengetahuan baru yaitu hasil pembelajaran berdasarkan matakuliah. Terdapat 15 Reduction dengan 90 Generate Rule. Namun secara keseluruhan, atribut yang berpengaruh terhadap tingkat pemahaman mahasiswa terhadap matakuliah adalah komunikasi dan media pembelajaran.	atribut yang digunakan komunikasi, suasana belajar, media pembelajaran, penampilan dan metode pengajaran.

Tabel 2.1 Penelitian Terkait (*Lanjutan*)

Author	Metode	Hasil	Keterangan
M. Arif Rahman (2020)	Rough set	Dari 13 data sampel penjualan proses Data Mining dengan metode rough set dihasilkan 5 Reduct yang diekstrak menjadi pengetahuan 11 Generate Rules sehingga menghasilkan sebuah keputusan yang diwakili dari rules-rule yang dihasilkan.	Data analisis menggunakan aplikasi Rosetta 1.4.4.1. Penelitian ini menunjukkan keputusan yang menghasilkan pengetahuan baru berupa rules sebagai pendukung keputusan nantinya.
Chairul Fadlan, Selfia Ningsih, Agus Perdana Windarto (2018)	Naïve Bayes	Hasil dari perhitungan menggunakan algoritma naïve bayes dapat di ambil kesimpulan bahwa nilai $P(C0 X) >$ dari nilai $P(C1 X)$ . maka diklasifikasikan kedalam kelas Tidak Layak	Dengan mengetahui masyarakat yang berhak menerima beras rastra, dapat meminimalisir kesalahan masyarakat yang berhak atau layak dalam menerima beras rastra.

Tabel 2.1 Penelitian Terkait (*Lanjutan*)

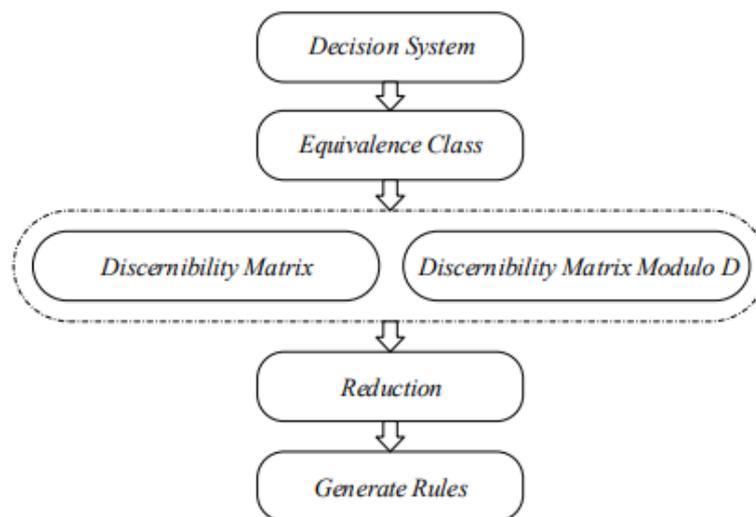
Author	Metode	Hasil	Keterangan
R. Annisa, A. Sasongko (2020)	Naïve Bayes	Hasil penelitian menunjukkan accuracy 96,24%, precision 95,76%, dan recall 100%.	algoritma Naïve Bayes menunjukkan hasil prediksi berdasarkan mahasiswa yang kuliah sambil bekerja, jadwal kerja mahasiswa, dan berdasarkan waktu kuliah.
Rofilde Hasudungan, Wawan Joko Pranoto (2021)	Naïve Bayes	Hasil penelitian menampilkan dimana algoritma Naïve Bayes dapat memprediksi prestasi akademik mahasiswa dengan tingkat akurasi sebesar 77,5%.	Penelitian ini mengungkap faktor keluarga terhadap prestasi menggunakan algoritma Naïve Bayes dan enam belas parameter yang terlibat.
Naisha Rahma Indraswari, Yogiek Indra Kurniawan (2018)	Naïve Bayes	nilai akurasi aplikasi tertinggi pada angka 78,69%, nilai precision tertinggi ada pada angka 70.14% dan nilai recall tertinggi ada pada angka 63.64%.	Dengan variable inputan faktor-faktor yang dialami oleh ibu hamil, diantaranya: usia ibu, tekanan darah, jumlah bayi, riwayat persalinan, riwayat abortus/ kuretase, malnutrisi, penyakit bawaan

Tabel 2.1 Penelitian Terkait (*Lanjutan*)

Author	Metode	Hasil	Keterangan
			sebelum hamil dan masalah saat kehamilan.
Tia Imandasari, Eka Irawan, Agus Perdana Windarto, Anjar Wanto (2019)	Naïve Bayes	Hasil dari perhitungan menggunakan algoritma Naive Bayes, diperoleh hasil klasifikasi dari 19 alternatif yang digunakan, dimana terdapat 8 kelas Layak dan 11 Tidak Layak dengan total Accuracy yang diperoleh sebesar 78,95%.	Diharapkan penelitian ini dapat membantu pihak PDAM Tirta Lihou dalam menentukan lokasi yang layak dilakukan pembangunan sumber air sehingga dapat memenuhi kebutuhan masyarakat.
Haditsah Annur (2018)	Naïve Bayes	Diperoleh tingkat akurasi sebesar 73% atau termasuk dalam kategori Good. Sementara nilai Precision sebesar 92% dan Recall sebesar 86%.	Atribut yang akan digunakan dalam melakukan klasifikasi penduduk adalah Umur, Pendidikan, Pekerjaan, Penghasilan, Tanggungan, Status (Kawin/Belum Kawin).

## 2.2 Teori Rough Set

Rough set merupakan salah satu metode matematika untuk menangani data yang tidak konsisten dan samar serta tidak lengkap. Tujuan dari metode rough set adalah untuk mendapatkan perkiraan rule yang singkat dari suatu tabel. Selain itu, keunggulan dari metode ini ialah tidak memerlukan parameter atau masukan karena informasi terkait dengan data diambil dari data itu sendiri (Aswanda Putra et al., 2018). Berikut adalah skema penyelesaian menggunakan algoritma Rough Set berdasarkan Salvina Samaray (2021).



Gambar 2.1 Skema Algoritma Rough Set

Berikut adalah keterangan dari skema penyelesaian menggunakan algoritma Rough Set :

1. Decision System (DS) adalah Information System (IS) dengan atribut tambahan berupa atribut keputusan. Decision system merupakan fungsi yang mendeskripsikan information system, dituliskan menjadi:  $DS = (U, \{A, C\})$ .
2. Equivalence class fungsinya adalah mengelompokkan objek-objek yang sama untuk atribut A (U, A).
3. Discernibility Matrix didefinisikan sebagai berikut, diberikan sebuah IS  $A = (U, A)$  dan B gabung A. Discernibility Matrix dari A adalah MB, di mana tiap-tiap

entry MB (i,j) terdiri dari sekumpulan atribut yang berbeda antara objek  $x_i$  dan  $x_j$ .

4. Discernibility Matrix Modulo D didefinisikan sebagai berikut, di mana Modulo (i,j) adalah sekumpulan atribut yang berbeda antara objek  $x_i$  dan  $x_j$  dan juga berbeda atribut keputusan. Diberikan sebuah DS  $A = (U, A \cup \{d\})$  dan subset dari atribut B gabung A. Discernibility Matrix Modulo D dari A,  $MB_d$ , 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.
5. Reduction merupakan teknik pencarian kombinasi atribut.
6. Generate Rules adalah proses menemukan pengetahuan dalam database, yaitu dengan ekstraksi aturan dari sistem pengambilan keputusan.

### 2.2.1 Information Table

Rough set merepresentasikan data dalam bentuk tabel, yang disebut dengan *information system* (sistem informasi) atau *information table* (tabel informasi) (Aswanda Putra, Zekson A Matondang & Program, 2018). Pada penelitian ini akan menggunakan istilah *information table* (tabel informasi). Pada tabel informasi, data dibentuk dalam baris dan kolom, dimana setiap baris mereferensikan sebuah objek atau entitas, sedangkan setiap kolom merepresentasikan atribut atau fitur dari objek tersebut. Suatu table informasi dinotasikan dengan  $S = (U, A, V, f)$ , dimana  $U$  ialah himpunan semesta,  $A$  ialah himpunan attribute,  $V = \bigcup_{a \in A} V_a$ ,  $V_a$  ialah domain atribut  $a$ ,  $f: U \times A \rightarrow V$  ialah fungsi yang memetakan suatu objek  $i$  terhadap atribut  $a$  sehingga  $f(i, a) \in V_a$ . Kemudian, jika suatu tabel terdapat label atau kelas maka disebut dengan table informasi keputusan (*decision information table*), dimana table ini dinotasikan dengan  $S' = (U, A \cup \{d\}, V, f)$ , dimana  $U, A, V$ , dan  $f$  sesuai dengan  $S$  dan  $\{d\}$  ialah atribut keputusan dimana  $\{d\} \cap A \neq \emptyset$  (Aswanda Putra, Zekson A Matondang & Program, 2018).

### 2.2.2 Indiscernible Relation

**Definisi 2.1.** Misalkan  $S = (U, A, V, f)$  merupakan tabel informasi dan misalkan  $B$  merupakan subset dari  $A$ . Dua buah objek  $x, y \in U$  dikatakan  $B$ -

*indiscernible* (tidak dapat dibedakan/ekuivalen) jika dan hanya jika  $a \in B \rightarrow (f(x, a) = f(y, a))$ . Dari definisi 2.1 maka untuk setiap subset  $A$  akan membentuk relasi *indiscernibility*. Sebuah relasi *indiscernibility* yang dibentuk oleh atribut  $B \subseteq A$  dinotasikan dengan  $IND(B)$ , dimana relasi tersebut ialah relasi ekuivalen. Kemudian partisi terhadap  $U$  akibat  $IND(B)$  pada  $S$  dinotasikan dengan  $U/B$  (M. Arif Rahman, 2020).

### 2.2.3 Set Approximation

**Definisi 2.2.** Untuk menentukan approximation (perkiraan) yang ada dalam tabel information, Misalkan  $S = (U, A, V, f)$  merupakan tabel informasi dan misalkan  $B$  merupakan sebarang subset dari  $A$ . *B-lower Approximation* dari  $X$  (dinotasikan dengan  $\underline{B}(X)$ ) dan *B-upper Approximation dari X* (dinotasikan dengan  $\overline{B}(X)$ ) berturut-turut didefinisikan :

$$\underline{B}(X) = \{x \in U | [x]_B \subseteq X\} \text{ dan } \overline{B}(X) = \{x \in U | [x]_B \cap X \neq \emptyset\} \quad (2.1)$$

### 2.2.4 Quick Reduct

Untuk membentuk relasi *indiscernibility* dari objek-objek, maka menggunakan atribut-atribut yang terdapat pada tabel informasi. Namun, kadang tidak perlu menggunakan semua atribut untuk membentuk relasi tersebut. Hal ini karena kadang terdapat atribut yang berlebihan, sehingga eliminasi atribut tersebut tidak berpengaruh terhadap relasi. Untuk data yang jumlah variabelnya besar sangat tidak mungkin mencari seluruh kombinasi variabel yang ada, sebab jumlah *indiscernibility* yang dicari sama dengan  $(2^n - 1)$ . Oleh karena itu diciptakan satu teknik pencarian kombinasi atribut yang dikenal dengan Quick Reduct (Samaray, 2021), yaitu dengan cara :

1. Nilai *indiscernibility* yang pertama dicari adalah *indiscernibility* untuk kombinasi atribut yang terkecil yaitu 1.
2. Kemudian lakukan proses pencarian *dependency attributes*. Apabila nilai *dependency attributes* yang didapatkan sama dengan 1 maka *indiscernibility* untuk himpunan minimal variabel adalah variabel tersebut.
3. Jika pada proses pencarian kombinasi atribut tidak ditemukan *dependency attributes* sama dengan 1, maka lakukan pencarian kombinasi yang lebih besar,

di mana kombinasi variabel yang dicari merupakan kombinasi dari variabel ditahap sebelumnya yang nilai dependency attributes paling besar. Lakukan proses (3), hingga didapat nilai dependency attributes sama dengan 1.

### 2.3 Naïve Bayes

Naive Bayes merupakan pengklasifikasian probalistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi serta kombinasi nilai dari dataset yang diberikan (Jeprianto & Aziz, 2020). Klasifikasi Naïve Bayes diasumsikan bahwa ada atau tidaknya ciri tertentu dari sebuah kelas yang tidak ada hubungannya dengan kelas lainnya. Jumlah data pelatihan (training) yang diperlukan oleh sistem Naïve Bayes sedikit, dimana hal ini merupakan keunggulan dari sistem ini (Hasudungan & Pranoto, 2021). Naïve Bayes juga didefinisikan sebagai pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuan inggis Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya (Jeprianto & Aziz, 2020). Persamaan dibawah ini adalah persamaan dari metode Naïve Bayes (Hasudungan & Pranoto, 2021).

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \quad (2.2)$$

Keterangan :

$P(H|E)$  : Probabilitas akhir bersyarat (conditional probability) suatu hipotesis H terjadi jika diberikan bukti (evidence) 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 yang lain.

Tujuan utama Teorema Bayes adalah untuk menghitung probabilitas bersyarat (Annisa & Sasongko, 2020). Aturan Bayes dapat diturunkan dari dua persamaan berikut:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2.3)$$

Persamaan di bawah ini mewakili probabilitas bersyarat dari A, mengingat B:

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \quad (2.4)$$

Oleh karena itu, pada penggabungan dua persamaan di atas kita menerima Teorema Bayes. Persamaan di atas merupakan variabel untuk prediktor tunggal, tetapi dalam aplikasi dunia nyata, ada lebih dari satu variabel prediktor dimana untuk keadaan masalah klasifikasi, ada lebih dari satu kelas output (Annisa & Sasongko, 2020).

## 2.4 Python

Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Python diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif (Syahrudin & Kurniawan, 2018). Python menyediakan dukungan yang kuat untuk integrasi dengan bahasa pemrograman lainnya. Python hadir dengan pustaka-pustaka standar yang bisa diperluas serta bisa dipelajari cuma dalam beberapa hari. Bahasa pemrograman yang interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode (Fitri et al., 2017). Python didistribusikan dengan beberapa lisensi yang berbeda dari beberapa versi. Tapi pada prinsipnya, Python dapat dan bisa dipergunakan secara bebas, misalnya untuk kepentingan komersial. Lisensi Python tidak bertentangan baik berdasarkan definisi Open Source maupun General Public License (GPL) (Syahrudin & Kurniawan, 2018).

## 2.5 Evaluasi

Evaluasi ialah sebuah agenda pekerjaan yang melakukan pengumpulan data atau sebuah informasi tentang kinerja sesuatu (mekanisme atau cara kerja, individu manusia, fasilitas peralatan), data terkait mampu dipergunakan teruntuk melakukan penentuan alternatif terbaik untuk memberikan putusan. Evaluasi ialah bentuk dari pengukuran ataupun upaya memperbaiki pada sebuah aktivitas

pekerjaan yang dilakukan, misalnya melakukan perbandingan perolehan aktivitas yang sudah dilakukan pencanangan. Oleh karenanya maksud serta tujuan aktivitas evaluasi ini ialah supaya perencanaan serta pencanangan yang sudah disusun mampu memperoleh tujuan yang sudah dilakukan penetapan serta mampu dilaksanakan (Syafnidawaty, 2020).

Adapun berbagai tujuan serta maksud dilaksanakannya aktivitas evaluasi ini yakni diantaranya:

1. Mengetahui tingkatan pemahaman serta mampu menguasai individu pada sebuah kompetensi.
2. Melakukan identifikasi berbagai hambatan yang dilalui pada sebuah aktivitas, kemudian evaluasi dilakukan untuk melakukan pemecahan persoalan serta kesukaran yang dijalannya pada sebuah aktivitas.
3. Teruntuk mengetahui tingkatan efisiensi serta optimalisasi pada sebuah mekanisme, media, serta sumber daya yang mendukung untuk memperlancar aktivitas evaluasi.
4. Menjadi bahan serta acuan data yang mempunyai peranan penting teruntuk melaksanakan aktivitas evaluasi pada proses perbaikan hal yang masih kurang.

### **2.5.1 Akurasi**

Akurasi ialah pengukuran bentuk pendekatan dari perolehan alat ukur dengan nilai sebenarnya. Tahapan ini dilakukan penentuan dengan melakukan perhitungan nilai rerata hasil analisis dari blanko yang dilakukan penambahan analit, bahan referensi standar, serta larutan standar. Kemudian, sampel spike matriks dilakukan pengukuran, hal tersebut mampu memperlihatkan akurasi yang akurat ataupun bias pada matriks sampel aktual. Akurasi mampu dinyatakan menjadi persentase dari hasil kembali (% recovery) dari sebuah nilai yang dilakukan pengukuran pada nilai sesungguhnya ataupun perolehan target sasaran. Apabila tahapan pengukuran memperoleh nilai rerata memiliki selisih dari nilai sesungguhnya ataupun perolehan target sasaran, tahapan tersebut mampu dinyatakan bias. Bias adalah kesalahan sistematis (systematic error) baik yang melekat dalam metode analisis (misalnya efisiensi ekstraksi) atau disebabkan oleh

artefak sistem pengukuran (misalnya kontaminasi). Karena bias dapat positif atau negatif, dan karena beberapa jenis bias dapat terjadi secara bersamaan, hanya total bias dapat dievaluasi dalam sebuah pengukuran (HOLIDIN, 2019).

### **2.5.2 Presisi**

Presisi ialah keahlian pada metode ataupun instrumen yang melakukan proses analitis teruntuk melakukan produksi pengukurannya sendiri. Tahapan ini ialah pengukuran variabilitas, ataupun kesalahan pada tahapam pengacakan, pada proses mengambil sampel, proses dalam menangani sampel serta pada penganalisan laboratorium (HOLIDIN, 2019). Presisi melakukan penggambaran pada keberagaman serta pengulangan dalam proses mengukur. Ini mengukur sejauh mana perolehan dekat satu dengan yang lainnya, yakni proses dalam mengukur dikelompokkan serentak. Tahapan ini ialah derajat keunggulan dari sebuah mekanisme strategi yang dipergunakan teruntuk memperoleh hasil yang diinginkan.(Arkundato, 2018).