

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Penelitian Terdahulu

Penelusuran referensi karya ilmiah terdahulu yang digunakan sebagai bahan rujukan dan pembanding dalam penelitian ini antara lain sebagai berikut:

**Tabel 2.1 Penelitian terdahulu**

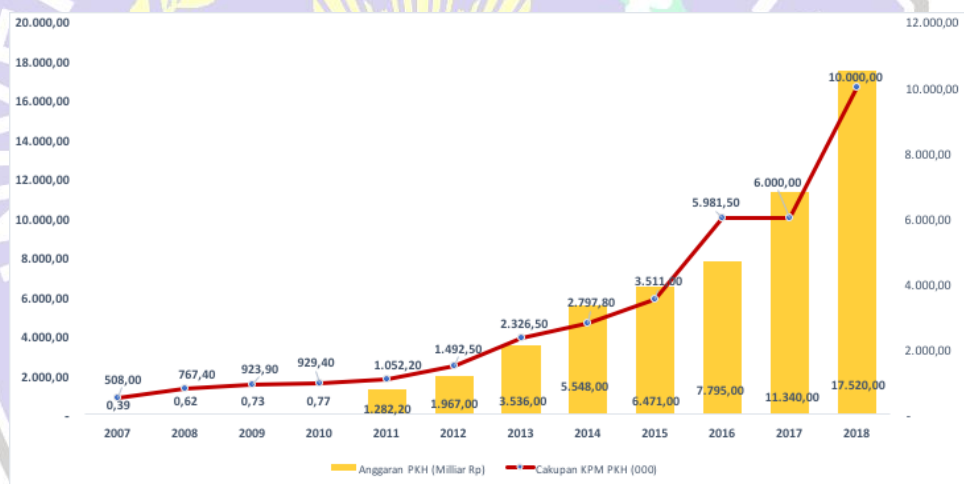
<b>Judul</b>	<b>Peneliti (Tahun)</b>	<b>Metode</b>	<b>Hasil Penelitian</b>
1 Data mining analisis untuk prediksi tingkat pendidikan remaja Ponorogo.	Shefira yosi meyharani (2017)	K-Means Clustering	Dengan menggunakan algoritma K-Means Clustering dapat memberikan hasil terhadap prediksi tingkat pendidikan remaja.
2 Prediksi penyewaan alat teater yakuzi di ponorogo	Yuda dwi hardianto (2018)	Fuzzy K-means	Penggunaan Fuzzy K-Means memberikan strategi terhadap penyewaan alat teater yakuzi dengan baik.
3 Penerapan Metode Naïve Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra	Chairul Fadlan Selfia Ningsih, Agus Perdana Windarto, (2018)	Naïve Bayes	Algoritma Naïve Byes memberikan prediksi terhadap kelayakan keluarga penerima beras rasta berdasarkan kriteria yang ditentukan.

#### 2.2 Program Keluarga Harapan (PKH)

##### a. Pengertian Program Keluarga Harapan (PKH)

Salah satu program bantuan kepada masyarakat kurang mampu adalah bantuan Program Keluarga Harapan yang merupakan salah satu bentuk bantuan yang ditujukan untuk memperkecil angka kemiskinan atau program bantuan bersyarat yang disalurkan untuk keluarga miskin yang diberlakukan sejak tahun 2007 oleh pemerintah Indonesia. Dengan fasilitas untuk layanan kesehatan dan layanan pendidikan, selain itu pula bantuan ini ditujukan untuk ibu hamil, anak balita dan juga anggota keluarga berusia lanjut dan penyandang disabilitas. [11]

Dengan adanya bantuan PKH ini keluarga kurang mampu dapat mendapatkan kehidupan yang lebih layak dengan pelayanan sosial dari pendidikan, kesehatan, gizi dan lain sebagainya. Bantuan ini bertujuan untuk menjadi penanggulangan angka kemiskinan Negara Indonesia yang semakin hari semakin naik dengan cara memberikan perlindungan sosial dan pemberdayaanya [11].



Gambar 2.1 Jumlah Penerima Bantuan PKH Tahun 2007 s.d. 2018

Sumber : Kemensos.go.id

#### b. Jenis Bantuan PKH

Terdapat dua jenis bantuan tetap pada program pkh, antara lain adalah:

1. Bantuan Reguler dengan anggaran Rp. 550.000. satu keluarga per-tahun
2. Bantuan PKH AKSES dengan anggaran Rp. 1.000.000. satu keluarga per-tahun

Selain bantuan tetap juga terdapat jenis bantuan dalam setiap jiwa yang terdapat dalam keluarga penerima bantuan, antara lain:

1. Ibu hamil = Rp. 2.400.000.
2. Anak usia dini = Rp. 2.400.000.
3. SD = Rp. 900.000.
4. SMP = Rp. 1.500.000.
5. SMA = Rp. 2.000.000.
6. Disabilitas berat = Rp. 2.400.000.
7. Lanjut usia = Rp. 2.400.000.

### 1.3 Data Mining

#### a. Pengertian Data Mining.

*Data mining* salah satu bidang dalam ilmu komputer yang menggunakan pembelajaran komputer sebagai media untuk menganalisis dan mendapatkan salah satu pengetahuan secara otomatis, Data mining sering diartikan sebagai model pembelajaran berbasis induksi (*induction-based learning*) dengan menggunakan cara observasi untuk mendapatkan contoh spesifik dan konsep yang untuk dipelajari. *Knowledge discovery in databases* (KDD) merupakan sebuah penerapan di dalam data mining. [12]

Data mining adalah sebuah metode yang digunakan dalam pencarian suatu pola atau model dengan tujuan untuk menyempurnakan sebuah pengetahuan yang bermanfaat dalam sebuah big data. Data mining memiliki cara untuk pengambilan keputusan dengan pencarian pola yang terdapat pada sekumpulan data dalam sebuah database yang digunakan dengan tujuan untuk membantu dalam pengambilan keputusan di masa yang akan datang yang digunakan untuk suatu analisa yang berguna dan memiliki wawasan untuk dipelajari lebih lanjut untuk pendukung keputusan lainnya. [13]

#### b. Teknik Data Mining.

##### 1. Klasifikasi

Klasifikasi merupakan sebuah cara untuk menentukan sebuah data baru dengan menggunakan data record yang telah didefinisikan

sebelumnya hal ini biasa disebut dengan *supervised learning*. [14]  
Pengaplikasian teknik ini antara lain adalah sebagai berikut :

- a. Penjualan langsung (*direct marketing*)
- b. Fraud detection

Tujuan: Menganalisa kasus kecurangan dalam sebuah transaksi kartu kredit.

- c. Customer attrition/churn

Tujuan: mencari peluang pelanggan keluar dari sebuah produk.

## 2. Regresi

Teknik Regresi merupakan salah satu teknik data mining yang digunakan untuk memprediksi nilai sebuah variable dengan nilai variable lain, istilah lain dari teknik ini adalah jaringan syaraf tiruan (*neural network*). [14] Contoh dalam penerapannya:

- a. Digunakan dalam prediksi jumlah sebuah produk.
- b. Prediksi tekanan udara, suhu dan sebagainya
- c. Memprediksi stok sebuah produk.

## 3. Kaidah Asosiasi (*Association Rules*)

Kaidah asosiasi merupakan teknik data mining yang dimana proses kerjanya menggunakan kumpulan dari atribut-atribut yang sering muncul dalam suatu kasus. [14]

Contoh penerapannya antara lain:

- a. Digunakan dalam pemasaran dan promosi penjualan.
- b. Pengelolaan barang supermarket

Tujuan: digunakan untuk mengenali barang yang laris dipasaran.

- c. Mengelola persediaan dalam sebuah perusahaan.

## 4. Pencarian Pola Skekuensial (*Sequence Mining*)

Pencarian Pola Skekuensial atau biasa disebut dengan *Sequence Mining* merupakan teknik data mining dengan cara kerja mencari

kejadian yang terjadi bersamaan dalam sebuah kasus, contohnya dalam urutan DNA. [14]

## 2.4 Naïve Bayes

### a. Pengertian Naïve Bayes

Naïve Bayes adalah salah satu algoritma *supervised learning* metode klasifikasi yang merupakan turunan dari Teorama Bayes. Algoritma ini kurang lebih ditemukan pada pertengahan abad ke-18, yang ditemukan oleh Reverend Thomas Bayes, akan tetapi banyak yang mengatakan bahwa teorama bayes telak ditemukan orang lain terlebih dahulu oleh Reverend Thomas Bayes. [15]

Algoitma naïve bayes ini biasa dikenal sebagai salah satu metode yang digunakan untuk pengelompokkan teks dan pengkategorian dengan cara menggunakan frekuensi kata-kata dari sebuah data. Ciri dari algoritma ini yaitu “bayes” dimana algoritma ini mengansumsikan setiap variabelnya tidak memiliki hubungan antara satu dengan yang lain yang dapat mempengaruhi dalam proses untuk memperoleh hasilnya. [16]

Algoritma ini sering digunakan dalam kasus-kasus yang dalam pemecahanya menggunakan metode probabilitas dan statistik, salah satu ciri dari penggunaan metode ini adalah dimana algoritma ini mencari sebuah peluang yang dimana didapatkan dari pembelajaran kejadian di masa lalu. [17]

Algoritma Naïve Bayes menggunakan dua bentuk data dalam memulai prediksinya, dimana terdapat sebuah dataset dan datatest. Dataset adalah dimana sekumpulan data sebagai data latih yang digunakan untuk menentukan peluang yang akan terjadi, sedangkan datatest adalah sekumpulan data yang akan di prediksi dengan datalatih yang sebelumnya telah dibuat dengan perhitungan probabilitas dan statistic. [18]

Urutan kerja dari algoritma ini dimulai dari pembacaan data training atau dataset, dimana data ini digunakan sebagai patokan dalam perhitungan selanjutnya. Kemudian menghitung jumlah dan

probabilitas dalam data set, akan tetapi berbeda dengan data yang berjenis numerik. Tahap selanjutnya akan memberikan data dalam tabel mean, standart deviasi dan nilai probabilitas kemudian yang terakhir adalah menampilkan sebuah solusi yang didapatkan dari proses perhitungan. [18]

Rumus dalam Algoritma Naiva bayes adalah sebagai berikut

$$[19]: P(H|X) = \frac{P(X|H) P(H)}{P(X)}$$

Dimana :

X = Data dengan class yang belum diketahui.

H = Hipotesis Data X merupakan suatu class spesifik

$P(H|X)$  = probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

$P(H)$  = Probabilitas hipotesis H (prior prob.)

$P(X|H)$  = probabilitas X berdasarkan kondisi tersebut

$P(X)$  = probabilitas dari X Kegunaan Algoritma Naïve Bayes

Contoh dari penggunaan Algoritma Naïve Bayes antara lain:

1. Mengklasifikasi sebuah data atau dokumen untuk keperluan tertentu, atau pun untuk mengklasifikasi sebuah berita.
  2. Digunakan untuk pembelajaran komputer yang menggunakan teknik probabilitas.
  3. Membuat sebuah diagnosa dalam dunia medis.
  4. Menyaring berita hoax, spam dan sebagainya.
- b. Karakteristik Naïve Bayes

Metode klasifikasi *Naive Bayes* memiliki cara kerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas. Hal ini memberikan karakteristik *Naive Bayes* sebagai berikut:[20]

1. Metode *Naive Bayes* bekerja teguh (robust) terhadap data-data yang terisolasi yang biasanya merupakan data dengan

karakteristik berbeda (outliner). Naive Bayes juga bisa menangani nilai atribut yang salah dengan mengabaikan data latih selama proses pembangunan model dan prediksi.

2. Tangguh menghadapi atribut yang tidak relevan.
3. Atribut yang mempunyai korelasi bisa mendegradasi kinerja klasifikasi *Naïve Bayes* karena asumsi independensi atribut tersebut sudah tidak ada.

c. Kelebihan Algoritma Naïve Bayes

1. Alur kerja dari algoritma yang mudah dipahami oleh pengguna.
2. Memiliki struktur kode yang tidak rumit / sederhana.
3. Memiliki rentang waktu pendek dalam perhitungan.
4. Dapat menangani masalah dengan data kuantitatif maupun data diskrit.
5. Memiliki titik noise yang kokoh dan kuat terhadap atribut yang tidak relevan.
6. Algoritma naïve bayes hanya memerlukan data kecil untuk melakukan proses klasifikasi.
7. Nilai yang hilang tidak termasuk dalam perhitungan dalam prediksi.

d. Kekurangan Algoritma Naïve Bayes

1. Perhitungan algoritma yang mengasumsikan variable secara bebas.
2. Perhitungan pada prediksi akan bermasalah jika terdapat 0 dalam nilai probabilitas, sehingga hasil prediksi juga akan bernilai 0.

## 2.5 Bahasa R

a. Pengertian

Bahasa R adalah sebuah bahasa pemrograman atau sistem perangkat lunak yang dimana pengaplikasiannya sering digunakan untuk menangani masalah-masalah untuk memecahkan terkait statistika, Bahasa R pertama kali dikembangkan oleh Ross Ihaka dan Robert Gentleman di Auckland University New Zealand pada tahun

1993. [21] Bahasa ini terung berkembang pesat seiring dengan populernya sekumpulan data besar atau juga disebut “*Big Data*”, selain itu bahasa ini juga banyak digunakan untuk kebutuhan analisa data pada suatu perusahaan tertentu oleh *data scientist* untuk memecahkan suatu permasalahan yang di hadapi. [22]

Rank	Language	Type	Score
1	Python	🌐 🗨️ ⚙️	100.0
2	Java	🌐 📱 🗨️	96.3
3	C	📱 🗨️ ⚙️	94.4
4	C++	📱 🗨️ ⚙️	87.5
5	R	🗨️	81.5
6	JavaScript	🌐	79.4
7	C#	🌐 📱 🗨️ ⚙️	74.5
8	Matlab	🗨️	70.6
9	Swift	📱 🗨️	69.1
10	Go	🌐 🗨️	68.0

Gambar 2.2 bahasa R masuk dalam 10 besar bahasa pemrograman populler 2019

(Sumber : <https://spectrum.ieee.org/>)

Dari data diatas bahasa R merupakan bahasa yang masuk dalam 10 besar bahasa pemograman yang diterbitkan oleh *IEEE Spectrum ranks languages* pada tahun 2019, Bahasa R masuk dalam posisi 5 dalam perengkingan tersebut.

b. Kelebihan Dan Kekurangan

Kelebihan:



- a. R merupakan bahasa pemrograman yang dapat berjalan di berbagai Sistem Operasi antara lain: Windows, Linux dan MacOS X. Kelebihan ini menjamin fleksibilitas pemanfaatan program yang dibuat berdasar bahasa R.
- b. Bahasa R sangat mudah untuk dikembangkan secara individu, selain R ini merupakan salah satu bahasa pemrograman dalam bahasa R juga terdapat banyak package yang banyak dibutuhkan untuk penelitian-penelitian tertentu.
- c. Komunitas pengguna bahasa ini mengembangkan *package* statistika secara berlanjut dan konsisten sehingga dapat membangun jaringan pada setiap user bahasa ini.
- d. Project-project terdokumentasi dengan baik pada situs resmi bahasa R yaitu <http://www.r-project.org/> sehingga dapat memudahkan user pada saat pengoprasian bahasa ini.
- e. Kemampuan Bahasa R untuk menampilkan grafis dengan hasil keakuratan yang baik.

Kekurangan:

- a. Terdapat banyak command window pada bahasa R.
- b. Pengendalian dalam penyusunan variabel yang kurang terstruktur.
- c. Terjadi kemungkinan banyak perulangan dalam script.
- d. Fungsi dalam bahasa ini berjalan berdasarkan prosedur yang di perintahkan.
- e. Untuk memberi dan mengganti nama pada variable harus menggunakan langkah-langkah yang rumit.
- f. R tidak bisa digunakan untuk menganalisa multivariable.
- g. Kemampuan yang rendah untuk memilih himpunan variable pada suatu kasus.
- h. Kurangnya GUI (*Graphical Users Interface*)

## 2.6 R studio



Gambar 2.3 R studio

RStudio adalah sebuah (*IDE*) khusus bagi bahasa pemrograman R, didalam Rstudio terdapat beberapa fitur, antara lain: *R console*, *code editor* dengan *syntax highlighting*, *code completion* dan *direct execution*, *environment*, *history*, *connections*, dan terdapat fitur-fitur tambahan lainnya seperti *file manager*, *packages manager*, *help*, *plot viewer*, hingga *project versioning* menggunakan git. RStudio memiliki dua versi, yaitu *open source* (gratis) dan *commercial edition* (berbayar). RStudio juga tidak hanya terbatas dalam bentuk aplikasi dekstop, melainkan terdapat versi RStudio Server, yaitu RStudio yang dapat diakses melalui browser yang terhubung dengan suatu jaringan komputer.[23]

## 2.7 Flowchart

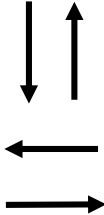









### a. Pengertian Flowchart

*Flowchart* atau kata lain disebut dengan diagram alir merupakan sebuah symbol grafis yang menggambarkan urutan sebuah algoritma atau langkah kerja maupun proses berjalanya langkah-langkah untuk memecahkan suatu masalah, *Flowchart* memiliki alur kerja langkah demi langkah untuk memecahkan suatu masalah dengan proses yang telah dibuat sebelumnya. [24]

Flowchart digunakan untuk menyusun alur kerja suatu program atau penelitian agar memiliki tahapan yang sesuai dengan yang direncanakan. Flowchart sering diartikan sebagai suatu alat atau sarana yang menunjukkan langkah-langkah yang harus dilaksanakan dalam menyelesaikan suatu permasalahan dengan membuat proses-proses dengan symbol. [25]

b. Simbol – Simbol flowchart

**Tabel 2.2 Simbol Flowchart**

Simbol Flowchart	Fungsi	Simbol Flowchart	Fungsi
	<b>Flow Direction symbol</b> digunakan untuk menghubungkan symbol-symbol pada flowchart		<b>Simbol Manual Input</b> Digunakan sebagai penanda input.
	<b>Terminator Symbol</b> Digunakan sebagai penanda Start dan Stop pada sebuah kasus.		<b>Simbol Preparation</b> Digunakan untuk menggambarkan persiapan.
	<b>Connector Symbol</b> Menggambarkan proses keluar masuk dari symbol-symbol yang dibuat.		<b>Simbol Predefine Proses</b> Digunakan untuk menggambarkan jalanya sub bagian.
	<b>Connector Symbol</b> Simbol keluar masuk atau penggabungan halaman kerja.		<b>Simbol Display</b> Digunakan untuk menyatakan output.
	<b>Processing Symbol</b> Digunakan untuk menampilkan pengolahan pada komputer		<b>Simbol disk</b> Digunakan untuk menggambarkan input atau output dari sebuah penyimpanan.

	<b>Simbol Manual Operation</b>		<b>Simbol Magnet tape unit</b>
	Digunakan untuk menampilkan proses yang tidak dilakukan komputer.		Digunakan untuk mempresentasikan inputan yang berasal dari pita magnetic
	<b>Simbol Decision</b>		<b>Simbol punch card</b>
	Digunakan untuk menggunakan kondisi.		Menyatakan sebuah input yang berasal dari kartu atau output ditulis ke kartu.
	<b>Simbol Input-Output</b>		<b>Simbol Dokumen</b>
	Digunakan untuk menyatakan proses input dan output.		Digunakan untuk menyatakan inputan yang berasal dari dokumen.

c. Jenis Flowchart

1. Flowchart Sistem

Flowchart sistem menggambarkan uraian atau sebuah langkah-langkah yang menggambarkan proses kerja dari aspek sistem keseluruhan, flowchart ini menjelaskan urutan dari proses yang bekerja di dalam sebuah sistem.

2. Flowchart Dokumen

Flowchart dokumen merupakan sebuah bagan alir yang didalamnya menunjukkan prosesjalanya arus dari laporan dan formulir termasuk semua tembusan-tembusannya didalamnya.

### 3. Flowchart skematik

Bagan alir skematik hampir mirip dengan diagram alir sistem dengan cara kerja menggambarkan prosedur yang berjalan di dalam sistem. Yang membedakan dari diagram alir sistem selain menggunakan simbol-simbol alir sistem juga terdapat gambar komputer dan peralatan penunjang.

### 4. Program Flowchart

Bagan alir program merupakan suatu bagan yang menjelaskan secara rinci tentang langkah dari proses suatu program, yang dibuat dari derivikasi bagan alir sistem yang terdiri dari dua macam, yaitu bagan alir logika program dan bagan alir program komputer terinci.

### 5. Process Flowchart

Bagan alir proses adalah bagan yang biasa digunakan pada teknik industri. Bagan alir yang digunakan untuk menganalisa sistem dan menggambarkan proses dalam suatu prosedur kerja.

