

## IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA NAÏVE BAYES UNTUK KLASIFIKASI PENERIMA PROGRAM INDONESIA PINTAR

<sup>1)</sup> Fahmi Dian Pratama, <sup>2)</sup> Ilka Zufria, <sup>3)</sup> Triase

<sup>1,2,3)</sup> Sistem Informasi, Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara

<sup>1,2,3)</sup> Dusun VI PKS Pagar Merbau – Lubuk Pakam - Indonesia

E-mail : [Fahmi.dian@uinsu.ac.id](mailto:Fahmi.dian@uinsu.ac.id), [ilkazufria@uinsu.ac.id](mailto:ilkazufria@uinsu.ac.id), [triase@uinsu.ac.id](mailto:triase@uinsu.ac.id)

### ABSTRAK

Peran pendidikan sangat penting dalam rangka meningkatkan kecerdasan sumber daya manusia baik secara intelektual, emosional maupun spiritual. Namun seringkali pendidikan tidak berjalan dengan baik karena kurangnya faktor ekonomi yang mengakibatkan banyak anak putus sekolah. Untuk mengatasi hal tersebut, pemerintah telah membuat sebuah program agar masyarakat yang kurang mampu tetap dapat bersekolah secara gratis. Program ini dinamakan Program Indonesia Pintar (PIP). Komputerisasi pada proses pemberian bantuan ini tak dapat dihindarkan. Agar tidak terjadi kesalahan dalam menentukan penerima bantuan maka diperlukan penerapan data mining menggunakan algoritma Naïve Bayes yang dapat mengklasifikasikan tingkat kelayakan masyarakat dalam menerima bantuan sehingga diperoleh hasil yang lebih akurat dalam menentukan penerima bantuan Program Indonesia Pintar. Hasil penelitian ini diharapkan terciptanya sebuah sistem data mining yang mampu mendapatkan hasil yang akurat dalam menentukan penerima Program Indonesia Pintar.

**Kata Kunci:** Program Indonesia Pintar, Algoritma, Naïve Bayes.

### ABSTRACT

*The role of education is very important in order to increase intelligent human resources both intellectually, emotionally and spiritually. But often education does not run properly due to a lack of economic factors that result in many children dropping out of school. To overcome this, the government has created a program so that the underprivileged can still go to school for free. This program is called Program Indonesia Pintar (PIP). Computerization in the process of providing assistance is unavoidable. In order to avoid errors in determining beneficiaries, it is necessary to apply data mining using the Naïve Bayes algorithm which can classify the level of the community eligibility in receiving assistance so as to obtain more accurate results in determining the recipients of Program Indonesia Pintar. The results of this research are expected to create a data mining system that is able to get accurate results in determining the recipients of Program Indonesia Pintar.*

**Keyword:** Program Indonesia Pintar, Algorithm, Naïve Bayes

### PENDAHULUAN

Pendidikan tinggi memiliki potensi yang sangat penting guna menambah sumber daya manusia yang berkualitas[1]. Pendidikan merupakan wadah yang digunakan untuk membentuk sumberdaya manusia yang cerdas baik secara emosional, intelektual, maupun spiritual . Namun seringkali proses pendidikan tidak berjalan lancar dikarenakan faktor ekonomi, untuk itu perlunya campur tangan pemerintah agar masalah tersebut dapat teratasi dengan membuat sebuah program dimana masyarakat kurang mampu tetap dapat bersekolah dengan gratis, program itu dinamakan Program Indonesia Pintar.

Program Indonesia Pintar

merupakan sebuah program bantuan tunai dari pemerintah untuk masyarakat yang kurang mampu, dimana anak yang berusia 6-21 tahun tetap bisa bersekolah dengan gratis[2].

Data mining atau sering juga disebut penemuan pengetahuan dalam basis data (KDD) merupakan sebuah kegiatan yang meliputi pengumpulan, penggunaan data historis untuk menemukan keteraturan pola dengan jumlah data yang besar. Hasil data mining bisa digunakan dalam membantu pengambilan keputusan di masa depan[3]. Data Mining juga sering disebut campuran data statistic , kecerdasan buatan, dan riset basis data yang masih berkembang [4]. Data Mining bisa digunakan sebagai Analisa data

yang berukuran besar untuk menemukan hubungan yang jelas serta menyimpulkan sesuatu yang belum diketahui sebelumnya [5]. Ada beberapa metode dalam data mining yaitu asosiasi, pengklusteran, prediksi, estimasi dan klasifikasi [6].

Dalam menilai suatu objek data untuk memasukkan ke dalam kelas tertentu dari sejumlah kelas yang tersedia disebut klasifikasi. Dalam klasifikasi terdapat dua pekerjaan utama yang dilakukan yaitu : pertama, membangun sebuah prototype untuk disimpan sebagai memori, kedua penggunaan model tersebut untuk melakukan pengenalan pada suatu objek data lain agar diketahui dikelas mana objek data tersebut dalam model yang disimpan [7]. Proses Klasifikasi merupakan proses penemuan model (fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang tidak diketahui [8]. Klasifikasi ini sendiri mempunyai model yang sama dengan kotak hitam, karena dalam satu model yang menerima masukan bisa melakukan pemikiran terhadap masukan tersebut dan juga bisa memberikan jawaban sebagai keluaran dari hasil pemikiran [9].

Naïve Bayes merupakan sebuah algoritma pengklasifikasian dengan metode probabilitas dan statistik yang ditemukan oleh ilmuwan yang berasal dari Inggris yaitu Thomas Bayes, klasifikasi Bayesian merupakan klasifikasi statistik yang digunakan untuk memprediksi probabilitas atau peluang di setiap anggota pada kelas tertentu dan menghitung peluang yang akan terjadi pada atribut yang ada [10]. Algoritma ini memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal teorema Bayes [5]. Dalam sebuah buku karangan

Muhammad Arhani dan Muhammad Nasir tahun 2020, algoritma Naïve Bayes dapat digunakan untuk masalah klasifikasi biner dan multikelas. Naïve Bayes membuat dan menilai model dengan sangat cepat dan skala secara linier dalam jumlah prediksi dan baris. Ada beberapa hal penting dalam penggunaan Naïve Bayes untuk klasifikasi yaitu

- Sangat mudah untuk dibangun karena tidak memerlukan skema untuk estimasi parameter *iterative* yang rumit dan metode ini dapat langsung diimplementasikan ke dalam jumlah data dalam skala yang besar
- Mudah untuk ditafsirkan sehingga pengguna yang kurang terampil dalam teknik klasifikasi dapat dengan mudah memahami hasil akhir yang diperoleh.

## METODE

Metode pengembangan sistem yang digunakan dalam penelitian ini yaitu metode RAD (*Rapid Application Development*) yang mempunyai beberapa tahapan yaitu *requirement planning*, desain, dan implementasi. Metode RAD merupakan model proses pengembangan perangkat lunak yang bersifat incremental terutama untuk pengerjaan yang pendek.



Gambar 1. Metode RAD

### A. *Requirement Planning* (Perencanaan Kebutuhan

Langkah ini merupakan perencanaan terhadap kebutuhan sistem. Pengumpulan data dalam tahap

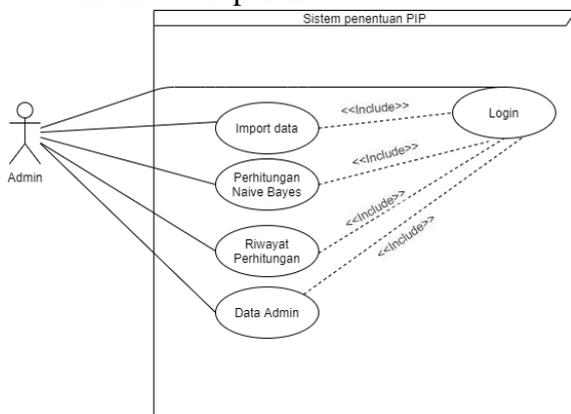
ini bisa melakukan sebuah penelitian, wawancara atau studi literatur. Sistem analis akan menggali informasi sebanyak-banyaknya dari user sehingga akan tercipta sebuah sistem komputer yang bisa melakukan tugas-tugas yang diinginkan oleh user tersebut. Tahapan ini akan menghasilkan dokumen user requirment atau bisa dikatakan sebagai data yang berhubungan dengan keinginan user dalam pembuatan sistem. Dokumen ini lah yang akan menjadi acuan sistem analis untuk menerjemahkan ke dalam bahasa pemrogram.

mengalami kesalahan karena perangkat lunak harus menyesuaikan dengan lingkungan (periperal atau sistem operasi baru) baru, atau karena pelanggan membutuhkan perkembangan fungsional.

### Perancangan Proses

Use Case Diagram memberikan gambaran seluruh elemen sistem. Terdapat satu entitas luar yaitu :

1. Admin ,pengguna sistem yang melakukan semua kegiatan yang ada pada sistem, mulai dari login, import data, perhitungan Naïve Bayes hingga membuat laporan.



Gambar 2. Use Case Diagram

### B. Desain Workshop (Permodelan)

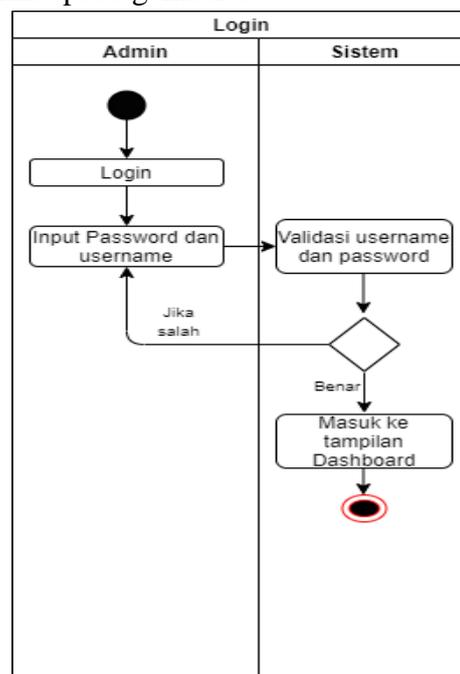
Tahapan dimana dilakukan penuangan pikiran dan perancangan sistem terhadap solusi dari permasalahan yang ada dengan menggunakan perangkat pemodelan sistem seperti alur kerja sistem yang akan digambarkan menggunakan diagram konteks, dan UML(Unified Modelling Language).

### C. Implementasi

Perangkat lunak yang sudah disampaikan kepada pelanggan pasti akan mengalami perubahan. Perubahan tersebut bisa karena

#### 1) Activity Diagram Login

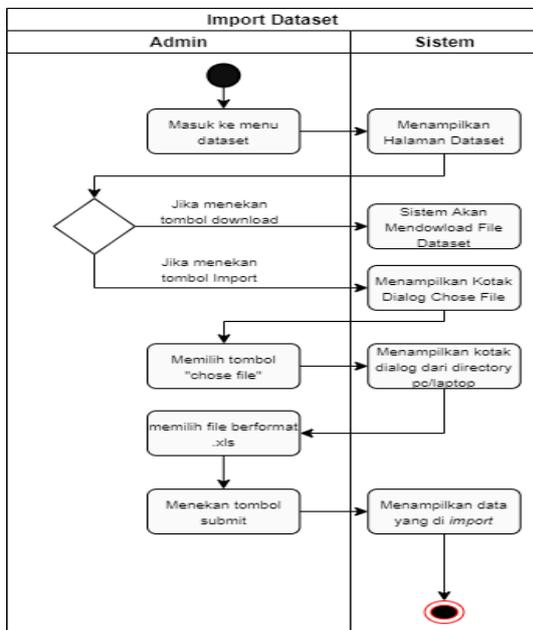
Pada activity diagram dibawah ini kita bisa melihat proses admin melakukan login ke dalam sistem dengan menggunakan username dan password yang tersedia, jika username dan password benar maka akan di alihkan ke dashboard, namun jika username atau password salah akan tetap di form login. Dapat kita lihat pada gambar 3 :



Gambar 3. Activity Login

### 2) Activity Diagram Import Dataset

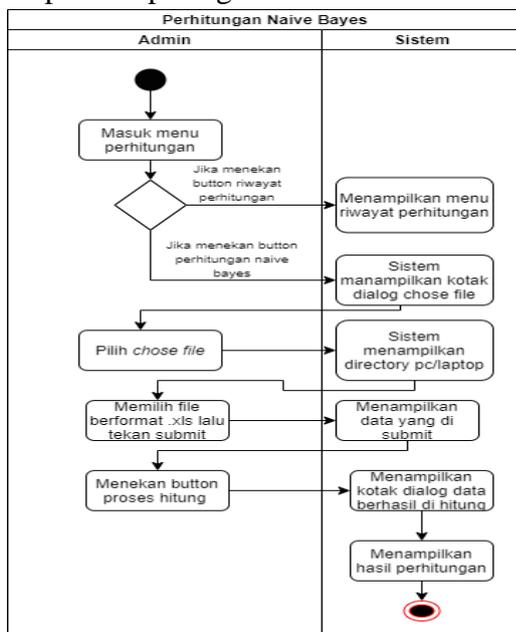
Pada activity ini admin akan melakukan import dataset sebelum melakukan perhitungan Naïve Bayes, dapat kita lihat pada gambar 4 :



Gambar 4. Activity Import Dataset

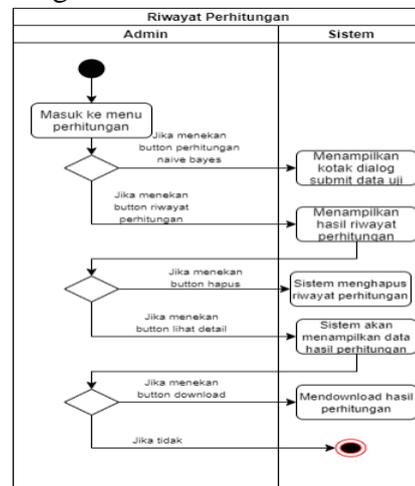
### 3) Activity Diagram Perhitungan Naïve Bayes

Pada activity ini admin akan melakukan perhtungan menggunakan sistem dengan memasukkan data uji dan bantuan algoritma Naïve Bayes untuk melakukan perhitungan. Dapat kita pada gambar 5 :



### 4) Activity Diagram Riwayat Perhitungan

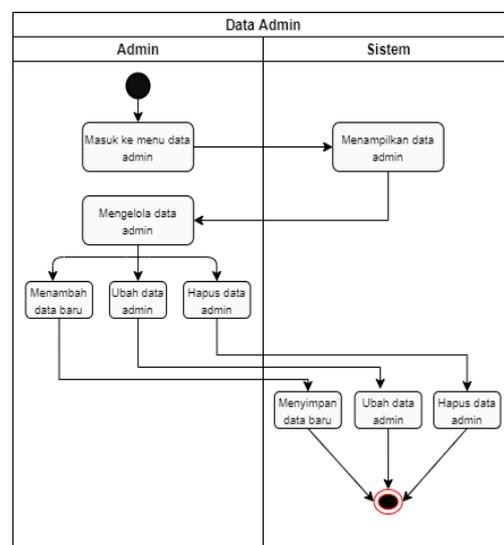
Pada activity ini admin dapat melihat riwayat perhitungan atau histori perhitungan yang pernah dilakukan oleh admin. Dapat kita lihat pada gambar 6 :



Gambar 6. Activity Riwayat Perhitungan

### 5) Activity Diagram Data Admin

Pada activity ini admin dapat menambah, mengubah dan menghapus user yang telah ada di dalam sistem. Dapat kita lihat pada gambar 7 :



Gambar 7. Activity Data Admin

## IMPLEMENTASI

Berdasarkan dari hasil pengumpulan data yang dilakukan oleh peneliti maka di dapatkan 100 data uji namun hanya 6 data yang ditampilkan beserta perhitungan untuk menentukan layak atau tidak layak nya

masyarakat menerima bantuan. Data uji dapat kita lihat pada tabel 1.

Tabel 1. Data Uji

Nama pd	Tgl Lahir	NISN	JK	N.Ayah	N. Ibu	SK TM	KKS	Kondisi Rumah	Penghasilan	Pekerjaan	Tanggungungan
Aditya Azhar	2013-04-06	3138270240	L	Azman	Usrik	Ya	Ya	Papan	1500000	Tukang Becak	3
Aqila Syafira	2012-11-13	3125414663	P	Sahrudin	Rodia	Ya	Ya	Papan	1500000	Tukang Becak	3
Arifin Syahputra	2012-12-11	0122772378	L	Bayu Santoso	Hindun	Ya	Ya	Papan	1500000	Supir Angkot	2
Jildan	2013-03-14	0134429479	L	Rasimin	Setiani	Ya	Ya	Papan	1500000	Supir Angkot	3
Nindi Dwi Akila	2013-03-15	0136747532	P	Jainal Arifin	Lela Ernawati	Ya	Ya	Semi Batu	2500000	Pegawai BUMN	3
Rahmadhani	2012-07-14	3128808708	L	Mahyudin	Arina	Tidak	Ya	Semi Batu	2000000	Pegawai Swasta	4

**Data record SKTM “YA”, KKS “YA”, Kondisi rumah “Papan”, Penghasilan “1500000” dan Tanggungan “3”).**

Penyelesaian :

- Langkah pertama hitung total keseluruhan record yang ada di data training atau data penerima sebelumnya

- $P(C \text{ “Layak”}) = 22/50$
- $P(C \text{ “Tidak Layak”}) = 28/50$

- Selanjutnya hitung  $P(X_k | C_i)$  untuk setiap class  $i$

- $P(\text{Record SKTM “YA”} | \text{Status “Layak”}) = 20/22$
- $P(\text{Record SKTM “YA”} | \text{Status “Tidak layak”}) = 4/28$
- $P(\text{Record KKS “YA”} | \text{Status “Layak”}) = 22/22$
- $P(\text{Record KKS “YA”} | \text{Status “Tidak layak”}) = 13/28$
- $P(\text{Kondisi rumah “Papan”} | \text{Status “Layak”}) = 10/22$

$$f) P(\text{Kondisi rumah “Papan”} | \text{Status “Tidak layak”}) = 10/28$$

$$g) P(\text{Penghasilan “1500000”} | \text{Status “Layak”}) = 10/22$$

$$h) P(\text{Penghasilan “1500000”} | \text{Status “Tidak layak”}) = 0/28$$

$$i) P(\text{Tanggungan “3”} | \text{Status “Layak”}) = 7/22$$

$$j) P(\text{Tanggungan “3”} | \text{Status “Tidak layak”}) = 6/28$$

- Dari hasil diatas langkah terakhir yaitu menghitung class layak dan tidak layak dengan cara :

$$P(X|C_i) = \prod_{k=1}^n P(X_k | C_i) * P(C_i)$$

$$P(\text{Layak}) = (20/22 * 22/22 * 10/22 * 10/22 * 7/22) * 22/50 = 0,0263$$

$$P(\text{Tidaklayak}) = (4/28 * 13/28 * 10/28 * 0/28 * 6/28) * 28/50 = 0$$

Dari hasil  $P(\text{Layak})$  dan  $P(\text{Tidak layak})$  maka dibandingkan dan mendapatkan hasil “Layak” (**Data record SKTM “YA”, KKS “YA”,**

**Kondisi rumah “Papan”, Penghasilan “150000” dan Tanggungan “2”).**

Penyelesaian :

Dikarenakan kondisi sama kecuali tanggungan, maka di hitung tanggungan :

$$P(\text{Tanggungan “2”} | \text{Hasil “Layak”}) = 2/22$$

$$P(\text{Tanggungan “2”} | \text{Hasil “Tidak layak”}) = 7/28$$

Jadi hasil P(Layak) adalah :

$$P(\text{Layak}) = (20/22 * 22/22 * 10/22 * 10/22 * 2/22) * 2/50 = 0,0075$$

$$P(\text{Tidaklayak}) =$$

$$(4/28 * 13/28 * 10/28 * 0/28 * 7/28) * 28/50 = 0$$

Maka jika dibandingkan di dapat hasil “Layak”.

**(Data record SKTM “YA”, KKS “YA”, Kondisi rumah “Semi Batu”, Penghasilan “250000” dan Tanggungan “3”).**

Penyelesaian :

$$a) P(\text{Record SKTM “YA”} | \text{Status “Layak”}) = 20/22$$

$$b) P(\text{Record SKTM “YA”} | \text{Status “Tidak layak”}) = 4/28$$

$$28/50 = 0,0004$$

Dari hasil diatas jika dibandingkan akan mendapatkan hasil “Layak”

**(Data record SKTM “Tidak”, KKS “YA”, Kondisi rumah “Semi Batu”, Penghasilan “200000” dan Tanggungan “4”).**

Penyelesaian :

$$a) P(\text{Record SKTM “Tidak”} | \text{Status “Layak”}) = 2/22$$

$$b) P(\text{Record SKTM “Tidak”} | \text{Status “Tidak layak”}) = 24/28$$

$$c) P(\text{Record KKS “YA”} | \text{Status “Layak”}) = 22/22$$

$$d) P(\text{Record KKS “YA”} | \text{Status “Tidak layak”}) = 13/28$$

$$e) P(\text{Kondisi rumah “Semi Batu”} | \text{Status “Layak”}) = 2/22$$

$$c) P(\text{Record KKS “YA”} | \text{Status “Layak”}) = 22/22$$

$$d) P(\text{Record KKS “YA”} | \text{Status “Tidak layak”}) = 13/28$$

$$e) P(\text{Kondisi rumah “Semi Batu”} | \text{Status “Layak”}) = 2/22$$

$$f) P(\text{Kondisi rumah “Semi Batu”} | \text{Status “Tidak layak”}) = 5/28$$

$$g) P(\text{Penghasilan “2500000”} | \text{Status “Layak”}) = 2/22$$

$$h) P(\text{Penghasilan “2500000”} | \text{Status “Tidak layak”}) = 9/28$$

$$i) P(\text{Tanggungan “3”} | \text{Status “Layak”}) = 7/22$$

$$j) P(\text{Tanggungan “3”} | \text{Status “Tidak layak”}) = 6/28$$

Jadi hasil untuk :

$$P(\text{Layak}) = (20/22 * 22/22 * 2/22 * 2/22 * 7/22) * 22/50 = 0,0010$$

$$P(\text{Tidaklayak}) = (4/28 * 13/28 * 5/28 * 9/28 * 6/28) *$$

$$f) P(\text{Kondisi rumah “Semi Batu”} | \text{Status “Tidak layak”}) = 5/28$$

$$g) P(\text{Penghasilan “2000000”} | \text{Status “Layak”}) = 10/22$$

$$h) P(\text{Penghasilan “2000000”} | \text{Status “Tidak layak”}) = 14/28$$

$$i) P(\text{Tanggungan “4”} | \text{Status “Layak”}) = 12/22$$

$$j) P(\text{Tanggungan “4”} | \text{Status “Tidak layak”}) = 10/28$$

Jadi hasil untuk :

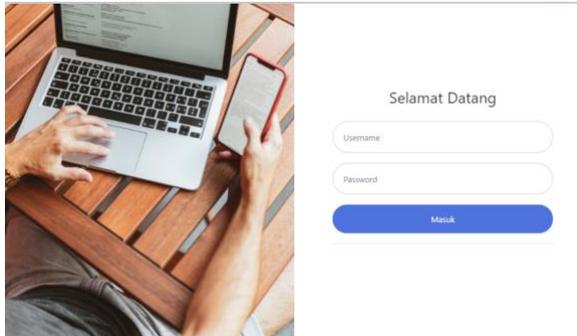
$$P(\text{Layak}) = (2/22 * 22/22 * 2/22 * 10/22 * 12/22) * 22/50 = 0,0009$$

$$P(\text{Tidaklayak}) = (24/28 * 13/28 * 5/28 * 14/28 * 10/28) * 28/50 = 0,0071$$

Dari hasil diatas jika dibandingkan kita akan mendapatkan hasil “Tidak Layak”

### Implementasi Tampilan Login

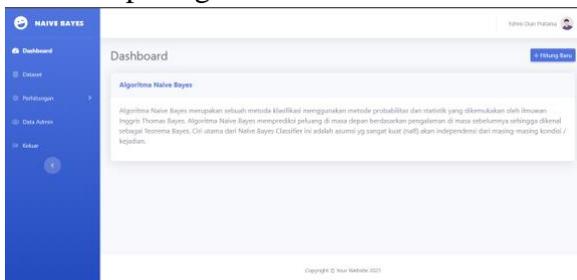
sistem sebelum kita masuk ke menu utama, pada menu login admin akan diarahkan untuk mengisi username dan password yang telah tersedia. Tampilan login dapat dilihat pada gambar 8.



Gambar 8 Halaman Login

### Implementasi Tampilan Dashboard

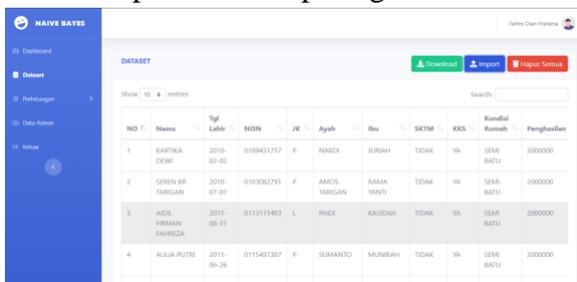
Tampilan dashboard merupakan menu utama yang ada pada sistem setelah admin melakukan login. Tampilan dashboard dapat kita lihat pada gambar 9.



Gambar 9. Halaman Dashboard

### Implementasi Tampilan Menu Dataset

Pada menu ini terdapat data-data penerima bantuan sebelumnya yang akan dijadikan acuan dalam perhitungan. Tampilan menu dataset dapat kita lihat pada gambar 10.

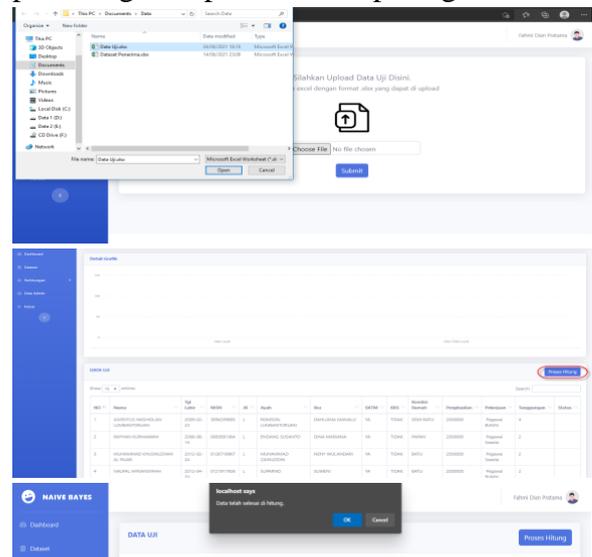


Gambar 10. Halaman Menu Dataset

### Implementasi Tampilan Menu Perhitungan

Menu perhitungan merupakan menu yang

Tampilan login merupakan tampilan awal disediakan untuk melakukan perhitungan Naïve Bayes dengan *import file* data uji yang telah tersedia, setelah file berhasil di submit admin melakukan perhitungan dengan menekan button proses hitung kemudian sistem akan mulai menghitung sampai sistem mengeluarkan kotak dialog data berhasil dihitung dan sistem akan menampilkan hasil perhitungannya. Tampilan menu perhitungan dapat kita lihat pada gambar 11.



Gambar 11. Tampilan Menu Perhitungan

### Implementasi Tampilan Menu Riwayat Perhitungan

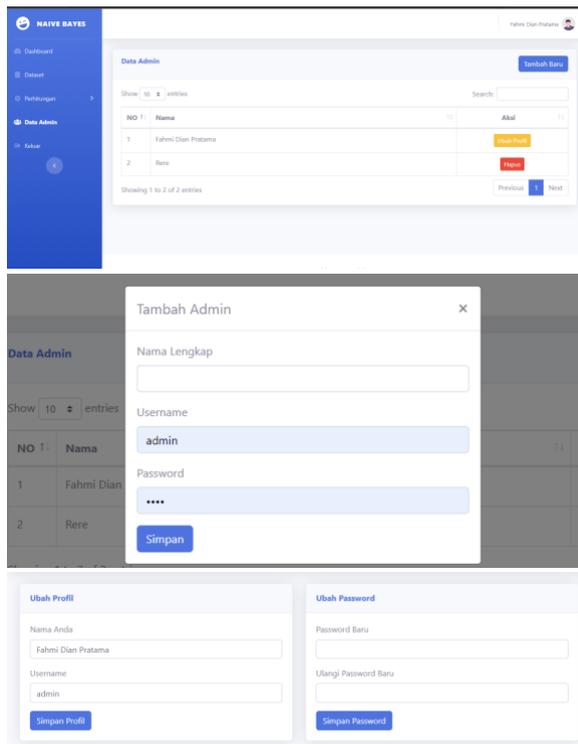
Pada tampilan ini dapat digunakan untuk melihat kembali data yang telah berhasil di hitung jika admin telah keluar dari sistem. Tampilan menu riwayat perhitungan dapat kita lihat pada gambar 12.



Gambar 12. Tampilan Menu Riwayat Perhitungan.

### Implementasi Tampilan Menu Admin

Pada menu ini admin bisa menambah, mengubah dan menghapus admin yang telah tersedia. Tampilan menu data admin dapat kita lihat pada gambar 13.



Gambar 13. Tampilan Menu Data Admin

## KESIMPULAN

Dalam mengklasifikasikan kelayakan penerima Program Indonesia Pintar menggunakan algoritma Naïve Bayes digunakan 5 kriteria. Peneliti menilai ini sudah kompleks. Dan harapan nya dapat mempermudah pihak instansi dalam menentukan masyarakat yang layak untuk menerima bantuan dari pemerintah.

## DAFTAR PUSTAKA

- [1] I. Ruhana, “PENGEMBANGAN KUALITAS SUMBER DAYA MANUSIA VS DAYA SAING,” *J. Profit*, vol. 6, no. 1, p. 52, 2012.
- [2] Elyana and A. I. Warnillah, “Sistem Informasi Seleksi Penerima Program Indonesia Pintar dengan Menggunakan Metode Maut,” *CKI Spot*, vol. 11, no. 2, pp. 96–105, 2018.

- [3] A. Ikhwan *et al.*, “KEBUTUHAN DATA MINING UNTUK MEMPROMOSIKAN PENDIDIKAN BERDASARKAN ALGORITMA FP-GROWTH,” vol. 9, pp. 1660–1669, 2018.
- [4] R. Rahim *et al.*, “C4 . 5 Klasifikasi Data Mining untuk Kontrol Inventaris,” *J. Int. Tek. Teknol.*, vol. 7, pp. 68–72, 2018.
- [5] M. Idris, “IMPLEMENTASI DATA MINING DENGAN ALGORITMA NAÏVE BAYES UNTUK MEMPREDIKSI ANGKA KELAHIRAN,” *Pelita Inform.*, vol. 18, no. 1, pp. 160–167, 2019.
- [6] D. Nofriansyah, K. Erwansyah, and M. Ramadhan, “Penerapan Data Mining dengan Algoritma Naive Bayes Classifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL ( Studi Kasus di,” *SAINTIKOM*, vol. 15, no. 2, pp. 81–92, 2016.
- [7] H. M. M. Al Laroussi, “Implementasi algoritma naïve bayes sebagai proses seleksi penerima beasiswa libyan embassy berbasis web,” 2015.
- [8] H. Annur, “KLASIFIKASI MASYARAKAT MISKIN MENGGUNAKAN METODE,” *ILKOM*, vol. 10, no. 2, pp. 160–165, 2018.
- [9] A. Lasarudin and Purwanto, “KLASIFIKASI PENGADUAN MASYARAKAT MENGGUNAKAN NAIVE,” *J. Teknol. Inf.*, vol. 14, no. 2, p. 63, 2018.
- [10] Triase and Samsudin, “IMPLEMENTASI DATA MINING DALAM MENGLASIFIKASIKAN UKT ( UANG KULIAH TUNGGAL ) PADA UIN SUMATERA UTARA MEDAN,” *J. Teknol. Inf.*, vol. 4, no. 2, pp. 370–376, 2020.