

IMPLEMENTASI DATA MINING DENGAN ALGORITMA NAIVE BAYES UNTUK KLASIFIKASI KELAYAKAN PENERIMA BANTUAN SEMBAKO

Ayu Entini Lumban Raja¹, Koko Handoko²

¹mahasiswa Program Studi Teknik Informatika, Universitas Putera Batam

²dosen Program Studi Teknik Informatika, Universitas Putera Batam

email: pb190210031@upbatam.ac.id

ABSTRACT

The percentage of the population receiving assistance was 5.19%, this figure increased compared to March 2022, while the line of aid recipients was recorded in 2022 of 783,730 people per capita per month. The number of poor households in Batu Aji District is Bukit Tempayang Village with 15,857 inhabitants, 300 poor households, 37,531 inhabitants in Buliang Village, 289 poor households, 28,693 The beneficiary of assistance is an inability to meet basic needs including food, clothing, education and housing. Thus it is necessary to carry out a strategy in dealing with the level of beneficiaries, namely by providing accurate and targeted data on beneficiaries. Naive Bayes is an algorithm that exists in data mining and is part of the data mining classification technique by using probability and statistical techniques to estimate or predict opportunities that will occur based on previous opportunities, namely there are two feasible and inappropriate classes. The results of the classification that will be carried out later will help in the processing of assistance to help make decisions regarding the classification of determining basic food recipients. And by testing calculations manually and using rapid minner software, you get an accuracy value of 80%

Keywords: Data mining; Naive Bayes method; Classification; Groceries

PENDAHULUAN

Batam memiliki jumlah penduduk sebanyak 1.196.396 jiwa perseptember 2020, namun terdapat penurunan laju pertumbuhan penduduk dalam kurun waktu 2010 hingga 2020 sebesar 2,32%. Dari jumlah keseluruhan penduduk Batam presentase jumlah penduduk penerima bantuan sebanyak 5,19 % angka ini meningkat dibandingkan maret 2022 sementara itu garis penerima bantuan tercatat tahun 2022 sebesar 783.730 jiwa perkapita perbulan. Sehingga penerima bantuan di Batam sudah menjadi persoalan yang cukup serius, ditambah lagi Pulau Batam

memiliki laju pertumbuhan penduduk yang cukup tinggi.

Dari hasil sensus penduduk Kecamatan Batu Aji sebagian besar penduduk pendatang yang nota bene mempunyai *skill* bermacam bidang. Wilayah Batu Aji berkembang dengan industri sebagai wilayah perdagangan dan jasa. Dan juga wilayah kelurahan yang berkembang sebatas perdagangan barang harian yang dikemas dalam usaha toko dan pedagang kecil lainnya. Jumlah rumah tangga miskin di Kecamatan Batu Aji dengan kelurahan Bukit Tempayang penduduk 15.857 jiwa, ruta miskin 300 jiwa, Kelurahan Buliang penduduk 37.531

jiwa, ruta miskin 289 jiwa, Kelurahan Kibling penduduk 28.693 jiwa, ruta miskin 1.057 jiwa, Kelurahan Tanjung Uncang penduduk 28.037 jiwa, ruta miskin 1.039 jiwa.

Penerima bantuan salah satu persoalan yang paling mendasar sehingga harus menjadi pusat perhatian pemerintah dari suatu negara sehingga menemukan kesejahteraan. Penerima bantuan merupakan suatu ketidakmampuan dalam memenuhi kebutuhan dasar termasuk makanan, pakaian, pendidikan dan perumahan. Dengan demikian di perlukan dilakukan sebuah strategi dalam penanggulangan tingkat penerima bantuan yaitu dengan menyediakan data penerima bantuan yang akurat dan tepat sasaran (Damuri et al., 2021). Dengan data yang akurat tersebut pemerintah berupaya agar tingkat penerima bantuan di suatu daerah dapat teratasi, dengan cara memberikan dorongan berupa program-program bantuan sembako yang di tujukan terhadap masyarakat yang mengalami penerima bantuan.

Pemerintah memberikan sembako (sembilan bahan pokok) yang terdiri atas makanan dan minuman yang digunakan untuk kebutuhan sehari-hari setiap bulannya melalui mekanisme perbankan. Namun dalam praktek lapangan pemberian bantuan sembako tidak tepat sasaran, sehingga membutuhkan data yang lebih valid lagi terhadap keluarga yang layak atau tidak layak menerima bantuan sembako. Dengan demikian untuk menentukan kelayakan penerima sembako dapat dengan memanfaatkan teknik data *mining*.

Data *mining* dapat diartikan sebagai sekumpulan proses yang digunakan untuk mengeksplorasi dan mencari nilai

berupa informasi terhadap relasi-relasi kompleks yang selama ini sudah tersimpan dalam basis data (Fajrin & Handoko, 2018). Dengan cara menggali informasi tersebut sehingga berbentuk pola yang menarik dari sumber data besar sehingga mendapatkan data informasi berupa pengetahuan statistik, matematika yang di inginkan dalam tindakan mengambil keputusan sehingga data *mining* dapat membantu tahapan perencanaan dalam memberikan informasi tepat untuk membuat prediksi berdasarkan pengalaman masa lalu untuk kondisi saat ini.

Fungsi prediksi dalam menemukan pola-pola tertentu dari suatu data dan dari pola tersebut dapat di ketahui berbagai variabel. Salah satu metode yang dapat digunakan dalam memprediksi adalah dengan menggunakan algoritma *Naive Bayes*. *Naive Bayes* adalah algoritma yang ada pada *data mining* dan bagian dari teknik klasifikasi *data mining* dengan menggunakan teknik probabilitas dan statistik untuk memperkirakan ataupun memprediksi peluang-peluang yang akan terjadi berdasarkan peluang sebelumnya yaitu terdapat dua kelas layak dan tidak layak. Hasil dari klasifikasi yang akan dilakukan nantinya akan membantu dalam pengolahan bantuan untuk membantu mengambil keputusan terkait klasifikasi penentuan penerima sembako.

KAJIAN TEORI

2.1 Knowledge Discovery In Database

KDD adalah bagian dari ilmu pengetahuan yang mencakup *database* serta dapat diuraikan sebagai proses pengenalan pola kepada pengetahuan yang memiliki sekelompok besar data dan menghasilkan informasi yang mudah

untuk dipahami. KDD memiliki beberapa proses untuk mendapatkan data yang ingin di capai seperti pemurnian data, *integrasi data*, dan pemilihan data dan presentasi pengetahuan (Wijaya & Dwiasnati, 2020)

2.2 Data mining

Data mining merupakan metode yang terdapat pada ilmu komputer yang dapat digunakan dalam proses *knowledge*. Tahapan didalam berguna untuk mencari pola-pola tertentu dari data yang ada pada *database*. Biasanya metode ini banyak ditemukan pada bidang *machine learning* dan statistika. Dengan cara menggali informasi tersebut sehingga berbentuk pola yang menarik dari sumber data besar sehingga mendapatkan data informasi berupa pengetahuan statistik, matematika yang di inginkan dalam tindakan mengambil keputusan sehingga *data mining* dapat membantu tahapan perencanaan dalam memberikan informasi tepat untuk membuat prediksi berdasarkan pengalaman masa lalu untuk kondisi saat ini

2.3 Naive Bayes

Naive Bayes adalah algoritma yang ada pada data *mining* dan bagian dari teknik klasifikasi data *mining* dengan menggunakan teknik probabilitas dan statistik untuk memperkirakan ataupun memprediksi peluang-peluang yang akan terjadi berdasarkan peluang sebelumnya yaitu terdapat dua kelas layak dan tidak layak (Wijaya & Dwiasnati, 2020).

2.4 Klasifikasi

Klasifikasi merupakan menemukan model atau fungsi yang menjelaskan dan membedakan kelas data atau ide agar dapat digunakan untuk memprediksi kelas objek yang label kelasnya tidak

diketahui adalah proses klasifikasi. Klasifikasi pada data *mining* adalah penempatan objek kedalam kategori atau kelas yang telah ditetapkan sebelumnya (Wijaya & Dwiasnati, 2020). Klasifikasi sering digunakan untuk memprediksi kelas.

2.5 Penerima Bantuan

Penerima bantuan adalah salah satu persoalan yang paling mendasar sehingga harus menjadi pusat perhatian pemerintah dari suatu negara sehingga menemukan kesejahteraan. penerima bantuan merupakan suatu ketidakmampuan dalam memenuhi kebutuhan dasar termasuk makanan, pakaian, pendidikan dan perumahan. Dengan demikian di perlukan dilakukan sebuah strategi dalam penanggulangan tingkat penerima bantuan yaitu dengan menyediakan data penerima bantuan yang akurat dan tepat sasaran (Sosial et al., 2023) .

2.6 Software Pendukung

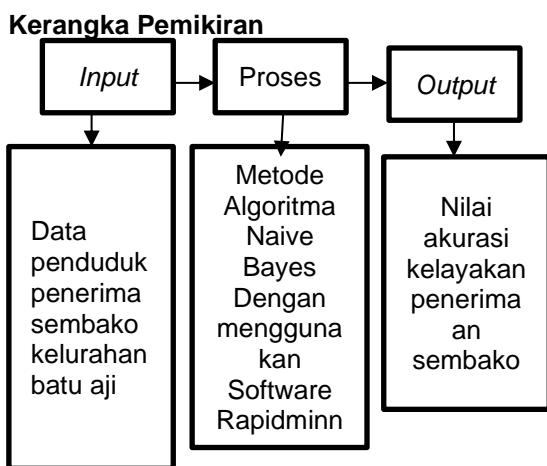
1. Rapidminer

Rapid miner merupakan salah satu cara yang sering digunakan dalam menemukan solusi pada data *mining* karena mampu di jalankan pada berbagai sistem operasi. *Rapid miner* adalah sebuah perangkat lunak yang memiliki sifat *open source*. Sebelum



Gambar 1 *RapidMiner*

Sumber: (Putri et al., 2021)

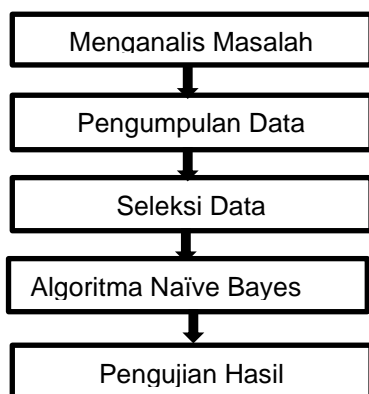


Gambar 2 Kerangka Pemikiran
Sumber Data penelitian 2023.

METODE PENELITIAN

3.1 Desain Penelitian

Desain penelitian adalah gambaran suatu tahap-tahap atau langkah yang digunakan oleh peneliti dalam menyelesaikan hasil penelitiannya. Gambar 3.1 merupakan desain penelitian pada penelitian ini:



Gambar 5 Desain Penelitian

Sumber: Data Penelitian 2023

3.2 Teknik Pengumpulan Data

Merupakan suatu cara yang digunakan oleh peneliti untuk mendapatkan data-data atau informasi mengenai data penelitian.

1. Wawancara

Merupakan teknik yang dilakukan peneliti dengan langsung bertanya kepada narasumber yaitu Bapak kelurahan batu aji dan meminta ijin data penerima sembako di jadikan data penelitian

2. Observasi

Merupakan pengumpulan data yang dilakukan peneliti dengan cara langsung terjun kelapangan yaitu kantor kelurahan batu aji, dan melihat kondisi yang ada didalam.

3. Studi Pustaka

Merupakan cara yang dilakukan peneliti dengan mencari informasi yang berkaitan dengan penelitian yang berasal dari jurnal dan buku-buku.

3.3 Operasional Variabel

Operasional variabel merupakan cara dalam menentukan variabel dan indikator yang memiliki hubungan variabel satu dengan variabel yang lain pada sebuah penelitian. Berikut ini merupakan variabel yang digunakan pada penelitian ini:

1. Kepala Rumah Tangga

Merupakan orang yang bertanggung jawab atas kebutuhan dan keperluan pada keluarga tersebut baik laki-laki atau perempuan

2. Jumlah Penghasilan

Merupakan suatu pendapatan yang diperoleh dari bekerja dengan kategori rendah, sedang dan tinggi.

3. Kondisi Rumah

Merupakan suatu bentuk kediaman yang ditempatin oleh keluarga

HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Penelitian ini menghasilkan nilai akurasi menggunakan metode algoritma naive bayes dalam memprediksi penjualan obat dengan melakukan perhitungan dan pengolahan data secara manual menggunakan *microsoft excel* dan menggunakan *software rapidminer*

1. Seleksi data

Merupakan data yang telah di seleksi guna untuk dilakukan pengujian set yang terdiri atas 4 atribut yaitu Kepala rumah tangga, Kondisi Rumah, Jumlah Penghasilan dan Status Milik pribadi. Data tersebut akan dilakukan pengolahan dengan melakukan proses data *mining*:

2. Analisa proses algoritma

Pada tahap ini dilakukan pemodelan data probability dengan menggunakan algoritma naive bayes. Data yang akan di uji akan dibagi menjadi dua bagian yaitu data *training* dan data *testing* kemudian di analisis kembali menggunakan software pendukung yaitu *rapidminer*. Untuk data *training* berjumlah 25 data record dan *testing* 5 data record.

3. Probabilitas Kepala keluarga

Berdasarkan data item set diatas dengan 25 data record maka di dapat diketahui bahwa terdapat 91 % dengan klasifikasi bantuan layak dan 93 % tidak layak dengan kepala keluarga laki-laki. Terdapat 9 % dengan klasifikasi bantuan sembako layak dan 7 % tidak layak dengan kepala keluarga perempuan.

4. Probabilitas kondisi rumah

Berdasarkan data item set diatas dengan 25 data record maka di dapat diketahui bahwa terdapat 64 % dengan klasifikasi bantuan layak dan 50 % tidak layak dengan kondisi rumah batu permanen. Terdapat 36 % dengan klasifikasi bantuan sembako layak dan 50 % tidak layak dengan kepala kondisi rumah anyam.

5. Probabilitas jumlah penghasilan

Berdasarkan data item set diatas dengan 25 data record maka di dapat diketahui bahwa terdapat 64 % dengan klasifikasi bantuan layak dan 0 % tidak layak dengan jumlah penghasilan rendah. Terdapat 36 % dengan klasifikasi bantuan sembako layak dan 50 % tidak layak dengan jumlah penghasilan sedang, terdapat 0 % dengan klasifikasi bantuan layak dan 50 % tidak layak dengan jumlah penghasilan tinggi.

6. Probabilitas pemilik rumah

Berdasarkan data item set diatas dengan 25 data record maka di dapat diketahui bahwa terdapat 55 % dengan klasifikasi bantuan layak dan 86 % tidak layak dengan status pemilik rumah. Terdapat 45 % dengan klasifikasi bantuan sembako layak dan 14 % tidak layak dengan status pemilik rumah.

7. Probabilitas kondisi rumah

Berdasarkan data item set diatas dengan 25 data record maka di dapat diketahui bahwa terdapat 64 % dengan klasifikasi bantuan layak dan 50 % tidak layak dengan kondisi rumah batu permanen. Terdapat 36 % dengan klasifikasi bantuan sembako layak dan 50 % tidak layak dengan kepala kondisi rumah anyam.

Tabel 1 Data *training*

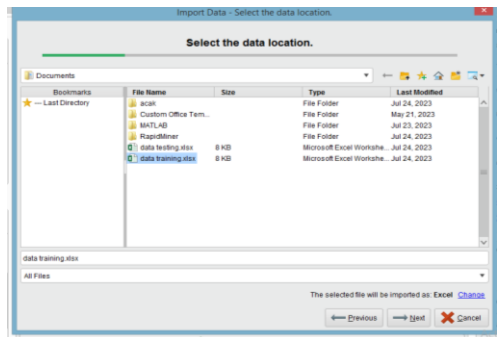
No	Kepala Rumah Tangga	Kondisi Rumah	Jumlah Penghasilan	Status pemilik rumah	Status Kelayakan
1	Laki-laki	Batu permanen	Sedang	Milik Pribadi	Tidak Layak
2	Perempuan	Batu permanen	Rendah	Sewa	Layak
3	Laki-laki	Batu permanen	Sedang	Milik Pribadi	Tidak Layak
4	Laki-laki	Batu permanen	Rendah	Milik Pribadi	Layak
5	Laki-laki	Batu permanen	Rendah	Milik Pribadi	Layak

Sumber: Data Penelitian 2023

4.2 Implementasi rapidminer

1. Import data

Data yang dipilih sebagai data *training* pada penelitian ini yaitu tentang klasifikasi kelayakan penerima bantuan sembako dengan 4 kriteria yaitu kepala keluarga, kondisi rumah, jumlah penghasilan dan status kepemilikan rumah. Maka data yang dipilih sebagai data *training* ada 25 *record* data dan 3 *sample* yang akan dijadikan sebagai data *testing*. Berikut di bawah ini merupakan data *training* dan data *testing* yang telah di import ke dalam software rapidminer.



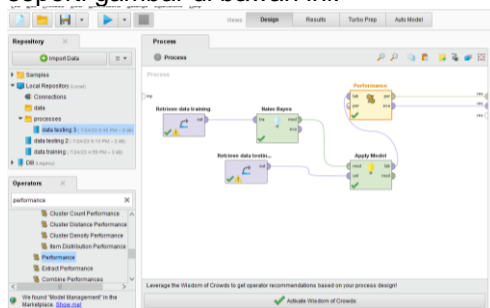
Gambar 2 Import data *training*

Sumber: Data Penelitian 2023

2. Menghubungkan Operator

Setelah langkah-langkah *import* data *training* dan data *testing* dilakukan maka langkah selanjutnya yang akan di lakukan adalah dengan mencari operator naive bayes pada tempat pencarian dengan cara drag and drop operator naive bayes lalu tepatkan pada lembar kerja, setelah

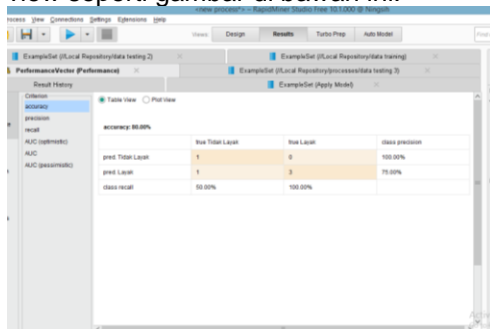
itu lakukan hal serupa pada apply model dan performance. Setelah seluruh operator yang akan di butuhkan ada berada lembar kerja maka selanjutnya langkah yang akan di lakukan adalah dengan menghubungkan setiap operator seperti gambar di bawah ini:



Gambar 3 menghungkan operator
Sumber: Data Penelitian 2023

3. Hasil akhir

Perhitungan yang telah selesai di proses pada perangkat lunak rapid minner dengan menggunakan metode naive bayes untuk klasifikasi kelayakan penerima bantuan sembako yang akan datang maka akan memberikan informasi mengenai data layak dan tidak layak Selanjutnya langkah yang akan di lakukan adalah menentukan tingkat keakuratan atau accuracy dengan cara klik tab *performance vector* yang ada di sebelah kanan maka akan muncul tabel view seperti gambar di bawah ini:



	Real Total Layak	Real Layak	Class precision
pred Total Layak	1	0	100.00%
pred Layak	1	0	75.00%
class recall	50.00%	100.00%	

Gambar 4 hasil akhir
Sumber: Data Penelitian 2023

SIMPULAN

Kesimpulan

Dari penelitian yang telah dilakukan maka dapat di tarik kesimpulan sebagai berikut:

1. Dari data yang telah di dapatkan dan diolah sebanyak 25 data sample maka dilakukan pengolahan data untuk menentukan kelayakan penerimaan bantuan sembako dengan cara data mining menggunakan metode naive bayes.
2. Data penerima bantuan 25 data set tersebut digunakan secara keseluruhan sebagai data penelitian. Dan terdapat 5 data *testing* untuk di jadikan sebagai data pengujian dengan menggunakan metode naive bayes yang dihitung secara manual dan menggunakan software rapid minner.
3. Dari data item set tersebut dengan 25 data record maka di dapat diketahui bahwa terdapat 91 % dengan klasifikasi bantuan layak dan 93 % tidak layak dengan kepala keluarga laki-laki. Terdapat 9 % dengan klasifikasi bantuan sembako layak dan 7 % tidak layak dengan kepala keluarga perempuan. Dan dengan melakukan pengujian perhitungan secara manual dan menggunakan software rapid minner mendapatkan nilai accuracy 80%.

DAFTAR PUSTAKA

- Damuri, A., Riyanto, U., Rusdianto, H., & Aminudin, M. (2021). Implementasi Data Mining dengan Algoritma Naive Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako.

- JURIKOM (Jurnal Riset Komputer)*, 8(6), 219.
<https://doi.org/10.30865/jurikom.v8i6.3655>
- Fajrin, A. A., & Handoko, K. (2018). Penerapan Data Mining Untuk Mengolah Association Rule Tata Letak Buku Dengan Metode. *Jurnal Ilmiah Informatika (JIF)*, 2, 60–65.
- Indra Borman, R., & Wati, M. (2020). Penerapan Data Mining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandar Lampung Dengan Algoritma Naïve Bayes. *Jurnal Ilmiah Fakultas Ilmu Komputer*, 09(01), 25–34.
- Pandie, E. S. Y. (2019). Implementasi Algoritma Data Mining Naive Bayes Pada Koperasi. *J-Icon*, 6(1), 15–20.
- Putri, Ucha, Sanni Irawan, E., & Rizky, F. (2021). Implementasi Data Mining Untuk Prediksi Penjualan Pestisida Pada CV MITRA ARTHA SEJATI Menggunakan Algoritma Naive Bayes. *KESATRIA (Jurnal Penerapan Sistem Informasi Dan Manajemen)*, 2(1), 39–46.
<http://eprosiding.ars.ac.id/index.php/pti>
- Riani, A., Susianto, Y., & Rahman, N. (2019). Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Menggunakan Metode Naive Bayes. *Journal of Innovation Information Technology and Application (JINITA)*, 1(01), 25–34.
<https://doi.org/10.35970/jinita.v1i01.64>
- Sosial, B., Kasus, S., Kemiskinan, K., Lurah, D., & Nurdyawan, O. (2023). IMPLEMENTASI ALGORITMA C4 . 5 UNTUK MENENTUKAN PENERIMA. 7(1), 373–377.
- Surahman, A., Hayati, U., Sosial, B., & Miner, R. (2023). IMPLEMENTASI ALGORITMA NAÏVE BAYES. 7(1), 347–352.
- Tyas, S. J. S., Febianah, M., Solikhah, F., Kamil, A. L., & Arifin, W. A. (2021). Analisis Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan. *Jurnal Teknologi Informasi Dan Komunikasi*, 8(1), 86–99.
<https://jurnal.plb.ac.id/index.php/tematik/article/view/576/365>
- Utomo, D. P., & Purba, B. (2019). Penerapan Datamining pada Data Gempa Bumi Terhadap Potensi Tsunami di Indonesia. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1(September), 846.
<https://doi.org/10.30645/senaris.v1i0.91>
- Wijaya, H. D., & Dwiasnati, S. (2020). Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat. *Jurnal Informatika*, 7(1), 1–7.
<https://doi.org/10.31311/ji.v7i1.6203>
- Damuri, A., Riyanto, U., Rusdianto, H., & Aminudin, M. (2021). Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako. *JURIKOM (Jurnal Riset Komputer)*, 8(6), 219.
<https://doi.org/10.30865/jurikom.v8i6.3655>
- Fajrin, A. A., & Handoko, K. (2018). Penerapan Data Mining Untuk Mengolah Association Rule Tata Letak Buku Dengan Metode. *Jurnal Ilmiah Informatika (JIF)*, 2, 60–65.
- Indra Borman, R., & Wati, M. (2020). Penerapan Data Mining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandar Lampung Dengan Algoritma Naïve Bayes. *Jurnal Ilmiah*

- Fakultas Ilmu Komputer, 09(01), 25–34.*
- Pandie, E. S. Y. (2019). Implementasi Algoritma Data Mining Naive Bayes Pada Koperasi. *J-Icon, 6(1), 15–20.*
- Putri, Ucha, Sanni Irawan, E., & Rizky, F. (2021). Implementasi Data Mining Untuk Prediksi Penjualan Pestisida Pada CV MITRA ARTHA SEJATI Menggunakan ALgoritma Naive Bayes. *KESATRIA(Jurnal Penerapan Sistem Informasi Dan Manajemen, 2(1), 39–46.*
<http://eprosiding.ars.ac.id/index.php/pt>
- Riani, A., Susianto, Y., & Rahman, N. (2019). Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Menggunakan Metode Naive Bayes. *Journal of Innovation Information Technology and Application (JINITA), 1(01), 25–34.*
<https://doi.org/10.35970/jinita.v1i01.64>
- Sosial, B., Kasus, S., Kemiskinan, K., Lurah, D., & Nurdiawan, O. (2023). *IMPLEMENTASI ALGORITMA C4 . 5 UNTUK MENENTUKAN PENERIMA. 7(1), 373–377.*
- Surahman, A., Hayati, U., Sosial, B., & Miner, R. (2023). *IMPLEMENTASI ALGORITMA NAIVE BAYES. 7(1), 347–352.*
- Tyas, S. J. S., Febianah, M., Solikhah, F., Kamil, A. L., & Arifin, W. A. (2021). Analisis Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan. *Jurnal Teknologi Informasi Dan Komunikasi, 8(1), 86–99.*
<https://jurnal.plb.ac.id/index.php/tematik/article/view/576/365>
- Utomo, D. P., & Purba, B. (2019). Penerapan Datamining pada Data Gempa Bumi Terhadap Potensi Tsunami di Indonesia. *Prosiding Seminar Nasional Riset Information Science (SENARIS), 1(September), 846.*
<https://doi.org/10.30645/senaris.v1i0>



Ayu Entini Lumbanraja merupakan mahasiswa Prodi Teknik Informatika dari Universitas Putera Batam yang aktif dalam mendalami bidang teknologi dan informasi.



Koko Handoko merupakan Dosen Prodi Teknik Informatika dari Universitas Putera Batam. Aktif sebagai tenaga kerja dan peneliti.