

# MODEL PENERIMAAN PINJAMAN NASABAH MENGGUNAKAN ALGORITMA NAÏVE BAYES DALAM DATASET BANK

## NASABAH LOAN APPROVAL MODEL USING NAIVE BAYES ALGORITHM IN BANK DATASET

Komang Wahyu Trisna<sup>1)</sup>

<sup>1)</sup>Program Studi Teknik Informatika, STMIK Primakara, Denpasar  
<sup>1)</sup>[kmayutrisna@gmail.com](mailto:kmayutrisna@gmail.com)

Diajukan Februari 14 Februari 2023 / Disetujui 20 Maret 2023

### Abstrak

Saat ini, setiap perusahaan perbankan memiliki sistem pengolahan data secara terkomputerisasi (database) menjadi suatu hal yang wajib. Hal ini dikarenakan database dapat membantu dalam pengolahan data yang ada pada database bank tersebut menjadi informasi yang dapat membantu bank untuk memecahkan berbagai masalah secara efektif, efisien, dan akurat. Pengolahan dataset yang dimiliki perusahaan perbankan adalah untuk menentukan probabilitas peminjaman yang dimiliki setiap nasabah menggunakan proses data mining tersebut. Model klasifikasi dapat dilakukan dengan algoritma Naive Bayes. Dalam penerapan algoritma tersebut, diperlukan sebuah software data mining untuk dijalankan. Pada jurnal ini, kami menggunakan software RapidMiner untuk memproses data yang ada sehingga didapat suatu gambar atau chart yang dapat membentuk berbagai parameter untuk mendukung representasi data yang ada agar lebih mudah untuk menentukan suatu keputusan. Terdapat hasil akurasi algoritma Naïve Bayes pada dataset ini mencapai 80.21%. Selain itu, terdapat *precision* dari dataset yang mencapai 78.95% dan *recall* sebesar 50% saja. Pada dataset ini, diperlukan tingkat *precision* yang lebih tinggi agar memungkinkan didapatkannya persetujuan yang diajukan oleh setiap nasabah kepada pihak bank (true positive) dan meminimalisir tidak disetujuinya pengajuan pinjaman yang dilakukan oleh setiap nasabah kepada pihak bank karena risiko kesalahan pengambilan keputusan yang dihasilkan dari proses data mining (*false positive*).

**Kata kunci:** Data Mining, Perbankan, Rapidminer, Klasifikasi, Naïve Bayes

### Abstract

Currently, every banking company has a computerized data processing system (database) which is mandatory. This is because the database can assist in processing the existing data in the bank's database into information that can help the bank to solve various problems effectively, efficiently and accurately. Processing of datasets owned by banking companies is to determine the probability of borrowing each nasabah using the data mining process. The classification model can be done with the Naive Bayes algorithm. In implementing this algorithm, a data mining software is required to run. In this journal, we use RapidMiner software to process existing data so that an image or chart is obtained that can form various parameters to support the representation of existing data to make it easier to make a decision. There is a result of the accuracy of the Naïve Bayes algorithm in this dataset reaching 80.21%. In addition, there is a precision of the dataset which reaches 78.95% and a recall of only 50%. In this dataset, a higher level of precision is required to enable obtaining approval submitted by each nasabah to the bank (true positive) and to minimize disapproval of loan applications made by each nasabah to the bank due to the risk of decision making errors resulting from data processing. mining (*false positives*).

**Keywords:** Data Mining, Banking, Rapidminer, Classification, Naïve Bayes

---

\*Korespondensi Penulis:  
E-mail: [kmayutrisna@gmail.com](mailto:kmayutrisna@gmail.com)

## Pendahuluan

### 1. LATAR BELAKANG MASALAH

Dalam mempertahankan suatu bisnis, perusahaan memerlukan komponen-komponen yang kuat agar nantinya dapat memiliki daya saing yang tinggi dan bertahan dalam jangka waktu yang panjang. Salah satu dari komponen tersebut, yaitu berupa data. Data adalah aset yang penting bagi kelangsungan suatu perusahaan karena mengandung berbagai macam fakta yang berhubungan dengan perusahaan dan mendapatkan informasi yang akurat. Perlu diketahui bahwa data akan menjadi kunci dasar dalam perkembangan perusahaan yang berkelanjutan di masa mendatang.

Salah satu perusahaan yang memakai data sebagai kunci dasar dalam perkembangan perusahaannya adalah perusahaan pada bidang perbankan. Perbankan memerlukan pengolahan data yang sangat teliti dan tidak dapat menerima adanya sebuah kekeliruan dalam pengolahan data tersebut. Pada umumnya, pengumpulan data, khususnya pada sebuah bank, masih dilakukan secara manual, yaitu dengan cara pencatatan tertulis dan pengolahannya pun membutuhkan ketelitian serta menghabiskan sumber daya manusia yang cukup banyak di dalam mengolah data-data tersebut. Namun, pada masa ini, di mana ilmu pengetahuan dan teknologi (IPTEK) telah semakin berkembang pesat, pengumpulan dan pengolahan data dapat dilakukan dengan cara yang lebih sederhana, efisien, akurat, serta mengurangi pemakaian sumber daya manusia dengan adanya sebuah sistem pengolahan data.

Pada saat ini, setiap perusahaan perbankan memiliki sistem pengolahan data secara terkomputerisasi (database) menjadi suatu hal yang wajib. Hal ini dikarenakan database dapat membantu dalam pengolahan data yang ada pada database bank tersebut menjadi informasi yang dapat membantu bank untuk memecahkan berbagai masalah secara efektif, efisien, dan akurat. Dalam sistem pengolahan data ini, terdapat juga penerapan metode data mining untuk mendapatkan hasil yang maksimal dalam mencerna informasi-informasi apa saja yang ada pada dataset bank tersebut.

Dengan adanya data mining, perusahaan perbankan akan sangat terbantu dalam sistem pengolahan dan jalannya operasional yang dimiliki sehingga pemilihan metode yang digunakan juga sangat penting dalam implementasi data mining. Seperti yang kita tahu, dalam data mining, terdapat berbagai macam metode yang dapat membantu sebuah sistem dalam pengolahan dataset yang dimiliki oleh perbankan tersebut. Salah satu metode yang sering digunakan dalam pengolahan dataset, yaitu klasifikasi, di mana pada metode ini, terdapat teknik data mining Naïve Bayes yang menggunakan teori probabilitas untuk menemukan peluang terbesar dari kemungkinan klasifikasi dengan memeriksa frekuensi setiap klasifikasi pada dataset yang ada.

Salah satu penerapan Naïve Bayes dalam pengolahan dataset yang dimiliki perusahaan perbankan adalah untuk menentukan probabilitas peminjaman yang dimiliki setiap nasabah menggunakan proses data mining tersebut. Oleh karena itu, dengan adanya penerapan metode ini, bank tidak perlu lagi menghitung secara manual untuk menentukan probabilitas tersebut. Selain itu, bank juga dapat mengalokasikan sumber daya manusia kepada divisi lain yang membutuhkan sumber daya tersebut.

Adanya penerapan Naïve Bayes ini, kesalahan yang disebabkan oleh kelalaian manusia dapat diminimalisir karena semua perhitungan terhadap probabilitas peminjaman setiap nasabah telah dihitung dan diatur melalui otomatisasi sistem pengolahan data yang telah ditetapkan sebelumnya. Untuk itu, kemungkinan adanya human error juga tentunya akan menurun. Dari fakta di lapangan yang dialami pihak bank, akibat analisa peminjaman kredit secara manual tersebut mengakibatkan 72.36% peminjam gagal bayar yang merugikan pihak bank kurang lebih \$32.800 rata-rata tiap tahunnya. Tentunya ini dihindari oleh pihak bank untuk terjadi yang menuntun pihak bank ingin melakukan data mining untuk menentukan siapa saja yang berhak mendapat peminjaman kredit untuk nasabah yang *eligible* berdasarkan data *real* yang terjadi sehingga meminimalisir fenomena gagal bayar yang terjadi.

Dalam penerapan algoritma tersebut, diperlukan sebuah *software* data mining untuk dijalankan. Pada jurnal ini, kami menggunakan *software* RapidMiner untuk memproses data yang ada sehingga didapat suatu gambar atau chart yang dapat membentuk berbagai parameter untuk mendukung representasi data yang ada agar lebih mudah untuk menentukan suatu keputusan.

## 2. RUMUSAN MASALAH

Rumusan masalah dari penelitian ini antara lain:

1. Bagaimana data mining dengan klasifikasi Naive Bayes dapat menemukan dan menafsirkan pola dari data yang dimiliki oleh bank?
2. Bagaimana membagi data menjadi *training set* dan *testing set* dalam algoritma Naive Bayes sehingga menghasilkan akurasi yang baik?

## 3. TUJUAN DAN MANFAAT

Tujuan dari studi kasus ini adalah untuk :

1. Mengetahui peran data mining dalam membantu perusahaan perbankan sehingga dapat mengurangi adanya *human error*.
2. Mengimplementasikan metode Naïve Bayes pada dataset perusahaan perbankan dengan menggunakan RapidMiner.
3. Mengetahui peran metode Naïve Bayes dalam membantu perusahaan perbankan untuk menentukan status peminjaman seorang nasabah.
4. Mengetahui jumlah *split* data antara *training set* dan *testing set* yang baik.

## Tinjauan Pustaka

### 1. DATA MINING

Data mining merupakan proses ekstraksi data sehingga didapatkan suatu informasi penting yang implisit dan sebelumnya tidak diketahui dari data tersebut (Witten, et al., 2011). Sedangkan, menurut Han, dkk (2011), data mining adalah ekstraksi *knowledge pattern* yang menarik dari sejumlah data yang besar. Jadi, dapat disimpulkan bahwa data mining merupakan disiplin ilmu yang mempelajari metode dalam ekstraksi data sehingga ditemukannya *knowledge pattern* dari suatu data. Untuk dapat lebih memahami ekstraksi *knowledge* dari suatu data, maka perbedaan antara data, informasi, dan pengetahuan adalah sebagai berikut :

1. Data adalah fakta yang tercatat dan tidak memiliki arti.
2. Informasi adalah rekapan, ringkasan, deskripsi, dan statistik dari data.
3. Pengetahuan adalah pola, rumus, aturan, atau model yang dihasilkan dari data.

Dalam proses data mining, terdapat beberapa operasi umum yang dapat dilakukan, di antaranya :

1. *Estimation*, yaitu metode yang menggunakan angka untuk tipe variabel target, di mana metode ini dibangun dengan data yang memberikan nilai-nilai variabel target dan prediktor serta nilai variabel target akan diestimasi berdasarkan nilai prediktor pada observasi baru.
2. *Prediction*, yaitu metode yang hasilnya terletak di masa depan.
3. *Classification*, yaitu metode yang menentukan kumpulan data baru ke salah satu dari beberapa kategori yang telah ditentukan sebelumnya, di mana dikenal dengan *supervised learning*.
4. *Clustering*, yaitu metode yang mengelompokkan dataset, observasi, atau kasus ke dalam kelas fitur yang serupa, di mana memiliki kumpulan *record* yang identik satu dengan yang lain dan berbeda dengan data di *record* pada *cluster* lain.
5. *Association*, yaitu metode yang menemukan relasi di antara atribut-atribut dari dataset yang ada, di mana sering disebut dengan *market basket analysis*.

### 2. CRISP-DM (CROSS-INDUSTRY STANDARD PROCESS FOR DATA MINING)

Ada beberapa metode yang dapat digunakan untuk memaksimalkan peluang keberhasilan dalam melakukan data mining. Salah satu metode yang terpopuler adalah CRISP-DM. Metode ini terdiri dari 6 tahapan, di antaranya adalah :

#### 1. *Business understanding*

Tahap ini dimulai dengan melakukan pemahaman secara menyeluruh mengenai kebutuhan manajerial dan spesifikasi eksplisit dari tujuan bisnis mengenai proyek yang dilakukan. Pada tahap ini, anggaran juga perlu ditetapkan; setidaknya dalam jumlah tinggi, walaupun dengan kisaran angka yang kasar.

## 2. Data understanding

Tahap ini dilakukan proses identifikasi data yang relevan dari database yang tersedia. Dalam hal ini, analis harus membuat deskripsi proyek data mining dengan jelas dan ringkas sehingga data yang paling relevan dapat diidentifikasi. Kemudian, analis harus memahami sumber data secara mendalam. Identifikasi dan pemilihan sumber data yang cermat dan variabel yang paling relevan dapat mempermudah algoritma data mining untuk menemukan *knowledge pattern* dengan cepat.

## 3. Data preparation

Tahap ini bertujuan untuk mengambil data yang diidentifikasi pada langkah sebelumnya dan mempersiapkannya untuk dianalisis dengan menggunakan metode data mining. Tahap ini dianggap sebagai tahap yang paling banyak menghabiskan waktu dan tenaga karena sering kali data yang didapat kurang lengkap, mengandung kesalahan, dan tidak konsisten.

## 4. Model building

Tahap ini memilih dan menerapkan berbagai teknik pemodelan pada dataset yang telah disiapkan sesuai dengan kebutuhan bisnisnya, di mana mencakup penilaian dan analisis komparatif dari berbagai model yang dibangun. Dalam hal ini, seseorang harus mengidentifikasi metode apa yang terbaik untuk digunakan sehingga mendapatkan hasil yang optimal.

## 5. Testing & evaluation

Tahap ini akan menilai dan mengevaluasi akurasi dari model yang telah dikembangkan. Tahap ini dikatakan sebagai tahap yang kritis dan menantang karena tidak ada value yang ditambahkan dalam data mining hingga nilai bisnis yang diperoleh dari *knowledge* yang ditemukan dapat diidentifikasi.

## 6. Deployment

Tahap ini mencakup aktivitas pemeliharaan untuk model yang diterapkan. Data yang mencerminkan aktivitas bisnis dapat berubah seiring berjalannya waktu sehingga model yang dibangun dari data yang sudah lama dapat menjadi usang, tidak relevan, atau menyesatkan. Oleh karena itu, dilakukanlah pemantauan dan pemeliharaan model, di mana persiapan strategi pemeliharaan yang cermat dapat membantu menghindari penggunaan hasil data mining yang salah.

## 3. CLASSIFICATION

*Classification* merupakan metode data mining yang paling sering digunakan, di mana metode tersebut mempelajari pola dari data yang lampau dan membedakan kelas data sehingga dapat digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Perlu diketahui bahwa metode ini berbeda dengan *clustering*. Perbedaan *classification* dengan *clustering*, yaitu *classification* mempelajari suatu fungsi karakteristik melalui *supervised learning*, sedangkan *clustering* mempelajari karakteristik melalui *unsupervised learning*. Terdapat beberapa teknik yang digunakan dalam pemodelan *classification*, di antaranya *decision tree analysis*, *statistical analysis*, *neural networks*, *case-based reasoning*, *bayesian classifiers*, *genetic algorithms*, dan *rough sets*.

Metode *classification* terdiri dari 2 langkah proses yang melibatkan development model (training) dan testing model (deployment). Dalam development model, kumpulan data input, termasuk label kelas aktual, akan digunakan. Kemudian, model akan diuji untuk dilihat nilai akurasi dan digunakan untuk memprediksi kelas dari data baru yang tidak diketahui label kelasnya. Berikut ini, beberapa faktor yang dijadikan pertimbangan dalam menilai model, yaitu:

1. *Predictive accuracy*, kemampuan model untuk memprediksi dengan benar mengenai label kelas dari data baru yang sebelumnya tidak diketahui.
2. *Speed*, biaya komputasi yang terlibat dalam menghasilkan dan menggunakan model yang lebih cepat dan lebih baik.
3. *Robustness*, kemampuan model untuk memprediksi secara akurat dengan data noisy atau data yang memiliki nilai missing dan erroneous.
4. *Scalability*, kemampuan untuk membangun model prediksi secara efisien mengingat jumlah data yang agak besar.
5. *Interpretability*, tingkat pemahaman dan insight yang diberikan oleh model.

## 4. NAÏVE BAYES

Naïve Bayes adalah salah satu algoritma dalam teknik data mining yang menerapkan teori Bayesian dalam klasifikasi. Teorema keputusan algoritma ini, yaitu pendekatan statistik mendasar terhadap pengenalan pola (*pattern recognition*) dan didasarkan pada penyederhanaan nilai atribut

ketika nilai output diberikan. Seperti yang telah dikatakan sebelumnya, Naïve Bayes salah satu bentuk klasifikasi data yang menggunakan metode probabilistik dan statistik sehingga algoritma ini dapat dikatakan sebagai metode tanpa aturan. Dalam algoritma Naïve Bayes, digunakan teori probabilitas untuk menemukan peluang terbesar dari kemungkinan klasifikasi dengan memeriksa frekuensi setiap klasifikasi pada data training.

Penggunaan algoritma Naïve Bayes memiliki keuntungan, yaitu metode ini hanya membutuhkan jumlah data training yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasiannya. Selain itu, algoritma ini dapat digunakan untuk data kuantitatif maupun data kualitatif dan jika terdapat nilai yang hilang, maka dapat diabaikan dalam proses perhitungannya. Di sisi lain, nilai akurasi dari Naïve Bayes tidak dapat diukur dengan 1 probabilitas saja sehingga dibutuhkan bukti-bukti lainnya dan jika probabilitas kondisionalnya memiliki nilai nol, maka probabilitas prediksi juga akan bernilai nol.

## 2.5 RAPIDMINER

RapidMiner adalah *tools* yang digunakan untuk menganalisis data mining, text mining, dan analisis prediksi, di mana menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang baik. Adapun fitur-fitur yang tersedia dalam RapidMiner, yaitu:

1. Terdapat berbagai macam algoritma data mining, seperti Naïve Bayes, *decision tree*, *self-organization map*, dan lain-lain.
2. Memiliki bentuk grafis yang canggih, seperti diagram histogram, *tree chart*, dan 3D *scatter plots*.
3. Memiliki beragam variasi *plugins*, seperti *text plugin* untuk analisis teks.
4. Menyediakan prosedur data mining dan machine learning, seperti ETL, *data preprocessing*, visualisasi, *modeling*, dan evaluasi.
5. Proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI.
6. Mengintegrasikan proyek data mining WEKA dan statistika R.

## Metode Penelitian

### 1. LATAR BELAKANG PERUSAHAAN

Secara umum, bank merupakan lembaga intermediasi keuangan yang diatur dengan wewenang untuk menerima setoran tunai, meminjamkan uang, dan menerbitkan promes (*banknote*). Perlu diketahui, kata 'bank' berasal dari bahasa Italia, yaitu 'banco', di mana berarti bangku tempat terjadinya proses pelayanan yang dilakukan oleh bankir. Di sisi lain, menurut UU Nomor 10 Tahun 1998 tentang Perbankan, bank adalah badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkannya kepada masyarakat dalam bentuk kredit dan atau bentuk-bentuk lainnya dalam rangka meningkatkan taraf hidup rakyat banyak.

Seperti yang kita tahu, bank memiliki beberapa jenis yang dikategorikan berdasarkan berbagai segi, di antaranya :

1. Berdasarkan fungsi, contohnya adalah bank sentral, bank umum, dan bank perkreditan rakyat.
2. Berdasarkan kepemilikan, contohnya adalah bank pemerintah, bank swasta, bank asing, dan bank campuran.
3. Berdasarkan operasional, contohnya adalah bank konvensional dan bank syariah.
4. Berdasarkan bentuk badan usaha, contohnya adalah koperasi, perusahaan perseorangan, perseroan terbatas, dan firma.

Adapun fungsi bank bagi masyarakat adalah sebagai berikut :

1. *Financial intermediary*, berperan sebagai perantara keuangan dalam menyalurkannya kepada masyarakat sehingga pemanfaatan keuangan dapat merata di seluruh kalangan.
2. *Agent of trust*, berperan sebagai pihak yang layak dipercaya oleh pengguna jasanya dalam menjaga dan memelihara nilai uang.

3. *Agent of development*, berperan sebagai pihak yang mampu memberikan aktivitas dan layanan dalam upaya meningkatkan dan mengembangkan pendapatan dengan investasi, konsumsi, distribusi, dan lainnya kepada masyarakat sehingga nantinya pembangunan ekonomi negara dapat semakin maju
4. *Agent of service*, berperan sebagai pihak yang melayani berbagai kepentingan keuangan masyarakat dengan menyediakan layanan keuangan semaksimal mungkin sesuai dengan kepentingan para pengguna jasanya.

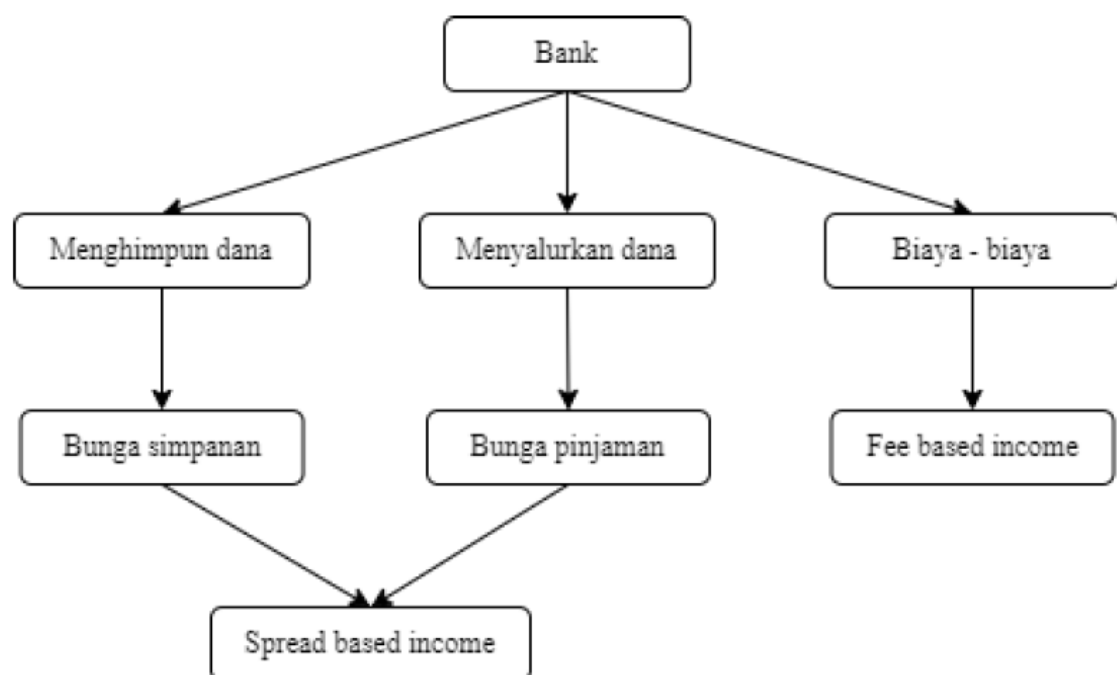
Dalam kehidupan sehari-hari, bank memiliki beberapa manfaat, yaitu :

1. Sebagai model investasi, di mana biasanya jenis investasinya ini dalam jangka pendek (*yield enhancement*).
2. Sebagai cara lindung nilai, di mana berfungsi sebagai salah satu cara untuk mencegah risiko (*risk management*).
3. Sebagai informasi harga, di mana berfungsi untuk memberikan informasi mengenai harga tertentu (*price discovery*).
4. Sebagai fungsi spekulatif, di mana memberikan kesempatan spekulasi terhadap perubahan nilai pasar.
5. Sebagai fungsi manajemen produksi yang berjalan dengan baik dan efisien, di mana memberikan gambaran dalam menilai suatu permintaan dan kebutuhan pasar di masa yang akan datang.

## 2. BISNIS PROSES PERUSAHAAN

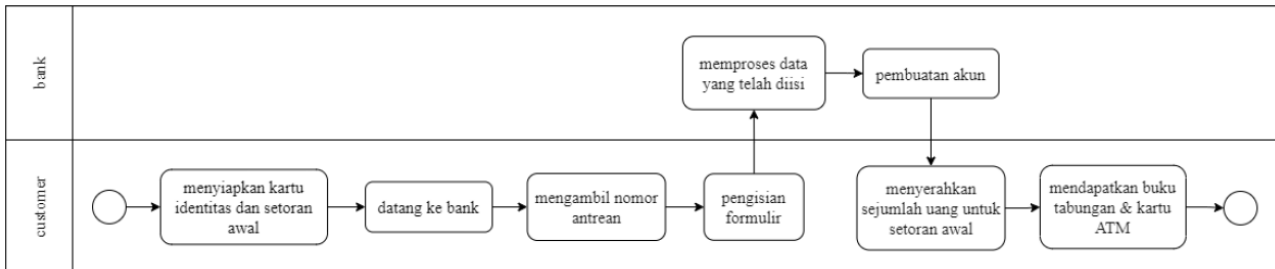
Pada gambar 3.1, terdapat 3 model bisnis proses yang berjalan, di antaranya adalah penghimpunan dana, penyaluran dana, dan pembayaran lainnya, di mana terbagi menjadi 2 model income, yaitu *spread based income* (pendapatan primer) dan *fee based income* (pendapatan sekunder).

*Spread based income* adalah keuntungan dari penghimpunan dan penyaluran dana, di mana pendapatan yang diperoleh berasal dari bunga yang ditentukan oleh bank. Pada dasarnya, bank akan memberikan bunga kepada kreditur (peminjam uang) dan juga kepada pemilik uang tersebut. Selisih dari bunga yang dipinjam dan diberikan itulah yang akan menjadi *spread based income*. Sedangkan, *fee based income* adalah keuntungan yang didapat dari transaksi yang diberikan pada saat nasabah menggunakan jasa-jasa bank lainnya di luar jasa *spread based*.



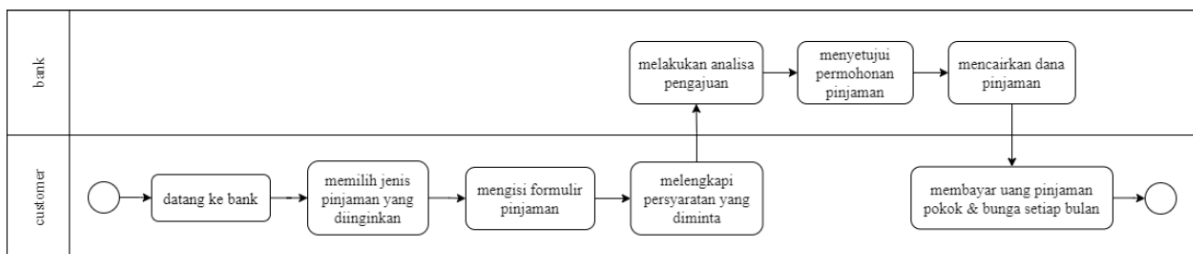
Gambar 3.1 Bisnis Proses Perusahaan Perbankan

Gambar 3.2 merupakan proses bisnis yang dapat dilakukan oleh nasabah untuk menyimpan uangnya di bank. Hal pertama yang dapat dilakukan oleh nasabah tersebut, yaitu menyiapkan kartu identitas (KTP) dan membawa jumlah uang yang ingin disetor saat ia datang ke bank. Saat tiba di bank, nasabah perlu mengambil nomor antrean terlebih dahulu. Nasabah akan melakukan pengisian formulir saat nomor antrean yang dimilikinya tiba. Lalu, pihak bank akan memproses data-data yang telah diisi sebelumnya dan akan membuatkan akun rekening untuk nasabah jika data-data tersebut disetujui. Setelah pembuatan akun selesai, nasabah dapat menyerahkan uang yang telah dibawanya untuk setoran awal. Sebagai buktinya, nasabah akan mendapatkan buku tabungan dan kartu ATM.



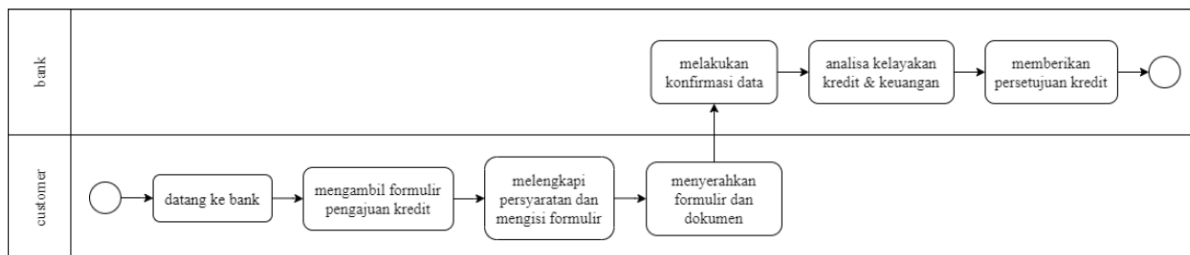
Gambar 3.2 BPMN Penyimpanan Uang di Bank

Gambar 3.3 merupakan proses bisnis yang dapat dilakukan oleh nasabah untuk meminjam uang di bank. Saat nasabah datang ke bank, ia memilih terlebih dahulu jenis pinjaman apa yang dibutuhkannya. Setelah itu, ia dapat melakukan pengisian formulir pinjaman dan melengkapi persyaratan-persyaratan apa saja yang diminta oleh pihak bank. Lalu, pihak bank akan melakukan analisa pengajuan dan jika permohonan pengajuan pinjaman disetujui, maka dana pinjaman akan dicairkan. Analisa pengajuan disini biasa dilakukan pihak bank dengan survey langsung ke lapangan yaitu rumah atau tempat kerja nasabah dan melakukan interview terkait kebenarannya. Dan hanya difokuskan ke atribut penghasilan nasabah dan kredit rating saja tanpa memikirkan atribut lain sehingga pihak bank bisa kecolongan untuk nasabah yang nakal. Jika benar, nasabah dapat membayar kembali uang pinjaman tersebut dengan bunganya di setiap bulan.



Gambar 3.3 BPMN Peminjaman Uang di Bank

Gambar 3.4 merupakan proses bisnis yang dapat dilakukan oleh nasabah untuk mengajukan kredit di bank. Nasabah dapat mengambil formulir pengajuan kredit terlebih dahulu saat ia telah tiba di bank. Lalu, formulir tersebut dapat diisi dan persyaratan yang diminta oleh pihak bank dapat dilengkapi oleh nasabah. Setelah formulir dan persyaratan telah diserahkan, pihak bank akan melakukan konfirmasi data dan melakukan analisa kelayakan kredit serta keuangan yang dimiliki nasabah. Jika pengajuan kredit yang diajukan telah disetujui, maka pihak bank akan memberikan persetujuan kredit tersebut kepada nasabah.



Gambar 3.4 BPMN Pengajuan Kredit di Bank

### 3. KERANGKA PIKIR JURNAL

Pada dasarnya, kerangka pikiran adalah suatu arah penalaran untuk dapat memberikan jawaban sementara terhadap rumusan masalah yang telah disebutkan. Menurut Suriasumantri dalam Sugiyono (2016, hlm. 60), kerangka pemikiran ini merupakan penjelasan sementara terhadap gejala-gejala yang menjadi titik fokus permasalahan. Dengan demikian, dapat disimpulkan bahwa kerangka berpikir adalah suatu konsepsi atau keadaan pikiran yang menjelaskan hubungan antara masalah yang berkaitan dengan berbagai teori berbeda yang telah dijelaskan untuk dianalisis dan diberikan solusi sehingga dapat dibentuk sebuah hipotesis.

Dalam studi kasus ini, kerangka pikirnya adalah sebagai berikut:

Tabel 3.1 Kerangka Pikir Jurnal

<i>Input</i>	<ol style="list-style-type: none"> <li>1. Menentukan industri perusahaan</li> <li>2. Mencari dan mempelajari jurnal sebelumnya yang relevan sebagai sumber referensi</li> <li>3. Merumuskan masalah yang terjadi</li> <li>4. Mencari <i>dataset</i> yang sesuai</li> <li>5. Melakukan <i>data understanding</i></li> </ol>
<i>Process</i>	<ol style="list-style-type: none"> <li>1. Melakukan persiapan terlebih dahulu</li> <li>2. Menentukan metode <i>data mining</i> yang tepat</li> <li>3. Melakukan pembuatan model dan pengolahan <i>data mining</i> dengan <i>dataset</i> yang telah dicari sebelumnya</li> </ol>
<i>Output</i>	<ol style="list-style-type: none"> <li>1. Mengevaluasi hasil <i>data mining</i> sesuai dengan <i>dataset</i> yang telah diolah</li> <li>2. Menganalisa hasil yang didapatkan</li> <li>3. Menulis laporan hasil</li> </ol>

### Hasil Dan Pembahasan

#### 1. ANALISIS & PEMBAHASAN

##### A. ANALISIS PERMASALAHAN & DATASET

Dataset pada penelitian ini diambil dari data nasabah kartu kredit di salah satu bank dunia terkenal yang telah terdaftar selama dua tahun dari tahun Januari 2020 sampai dengan Desember 2022. Total *record* dalam dataset ini adalah 8320 data yang tersebar di beberapa daerah dan gender.



Tabel 4.1 Atribut Dataset

Row No.	Gender (polynomial) regular	Married (binominal) regular	Dependents (polynomial) regular	Education (binominal) regular	Self_Employed (binominal) regular	ApplicantIncome (integer) regular	CoapplicantIncome (integer) regular	LoanAmount (integer) regular	Loan_Amount_Term (integer) regular	Credit_History (integer) regular	Loan_Status (binominal) label
1	Male	Yes	1	Graduate	No	4583	1508	128	360	1	N
2	Male	Yes	0	Graduate	Yes	3000	0	66	360	1	Y
3	Male	Yes	0	Not Graduate	No	2583	2358	120	360	1	Y
4	Male	No	0	Graduate	No	6000	0	141	360	1	Y
5	Male	Yes	2	Graduate	Yes	5417	4196	267	360	1	Y
6	Male	Yes	0	Not Graduate	No	2333	1516	95	360	1	Y
7	Male	Yes	3+	Graduate	No	3036	2504	158	360	0	N
8	Male	Yes	2	Graduate	No	4006	1526	168	360	1	Y
9	Male	Yes	1	Graduate	No	12841	10968	349	360	1	N
10	Male	Yes	2	Graduate	No	3200	700	70	360	1	Y
11	Male	Yes	2	Graduate	No	3073	8106	200	360	1	Y
12	Male	No	0	Graduate	No	1853	2840	114	360	1	N
13	Male	Yes	2	Graduate	No	1299	1086	17	120	1	Y
14	Male	No	0	Graduate	No	4950	0	125	360	1	Y

Pada tabel 4.1, terdapat 11 atribut dalam dataset ini, antara lain:

1. Row No. Merupakan penomoran auto increment untuk setiap data yang ada pada dataset.
2. Gender. Atribut ini memiliki tipe nilai polynominal karena memiliki lebih dari 2 kategori. Atribut ini tergolong sebagai regular karena nilainya ini akan menjadi penentu dalam perhitungan data mining.
3. Married. Atribut ini memiliki tipe nilai binominal karena memiliki 2 kategori, yaitu 'Yes' dan 'No'. Atribut ini tergolong sebagai regular karena nilainya ini akan menjadi penentu dalam perhitungan data mining.
4. Dependents. Atribut ini memiliki tipe nilai polynominal karena memiliki lebih dari 2 kategori. Atribut ini tergolong sebagai regular karena nilainya ini akan menjadi penentu dalam perhitungan data mining.
5. Education. Atribut ini memiliki tipe nilai binominal karena memiliki 2 kategori, yaitu 'Graduate' dan 'Not Graduate'. Atribut ini tergolong sebagai regular karena nilainya ini akan menjadi penentu dalam perhitungan data mining.
6. Self\_Employed. Atribut ini memiliki tipe nilai binominal karena memiliki 2 kategori, yaitu 'Yes' dan 'No'. Atribut ini tergolong sebagai regular karena nilainya ini akan menjadi penentu dalam perhitungan data mining.
7. ApplicantIncome. Atribut ini memiliki tipe nilai integer karena merupakan bilangan bulat. Atribut ini tergolong sebagai regular karena nilainya ini akan menjadi penentu dalam perhitungan data mining.
8. CoapplicantIncome. Atribut ini memiliki tipe nilai integer karena merupakan bilangan bulat. Atribut ini tergolong sebagai regular karena nilainya ini akan menjadi penentu dalam perhitungan data mining.
9. LoanAmount. Atribut ini memiliki tipe nilai integer karena merupakan bilangan bulat. Atribut ini tergolong sebagai regular karena nilainya ini akan menjadi penentu dalam perhitungan data mining.
10. Loan\_Amount\_Term. Atribut ini memiliki tipe nilai integer karena merupakan bilangan bulat. Atribut ini tergolong sebagai regular karena nilainya ini akan menjadi penentu dalam perhitungan data mining.
11. Credit\_History. Atribut ini memiliki tipe nilai integer karena merupakan bilangan bulat. Atribut ini tergolong sebagai regular karena nilainya ini akan menjadi penentu dalam perhitungan data mining.
12. Loan\_Status. Atribut ini memiliki tipe nilai binominal karena memiliki 2 kategori, yaitu 'Y' dan 'N'. Atribut ini tergolong sebagai regular karena nilainya ini akan menjadi penentu dalam perhitungan data mining.

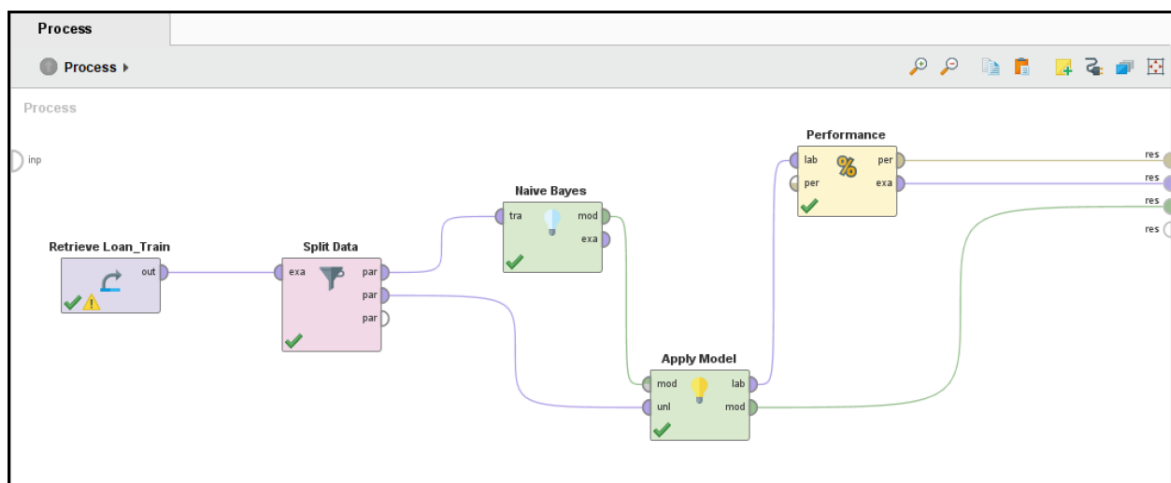
Melalui analisis dataset dengan pengimplementasian metode Naïve Bayes menggunakan RapidMiner, dapat diketahui faktor-faktor keputusan akhir terhadap peminjaman yang akan dilakukan oleh setiap nasabah. Dalam hal ini, RapidMiner memberikan informasi-informasi mengenai probabilitas peminjaman nasabah melalui faktor-faktor tersebut. Dengan adanya faktor-faktor ini, hal

tersebut dapat membantu dalam mengatasi permasalahan yang telah dijelaskan sebelumnya. Oleh karena itu, bank dapat menentukan nasabah mana saja yang dapat diberikan persetujuan dalam peminjaman yang diajukan. Ciri-ciri nasabah yang memiliki probabilitas tinggi untuk mengajukan pinjaman dari bank adalah sebagai berikut:

1. Nasabah yang memiliki jenis kelamin pria, belum menikah, tidak memiliki tanggungan, menempuh pendidikan hingga gelar magister, bukan merupakan self-employed, dan memiliki pendapatan sebesar 6000 USD dengan jumlah pinjaman sebesar 141 USD dan cicilan selama 30 tahun akan mendapatkan persetujuan peminjaman dari bank dengan kemungkinan sebesar 95.9%.
2. Nasabah yang memiliki jenis kelamin wanita, sudah menikah, tidak memiliki tanggungan, menempuh pendidikan hingga gelar magister, bukan merupakan self-employed, dan memiliki pendapatan sebesar 4583 USD dengan jumlah pinjaman sebesar 128 USD, cicilan selama 30 tahun, dan membagi pelunasan peminjaman dengan seseorang yang memiliki pendapatan sebesar 1508 USD akan mendapatkan persetujuan peminjaman dari bank dengan kemungkinan sebesar 96.8%
3. Nasabah yang memiliki jenis kelamin pria, sudah menikah, memiliki 2 tanggungan, menempuh pendidikan hingga gelar magister, bukan merupakan self-employed, dan memiliki pendapatan sebesar 3800 USD dengan jumlah pinjaman sebesar 216 USD, cicilan selama 30 tahun, dan membagi pelunasan peminjaman dengan seseorang yang memiliki pendapatan sebesar 3600 USD tidak akan mendapatkan persetujuan peminjaman dari bank.

## B. PEMBAHASAN DATA MINING

Pada proses data mining, kami menggunakan menggunakan metode Naïve Bayes, di mana dataset tersebut akan dimasukkan ke dalam proses *split* data dan terbagi menjadi 2 bagian, yaitu *training set* (rasio pembagian 0.8) dan *testing set* (rasio pembagian 0.2). Setelah dilakukan proses *split data*, data pada *training set* akan diuji dalam metode Naïve Bayes untuk dimasukkan ke dalam model. Sedangkan, data pada *testing set* akan dimasukkan secara langsung ke dalam model tanpa diuji dalam metode Naïve Bayes terlebih dahulu. Hasil dari training set dan testing set tersebut akan dibandingkan satu dengan yang lain. Selain itu, *operator performance* akan dijalankan, di mana berguna untuk mengetahui area penyebaran data dan akurasi dari *data training* yang telah dibandingkan dengan *data testing*.



Gambar 4.1 Proses *Split Data*

Tabel 4.2 *Example Set*

RowNo. ↑	prediction(Loan_Status)	confidence(N)	confidence(Y)	Gender	Married	Dependents	Education	Self_Employ...	ApplicantInc...
1	Y	0.041	0.959	Male	No	0	Graduate	No	6000
2	Y	0.256	0.744	Male	Yes	1	Graduate	No	5955
3	N	1.000	0.000	Male	Yes	0	Not Graduate	No	2600
4	Y	0.063	0.937	Male	No	0	Not Graduate	No	1442
5	Y	0.028	0.972	Male	Yes	0	Graduate	No	5821
6	Y	0.047	0.953	Female	No	0	Graduate	No	4230
7	Y	0.023	0.977	Male	Yes	0	Graduate	No	2132
8	Y	0.033	0.967	Male	Yes	2	Not Graduate	No	3357
9	Y	0.034	0.966	Male	Yes	3+	Graduate	No	3029
10	N	1.000	0.000	Female	No	0	Graduate	No	4166
11	N	0.762	0.238	Male	Yes	3+	Not Graduate	Yes	7100
12	Y	0.037	0.963	Male	Yes	2	Not Graduate	Yes	1875
13	Y	0.096	0.904	Male	No	0	Graduate	No	3750
14	Y	0.041	0.959	Male	No	0	Graduate	No	4133
15	N	1.000	0.000	Male	Yes	2	Graduate	No	3800
16	Y	0.111	0.889	Male	Yes	1	Graduate	No	14583
17	Y	0.036	0.964	Male	No	0	Graduate	No	2014
18	Y	0.041	0.959	Male	No	0	Graduate	No	2718
19	Y	0.308	0.692	Male	Yes	0	Graduate	Yes	3459
20	Y	0.032	0.968	Female	Yes	0	Graduate	No	4583
21	Y	0.044	0.956	Male	No	0	Graduate	No	5417

Setelah proses data mining dilakukan, akan didapatkan tampilan *Example Set*, di mana berisi informasi yang menunjukkan prediksi sehingga dapat menentukan keputusan apakah nasabah yang mengajukan pinjaman diizinkan atau tidak oleh pihak bank. Jika prediksi tertulis ‘Y’, maka kemungkinan pinjaman yang diajukan oleh nasabah tersebut akan disetujui pihak bank. Sebaliknya, jika prediksi tertulis ‘N’, maka kemungkinan pinjaman yang diajukan oleh nasabah tersebut tidak akan disetujui pihak bank. Seperti yang kita ketahui bahwa selain prediksi, terdapat juga *confidence*, di mana dapat menentukan besar persentase keyakinan bank dalam persetujuan peminjaman yang diajukan oleh nasabah.

Tabel 4.3 Performance Vector

	true N	true Y	class precision
pred. N	15	4	78.95%
pred. Y	15	62	80.52%
class recall	50.00%	93.94%	

Tabel 4.3 menampilkan *confusion matrix*, di mana terdapat penjelasan mengenai akurasi algoritma Naïve Bayes pada dataset ini mencapai 80.21%. Selain itu, terdapat *precision* dari dataset yang mencapai 78.95% dan *recall* sebesar 50% saja. Pada dataset ini, diperlukan tingkat *precision* yang lebih tinggi agar memungkinkan didapatkannya persetujuan yang diajukan oleh setiap nasabah kepada pihak bank (*true positive*) dan meminimalisir tidak disetujuinya pengajuan pinjaman yang dilakukan oleh setiap nasabah kepada pihak bank karena risiko kesalahan pengambilan keputusan yang dihasilkan dari proses data mining (*false positive*).

## Simpulan

### 1. KESIMPULAN

Dari hasil analisis dengan menggunakan dataset yang dipilih, dapat ditarik kesimpulan, yaitu sebagai berikut:

3. Dengan adanya data mining, perusahaan dapat menemukan dan menafsirkan pola dari data yang dimilikinya. Hal ini tentunya membantu perusahaan dalam membuat keputusan dan melayani nasabah dengan lebih baik melalui informasi yang telah diolah dari data tersebut. Seperti yang telah dijelaskan, data mining memiliki berbagai macam algoritma, di mana memiliki fungsi dan tujuannya masing-masing. Dalam jurnal ini, kami menggunakan algoritma Naive Bayes dengan RapidMiner dalam mengolah data mining.
4. Pada algoritma Naive Bayes, *split* data membagi data menjadi *training set* dan *testing set*, di mana keduanya ini akan dibandingkan antara satu dengan yang lain.

### 2. SARAN

Berdasarkan hasil analisis yang dilakukan, saran yang dapat diberikan, yaitu data mining dapat menimbulkan kekhawatiran dalam hal privasi. Hal ini dikarenakan data legal (sah) yang digunakan dalam proses data mining tidak selamanya etis. Oleh karena itu, pemahaman terhadap data mining tidak dapat hanya didasarkan pada teori dan standar proses tertentu saja, tetapi juga harus disertai dengan pemahaman terhadap kode etikanya. Cara untuk melindungi privasi data ini, yaitu dengan menggabungkan data, menganonimkan pengamatan melalui penghapusan nama dan informasi yang bersifat privasi, dan menyimpannya di lingkungan yang aman dan terlindungi

## Daftar Pustaka

- Act of the Republic of Indonesia (1998) ‘Undang Undang Nomor 10 Tahun 1998 tentang Perbankan’, p. 63. Available at: [http://www.komnasham.go.id/sites/default/files/dokumen/UU\\_NO\\_39\\_TAHUN\\_1999\\_HAM\\_0.pdf](http://www.komnasham.go.id/sites/default/files/dokumen/UU_NO_39_TAHUN_1999_HAM_0.pdf).
- Admin (2020) RapidMiner : Mengetahui Aplikasi Data Mining Terkemuka di Dunia, doavers.com. Available at: <https://www.doavers.com/blog/rapidminer-mengenal-aplikasi-data-mining-terkemuka-di-dunia> (Accessed: 9 June 2022).
- Ahmad (2021) Pengertian Bank: Fungsi, dan Jenis-Jenis Bank di Indonesia, gamedia.com. Available at: <https://www.gamedia.com/literasi/pengertian-bank/> (Accessed: 30 March 2022).
- Alliance, D. S. P. (2022) What is CRISP DM?, datascience-pm.com. Available at: <https://www.datascience-pm.com/crisp-dm-2/> (Accessed: 29 March 2022).
- Defiyanti, S. (2015) ‘Integrasi Metode Klasifikasi Dan Clustering dalam Data Mining’, (March), pp. 39–44.
- NISP, R. O. (2021) Pengertian Bank, Jenis-Jenis, dan Fungsinya Bagi Masyarakat, ocbcnisp.com. Available at: <https://www.ocbcnisp.com/id/article/2021/07/15/pengertian-bank> (Accessed: 30 March 2022).
- Nuraeni, N. (2021) ‘Klasifikasi Data Mining untuk Prediksi Potensi Nasabah dalam Membuat Deposito Berjangka’, *Jurnal Ilmiah Intech: Information Technology Journal ...*, 3(01), pp. 65–74. Available at: <http://jurnal.umus.ac.id/index.php/intech/article/view/418%0Ahttp://jurnal.umus.ac.id/index.php/intech/article/download/418/281>.
- Retnosari, R. et al. (2021) ‘Analisa kelayakan kredit usaha mikro berjalan pada perbankan dengan metode naive bayes’, *PROSISKO: Jurnal Pengembangan Riset dan Observasi Sistem Komputer*, 8(1), pp. 53–59. Available at: <https://ejournal.lppmunsera.org/index.php/PROSISKO/article/view/2848>.

- Romadhon, R. (2021) Inilah Pentingnya Data Dan Informasi Bagi Kemajuan Perusahaan, [softwareseni.co.id](https://www.softwareseni.co.id). Available at: [https://www.softwareseni.co.id/blog/pentingnya-data-dan-informasi#:~:text=Data memiliki peran krusial dalam,bagi informasi yang akurat pula](https://www.softwareseni.co.id/blog/pentingnya-data-dan-informasi#:~:text=Data%20memiliki%20peran%20krusial%20dalam%20bagi%20informasi%20yang%20akurat%20pula.). (Accessed: 28 March 2022).
- Setiawan, R. (2021) Apa itu Data Mining dan Bagaimana Metodenya?, [dicoding.com](https://www.dicoding.com). Available at: <https://www.dicoding.com/blog/apa-itu-data-mining/> (Accessed: 29 March 2022).
- Sharda, R., Delen, D. and Turban, E. (2021) *Analytics, Data Science, & Artificial Intelligence for Decision Support* 11th Edition. Pearson.
- Alifia, N., & Rikumahu, B. (2020). Prediksi Financial Distress Perusahaan Pertambangan Batubara Di Bursa Efek Indonesia. *Universitas Telkom*, 2(4), 3. <http://ejournalmitramanajemen.com/index.php/jmm/article/view/125/69>.
- North, M. (2021). *Data Mining for the Masses*. Athens: Global Text Project.
- Jovansgha Avegad1, A. W. (2019). *Data Mining Klasifikasi Untuk Memprediksi Status Keberlanjutan Polis Asuransi Kesehatan Dengan Algoritma Naïve Bayes*. JUSTIN.