

BAB II

TINJAUAN PUSTAKA

2.1. Deposito

2.1.1. Pengertian Deposito

Menurut Undang – Undang Nomor 10 tahun 1998, yaitu deposito adalah simpanan yang penarikannya hanya dapat dilakukan pada waktu tertentu berdasarkan perjanjian nasabah penyimpan dengan bank [1]. Deposito merupakan salah satu tempat bagi nasabah untuk melakukan investasi dalam bentuk surat-surat berharga. Pemilik deposito disebut deposan. Kepada setiap deposan akan diberikan imbalan bunga atas depositonya. Bagi bank, bunga yang diberikan kepada para deposan merupakan bunga yang tertinggi, jika dibandingkan dengan simpanan giro atau tabungan, sehingga deposito oleh sebagian bank dianggap sebagai dana mahal.

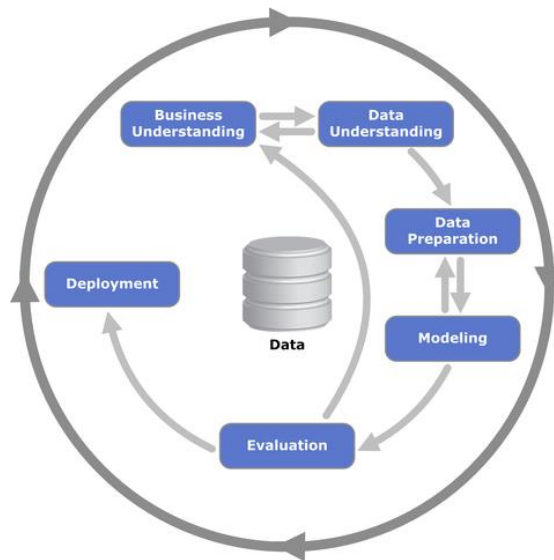
2.1.2. Jenis Deposito

Menurut [2] deposito ini memiliki 3 jenis yaitu Deposito Berjangka, Sertifikat Deposito dan *Deposito On-Call*.

1. Deposito Berjangka, yaitu deposito yang sangat umum dikenal masyarakat luas. Deposito berjangka adalah jenis tabungan berjangka yang penarikannya hanya dapat dilakukan pada waktu tertentu. Deposito berjangka diterbitkan bisa dengan atas nama perorangan maupun lembaga. Pihak bank akan memberikan bunga ke tabungan deposito kita. tingkat bunga yang ditawarkan oleh bank lebih tinggi dibandingkan dengan tabungan biasa. Namun deposito berjangka ini mengharus kita untuk menyimpan uang pada waktu tertentu. Uang yang disimpan hanya bisa diambil ketika jatuh tempo oleh pihak yang tertera pada bilyetnya.
2. Sertifikat Deposito, Merupakan sertifikat yang tidak mengacu pada nama seseorang atau lembaga tertentu, sehingga dapat dipindahtangankan dan sangat mungkin untuk diperjualbelikan.
3. Deposito On Call, yaitu tabungan berjangka dengan waktu penyimpanan yang relatif singkat, minimal 7 hari dan paling lama hanya kurang dari 1 bulan. Deposito ini dikhususkan dalam jumlah yang besar.

2.2. CRISP-DM

Cross-Industry Standard Process untuk Data Mining atau CRISP-DM dikembangkan tahun 1996 oleh analis dari beberapa industri seperti Daimler Chrysler, SPSS dan NCR [14]. CRISP-DM merupakan standarisasi proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian [14]. Dalam CRISP-DM sebuah proyek data mining memiliki siklus hidup yang terbagi dalam enam fase.



Gambar 2.1 Siklus hidup dalam CRISP-DM

Sumber: [3]

Keseluruhan fase yang ada pada CRISP-DM bersifat adaptif. Fase berikutnya dalam urutan tergantung pada fase sebelumnya. Hubungan penting antar fase yaitu dihubungkan dengan panah. Sebagai contoh yaitu jika sebuah fase *modeling* berdasarkan pada perilaku dan karakteristik data [14].

Dari gambar 2.1 dapat dilihat terdapat 6 siklus hidup dari CRISP-DM yaitu:

1. *Business Understanding*
2. *Data Understanding*
3. *Data Preparation*
4. *Modeling*
5. *Evaluation*
6. *Deployment*

2.3. Classification

Klasifikasi merupakan proses pembelajaran suatu fungsi yang menetapkan suatu atribut x ke satu dari label kelas y yang didefinisikan sebelumnya. Untuk mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan sebuah data kedalam kelas yang berbeda sesuai dengan aturan dan fungsi tertentu. Model fungsi dan aturan tersebut bisa dalam bentuk hitungan sistematis atau pohon keputusan [4].

2.4. Naive Bayes

Algoritma Naive Bayes menurut ahli matematika Inggris Thomas Bayes. Algoritma Bayesian adalah salah satu teknik klasifikasi statistik dan didasarkan pada teorema statistik Bayesian. Pengklasifikasi Bayes adalah model prediktif, lebih mudah diterapkan. Naive Bayes adalah algoritma klasifikasi yang menunjukkan hubungan antara variabel independen dan variabel target [15].

2.5. Decision Tree

Algoritma C4.5 telah dikembangkan oleh Ross Quinlan. Algoritma ini membuat rasio Penguatan digunakan dalam pohon keputusan C4.5. Algoritma C4.5 dapat digunakan pada dataset yang memiliki tipe data kategorikal dan numerik. Pohon Keputusan yang dihasilkan oleh C4.5 dapat digunakan untuk klasifikasi oleh karena itu, C4.5 umumnya disebut statistic penggolong [16].

2.6. Neural Network

Dalam studi ini, salah satu model JST yang banyak digunakan, *feed forward neural network* digunakan untuk peramalan. Biasanya, model NN terdiri dari satu lapisan masukan, satu lapisan keluaran, dan satu atau lebih lapisan tersembunyi. Lapisan tersembunyi dapat menangkap hubungan non-linier antar variabel. Setiap lapisan terdiri dari beberapa neuron yang terhubung ke neuron di lapisan yang berdekatan. Jaringan saraf dapat dilatih oleh data historis dari deret waktu untuk menangkap karakteristik non-linier dari deret waktu tertentu [5].

2.7. K-NN

Algoritma K-NN ini adalah salah satu algoritma paling dasar berbasis sampel algoritma pembelajaran. Dalam proses pembelajaran algoritma ini adalah dilakukan dengan data dalam set data training. Sampel baru adalah

diklasifikasikan menurut kesamaan dalam sampel di perlengkapan data training [14]. Algoritma k-terdekat tetangga untuk menemukan k sampel yang paling dekat dengan data yang tidak diketahui dengan cara melihat ruang pola untuk menemukan kelas dari data yang tidak diketahui. Jarak dihitung dengan metode penghitungan seperti Euclidean dan Manhattan, dan jarak antar tetangga ditemukan. Data yang tidak diketahui ditetapkan ke nilai kelas yang paling mirip dengan tetangga terdekat [17].

2.8. Random Forest

Pada prinsipnya, *random forest* terdiri dari banyak pohon keputusan dalam tetapi tidak berkorelasi yang dibangun di atas sampel data yang berbeda [18]. Proses membangun *random forest* itu sederhana. Untuk setiap pohon keputusan, pertama-tama kami membuat subset sebagai sampel secara acak dari kumpulan data asli. Kemudian, kami menumbuhkan pohon keputusan dengan sampel ini hingga kedalaman JRF maksimumnya. Sedangkan fitur mRF yang digunakan pada setiap split dipilih secara acak dari fitur p. Setelah mengulangi prosedur beberapa kali dengan dataset asli, pohon keputusan nRF dihasilkan. Hasil akhir adalah kumpulan dari semua pohon keputusan, dan klasifikasi dilakukan melalui pemungutan suara mayoritas. Kompleksitas komputasi

2.9. Feature Selection

Teknik *feature selection* relevan dengan pembuatan model prediktif. Banyak metode dan metrik peringkat fitur yang digunakan oleh para peneliti untuk memilih fitur yang relevan [19]. Pemilihan fitur sangat penting untuk prediksi Alzheimer dengan menggunakan data multi-modal juga. Relevansi atau kombinasi dari tipe ciri apa yang paling penting dalam membedakan DA dari demensia lain belum digeneralisasikan [19].

2.10. Accuracy

Akurasi didefinisikan sebagai persentase prediksi yang benar untuk data pengujian. Ini dapat dihitung dengan mudah dengan membagi jumlah prediksi yang benar dengan jumlah total prediksi [6].

$$Accuracy = \frac{\text{correct prediction}}{\text{all predictions}}$$

Rumus 2.1 Rumus validasi akurasi

2.11. Precision

Presisi didefinisikan sebagai bagian dari contoh yang relevan (positif benar) di antara semua contoh yang diperkirakan termasuk dalam kelas tertentu [6].

$$Precision = \frac{true\ positives}{true\ positives + false\ positives}$$

Rumus 2.2 Rumus validasi presisi

2.12. F-Measure

F-measure biasa disebut F1 yaitu secara efektif mereferensikan positif benar ke Rata-rata Aritmatika dari Positif yang Diprediksi dan Positif Nyata, menjadi tarif yang dikonstruksikan yang dinormalisasi ke nilai yang diidealkan, dan dinyatakan dalam bentuk ini, hal itu dikenal dalam statistik sebagai Proporsi dari Perjanjian Khusus karena itu adalah diterapkan ke kelas tertentu, sehingga diterapkan ke Kelas Positif [7].

$$F1\ Score = 2 * \frac{recall * precision}{recall + precision}$$

Rumus 2.3 Rumus validasi f-measure

2.13. Perbandingan Penelitian Terdahulu

Penelitian ini melakukan rujukan dari beberapa jurnal sebagai referensi metode dan algoritma yang digunakan. Dilihat dari tabel 2.1 berisi judul, penulis, dan algoritma serta metode penelitian yang digunakan. Untuk penerapan Teknik *feature selection* merujuk pada 2 jurnal yaitu dari [8] dan [9]. Untuk penerapan algoritma *Neural Network* merujuk pada jurnal [8]. Untuk penerapan algoritma *Decision Tree* merujuk pada jurnal [10]. Untuk penerapan algoritma *Naïve Bayes*, *K-NN*, dan *Random Forest* merujuk pada jurnal [9].

Tabel 2.1 Tabel Perbandingan Penelitian Terdahulu

| Jurnal / Penulis (Tahun) | Algoritma (Akurasi) | | | | | Metode |
|---|---------------------|-------|-------|-------|-------|---|
| | NB | DT | K-NN | NN | RF | |
| A Data Driven Approach to Predict the Success of Bank Telemarketing | - | 83.6% | - | 91.4% | - | Data mining dan feature selection |
| Classification Of a Bank Data Set on Various Data Mining Platforms | 86.9% | 88% | 84.6% | - | - | Data mining, holdout training testing |
| Using data mining for bank direct marketing: An application of the crisp-dm methodology | 87% | 86.8% | - | - | - | Data mining, Crisp DM |
| Perbandingan Teknik Klasifikasi Dalam Data Mining Untuk Bank Direct Marketing | 97.2% | 97.2% | 97.1% | - | 97.1% | Data mining, feature selection |
| Predicting Customer Response to Bank Direct Telemarketing Campaign | - | 84.7% | - | - | - | Data mining |
| Enhancing Bank Direct Marketing throught Data Mining | 82.70 | 79% | - | - | - | Data mining, Crisp DM, Confusion matrix |

2.14. Penelitian Terdahulu

Pada tabel 2.2 ditampilkan penelitian terdahulu yang memiliki hasil dan metode yang cukup sejenis dengan penelitian ini.

Tabel 2.2 Tabel Penelitian Terdahulu

| Nama Jurnal /Judul | Penulis/Tahun /Vol / Pages | Hasil | Kesimpulan |
|--|--|--|--|
| 2018 Electric Electronics / Classification of A Bank Data Set on Various Data Mining Platforms | Basarslan, M.S. (2018) / Vol 978 / Pages 1-4 | Decision Tree memiliki nilai yang paling baik setelah dikalkulasikan dengan validasi model. Memiliki akurasi sebesar 90% | Disarankan bahwa studi lebih lanjut juga diberlakukan untuk mendukung hasil model ini dengan bekerja sama menggunakan dataset selain data bank. |
| Procedia CIRP / DMME: Data mining methodology for engineering applications | Stefen Huber (2019) / Vol 79 / Pages 403-408 | Hasil pada pendekatan CRISP-DM dengan menggunakan model K-NN memiliki range akurasi sebesar 90% - 100% | Untuk pekerjaan di masa depan, kami bertujuan untuk mengembangkan DMME lebih lanjut untuk mendukung tingkat sub tugas serta tampilan tambahan termasuk data, kualitas data, sistem IT, teknik dan perspektif bisnis. |
| Using data mining for bank direct marketing | Sergio Moro and Raul M. S (2011) / Vol N/A / Pages 117-121 | Hasilnya yaitu SVM adalah prediksi terbaik karna memiliki AUC dan ROE tertinggi sebesar 90% | Hasil penting lainnya adalah konfirmasi teknologi sumber terbuka di bidang DM yang mampu menyediakan model berkualitas. |

Tabel 2.3 Tabel Penelitian Terdahulu (Cont.)

| Nama Jurnal /Judul | Penulis/Tahun/Vol / Pages | Hasil | Kesimpulan |
|--|---|--|---|
| <p>Jurnal ULTIMA InfoSys / Analisis Data Pembayaran Kredit Nasabah Bank Menggunakan Metode Data Mining</p> | <p>Ira Melissa dan Raymond S. Oetama (2013) / Vol 4 / Pages 18-27</p> | <p>Proses undersampling 50% paling banyak menebak jumlah nasabah dengan status kredit bad. Maka dapat disimpulkan metode yang paling baik untuk mendeteksi jumlah nasabah dengan status kredit bad pada dataset german credit adalah metode undersampling 50%.</p> | <p>Data german credit merupakan data well preprocessed yang berarti bahwa data german credit sudah melalui proses data cleaning jadi tidak perlu dilakukan proses pembersihan data.</p> |