

Analisa Kelayakan Kredit Menggunakan Artificial Neural Network dan Backpropogation (Studi Kasus German Credit Data)

Achmad Yani¹ dan Ega Hegarini²

¹Magister Manajemen Sistem Informasi, Universitas Gunadarma

²Sistem Informasi, Universitas Gunadarma

Jalan Margonda Raya 100, Depok

E-mail : achmadyani308@gmail.com, ega@staff.gunadarma.ac.id

Abstrak

Artificial Neural Network (ANN) merupakan cabang dari ilmu kecerdasan buatan. Saat ini banyak algoritma yang digunakan pada ANN, salah satunya adalah backpropogation. Manfaat ANN dan backpropogation dapat digunakan untuk menganalisa kelayakan kredit. Penelitian ini menggunakan German Credit Dataset dengan metode ANN dan backpropogation. Pada penelitian ini Arsitektur ANN menggunakan 24 input layer, 1 hidden layer, 20 hidden neuron, 1 output layer, iterasi sebanyak 10000, fungsi aktivasi menggunakan sigmoid biner dan learning rate sebesar 0.1. Hasil dari pengujian ditampilkan dalam bentuk confusion matrix yang akan digunakan untuk perhitungan akurasi dan perhitungan performa. Hasil analisa ANN dan backpropogation berhasil dilakukan dan mendapatkan nilai akurasi sebesar 0.71333 (71.333%) dan nilai AUC sebesar 0.72 (72%) dan termasuk ke dalam Fair Model.

Kata Kunci : Artificial Neural Network, Backpropogation, German Credit Dataset.

Pendahuluan

Kecerdasan buatan (AI) adalah teknologi penting yang mendukung kehidupan sosial sehari-hari dan kegiatan ekonomi. Saat ini pengembangan dan pemanfaatan perangkat AI telah mendapatkan momentum selama beberapa tahun terakhir [2]. Artificial Neural Network (ANN) merupakan cabang dari ilmu dari ilmu kecerdasan buatan. ANN dibentuk sebagai generalisasi model matematika dari jaringan syaraf manusia [3] [7]. ANN bekerja dengan mempelajari data masa lalu sehingga mempunyai kemampuan untuk memberi keputusan terhadap data yang belum pernah dipelajari [8]. Algoritma yang digunakan dalam ANN, salah satunya adalah backpropogation. Algoritma backpropogation adalah algoritma yang sangat baik untuk menangani masalah pengenalan pola yang kompleks. Manfaat ANN dan backpropogation dapat digunakan untuk menganalisa kelayakan kredit.

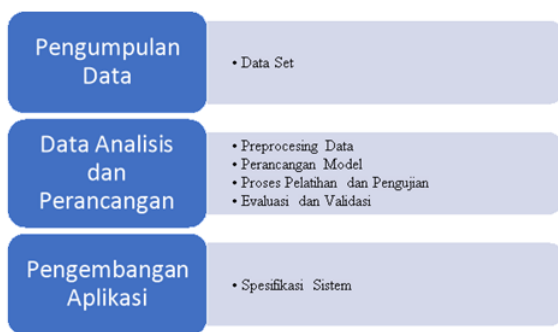
Penelitian Terdahulu

Penelitian terdahulu adalah acuan penulis dalam melakukan penelitian ini sehingga penulis dapat memperkaya teori yang digunakan untuk mengkaji penelitian yang dilakukan. Penelitian pertama oleh Aldi Nurzahputra dan Much Aziz Muslim dengan judul “Peningkatan Akurasi Pada Algoritma C4.5 Menggunakan Adaboost Untuk Meminimalkan Resiko Kredit”. Penelitian ini menyimpulkan Algoritma C4.5 mendapatkan Akurasi Sebesar 70.5% sedangkan untuk algoritma C4.5 dengan penambahan adaboost mendapatkan akurasi sebesar 74.2% [12]. Penelitian kedua oleh Heru Mardiansyah dengan judul “Penanganan Masalah Data Kredit Untuk Kelas Tidak Seimbang Menggunakan SmotexBoost”. Penelitian ini menyimpulkan metode SMOTEXGBoost memiliki tingkat akurasi yang paling tinggi, algoritma (LR,RF,SVM) memiliki kelemahan dalam prediksi dalam data kelas tidak seimbang, nilai F1-Measure > 0.5 dan nilai AUC > 90 [9]. Penelitian ketiga oleh Rinawati den-

gan judul “Penentuan Penilaian Kredit Menggunakan Metode Naive Bayes Berbasis Particle Swarm Optimization”. Penelitian ini menyimpulkan nilai akurasi dan AUC menggunakan Naive Bayes adalah 72.40% dan 0.765 sedangkan menggunakan Naive Bayes berbasis Particle Swarm Optimization adalah 75.90% dan 0.773 [14]. Penelitian keempat oleh Ni Luh Gede Pivin Suwirmayanti dengan judul “Penerapan Metode Fuzzy C-Means Untuk Pengelompokan Data Kredit”. Penelitian ini menyimpulkan perancangan sistem ini menggunakan Use Case Diagram dan telah berhasil diterapkan metode Fuzzy C-Means untuk mengelompokkan data kredit [16]. Penelitian kelima oleh Dwiza Riana Nurajijah dengan judul “Algoritma Naive Bayes, Decision Tree, dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah”. Penelitian ini menyimpulkan Klasifikasi kredit macet dan lancar, metode SVM memiliki akurasi paling baik dengan nilai 89.86% dan nilai AUC adalah 0.939 [11].

Metode Penelitian

Metode penelitian diuraikan dalam skema penelitian untuk memberikan gambaran yang jelas, teratur dan sistematis.



Gambar 1: Skema Penelitian

Penelitian ini menggunakan dataset. Dataset dapat terdiri dari kumpulan file atau dokumen [15]. Dataset yang digunakan pada penelitian ini adalah german credit data yang disediakan oleh Professor Dr. Hans Hofmann Institut f’ur Statistik und ”Okonometrie Universit”at Hamburg [6]. Dataset ini berjumlah 1000 data dan termasuk dalam jenis data kelas tidak seimbang. Pada penelitian ini penulis

menggunakan perangkat lunak R karena paling populer dan banyak digunakan untuk statistik, data mining, dan machine learning [1]. Selanjutnya penulis melakukan preprocessing data. Preprocessing data merupakan langkah penting dalam data mining [4]. Preprocessing data menggunakan metode transformation dan model normalisasi. Transformation mengubah tipe atribut input int menjadi numeric dan atribut target menjadi factor. Normalisasi menggunakan metode min max.

$$New = (Data - MinData)/(MaxData - MinData) \quad (1)$$

New = Hasil normalisasi data

Data = Data yang akan dinormalisasi

MinData = Nilai minum data

MaxData = Nilai maximum data

Arsitektur ANN yang digunakan pada penelitian ini terlihat pada tabel 1.

Tabel 1: Arsitektur ANN

| Karakteristik | Spesifikasi |
|-----------------------------|---------------|
| Jumlah <i>Input</i> | 24 |
| Jumlah <i>hidden layer</i> | 1 |
| Jumlah <i>hidden neuron</i> | 20 |
| Jumlah <i>Output</i> | 1 |
| Jumlah Maximal Iterasi | 10000 |
| Fungsi Aktivasi | Sigmoid Biner |
| <i>Learning Rate</i> | 0.1 |

Untuk mengukur evaluasi dan validasi dari model yang dibuat menggunakan beberapa teknik sebagai berikut.

A. Confusion Matrix

Untuk melihat akurasi dari hasil yang diklasifikasikan dengan benar Confusion matrix adalah metode yang menggunakan table matriks. Jika data set terdiri dari dua kelas, maka yang satu akan dianggap positif dan lainnya akan dianggap negatif [13]. Pada grafis antara False Positive terhadap True Positive disebut dengan ROC [10]. ROC akan menampilkan grafik dua dimensi dengan false positive sebagai

garis horizontal dan true sebagai garis vertical [17].

Tabel 2: Confusion Matrix

| | | True Class | |
|-----------------|----------|----------------------------|----------------------------|
| | | Positive | Negative |
| Predicted Class | Positive | true positives count (TP) | false negatives count (FP) |
| | Negative | false positives count (FN) | true negatives count (TN) |

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (2)$$

B. AUC

AUC digunakan menghitung performa metode yang digunakan.

$$\frac{1}{2} \sum_{i=1}^n (X_{i+1} - X_i)(Y_{i+1} - Y_i) \quad (3)$$

Tabel 3: Nilai AUC

| AUC | Interpretation |
|------------------------|------------------------|
| 1.0 (100%) | <i>Perfect Model</i> |
| 0.9 – 0.99 (90 - 99%) | <i>Excellent Model</i> |
| 0.8 – 0.89 (80 – 89%) | <i>Very Good Model</i> |
| 0.7 – 0.79 (70 – 79%) | <i>Fair Model</i> |
| 0.51 – 0.69 (51 – 69%) | <i>Poor Model</i> |
| < 0.5 (50%) | <i>Worthless Model</i> |

Analisis dan Pembahasan

Data yang digunakan untuk pelatihan sebanyak 700 dan pengujian sebanyak 300. Data yang akan dilakukan pelatihan perlu dilakukan preprocessing data agar mendapatkan hasil yang lebih akurat. Preprocessing data pada penelitian ini adalah transformasi dan normalisasi menggunakan metode min max.

```
'data.frame': 1000 obs. of 25 variables:
 $ V1 : num 1 2 4 1 1 4 4 2 4 2 ...
 $ V2 : num 6 48 12 42 24 36 24 36 12 30 ...
 $ V3 : num 4 2 4 2 3 2 2 2 2 4 ...
 $ V4 : num 12 60 21 79 49 91 28 69 31 52 ...
 $ V5 : num 5 1 1 1 1 5 3 1 4 1 ...
 $ V6 : num 5 3 4 4 3 3 5 3 4 1 ...
 $ V7 : num 3 2 3 3 3 3 3 3 1 4 ...
 $ V8 : num 4 2 3 4 4 4 4 2 4 2 ...
 $ V9 : num 1 1 1 2 4 4 2 3 1 3 ...
 $ V10: num 67 22 49 45 53 35 53 35 61 28 ...
 $ V11: num 3 3 3 3 3 3 3 3 3 3 ...
 $ V12: num 2 1 1 1 2 1 1 1 1 2 ...
 $ V13: num 1 1 2 2 2 2 1 1 1 1 ...
 $ V14: num 2 1 1 1 2 1 2 1 1 ...
 $ V15: num 1 1 1 1 1 1 1 1 1 1 ...
 $ V16: num 0 0 0 0 1 0 0 0 0 1 ...
 $ V17: num 0 0 0 0 0 0 0 1 0 0 ...
 $ V18: num 1 1 1 0 1 1 1 1 1 1 ...
 $ V19: num 0 0 0 0 0 0 0 0 0 0 ...
 $ V20: num 0 0 0 0 0 0 0 1 0 0 ...
 $ V21: num 1 1 1 0 0 0 1 0 1 1 ...
 $ V22: num 0 0 0 0 0 0 0 0 0 0 ...
 $ V23: num 0 0 1 0 0 1 0 0 1 0 ...
 $ V24: num 1 1 0 1 1 0 1 0 0 0 ...
 $ V25: Factor w/ 2 levels "0","1": 1 2 1 1 2 1 1 1 1 2 ...
```

Gambar 2: Hasil Transformasi

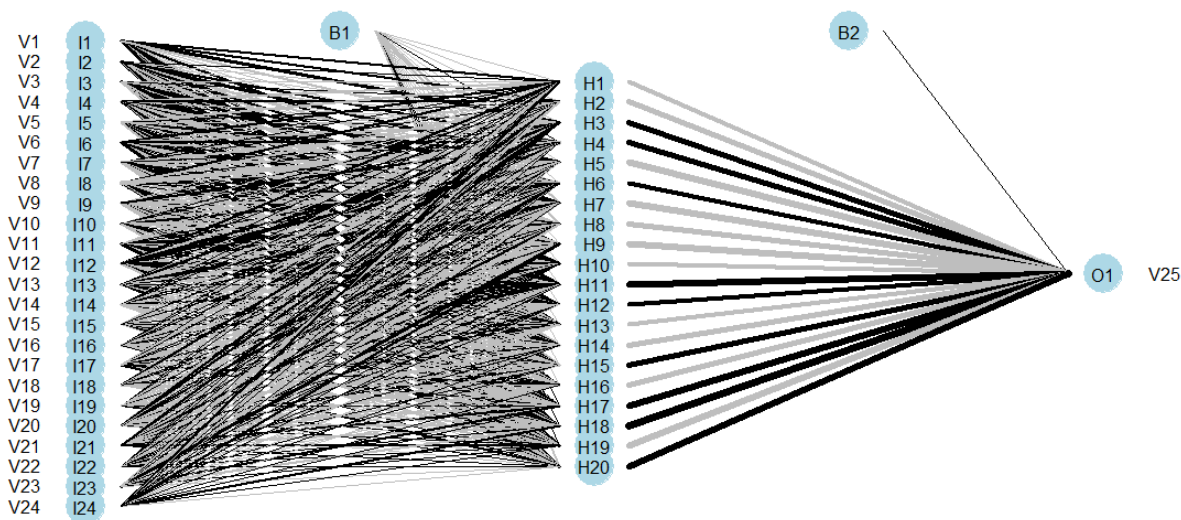
Setelah preprocessing data dilakukan selanjutnya adalah pelatihan dan pengujian. Pelatihan dan pengujian dilakukan dengan asitektur yang telah dijelaskan pada Tabel 1. Proses pelatihan berhasil dibuat dengan 24 input, 20 hidden neuron dan 1 output dengan bobot 521.

Tabel 4: Nilai AUC

| No | V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 | ... | V24 |
|------|-----------|------------|------|------------|------|------|-----------|-----------|-----|-----|
| 1 | 0.0000000 | 0.02941176 | 1.00 | 0.05494505 | 1.00 | 1.00 | 0.6666667 | 1.0000000 | ... | 1 |
| 2 | 0.3333333 | 0.64705882 | 0.50 | 0.31868132 | 0.00 | 0.50 | 0.3333333 | 0.3333333 | ... | 1 |
| 3 | 1.0000000 | 0.11764706 | 1.00 | 0.10439560 | 0.00 | 0.75 | 0.6666667 | 0.6666667 | ... | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1000 | 0.3333333 | 0.60294118 | 1.00 | 0.24175824 | 0.25 | 0.00 | 0.6666667 | 1.0000000 | ... | 1 |

```
a 24-20-1 network with 521 weights
inputs: v1 v2 v3 v4 v5 v6 v7 v8 v9 v10 v11 v12 v13 v14 v15 v16 v17 v18 v19 v20 v21 v22 v23 v24
output(s): v25
options were - entropy fitting decay=0.001
```

Gambar 3: Hasil Pelatihan

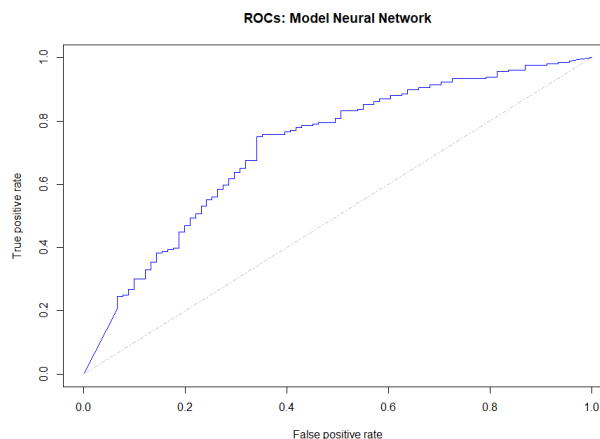


Gambar 4: Arsitektur ANN

Model pelatihan yang telah dibuat akan digunakan untuk pengujian. Pada pengujian menggunakan 300 data mendapatkan table confusion matrix dengan nilai sebagai berikut :

Tabel 5: Hasil Confusion Matrix

| | Good | Bad |
|------|------|-----|
| Good | 48 | 43 |
| Bad | 43 | 166 |



Gambar 5: Hasil ROC

$$Accuracy = \frac{48+166}{48+166+43+43} = 0,7133$$

Setelah proses pengujian selesai maka akan dihitung performa dengan menggunakan AUC. AUC menghitung nilai yang berada dibawah kurva ROC dengan persamaan 3. Hasil AUC adalah sebagai berikut:

```

| Model          | AUC |
| :-----:    | :---:|
| Neural net    | 0.72 |
>
    
```

Penutup

Analisa kelayakan kredit menggunakan Artificial Neural Network dan Backpropogation dengan arsitektur 24 input layer, 1 hidden layer, 20 hidden neuron, 1 output layer, iterasi sebanyak 10000, fungsi aktivasi menggunakan sigmoid biner dan learning rate sebesar 0.1 berhasil dilakukan. Hasil dari pengujian ditampilkan dalam bentuk confusion matrix dengan nilai True Positives 48, False Negatives 43, False Positives 43 dan True Negatives 166. Berdasarkan nilai Confusion Matirx dapat di-

hitung nilai akurasi sebesar 0.71333 (71.333%) dan nilai performa nilai sebesar 0.72 (72%). Nilai performa terdapat pada range 70% sampai dengan 79% sehingga termasuk kedalam Fair Model.

Saran dari hasil penelitian antara lain:

1. Bagi pembaca, hasil penelitian ini diharapkan dapat menambah wawasan pengetahuan terkait ANN Backpropogation.
2. Bagi peneliti selanjutnya, agar dapat mengembangkan hasil penelitian ini dengan metode yang lain agar mendapatkan nilai accuracy yang lebih tinggi dan performa yang baik.

Daftar Pustaka

- [1] B. Bischl, M. Lang, L. Kottho, J. Schiner, J. Richter, E. Studerus, G. Casalicchio and Z. M. Jones, “Mlr : Machine Learning in R”, *Journal of Machine Learning Research*, Vol 17(2016) : 1-5, 2016.
- [2] Y. Duan, J.S. Edwards & Y.K. Dwivedi, “Artificial intelligence for decision making in their a of Big Data – evolution, challenges and research agenda”, *International Journal of Information Management*. 48(9) : 63–71, 2019.
- [3] Laurene Fausett, “Fundamentals of Neural Networks: Architectures, Algorithms, and Applications”, London: Prentice Hall, Inc., 1994.
- [4] Jiawei Han, Micheline Kamber, & Jian Pei, “Data Mining Concepts and Techniques”, Elsevier Inc, 2012.
- [5] S. Haykin, “Neural Networks and Learning Machines”, 3rd edition, New Jersey: Pearson International Edition, 2009.
- [6] H. Hofmann, “Statlog (German Credit Dataset)”, akses daring di <https://archive.ics.uci.edu/ml/machine-learning-databases/statlog/german> diakses tanggal 15 September 2019 pukul 13.56 WIB, 2000.
- [7] Sihem Khemakhem & Younes Boujelbene, “Credit Risk Prediction : A Comparative Study Between Discriminat Analysis And The Neural Netwok Approach”, *Accounting And Management Information System*. 14(1) : 60–78, 2015.
- [8] Sri Kusumadewi, “Artificial Intelligence (Teknik dan Aplikasinya)”, Yogyakarta : Graha Ilmu, 2003.
- [9] Heru Mardiansyah, “Penanganan Masalah Data Kredit Untuk Kelas Tidak seimbang Menggunakan Smotexgboost”, Tesis Magister Universitas Sumatera Utara, 2019.
- [10] Naser Mohammadi dan Maryam Zangeneh, “Customer Credit Risk Assessment Using Artificial Neural Networks”, *I.J. Information Technology and Computer Science*. 8(3): 58–66, 2016.
- [11] Dwiza Riana Nurajijah, “Algoritma Naïve Bayes, Decision Tree, dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah”, *Jurnal Teknologi dan Sistem Komputer*. 7(2): 77-82, 2019.
- [12] Aldi Nurzahputra, Much Aziz Muslim, “Peningkatan Akurasi Pada Algoritma C4.5 Menggunakan Adaboost Untuk Meminimalkan Resiko Kredit”, SNATIF. Kudus: FT Universitas Muria Kudus, 2017.
- [13] F David Olson and Delen Dursun, “Advance Data Mining Techniques”, German : Springer, 2008.
- [14] Rinawati, “Penentuan Penilaian Kredit Menggunakan Metode Naive Bayes Berbasis Particle Swarm Optimization”, *Jurnal Sains Komputer & Informatika (J-SAKTI)*. 1(1): 48-58, 2017.
- [15] C. Snijders, U. Matzat and U.-D. Reips, U.-D, “Big Data: Big gaps of knowledge in the field of Internet”, *International Journal of Internet Science*. 7(1): 1–5, 2012.
- [16] Suwirmayanti, Ni Luh Gede Pivin, “Penerapan Metode Fuzzy C-Means Untuk Pengelompokan Data Kredit”, Seminar Nasional Sistem Informasi dan Teknologi Informasi 2018, Pontianak:12 Juli 2018, Hal.390-395, 2018.
- [17] C. Vercellis, “Business Intelligent: Data Mining and Optimization for Decision Making Southern Gate”, West Sussex: John Willey & Sons, Ltd, 2009.

Halaman ini sengaja dikosongkan.