

## ***Feature Selection Correlation-Based pada Prediksi Nasabah Bank Telemarketing untuk Deposito***

### ***Feature Selection Correlation-Based on Bank Telemarketing Customer Predictions for Time Deposits***

Annisa Nurul Puteri<sup>1</sup>, Arizal<sup>2</sup>, Andini Dani Achmad<sup>3</sup>

<sup>1</sup>STMIK AKBA

<sup>2</sup>Politeknik Siber dan Sandi Negara

<sup>3</sup>Universitas Hasanuddin

#### **Article Info**

##### **Article history:**

Received, 24 April 2021

Revised, 6 Mei 2021

Accepted, 19 Mei 2021

##### **Kata Kunci:**

Seleksi Fitur

Correlation Based

Klasifikasi

Prediksi

Multilayer Perceptron

#### **ABSTRAK**

*Pre-processing* merupakan tahap yang penting dalam melakukan klasifikasi data. *Pre-processing* berguna untuk mempersiapkan data sehingga teknik klasifikasi yang diterapkan menghasilkan pola yang berkualitas dan akurat. Salah satu teknik data *pre-processing* yang sering digunakan untuk mengetahui atribut yang paling berpengaruh pada sebuah dataset adalah *feature selection*. Data yang digunakan dalam penelitian ini adalah *customer data collection* dari a Portuguese banking institution in UCI Machine Learning Repository. Penelitian ini menggunakan metode *feature selection correlation-based* yang dikombinasikan dengan metode klasifikasi *Multilayer Perceptron Neural Networks*. Tujuan penelitian ini untuk mengidentifikasi atribut yang paling relevan dan berpengaruh dari dataset dalam memprediksi nasabah yang potensial untuk penawaran deposito berjangka. Penelitian ini menghasilkan 10 atribut yang memiliki ranking teratas. Atribut-atribut tersebut adalah *duration*, *previous*, *contact*, *cons.price.idx*, *month*, *cons.cof.idx*, *age*, *job*, *marital*, dan *housing*. Hasil klasifikasi dari atribut yang terpilih memiliki tingkat akurasi tertinggi sebesar 80.5% dan tingkat akurasi terendah 79.1%.

#### **ABSTRACT**

*Pre-processing* is an important step in classifying data. This is useful for preparing data so that the classification technique applied can generate quality and accurate patterns. One of the data *pre-processing* techniques that are often used to determine the most influential attributes on a dataset is *feature selection*. The data used in this study is *customer data collection* from a Portuguese banking institution in UCI Machine Learning Repository. This study employs the *correlation-based feature selection* method combined with the *Multilayer Perceptron Neural Networks* classification method. The purpose of this study is to identify the most relevant and influential attributes of the dataset in predicting potential customers for time deposit offers. This study generates 10 attributes that have the highest ranking. These attributes are *duration*, *previous*, *contact*, *cons.price.idx*, *month*, *cons.cof.idx*, *age*, *job*, *marital*, and *housing*. The classification results of the selected attributes have the highest accuracy rate of 80.5% and the lowest accuracy rate of 79.1%.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



#### **Penulis Korespondensi:**

Arizal,

Program Studi Rekayasa Perangkat Keras Kriptografi,

Politeknik Siber dan Sandi Negara,

Email: arizal@poltekssn.ac.id

## 1. PENDAHULUAN

*Telemarketing* merupakan salah satu cara yang digunakan oleh bank dalam melakukan proses pemasaran. *Telemarketing* adalah proses interaktif antara perusahaan dan nasabah melalui telepon. Akan tetapi, *telemarketing* dapat menimbulkan dampak negatif ketika telemarketer bank melakukan interaksi dengan nasabah yang tidak potensial. Pelanggan merasa bahwa produk dan layanan yang ditawarkan tidak relevan dengan kebutuhan mereka [1]. Akibatnya, penawaran produk baru dari bank sering kali ditolak oleh nasabah. Bank sebagai salah satu perusahaan yang memiliki banyak nasabah dan melakukan transaksi setiap hari tentu juga memiliki data yang terus meningkat. Data-data tersebut dapat diolah dan dianalisis untuk mengidentifikasi faktor yang berpengaruh dalam keberhasilan pemasaran produk jasa. Hal ini dapat membantu dalam proses pembuatan keputusan.

Dalam mengolah data, salah satu metode komputasi yang dapat digunakan adalah data mining. Data *mining* dapat memprediksi nasabah yang potensial bagi pemasaran sebuah produk jasa. Dalam data *mining*, data *pre-processing* merupakan tahapan yang sangat penting. *Pre-processing* berguna untuk mempersiapkan data sehingga teknik data *mining* yang diterapkan menghasilkan pola yang berkualitas dan akurat [2]. Salah satu teknik data *pre-processing* yang sering digunakan untuk mengetahui atribut yang paling berpengaruh pada sebuah dataset adalah seleksi fitur. Teknik ini digunakan untuk mengurangi kompleksitas atribut yang akan diolah dan dianalisis [3]. Ada dua jenis algoritma seleksi fitur berdasarkan kriteria evaluasi, yaitu *Filter based* dan *Wrapper based approaches*. Metode Filter melakukan pengukuran korelasi, konsistensi, pengukuran berbasis teori informasi, jarak, dan lain-lain untuk menentukan ranking masing-masing fitur. Sedangkan metode *Wrapper* memungkinkan pembelajaran algoritma untuk *feature subset evaluation* [4].

Seleksi fitur sering digunakan dalam berbagai bidang, seperti pada penelitian [5]. Penelitian ini memprediksi hasil panggilan *telemarketing* dalam penjualan produk jangka panjang guna mendukung keputusan seleksi klien bagi manajer pemasaran bank menggunakan model *data-driven*. Peringkat pada atribut *input* dipilih menggunakan *Sensivity analysis*. Selain itu, *Sensitivity analysis* juga digunakan untuk menunjukkan pengaruh rata-rata dari fitur yang relevan respon dari *Neural Networks*. Hasilnya adalah *Euribor* tiga bulan terakhir dianggap sebagai atribut yang paling relevan dengan *Sensivity analysis*, kemudian atribut *direction call (outbound and inbound)*, *bank agent experience*, *different between the best possible rate*, dan *the duration of previous calls* yang perlu dijadwalkan kembali untuk mendapatkan jawaban akhir dari klien.

Penelitian untuk mengidentifikasi kelompok pelanggan dilakukan [6] pada tahun 2017. Tujuan penelitian ini adalah memprediksi respon nasabah terhadap *bank direct marketing* dengan menerapkan empat metode klasifikasi yaitu *Multilayer Perceptron Neural Network (MLPNN)*, *Decision Tree (C4.5)*, *Logistic Regression*, dan *Random Forest (RF)*. Hasil penelitian menunjukkan bahwa *Random Forest Classifier* dengan akurasi 87% merupakan klasifikasi paling produktif dalam hal kemampuan prediktif. Selain itu, penelitian ini juga mengidentifikasi fitur utama dari pelanggan yang berlangganan dan kemungkinan besar akan berlangganan deposito berjangka. Tujuan ini dicapai dengan menerapkan analisis *cluster*. Hasil dari analisis *cluster* mengungkapkan bahwa, pelanggan yang durasi panggilannya lebih lama memiliki kemungkinan yang lebih tinggi untuk berlangganan deposito berjangka. Hasil tersebut juga menunjukkan bahwa nasabah dengan minimal pendidikan sekolah menengah merupakan prospek yang baik untuk dijadikan sasaran langganan deposito berjangka.

Penelitian [7] menggunakan *predictive Artificial Neural Networks models (ANNs)* untuk membantu memprediksi apakah klien baru akan lebih cenderung menyetujui deposito berjangka berdasarkan catatan bank dari klien sebelumnya. Penggunaan *Algorithms Reduced the size of the bank dataset dramatically*; dari kumpulan data besar dengan 16 atribut, seleksi fitur menghasilkan 9 atribut yang akan digunakan. Namun, hasil prediktif dari kedua dataset dalam hal *Mean Square Error (MSE)* dan parameter regresi hampir sama. Maka dapat disimpulkan bahwa manfaat menggunakan algoritma Feature Selection adalah tingkat akurasi yang hampir sama dan waktu yang digunakan saat memproses data berkurang.

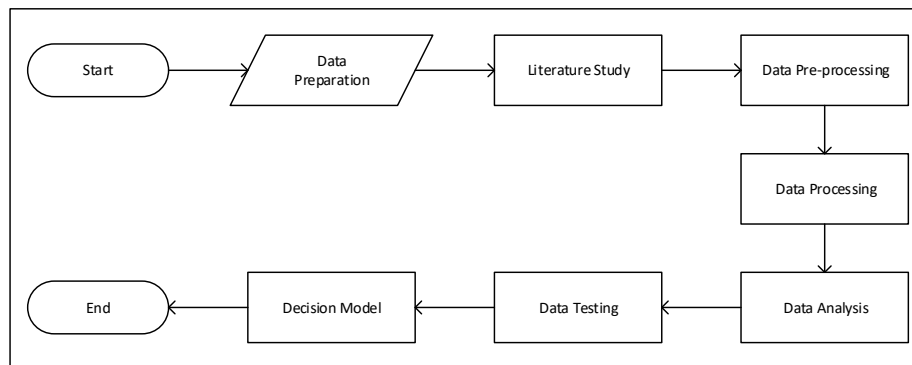
Perbandingan metode seleksi fitur dilakukan oleh [8] pada tahun 2018. Penelitian ini membandingkan *Correlation-based feature selection* dengan *Information Gain*. Data aktivitas pembelajaran siswa dalam sistem manajemen *e-learning* digunakan dalam penelitian ini. Pengujian data menggunakan *10 folds cross validation* dengan metode C4.5. Selanjutnya, *confusion matrix* digunakan untuk mengevaluasi data. Hasil dari pengujian data menunjukkan bahwa *Correlation-based feature selection* yang dikolaborasikan dengan algoritma C4.5 menghasilkan nilai akurasi yang lebih tinggi, yakni sebesar 76.92%. Penelitian [9] juga melakukan penelitian untuk memperoleh informasi baru yang memiliki makna dari data perbankan. Seleksi fitur dilakukan menggunakan *Information Gain* untuk menyeleksi data yang tidak relevan dan redundan. Hasilnya dari 21 atribut, terpilih 14 atribut yang digunakan dalam penelitian. *Feedforward Neural Networks* digunakan untuk menguji kinerja metode seleksi fitur. Peneliti membandingkan kinerja antara *Multilayer Perceptron Neural Networks (MLPNN)* dan *Radial Basis Function Neural Networks (RBFNN)*. Dari hasil pengujian diperoleh RBFNN memiliki akurasi 95.3% dan *sensitivity* 86.1%, lebih tinggi daripada MLPNN yang memiliki akurasi 88.0% dan *sensitivity* 67.4%.

Penelitian [10] menggunakan model matematika dalam menentukan kebijakan untuk memilih pelanggan yang akan ditargetkan bersama dengan jumlah panggilan yang optimal untuk digunakan untuk setiap pelanggan yang dipilih. Algoritma *K-means* digunakan dalam memilih fitur yang memengaruhi metrik untuk menentukan pelanggan mana yang harus didekati dan jumlah panggilan yang harus dilakukan kepada pelanggan. Peneliti menggunakan *5-Fold cross-validation* guna mengevaluasi metode yang diusulkan pada dataset. Hasil penelitian ini menunjukkan bahwa menerapkan metode yang diusulkan akan meningkatkan keberhasilan kampanye pemasaran jarak jauh dengan anggaran panggilan terbatas. Berdasarkan penelitian yang telah dilakukan [8] di mana *feature selection Correlation-based* menghasilkan performa lebih baik dari *Information Gain*, maka

penelitian ini menggunakan metode *feature selection Correlation-based* yang dikombinasikan dengan metode klasifikasi *Artificial Neural Networks* (ANNs). Penelitian ini bertujuan untuk mengidentifikasi atribut yang paling relevan dan berpengaruh dari dataset dalam memprediksi nasabah yang potensial untuk penawaran deposito berjangka. Pemilihan atribut pada dataset nasabah diharapkan dapat membantu pihak-pihak terkait seperti manager pemasaran dalam mengambil suatu keputusan untuk pemasaran produk deposito.

## 2. METODE PENELITIAN

Metode eksperimen adalah jenis metode yang digunakan dalam penelitian ini. Secara garis besar tahapan penelitian terdiri dari data preparation, literature study, data *pre-processing*, data *processing*, data *analysis*, data *testing*, and membuat *decision model* seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Data diperoleh dari *UCI Machine Learning Repository*, selanjutnya dilakukan studi literatur untuk memilih metode yang tepat dalam melakukan seleksi fitur dan klasifikasi data. Sebelum diproses, data diolah pada tahap *preprocessing*, yaitu pembersihan data dari data yang tidak relevan dan redundan. Setelah data dibersihkan, data diproses dengan metode klasifikasi *Multilayer Perceptron Neural Network*. Hasil klasifikasi dianalisis untuk melihat parameter-parameter yang berpengaruh pada metode yang digunakan. Tingkat keakuratan dari metode yang digunakan dilihat menggunakan *confusion matrix*. Beberapa percobaan dilakukan dengan mengubah beberapa parameter untuk menemukan *decision model* yang terbaik.

### 2.1. Dataset

Data yang digunakan dalam penelitian ini adalah *customer data collection* dari a *Portuguese banking institution* in *UCI Machine Learning Repository*. Data ini terdiri dari 21 atribut dengan total 41.118 data. Data telah dikumpulkan dari *campaigns* sebelumnya dan atribut data telah dipilih oleh bank Portages.

### 2.2. Pre-processing

Pada tahap *pre-processing*, *data cleaning* diperlukan untuk membersihkan *missing values* pada data. *Missing values* dapat berpengaruh negatif pada kinerja metode yang digunakan. Tahap *pre-processing* juga dilakukan untuk membersihkan data yang tidak relevan. Metode seleksi fitur diterapkan untuk menghapus fitur yang tidak relevan atau redundan. Karakteristik kumpulan data seperti tipe data, ukuran data, dan *noise* sangat menentukan dalam pemilihan metode seleksi fitur [11].

*Correlation Attribute Evaluation* merupakan salah satu metode seleksi fitur yang menggunakan metode pencarian *ranking*. *Correlation Attribute Evaluation* memperhatikan target *class*. Korelasi antar setiap atribut dan target class diukur menggunakan *Pearson's Correlation Method*. Setiap nilai bertindak sebagai indikator dengan mempertimbangkan nominal attributes dalam basis nilai [12].

Langkah pertama adalah menghitung korelasi fitur Pearson dalam dataset menggunakan persamaan (1). *Pearson's Correlation* dilambangkan sebagai parameter populasi. Metode ini digunakan jika kedua variabel dalam penelitian berdistribusi normal. Nilai koefisien dipengaruhi oleh nilai ekstrim yang dapat membesarkan atau mengurangi kekuatan hubungan. Oleh karena itu, tidak sesuai jika satu atau kedua variabel tidak terdistribusi secara normal.

$$\rho_i = \frac{cov(X_i, Y)}{\sigma(X_i)\sigma_Y} \quad (1)$$

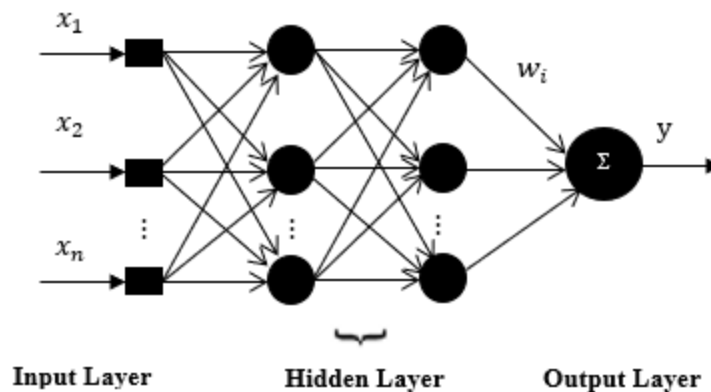
Yang paling berdampak pada target, koefisien korelasi Pearson antara Fitur dan target dituliskan dalam (1), di mana *covariance* merupakan standar deviasi; memiliki kisaran antara 0 dan 1. Langkah selanjutnya adalah memberi peringkat fitur berdasarkan koefisien korelasi absolut dengan target seperti persamaan (2) [13].

$$l_i = \begin{cases} 1, & \text{if } \rho_i > c \\ 0, & \text{if other} \end{cases} \quad (2)$$

Rumus pemilihan fitur dalam dataset baru diberikan dalam (2), di mana daftar flag dari dataset tersebut memiliki dua kemungkinan hasil. Jika nilai flag adalah 0, maka diabaikan, tetapi jika flag adalah 1, maka itu dipilih sebagai fitur dalam dataset baru.

### 2.3. Klasifikasi

Klasifikasi merupakan teknik dalam data mining untuk mengelompokkan data berdasarkan keterikatan data terhadap data sampel. Tujuannya agar model yang dihasilkan dapat digunakan untuk memprediksi kelas dari suatu data yang tidak mempunyai label kelas. Klasifikasi mampu menemukan model dari sekumpulan data yang terdiri dari beberapa fitur dan kelas sebagai fungsi dari fitur-fitur lain [14]. Ada beberapa metode yang dapat digunakan untuk klasifikasi, salah satunya adalah *Artificial Neural Network*. *Artificial Neural Network* (ANN) adalah algoritma yang dapat belajar dari kejadian yang telah dialami sebelumnya serta mampu meningkatkan kinerjanya dan menyesuaikan diri dengan perubahan lingkungan. *Artificial Neural Network* mampu memproses sejumlah besar data dan menggeneralisasi hasilnya [15]. Metode *Multilayer Perceptron Neural Networks* (MLPNN) digunakan dalam penelitian ini.



Gambar 2. Arsitektur Multilayer Perceptron Neural Networks

*Multilayer Perceptron Neural Networks* merupakan suatu kelas yang fleksibel dari fungsi-fungsi nonlinear. Penelitian ini menggunakan tiga *hidden layer* dan dua puluh *neuron* pada masing-masing *layer*. MLPNN pada gambar 2 bekerja dengan menerima suatu vektor dari input  $X$  dan kemudian menghitung suatu respon atau output  $Y(X)$  dengan memproses (propagating)  $X$  melalui elemen-elemen proses yang saling terkait. Elemen-elemen proses tersusun dalam beberapa layer dan data,  $X$  mengalir dari satu layer ke layer berikutnya secara berurutan. Dalam tiap-tiap layer, input-input ditransformasikan ke dalam layer secara nonlinear oleh elemen-elemen proses dan kemudian diproses maju ke lapis berikutnya. Akhirnya, nilai-nilai output  $Y(X)$ , yang dapat berupa nilai-nilai scalar atau vector, dihitung pada output layer [9]. Nilai-nilai respon atau output  $Y(X)$  pada metode ini dihitung dengan persamaan (3).

$$Y(X) = b_0 + \sum_{j=1}^H b_j y \left( g_{j0} + \sum_{i=1}^I g_{ji} X_i \right) \quad (3)$$

dengan  $(b_0, b_1, \dots, b_H, g_{j0}, \dots, g_{HI})$  adalah besaran-besaran bobot (*weights*) atau parameter-parameter MLPNN. Bentuk nonlinear dari fungsi  $Y(X)$  terjadi melalui suatu fungsi yang disebut aktivasi  $y$ , yang biasanya fungsi halus atau *smooth* seperti fungsi logistic sigmoid seperti persamaan (4).

$$y(Z) = \frac{1}{1 + e^{-Z}} \quad (4)$$

MLPNN dengan fungsi linear pada *output layer* dapat memberikan pendekatan sebarang yang akurat pada sebarang fungsi dalam berbagai ruang fungsi norm jika dimensi ruang bobot cukup besar.

## 2.4. Evaluasi Kinerja

Tingkat kesalahan pada metode klasifikasi dilihat menggunakan evaluasi kinerja. Salah satu cara menghitung kesalahan yang sering digunakan dalam klasifikasi adalah *confusion matrix*. Hasil proses klasifikasi pada *confusion matrix* dapat direpresentasikan sebagai *False Positive* (FP), dan *False Negative* (FN), *True Positive* (TP), *True Negative* (TN) seperti ditunjukkan pada Tabel 1. Nilai *False Positive* (FP) merupakan data negatif yang terdeteksi sebagai data positif, sedangkan *False Negative* (FN) adalah data negatif yang terdeteksi dengan benar sebagai data negatif. *True Positive* (TP) merupakan data positif yang terdeteksi benar sebagai data positif, sedangkan *True Negative* (TN) adalah data positif yang terdeteksi sebagai data negatif.

Tabel 1. *Confusion Matrix*

	Prediction	True	False
Actual			
True		TP	TN
False		FP	FN

Tingkat kedekatan data antara data yang diprediksi dengan data sebenarnya atau rasio jumlah data yang diklasifikasikan dengan benar disebut tingkat akurasi [9]. Tingkat akurasi dapat didefinisikan dalam persamaan (5):

$$Akurasi = \frac{TP + TN}{TN + FP + FN + TP} \quad (5)$$

*Recall* adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi [8]. *Recall* didefinisikan pada persamaan (6).

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

Presisi adalah tingkat ketepatan antara informasi yang diminta dengan jawaban yang diberikan oleh sistem. Presisi menggunakan persamaan (7).

$$Presisi = \frac{TP}{TP + FP} \quad (7)$$

## 3. HASIL DAN ANALISIS

Penelitian ini menggunakan *Waikato Environment for Knowledge Analysis* (WEKA) pada tahap seleksi fitur. Evaluasi kinerja dilakukan dengan Matlab. Data yang dapat diolah adalah dataset dengan 21 atribut dan total 41.118 data. Akan tetapi, merujuk pada [9] penelitian ini menggunakan data dengan jumlah 15.713 data karena adanya selisih jumlah data yang signifikan antara data pada target deposit pelanggan yang berlangganan dengan data pada target pelanggan yang tidak berlangganan. Data ini dibagi menjadi 70% data latih dan 30% data uji.

### 3.1. Preprocessing Data

Tahap *preprocessing* dilakukan dengan membersihkan data dari *missing value* yang akan berpengaruh pada hasil klasifikasi. Selanjutnya atribut data diseleksi untuk menghilangkan data yang tidak relevan dan redundan. Pada penelitian ini, seleksi fitur dilakukan dengan metode *correlation-based* dan *correlation attribute evaluation* sebagai selektor. Seleksi fitur ini menghasilkan 10 atribut dengan ranking teratas dari threshold sebesar 0.01. Oleh karena itu, penelitian ini dilanjutkan menggunakan 10 atribut data yang dapat dilihat pada Tabel 2. Informasi terkait masing-masing atribut yang terpilih ditunjukkan pada Tabel 3.

Tabel 2. Hasil Seleksi Fitur

No.	Atributte	Correlation Coefficients
-----	-----------	--------------------------

1	Duration	0.39353
2	Previous	0.228
3	Contact	0.14386
4	Cons.price.idx	0.07113
5	Month	0.06302
6	Cons.conf.idx	0.06164
7	Age	0.04873
8	Job	0.04809
9	Marital	0.01683
10	Housing	0.01007
11	Y (output)	(output)

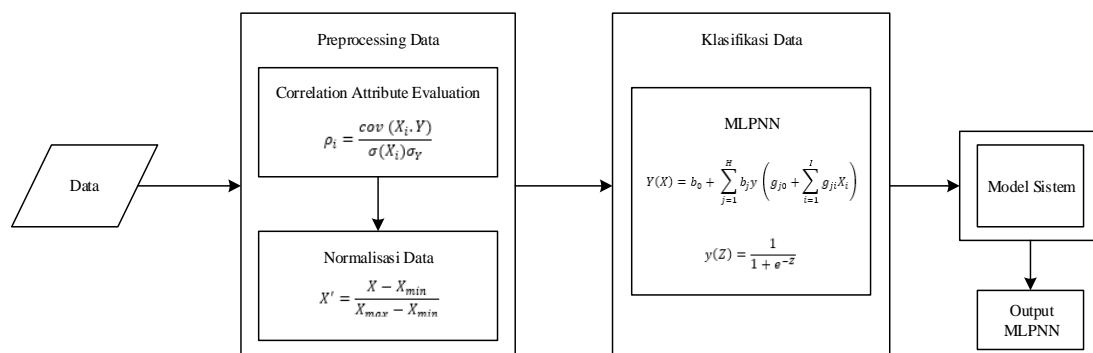
Tabel 3. Informasi Fitur

No.	Atributte	Type	Information
1	Duration	Numeric	Last contact duration
2	Previous	Numeric	Number of contacts performed before this campaign and for this client
3	Contact	Categorical	Contact communication type
4	Cons.price.idx	Numeric	Consumer price index - monthly indicator
5	Month	Categorical	Last contact month of year
6	Cons.conf.idx	Numeric	Consumer confidence index - monthly indicator
7	Age	Numeric	Consumer's age
8	Job	Categorical	Type of job
9	Marital	Categorical	Marital status
10	Housing	Categorical	Has housing loan?
11	Y (output)	Binary	Has the client subscribed a term deposit?

Kemudian tahap normalisasi dilakukan untuk mengubah setiap tipe atribut menjadi *numeric*. Hal ini dibutuhkan karena pemrosesan data hanya dapat dilakukan pada data dengan tipe *numeric*. Jenis target diubah menjadi 1 untuk kelas “Ya” dan 0 untuk kelas “Tidak”. Selain jenis target, nilai data pada setiap atribut diubah ke dalam kisaran 0 dan 1.

### 3.2. Klasifikasi

Metode klasifikasi untuk menguji hasil dari metode *feature selection correlation-based* yang digunakan dalam penelitian ini adalah *Multilayer Perceptron Neural Networks* (MLPNN). Data yang telah melalui tahap *preprocessing* akan diklasifikasi menggunakan MLPNN seperti pada Gambar 3.

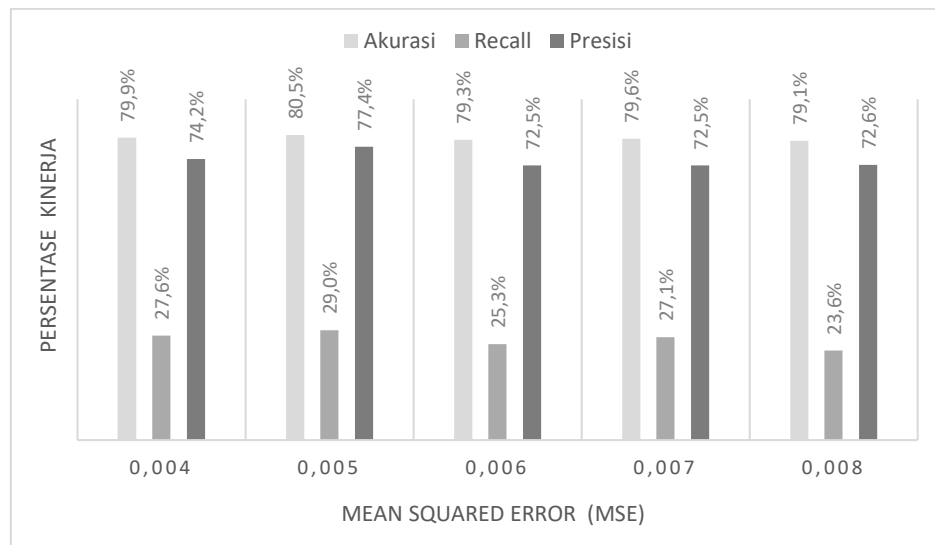


Gambar 3. Model Sistem

Klasifikasi dilakukan dengan tiga *hidden layer* dan dua puluh *neuron* pada masing-masing *layer*. Data latih digunakan untuk membentuk model jaringan dan data uji untuk menguji kinerja jaringan yang telah terbentuk. *Output* dari MLPNN dianalisis untuk melihat parameter-parameter yang berpengaruh pada metode yang digunakan. Evaluasi kinerja model sistem juga akan dilihat dari *output* dari MLPNN.

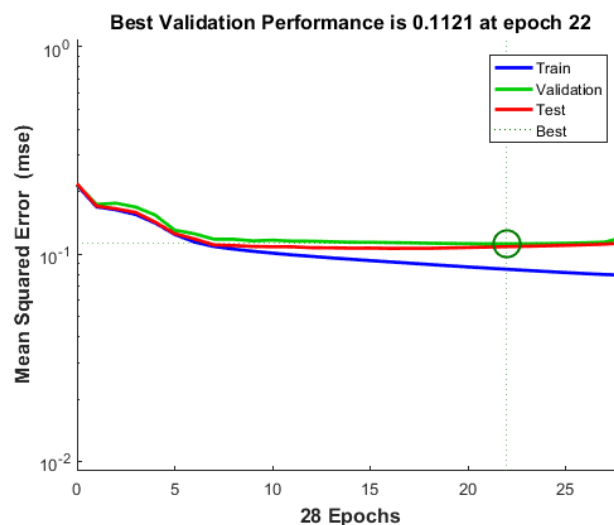
### 3.3. Analisis

Penelitian ini menggunakan beberapa percobaan untuk mendapatkan kinerja terbaik dari MLPNN. Hasil kinerja dari beberapa percobaan tersebut dapat dilihat pada Gambar 4.



Gambar 4. Persentase kinerja MLPNN dengan variasi nilai MSE

Gambar 4 menunjukkan bahwa kinerja terbaik dari hasil klasifikasi menggunakan MLPNN adalah pelatihan dengan batas *Mean Squared Error* (MSE) sebesar 0.005. Kinerja terbaik dari hasil klasifikasi MLPNN memiliki tingkat akurasi 80.5%, recall 29.0%, dan presisi 77.4%. Tingkat akurasi terendah 79.1%, *recall* 23.6%, dan presisi 72.6% berada pada batas MSE 0.008.



Gambar 5. Best Validation Performance

Tingkat akurasi tertinggi yang dihasilkan oleh batas *Mean Squared Error* (MSE) sebesar 0.005 memiliki *best validation performance* dengan nilai 0.1121. Nilai *best validation performance* ini diraih pada *epoch* ke 22 seperti yang terlihat pada Gambar 5.

## 4. KESIMPULAN

*Feature selection correlation-based* dengan *Correlation Attribute Evaluation* sebagai selektornya menghasilkan 10 atribut yang memiliki ranking teratas dari batasan *threshold* 0.01. Atribut-atribut tersebut adalah *duration*, *previous*, *contact*, *cons.price.idx*, *month*, *cons.cof.idx*, *age*, *job*, *marital*, dan *housing*. *Duration* sebagai atribut yang memiliki nilai korelasi koefisien yang tertinggi, kemudian diikuti dengan dua atribut lain, yaitu *Previous* dan *Contact*. Hal ini menjelaskan bahwa

*Duration* atau lamanya waktu yang digunakan oleh seorang telemarketer merupakan atribut yang paling relevan dan berpengaruh dalam memprediksi nasabah yang potensial untuk penawaran deposito berjangka. Selain itu, *Previous* dan *Contact* juga patut dipertimbangkan dalam memprediksi nasabah yang potensial untuk penawaran deposito berjangka.

Hasil klasifikasi dari atribut yang terpilih memiliki tingkat akurasi tertinggi sebesar 80.5% dengan nilai recall 29.0% dan nilai presisi 77.44%. Nilai akurasi tertinggi ini dihasilkan dari batasan MSE 0.005. Oleh karena itu, *feature selection correlation-based* dapat menjadi pilihan dalam mereduksi dan memilih atribut yang relevan dan berpengaruh dalam memprediksi nasabah yang potensial untuk penawaran deposito berjangka. MLPNN juga dapat digunakan untuk membuat model sistem karena menghasilkan tingkat akurasi yang tinggi dalam melakukan klasifikasi. Penelitian selanjutnya dapat dilakukan dengan mengkombinasikan metode klasifikasi lain dengan *Feature selection correlation-based* untuk mencapai tingkat akurasi yang lebih optimal.

## REFERENSI

- [1] A. E. Maulana and K. Nurulfirdausi, "Permissive, Aggressive or Apathetic? Indonesian Telemarketing Customer," in *Procedia - Social and Behavioral Sciences*, Jan. 2015, vol. 169, pp. 69–74.
- [2] J. Nalic and A. Svraka, "Importance of data pre-processing in credit scoring models based on data mining approaches," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2018, pp. 1046–1051.
- [3] I. made B. Adnyana, "Penerapan Feature Selection untuk Prediksi Lama Studi Mahasiswa," *Jurnal Sistem Dan Informatika*, vol. 13, no. 2, pp. 72–76, 2019.
- [4] N. K. Suchetha, A. Nikhil, and P. Hrudya, "Comparing the Wrapper Feature Selection Evaluators on Twitter Sentiment Classification," in *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*, 2019, pp. 1–6.
- [5] S. Moro, P. Cortez, and P. Rita, "A data-driven approach to predict the success of bank telemarketing," *Decision Support System*, vol. 62, pp. 22–31, 2014.
- [6] J. Asare-Frempong and M. Jayabalan, "Predicting customer response to bank direct telemarketing campaign," in *2017 International Conference on Engineering Technology and Technopreneurship (ICE2T)*, Sep. 2017, vol. 2017-Janua, pp. 1–4.
- [7] K. Morani, E. K. Ayana, and S. N. Engin, "Development of Prediction in Clients' Consent to a Bank Term Deposit Using Feature Selection," in *2018 6th International Conference on Control Engineering & Information Technology (CEIT)*, Oct. 2018, pp. 1–5.
- [8] A. S. B. Asmoro, W. S. G. Irianto, and U. Pujianto, "Perbandingan Kinerja Hasil Seleksi Fitur pada Prediksi Kinerja Akademik Siswa Berbasis Pohon Keputusan," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 4, no. 2, pp. 84–89, Dec. 2018.
- [9] A. N. Puteri, Dewiani, and Z. Tahir, "Comparison of Potential Telemarketing Customers Predictions with a Data Mining Approach using the MLPNN and RBFNN Methods," in *2019 International Conference on Information and Communications Technology (ICOIACT)*, Jul. 2019, pp. 383–387.
- [10] P. Hosein, S. Ramoudith, and I. Rahaman, "On the Optimal Allocation of Resources for a Marketing Campaign," in *Proceedings of the 10th International Conference on Operations Research and Enterprise Systems*, 2021, pp. 169–176.
- [11] M. Dash and H. Liu, "Feature Selection for Classification," *Intelligent Data Analysis.*, vol. 1, no. 4, pp. 131–156, 1997.
- [12] G. S, T. M, M. V.T, and G. V, "Classification Algorithms with Attribute Selection: An Evaluation Study using WEKA," *International Journal of Advanced Networking and Applications*, vol. 9, no. 6, pp. 3640–3644, 2018.
- [13] Y. Sugianela and T. Ahmad, "Pearson Correlation Attribute Evaluation-based Feature Selection for Intrusion Detection System," in *2020 International Conference on Smart Technology and Applications (ICoSTA)*, Feb. 2020, pp. 1–5.
- [14] A. Kustiyo, H. Firqiani, and E. Giri, "Seleksi Fitur Menggunakan Fast Correlation Based Filter pada Algoritma Voting Feature Intervals 5," *Jurnal Ilmu Komputer*, vol. 6, no. 2, pp. 1–12, 2008.
- [15] M. A. Halali, V. Azari, M. Arabloo, A. H. Mohammadi, and A. Bahadori, "Application of a radial basis function neural network to estimate pressure gradient in water–oil pipelines," *Journal of the Taiwan Institute Chemical Engineers*, vol. 58, pp. 189–202, Jan. 2016.