

## ANALISIS ALGORITMA KNN BERBASIS FEATURE SELECTION UNTUK MEMPREDIKSI NASABAH PENGGUNA DEPOSITO MELALUI PEMASARAN LANGSUNG

Ami Rahmawati<sup>1</sup>, Ita Yulianti<sup>2</sup>, Yuri Yuliani<sup>3</sup>, Nurhadianto<sup>4</sup>, Hafifah Bella Novitasari<sup>5</sup>

<sup>1,2,3</sup> Universitas Bina Sarana Informatika  
<sup>4,5</sup> STMIK Nusa Mandiri

E-mail: ami.amv@bsi.ac.id<sup>1</sup>, ita.iyi@bsi.ac.id<sup>2</sup>, yuri.yyi@bsi.ac.id<sup>3</sup>, dion.adiyanto@gmail.com<sup>4</sup>,  
hafifah.bella@gmail.com<sup>5</sup>

### Abstraksi

Sebuah bank menggunakan teknik pemasaran langsung dalam menargetkan segmen nasabah dengan cara menghubungi nasabah tersebut untuk memenuhi tujuan tertentu. Setelah menghubungi nasabah, bank mendapatkan informasi apakah nasabah tersebut sudah berlangganan produk yang ditawarkan oleh bank atau belum. Salah satu produk yang ditawarkan oleh bank antara lain yaitu deposito. Dari banyaknya informasi nasabah yang dikumpulkan, bank mampu menawarkan produk dan layanan kepada nasabah. Kemampuan tersebut dapat menggunakan teknologi data mining, seperti tujuan dibuatnya penelitian ini yaitu memprediksi nasabah yang berlangganan deposito dengan algoritma *K-Nearest Neighbor* (KNN) dan *feature selection* yang diproses menggunakan *tools* Anaconda dan bahasa pemrograman python. Dari hasil penelitian yang diperoleh, akurasi dari penggunaan algoritma *K-Nearest Neighbor* (KNN) sebesar 74,37% dengan nilai K=9, sedangkan akurasi algoritma klasifikasi *K-Nearest Neighbor* dengan menggunakan *feature selection* sebesar 89,72% dengan nilai K=3, sehingga didapat selisih peningkatan akurasi sebesar 15,35%.

Kata Kunci : Anaconda, Bank Marketing, KNN, Python

### Abstract

*A bank uses direct marketing techniques to target customer segments by contacting these customers to meet certain objectives. After contacting the customer, the bank gets information whether the customer has subscribed to the product offered by the bank or not. One of the products offered by banks is deposit. From the large amount of customer information collected, banks are able to offer products and services to customers. This ability can use data mining technology, such as the purpose of this research that is to predict customers who subscribe to deposits with the K-Nearest Neighbor (KNN) algorithm and feature selection that is processed using Anaconda tools and the python programming language. From the research results obtained, the accuracy of using the K-Nearest Neighbor (KNN) algorithm is 74.37% with a value of K = 9, while the accuracy of the K-Nearest Neighbor classification algorithm by using a feature selection is 89.72% with a value of K = 3, so that the difference is obtained an increase in accuracy of 15.35%.*

*Keywords: Anaconda, Bank Marketing, KNN, Python*

### 1. Pendahuluan

Pemasaran langsung merupakan cara efektif yang digunakan oleh bank untuk mempromosikan layanan atau produk (Ruangthong & Jaiyen, 2015). Sebuah bank menggunakan teknik pemasaran langsung dalam menargetkan segmen nasabah dengan cara menghubungi nasabah tersebut untuk memenuhi tujuan tertentu (Moro, Cortez, & Rita, 2014). Nasabah yang terpilih oleh bank biasanya akan dihubungi melalui kontak perorangan seperti telepon seluler ataupun email sehingga bank memperoleh informasi

langsung dari nasabah dengan mengetahui apakah nasabah tersebut sudah berlangganan produk yang ditawarkan oleh bank atau belum (Dhikhi, Kumar, Nadu, & Nadu, 2018).

Salah satu produk yang ditawarkan oleh bank antara lain deposito. Deposito merupakan simpanan yang penarikannya hanya dapat dilakukan pada waktu yang telah disepakati. Minat nasabah untuk dapat berlangganan deposito pada sebuah bank masih terbilang minim. Hal ini disebabkan karena penarikan uang deposito hanya dapat dilakukan pada waktu tertentu, padahal suku

bunga deposito lebih tinggi dibandingkan tabungan biasa (Darmawan, 2017), sehingga deposito juga dapat dijadikan sebagai sarana alternatif dalam berinvestasi (Ula, 2018) Maka dari itu pemasaran langsung kepada nasabah digunakan sebagai strategi pihak perbankan (Kosti, Miloš, Simi, & Kosti, 2018).

Dari banyaknya informasi nasabah yang dikumpulkan, bank mampu menawarkan produk dan layanan kepada nasabah potensial berdasarkan kebutuhan mereka, kemampuan seperti itu dapat menggunakan teknologi data mining (Kalid, Khor, Ng, & Ting, 2014). Oleh karena itu tujuan dibuatnya penelitian ini untuk memprediksi nasabah yang berlangganan deposito dengan memanfaatkan teknik data mining.

Adapun penelitian relevan yang sudah melakukan penelitian mengenai dataset bank marketing yaitu penelitian yang dilakukan oleh Sergio Moro, Paulo Cortez, Paulo Rita tentang Pendekatan berbasis data untuk memprediksi keberhasilan telemarketing bank dengan membandingkan empat metode *data mining* antara lain regresi logistic, *decision tree (DT)*, *neural network (NN)*, dan *support vector machine (SVM)* dengan nilai akurasi untuk regresi logistic 71,5%, untuk *decision tree* 75,7%, untuk *support vector machine* 76,7%, dan untuk *neural network* 79,4% (Moro et al., 2014)

## 2. Metode Penelitian

Model CRISP-DM (*Cross Industry Standard Process for Data Mining*) digunakan sebagai metode penelitian dalam penelitian ini. Model tersebut terdiri *business understanding, data understanding, data preparation, modeling, evaluation, dan deployment* (Rusdiawan et al., 2019).

### 1. Tahap *Business Understanding*

Tahapan ini difokuskan untuk penentuan tujuan dari penelitian yaitu menerapkan metode klasifikasi data mining menggunakan algoritma *K-Nearest Neighbor* untuk memprediksi nasabah yang berlangganan deposito.

### 2. Tahap *Data Understanding*

Setelah penentuan tujuan dilakukan, tahapan selanjutnya yaitu proses pengumpulan, analisa dan evaluasi kualitas data pada data Bank Marketing yang dibuat oleh sebuah lembaga perbankan Portugis yang diperoleh dari UCI Machine Learning Repository. Data tersebut memiliki 11.162 record dengan 16 atribut predictor dan 1 atribut hasil.

Tabel 1. Atribut dan Nilai Dataset Bank Marketing

Atribut	Nilai
Age	Usia subjek disesuaikan dengan nasabah
Job	('admin.', 'unknown', 'unemployed', 'management', 'housemaid', 'entrepreneur', 'student', 'blue-collar', 'self-employed', 'retired', 'technician', 'services')
Marital	(categorical: 'divorced', 'married', 'single')
Education	('unknown', 'secondary', 'primary', 'tertiary')
Default	apakah kredit dalam default? (binary: 'yes', 'no')
Balance	saldo tahunan menurut rata-rata
Housing	punya pinjaman perumahan? (binary: 'yes', 'no')
Loan	punya pinjaman pribadi? (binary: 'yes', 'no')
Contact	jenis komunikasi kontak (categorical: 'unknown', 'telephone', 'cellular')
Day	hari kontak terakhir dalam sebulan
Month	bulan kontak terakhir tahun ini (kategori: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
Duration	durasi kontak terakhir, dalam detik
Campaign	jumlah kontak yang dilakukan selama pemasaran ini dan untuk nasabah ini (termasuk kontak terakhir)
Pdays	jumlah hari yang berlalu setelah nasabah terakhir dihubungi dari pemasaran sebelumnya (- 1 berarti klien sebelumnya tidak dihubungi)
Previous	jumlah kontak yang dilakukan sebelum pemasaran ini dan untuk nasabah ini
Outcome	hasil dari pemasaran sebelumnya (categorical: 'unknown', 'other', 'failure', 'success')
Deposit	Variabel output (binary: 'yes', 'no')

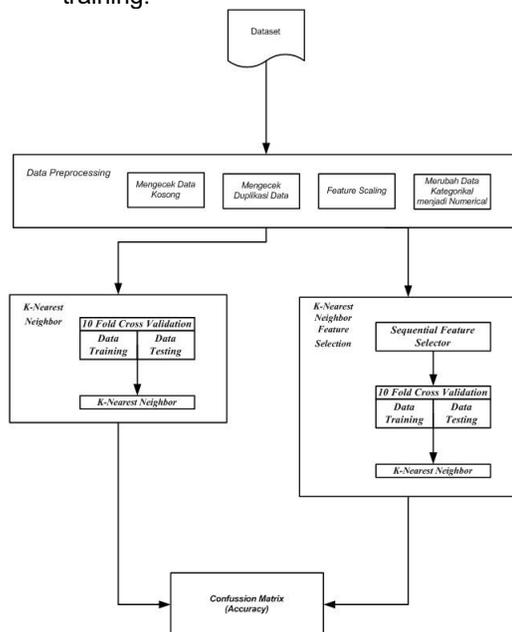
### 3. Tahap *Data Preparation*

Pada tahap *Data Preparation* dilakukan pengolahan data Bank Marketing, jumlah data yang diperoleh pada penelitian ini sebanyak 11.162 record, yang terdiri dari nasabah yang sudah dan tidak berlangganan deposit. Akan tetapi data

tersebut masih terdapat data anomali, oleh karena itu sebelum melakukan pemodelan perlu dilakukan *data preparation* yang mana dilakukan pengecekan data kosong dan duplikat data, melakukan *scaling feature*, serta melakukan normalisasi data dengan mengkonversi seluruh data menjadi numerik (Hadi, 2017).

#### 4. Tahap *Modelling*

Tahap ini dilakukan setelah melakukan *data preparation*, yaitu melakukan pemodelan dengan *tools* anaconda dan bahasa pemrograman python menggunakan metode klasifikasi algoritma *K-Nearest Neighbor* dan *feature selection* serta *10 fold cross validation* pada data testing dan data training.



Gambar 1. Model yang diusulkan

#### 5. Tahap *Evaluation*

Tahap ini melakukan evaluasi terhadap model-model yang telah terbentuk dengan menggunakan metode *Confusion Matrix* sehingga mendapatkan informasi model yang akurat.

#### 6. Tahap *Deployment*

Dari model yang telah dihasilkan maka perlu diuji dengan menggunakan data baru dan dilakukan kembali evaluasi untuk keakuratan data.

### 3. Hasil dan Pembahasan

#### 3.1. Dataset

Dataset yang digunakan dalam penelitian ini adalah dataset Bank Marketing. Pengolahan data menggunakan bahasa pemrograman *python*.

#### 3.2. *Data Preprocessing*

##### A. Mengecek Data Kosong

Dalam tahap *preprocessing* untuk menangani data kosong pada dataset Bank Marketing dengan memasukan perintah *isna*.

```
In [78]: #Menghitung total baris di setiap kolom
df.isna().sum()

Out[78]: age      0
job          0
marital      0
education    0
default      0
balance      0
housing      0
loan         0
contact      0
day          0
month        0
duration     0
campaign     0
pdays      0
previous     0
poutcome    0
deposit      0
dtype: int64
```

Gambar 2. Mengecek Data Kosong pada Dataset

Dalam dataset Bank Marketing semua atribut bernilai kosong yang berarti dalam data tersebut tidak memiliki data kosong.

##### B. Mengecek Duplikasi Data

Langkah *preprocessing* selanjutnya adalah mengecek duplikasi data dengan memasukkan perintah menghitung jumlah record duplikat.

```
In [79]: #menghitung jumlah duplikasi data
sum(df.duplicated())

Out[79]: 0
```

Gambar 3. Mengecek Duplikasi Data pada Dataset

Hasil yang didapatkan adalah data tersebut tidak memiliki duplikasi data.

##### C. *Feature Scaling*

*Feature Scaling* adalah suatu cara untuk membuat data numerik pada dataset supaya memiliki jangkauan nilai (*scale*) yang sama. Dalam penelitian ini *feature scaling* yang digunakan adalah *standard scaler* untuk menghindari salah satu atribut yang mendominasi.

##### D. Konversi Data Kategorikal menjadi Numerik

Pada dataset Bank Marketing terdapat beberapa *type* atribut yang kategorikal. Dalam

pengolahan menggunakan python, data kategorikal harus dikonversi kedalam data numerik baik integer maupun float agar dapat diproses.

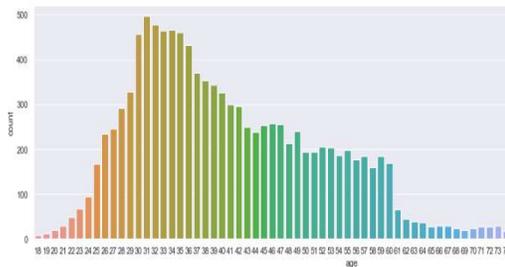
```
In [89]: print(X.dtypes)

age          float64
job          float64
marital      float64
education    float64
default      float64
balance      float64
housing      float64
loan         float64
contact      float64
day          float64
month        float64
duration     float64
campaign     float64
pdays       float64
previous     float64
poutcome    float64
dtype: object
```

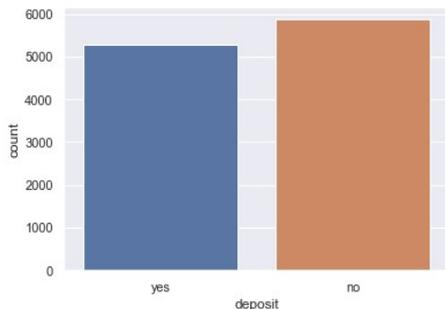
Gambar 4. Merubah Data Kategorikal menjadi Numerik

### 3.3. Explorasi Data

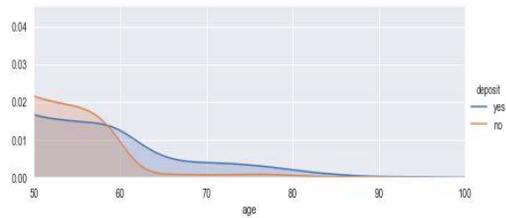
Langkah *explorasi data* dilakukan untuk mendapatkan informasi-informasi dan memahami dataset tersebut. *Explorasi data* yang dilakukan dengan menampilkan berbagai informasi atribut dalam bentuk grafik seperti berikut:



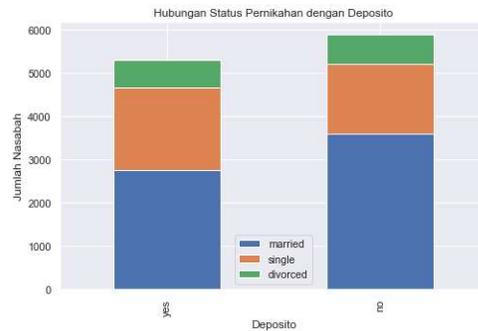
Gambar 5. Usia Nasabah pada Dataset Bank Marketing



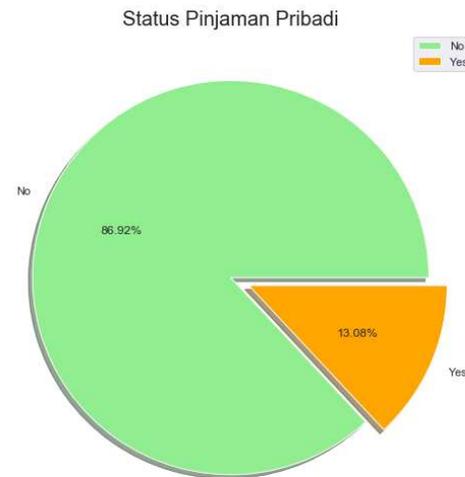
Gambar 6. Perbandingan Class Nasabah Deposit



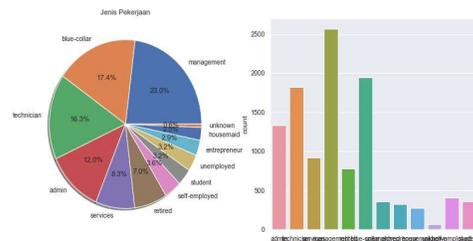
Gambar 7. Perbandingan Usia Nasabah terhadap data yang berlangganan deposit



Gambar 8. Hubungan antara Status Pernikahan dengan Class Deposito



Gambar 9. Persentase Status Pinjaman Pribadi pada Dataset Bank Marketing



Gambar 10. Persentase Jenis Pekerjaan pada Dataset Bank Marketing

### 3.4. Klasifikasi K-Nearest Neighbor

Tahap klasifikasi dilakukan dengan algoritma *K-Nearest Neighbor* dengan menggunakan *k-fold Cross Validation*. Terlebih dahulu menentukan *variable target* dan *variable predictor*.

```
In [86]: # Split dataset in features and target variable
Y = df['deposit'].values # Target for the model
X = df[['age', 'job', 'marital', 'education', 'default', 'balance', 'housing', 'loan',
        'contact', 'day', 'month', 'duration', 'campaign', 'pdays', 'previous', 'poutcome']]

print ("Feature: ", X.shape) # matrix
print ("Response: ", Y.shape) # series

Feature: (11162, 16)
Response: (11162,)
```

Gambar 11. Menentukan *Variabel Target* dan *Variabel Predictor*

Kemudian menginput perintah *k-Fold Cross Validation* dengan nilai *fold* sebanyak 10 kali dengan nilai K=3, K=5, K=7, K=9. Setelah dilakukan pengujian maka diperoleh hasil sebagai berikut :

#### A. Nilai K = 3

```
Confusion Matrix :
[[4784 1089]
 [1885 3404]]

precision  recall  f1-score  support
no         0.72    0.81    0.76    5873
yes        0.76    0.64    0.70    5289

micro avg  0.73    0.73    0.73    11162
macro avg  0.74    0.73    0.73    11162
weighted avg 0.74    0.73    0.73    11162

73.3560293854148
```

Gambar 12. Hasil Akurasi dengan Nilai K = 3

Nilai akurasi yang didapat dengan menggunakan algoritma *K-Nearest Neighbor* sebesar 73,35%, dengan jumlah *True Positive* sebanyak 4.784 data, *False Positive* sebanyak 1.089 data, *True Negative* sebanyak 3.404 data, dan *False Negative* sebanyak 1.885 data.

#### B. Nilai K = 5

```
Confusion Matrix :
[[4847 1026]
 [1877 3412]]

precision  recall  f1-score  support
no         0.72    0.83    0.77    5873
yes        0.77    0.65    0.70    5289

micro avg  0.74    0.74    0.74    11162
macro avg  0.74    0.74    0.74    11162
weighted avg 0.74    0.74    0.74    11162

73.99211610822434
```

Gambar 13. Hasil Akurasi dengan Nilai K = 3

Nilai akurasi yang didapat dengan menggunakan algoritma *K-Nearest Neighbor* sebesar 73,99%, dengan jumlah *True Positive* sebanyak 4.784 data, *False Positive* sebanyak 1.026 data, *True Negative* sebanyak 3.412 data, dan *False Negative* sebanyak 1.877 data.

#### C. Nilai K = 7

```
Confusion Matrix :
[[4859 1014]
 [1881 3408]]

precision  recall  f1-score  support
no         0.72    0.83    0.77    5873
yes        0.77    0.64    0.70    5289

micro avg  0.74    0.74    0.74    11162
macro avg  0.75    0.74    0.74    11162
weighted avg 0.74    0.74    0.74    11162
```

74.06378785163949

Gambar 14. Hasil Akurasi dengan Nilai K = 7

Nilai akurasi yang didapat dengan menggunakan algoritma *K-Nearest Neighbor* sebesar 74,06%, dengan jumlah *True Positive* sebanyak 4.859 data, *False Positive* sebanyak 1.014 data, *True Negative* sebanyak 3.408 data, dan *False Negative* sebanyak 1.881 data.

#### D. Nilai K = 9

```
Confusion Matrix :
[[4871 1002]
 [1858 3431]]

precision  recall  f1-score  support
no         0.72    0.83    0.77    5873
yes        0.77    0.65    0.71    5289

micro avg  0.74    0.74    0.74    11162
macro avg  0.75    0.74    0.74    11162
weighted avg 0.75    0.74    0.74    11162
```

74.37735172908081

Gambar 15. Hasil Akurasi dengan Nilai K = 9

Nilai akurasi yang didapat dengan menggunakan algoritma *K-Nearest Neighbor* sebesar 74,37%, dengan jumlah *True Positive* sebanyak 4.871 data, *False Positive* sebanyak 1.002 data, *True Negative* sebanyak 3.431 data, dan *False Negative* sebanyak 1.858 data.

Selanjutnya dilakukan pengujian dengan menggunakan *feature selection* dengan *k-fold Cross Validation* dengan nilai *fold* sebanyak 10 kali dengan nilai K=3, K=5, K=7, K=9. Setelah dilakukan pengujian maka didapat hasil sebagai berikut :

A. Nilai K = 3

```

confusion matrix :
[[5212 661]
 [ 486 4803]]
precision  recall  f1-score  support
no         0.91    0.89    0.90    5873
yes        0.88    0.91    0.89    5289

micro avg  0.90    0.90    0.90    11162
macro avg  0.90    0.90    0.90    11162
weighted avg 0.90    0.90    0.90    11162
    
```

89.72406378785163

Gambar 16. Hasil Akurasi dengan Nilai K = 3

Nilai akurasi yang didapat dengan menggunakan algoritma *K-Nearest Neighbor* dengan *feature selection* sebesar 89,72%, dengan jumlah *True Positive* sebanyak 5.212 data, *False Positive* sebanyak 661 data, *True Negative* sebanyak 4.803 data, dan *False Negative* sebanyak 486 data.

B. Nilai K = 5

```

confusion matrix :
[[5085 788]
 [ 532 4757]]
precision  recall  f1-score  support
no         0.91    0.87    0.89    5873
yes        0.86    0.90    0.88    5289

micro avg  0.88    0.88    0.88    11162
macro avg  0.88    0.88    0.88    11162
weighted avg 0.88    0.88    0.88    11162
    
```

88.17416233649884

Gambar 17. Hasil Akurasi dengan Nilai K = 3

Nilai akurasi yang didapat dengan menggunakan algoritma *K-Nearest Neighbor* dengan *feature selection* sebesar 88,17%, dengan jumlah *True Positive* sebanyak 5.085 data, *False Positive* sebanyak 788 data, *True Negative* sebanyak 4.757 data, dan *False Negative* sebanyak 532 data.

C. Nilai K = 7

```

confusion matrix :
[[5033 840]
 [ 658 4631]]
precision  recall  f1-score  support
no         0.88    0.86    0.87    5873
yes        0.85    0.88    0.86    5289

micro avg  0.87    0.87    0.87    11162
macro avg  0.87    0.87    0.87    11162
weighted avg 0.87    0.87    0.87    11162
    
```

86.57946604551155

Gambar 18. Hasil Akurasi dengan Nilai K = 7

Nilai akurasi yang didapat dengan menggunakan algoritma *K-Nearest Neighbor* dengan *feature selection* sebesar 86,57%, dengan jumlah *True Positive* sebanyak 5.033

data, *False Positive* sebanyak 840 data, *True Negative* sebanyak 4.631 data, dan *False Negative* sebanyak 658 data.

D. Nilai K = 9

```

confusion matrix :
[[4879 994]
 [ 937 4352]]
precision  recall  f1-score  support
no         0.84    0.83    0.83    5873
yes        0.81    0.82    0.82    5289

micro avg  0.83    0.83    0.83    11162
macro avg  0.83    0.83    0.83    11162
weighted avg 0.83    0.83    0.83    11162
    
```

82.70023293316609

Gambar 19. Hasil Akurasi dengan Nilai K = 9

Nilai akurasi yang didapat dengan menggunakan algoritma *K-Nearest Neighbor* dengan *feature selection* sebesar 82,70%, dengan jumlah *True Positive* sebanyak 4.879 data, *False Positive* sebanyak 994 data, *True Negative* sebanyak 4.352 data, dan *False Negative* sebanyak 937 data.

Nilai akurasi yang terbesar diperoleh menggunakan menggunakan algoritma *K-Nearest Neighbor* dengan *feature selection* pada nilai K=3 dengan nilai akurasi 89,72%.

Perhitungan evaluasi algoritma *K-Nearest Neighbor* dengan *confusion matrix* pada dataset Bank Marketing sebagai berikut:

Tabel 2. *Confusion Matrix K-Nearest Neighbor* Dataset Bank Marketing

Class	Diklasifikasikan sebagai	
	Yes	No
Yes	4871 (TP)	1002 (FP)
No	1858 (FN)	3431 (TN)

Berdasarkan *confusion matrix* pada tabel 1 dari 1116 data Bank Marketing diketahui sebanyak 4871 data positif diklasifikasikan sebagai positif dan sebanyak 3431 data negatif diklasifikasikan sebagai negatif. Sedangkan sebanyak 1002 data positif diklasifikasikan sebagai negatif dan sebanyak 1858 data negatif diklasifikasikan sebagai positif. Berikut merupakan perhitungan nilai *accuracy*:

$$Accuracy = \frac{4871+3431}{4871+1858+1002+3431} \times 100\% = 74,37\%$$

Perhitungan evaluasi algoritma *K-Nearest Neighbor Feature Selection* dengan *confusion*

matrix pada dataset Bank Marketing sebagai berikut:

Tabel 3. *Confusion Matrix K-Nearest Neighbor Feature Selection* Dataset Bank Marketing

Class	Diklasifikasikan sebagai	
	Yes	No
Yes	5212 (TP)	661 (FP)
No	486 (FN)	4803 (TN)

Berdasarkan *confusion matrix* pada tabel 2 dari 1.116 data Bank Marketing diketahui sebanyak 5.212 data positif diklasifikasikan sebagai positif dan sebanyak 4.803 data negatif diklasifikasikan sebagai negatif. Sedangkan sebanyak 661 data positif diklasifikasikan sebagai negatif dan sebanyak 486 data negatif diklasifikasikan sebagai positif. Berikut merupakan perhitungan nilai *accuracy*:

$$Accuracy = \frac{5212+4803}{5212+4803+661+486} \times 100\% = 89,72\%$$

Tabel 4. Pengujian Algoritma *K-Nearest Neighbor*

Nilai K	Accuracy
K=3	73,35%
K=5	73,99%
K=7	74,06%
K=9	74,37%

Tabel 5. Pengujian Algoritma *K-Nearest Neighbor* dan *Feature Selection*

Nilai K	Accuracy
K=3	89,72%
K=5	88,17%
K=7	86,57%
K=9	82,70%

#### 4. Kesimpulan

Dari hasil penelitian yang diperoleh, akurasi dari penggunaan algoritma klasifikasi *K-Nearest Neighbor* dalam prediksi Bank Marketing sebesar 74,37% dengan nilai K=9, sedangkan untuk akurasi algoritma klasifikasi *K-Nearest Neighbor* dengan menggunakan *feature selection* sebesar 89,72% dengan nilai K=3, sehingga didapat selisih peningkatan akurasi sebesar 15,35%.

#### Referensi

- Anam, C., & Santoso, H. B. (2018). *Perbandingan Kinerja Algoritma C4 . 5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa*. 8(1), 13–19.
- Ary, M. (2019). *SATIN – Sains dan Teknologi Informasi Ukuran Akurasi Klasifikasi Penyakit Mesothelioma Menggunakan Algoritma K-Nearest Neighbor dan Backward Elimination*. 5(1).
- Bode, A. (2017). *K-NEAREST NEIGHBOR DENGAN FEATURE SELECTION MENGGUNAKAN BACKWARD ELIMINATION UNTUK PREDIKSI HARGA KOMODITI KOPI ARABIKA*. 9, 188–195.
- Darmawan, H. (2017). pilih mana menabung di deposito bank atau menabung. Retrieved August 20, 2019, from <https://www.finansialku.com/pilih-mana-menabung-di-deposito-bank-atau-menabung-di-saham-bank/>
- Dhikhi, T., Kumar, V., Nadu, T., & Nadu, T. (2018). *Bank Marketing Analysis*. 8(11), 4–9.
- Hadi, A. F. (2017). *ANALISIS DATA MINING UNTUK MENENTUKAN VARIABEL – VARIABEL YANG MEMPENGARUHI*. 4(1), 108–116.
- Kalid, S. N., Khor, K. C., Ng, K. H., & Ting, C. Y. (2014). *Effective Classification for Unbalanced Bank Direct Marketing Data with Over-sampling*. i(August), 12–15.
- Kosti, S. M., Miloš, Đ., Simi, M. I., & Kosti, M. V. (2018). *Data Mining and Modeling Use Case in Banking Industry*. 1–4.
- Lestari, M. E. I. (2014). *PENERAPAN ALGORITMA KLASIFIKASI NEAREST NEIGHBOR ( K-NN ) UNTUK MENDETEKSI PENYAKIT JANTUNG*. 7(September 2010), 366–371.
- Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*. <https://doi.org/10.1016/j.dss.2014.03.001>
- Mustafa, M. S., & Simpen, I. W. (n.d.).

---

*Implementasi Algoritma K-Nearest Neighbor ( KNN ) Untuk Memprediksi Pasien Terkena Penyakit Diabetes Pada Puskesmas Manyampa Kabupaten Bulukumba. VIII(1), 1–10.*

Pengetahuan, J., & Komputer, D. A. N. T. (2017). *PENERAPAN METODE K-NEAREST NEIGHBOR PADA PENENTUAN GRADE DEALER. 2(2), 108–112.*

Ruangthong, P., & Jaiyen, S. (2015). *Bank Direct Marketing Analysis of Asymmetric Information Based on Machine Learning. 93–96.*

Rusdiawan, T. W., Alamsyah, A., Manajemen, M., Ekonomi, F., Bisnis, D., & Telkom, U. (2019). *EKSPLORASI DATA PELANGGAN , UNTUK KONTEKSTUAL MARKETING VOICE OVER LONG TERM EVOLUTION PT . TELKOMSEL MENGGUNAKAN METODE CLUSTERING K-MEANS CUSTOMER DATA EXPLORATION , FOR CONTEXTUAL MARKETING VOICE OVER LONG TERM EVOLUTION PT . TELKOMSEL USING K-MEANS CLUSTERING METHOD. 6(1), 537–544.*

Ula, R. (2018). *PENGARUH CAPITAL ADEQUACY RATIO ( CAR ), INFLASI , DAN SUKU BUNGA SERTIFIKAT BANK INDONESIA ( SBI ) TERHADAP TINGKAT SUKU BUNGA DEPOSITO BERJANGKA ( Studi Pada Perusahaan Bank Pembangunan Daerah di Indonesia Periode 2010-2015 ). 56(1).*