

## KLASIFIKASI STATUS PEMBAYARAN PREMI MENGGUNAKAN ALGORITMA NEIGHBOR WEIGHTED K-NEAREST NEIGHBOR (NWKNN) (STUDI KASUS: PT. BUMIPUTERA KOTA SAMARINDA)

### *Classification of Premium Payment Status Using Neighbor Weighted K-Nearest Neighbor (NWKNN) Algorithm (Case Study: PT. Bumiputera Kota Samarinda)*

Grassella<sup>1\*</sup>, Ika Purnamasari<sup>2</sup>, Fidia Deny Tisna Amijaya<sup>3</sup>

<sup>1</sup>Jurusan Matematika, Fakultas MIPA Universitas Mulawarman

<sup>2</sup>Lab. Statistika Ekonomi dan Bisnis, Fakultas MIPA Universitas Mulawarman

<sup>3</sup>Lab. Matematika Komputasi, Fakultas MIPA Universitas Mulawarman

<sup>1,2,3</sup>Jln. Barong Tongkok No. 4 Kampus Gunung Kelua, Samarinda - Kalimantan Timur 75123 Indonesia

*e-mail corresponding author:* grassellagn@gmail.com\*

---

**Abstrak:** Algoritma Neighbor Weighted K-Nearest Neighbor (NWKNN) merupakan pengembangan dari algoritma K-Nearest Neighbor (KNN), dengan memberikan bobot pada setiap kelas yang akan diklasifikasikan. Penelitian ini membahas tentang klasifikasi menggunakan algoritma NWKNN yang diaplikasikan pada data status pembayaran premi. Tujuannya untuk mengetahui nilai eksponen ( $E$ ) dan nilai ketetanggaan ( $K$ ) yang optimal, serta nilai akurasi dari klasifikasi data status pembayaran Premi di PT. Bumiputera Kota Samarinda. Tahapan dalam penelitian ini yaitu menentukan nilai  $E$  dan nilai  $K$  menggunakan  $k$ -fold cross validation, menghitung jarak euclidean, menghitung bobot dan skor setiap kelas, melihat nilai skor terbesar untuk menentukan hasil klasifikasi, kemudian menghitung nilai akurasi klasifikasi. Hasil penelitian menunjukkan bahwa nilai  $K$  dan nilai  $E$  yang optimal untuk klasifikasi status pembayaran premi di PT. Bumiputera Kota Samarinda menggunakan NWKNN sebesar  $K=3$  dan  $E=6$  dengan nilai akurasi sebesar 75 persen.

**Kata Kunci:** NWKNN, Nilai  $K$ , Nilai  $E$ .

**Abstract:** Neighbor Weighted K-Nearest Neighbor (NWKNN) Algorithm is the development of the K-Nearest Neighbor (KNN) algorithm, by giving a weight to each class that will be classified. This study discusses the classification using the NWKNN algorithm which is applied to the data payment premium status. The aim is to find out the optimal exponent value ( $E$ ) and neighborliness value ( $K$ ), and accuracy value of the data payment premium status classification at PT. Bumiputera Kota Samarinda. The stages in this research are determining the  $E$  value and  $K$  value using  $k$ -fold cross validation, calculating the euclidean distance, calculating the weight and score of each class, look at the largest score to determine the classification results, then calculating the classification accuracy. The results showed that the optimal  $K$  value and  $E$  value for the classification of payment premium status at PT. Bumiputera Kota Samarinda using NWKNN of  $K=3$  and  $E=6$  with an accuracy value of 75 percent.

**Keywords:** NWKNN,  $K$  value,  $E$  value.

---

## 1. PENDAHULUAN

Klasifikasi merupakan menilai suatu objek data untuk memasukkan ke dalam kelas tertentu dari kelas-kelas yang sudah tersedia [1]. Klasifikasi bertujuan memprediksi kelas dari objek data yang belum diketahui kelasnya. Banyak algoritma dalam klasifikasi yang sudah dikembangkan oleh para peneliti, seperti *K-Nearest Neighbor* (KNN) dan *Neighbor Weighted K-Nearest Neighbor* (NWKNN).

Algoritma yang menggunakan kedekatan jarak suatu data dengan data yang lainnya dalam melakukan klasifikasi disebut dengan KNN. KNN merupakan salah satu metode klasifikasi yang terkenal dan

sederhana, namun tidak dilengkapi dengan kemampuan bekerja pada data tidak seimbang. Data tidak seimbang merupakan kondisi dimana distribusi kelas data tidak seimbang atau tidak merata, artinya jumlah kelas data yang satu lebih sedikit atau lebih banyak dibanding dengan jumlah kelas data lainnya. Kelas yang jumlah datanya lebih sedikit disebut kelas minoritas, sedangkan kelas yang jumlah datanya lebih banyak disebut kelas mayoritas. Jika algoritma KNN bekerja pada data tidak seimbang, maka akan menghasilkan akurasi yang lebih baik untuk kelas mayoritas[3].

NWKNN merupakan pengembangan dari algoritma KNN digunakan untuk mengklasifikasikan data yang tidak seimbang [2]. Algoritma NWKNN menggunakan prinsip pembobotan pada kelas data, sehingga membedakan dengan algoritma KNN. Fungsi pembobot digunakan untuk menyeimbangkan data. Pemberian bobot pada data yang berasal dari kelas minoritas akan diberikan bobot besar, sedangkan pada data yang berasal dari kelas mayoritas akan diberikan bobot kecil [5].

Pendapatan premi yang diperoleh perusahaan asuransi ditentukan dari premi yang dibayarkan oleh nasabah. Salah satu permasalahan yang terjadi pada perusahaan asuransi adalah beberapa nasabah tidak lancar dalam membayar premi. Lancar atau tidak lancarnya nasabah dalam membayar premi pada suatu perusahaan asuransi dapat dilihat dari profil data nasabah yaitu meliputi umur, pendapatan, pekerjaan, cara pembayaran, masa pembayaran, dan besar pembayaran premi. Tujuan penelitian ini untuk mengkaji metode klasifikasi algoritma NWKNN dalam pembayaran premi di perusahaan asuransi dengan menentukan nasabah lancar atau tidak lancar dalam pembayaran premi. Dari klasifikasi yang diperoleh dapat membantu perusahaan asuransi dalam memprediksi nasabah dalam pembayaran premi.

## 2. METODOLOGI

### 2.1. Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian merupakan data sekunder yang diperoleh di PT. Bumiputera Kota Samarinda Tahun 2018 dengan jumlah nasabah yang diperoleh 116 nasabah. Adapun variabel dalam penelitian ini adalah sebagai berikut:

**Tabel 1. Variabel Penelitian**

Variabel Penelitian	Keterangan	Satuan
Umur	Umur tertanggung	Tahun
Masa pembayaran	Masa pembayaran premi tertanggung	Tahun
Besar pembayaran	Besar pembayaran premi tertanggung	Rupiah
Status pembayaran premi	Klasifikasi pembayaran premi yaitu: 1. Pembayaran premi lancar 2. pembayaran premi tidak lancar	-

### 2.2. Metode Analisis

Adapun prosedur analisis dalam penelitian ini sebagai berikut:

#### 1. Statistika Deskriptif

Melakukan analisis statistika deskriptif yang disajikan dalam bentuk diagram.

#### 2. Data *testing* dan data *training*

Membagi data menjadi dua bagian yaitu data *training* dan data *testing*.

#### 3. Mencari nilai $K$ optimal

Nilai  $K$  yang digunakan dalam penelitian ini adalah  $K=2,3,4,5,6,7,8,9$  [4].

#### 4. Menghitung jarak ketetangaan antara data *testing* terhadap data *training*

Menghitung jarak ketetangaan antara data *testing* terhadap data *training* menggunakan jarak *euclidean* sesuai Persamaan (2).

5. Mengurutkan nilai jarak ketetangaan

Mengurutkan nilai jarak ketetangaan dari nilai terkecil ke nilai terbesar.

6. Menghitung bobot setiap kelas

Pembobotan pada kelas data merupakan prinsip pada algoritma NWKNN. Untuk menghitung bobot setiap kelas menggunakan Persamaan (3). Sebelum melakukan perhitungan bobot ditentukan terlebih dahulu nilai  $E$ , dimana nilai  $E$  yang digunakan dalam penelitian ini adalah  $E=2,3,4,5,6,7,8,9$ .

7. Menghitung skor setiap kelas

Menghitung nilai skor pada setiap kelas menggunakan persamaan (4), kemudian lihat nilai skor terbesar untuk menentukan hasil klasifikasi.

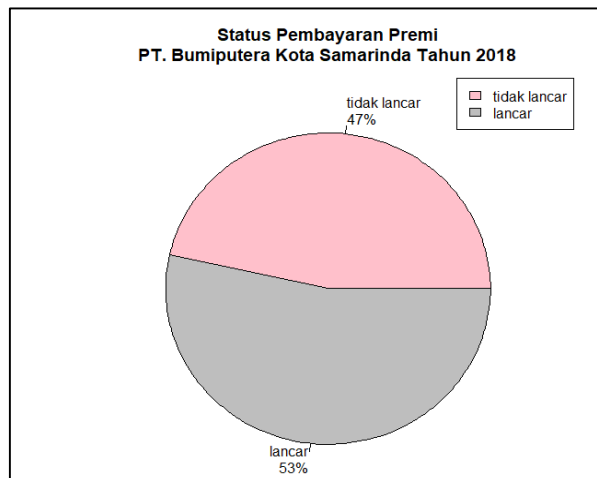
8. Prediksi akurasi

9. Kesimpulan

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Statistika Deskriptif

Analisis statistika deskriptif bertujuan untuk menggambarkan status pembayaran premi (lancar dan tidak lancar) yang dimiliki oleh PT. Bumiputera Kota Samarinda pada Tahun 2018.



Gambar 1. Status Pembayaran Premi

Berdasarkan Gambar 1. dapat diketahui bahwa nasabah PT. Bumiputera Kota Samarinda dalam pembayaran premi memiliki status pembayaran tidak lancar sebesar 47% dan status pembayaran premi lancar sebesar 53%. Tidak lancarnya nasabah yang membayar premi diakibatkan adanya keterlambat nasabah dalam pembayaran premi, dimana nasabah telah lewat waktu yang telah ditemukan dalam membayar premi.

#### 3.2. Data *testing* dan data *training*

Data *training* yang dimaksud adalah data yang digunakan untuk melakukan penentuan  $K$  optimal dari algoritma NWKNN dengan menggunakan teknik *k-fold cross validation*. Data *testing* digunakan untuk memperoleh hasil prediksi klasifikasi pembayaran premi nasabah dengan menggunakan NWKNN dengan nilai  $K$  yang sudah optimal. Untuk Status Pembayaran dengan kode 1 untuk nasabah lancar, dan kode

0 untuk nasabah yang tidak lancar. Untuk Variabel yang digunakan usia (U), masa pembayaran (MP), besar pembayaran (BP), dan status pembayaran (SP). Hasil pembagian data *training* dan *testing* dengan proporsi 90:10 yang diacak dengan *software* R dilihat pada Tabel 2 berikut.

**Tabel 2. Data *testing* dan data *training***

Data <i>training</i>				
No	U	MP	BP	SP
102	0,5	0	0,0031	1
14	0,6429	0	0,0038	1
⋮	⋮	⋮	⋮	⋮
29	0,3810	0	0,0139	1
Data <i>testing</i>				
10	0	0	0,0038	1
12	0,7381	0,4167	0,0406	0
⋮	⋮	⋮	⋮	⋮
114	0,9286	0	0,0031	1

### 3.3. Menentukan *Subset Data*

Penentuan nilai  $K$  optimal dengan *k-fold cross validation*, jumlah data dalam satu *subset* dapat dihitung dengan menggunakan Persamaan (1) sebagai berikut:

$$b = \frac{n}{k} \quad (1)$$

Tujuan dari *k-fold cross validation* adalah setiap *subset* memiliki kesempatan untuk menjadi sebagai data *testing*. Apabila satu *subset* sebagai data *testing* maka sembilan *subset* lainnya akan menjadi data *training*. Nilai  $k$  yang digunakan umumnya  $k=10$  karena mendapatkan akurasi konsisten dalam memperkirakan sebuah nilai [4]. Pada Penelitian ini terdapat 10 *subset* yang digunakan sebagai data *testing* dan data *training*, sebagai contoh hasil pembagian setiap *subset* sebagai berikut.

**Tabel 3. *Subset data***

Subset 1				
No	U	MP	BP	SP
102	0,5	0	0,0031	1
14	0,6429	0	0,0038	1
4	0,5714	0	0,0031	0
29	0,3810	0	0,0139	1
113	0,4762	0	0,0032	1
73	0,5	0,5	0,7076	1
58	0,5238	0	0,0031	0
90	0,1905	0,8333	0,0125	1
13	0,2381	0,4167	0,0094	0
116	0,2381	1	0,0407	1

Subset 10

No	U	MP	BP	SP
103	0,1429	0	0,0032	0
105	0,0952	0	0,0031	1
91	0,5238	0,1667	0,0031	0
92	0,7381	0	0,0088	1
108	0,0952	0,1667	0,0031	1
27	0,3571	0,0833	0,0089	0
72	0,8333	0	0,0032	0
2	0,3571	0	0,0031	0
98	0,4524	0	0,0031	1
37	0,5952	0	0,0094	1
17	0,2619	0,8333	0,0217	1

### 3.4. Euclidean Distance

KNN adalah metode yang menggunakan kedekatan nilai kelas dalam melakukan klasifikasi. Dalam perhitungan kedekatan ketetanggaan pada PT. Bumiputera Kota Samarinda metode KNN menggunakan persamaan *euclidean distance* antara data *training* dan data *testing* pada Persamaan (2) sebagai berikut:

$$d(x_j, y_i) = \sqrt{\sum_{k=1}^n (x_{jk} - y_{ik})^2} \quad (2)$$

Keterangan :

$x_{jk}$  = nilai data *training*  $x$  ke- $j$  pada variabel data ke- $k$

$y_{ik}$  = nilai data *testing*  $y$  ke- $i$  pada variabel data ke- $k$

Perhitungan *euclidean distance* antara data *training* dan data *testing* kemudian dilanjutkan dengan mengurutkan (*Rank*) *euclidean distance* dari nilai terkecil hingga nilai terbesar seperti Tabel 4. Untuk perhitungan data *testing* pada *subset 1* dan data *training subset* lainnya. Perhitungan yang dilakukan untuk setiap *subset* yang menjadi data *testing*.

**Tabel 4. Rank Euclidean Distance**

Rank	Euclidean Distance	Kelas
1	0,0238	1
2	0,0238	1
3	0,0238	1
4	0,0238	1
5	0,0476	0
6	0,0476	1
7	0,0476	0
8	0,0714	0
9	0,0867	0

### 3.5. Perhitungan Bobot dan Skor

Perhitungan *euclidean distance* yang telah dilakukan dilanjutkan dengan algoritma NWKNN. Algoritma NWKNN adalah algoritma yang yang dimodifikasi dari algoritma KNN. Algoritma NWKNN didasari dengan adanya data yang tidak seimbang. Algoritma NWKNN menggunakan pembobotan pada kelas data, yang membedakan dengan algoritma KNN. Setelah melakukan perhitungan jarak *euclidean* yang telah di urutkan dari terkecil, lalu dilakukan penghitung bobot pada kelas data pada status pembayaran premi di PT. Bumiputera Kota Samarinda. Untuk setiap data *training*  $d$ , pilih  $K$  tetangga terdekat dari data *training*

yang terdapat pada kategori  $K = \{C_1^d, C_2^d, \dots, C_{K^*}^d\}$ , perhitungan bobot dapat dilakukan dengan Persamaan (3) berikut [5].

$$Weight_i = \frac{1}{\left( \frac{Num(C_i^d)}{\text{Min}\{Num(C_j^d) | j = 1, \dots, K\}} \right)^{\frac{1}{E}}}, \tag{3}$$

Keterangan:

$Num(C_i^d)$  = jumlah data *training*  $d$  pada kelas  $i$ ,  $i=1, \dots, n$

$Num(C_j^d)$  = jumlah data *training*  $d$  pada kelas  $j$ , dimana  $j$  terdapat dalam himpunan  $K$

$E$  = eksponen ( $E$  bernilai lebih dari 1)

Setiap kelas data pada status pembayaran premi yang telah dihitung nilai bobot kelasnya akan digunakan untuk menghitung nilai skor, hasil dari skor dalam pengambilan keputusan dilihat skor terbesar pada kelas data status pembayaran premi. Perhitungan skor pada algoritma NWKNN dapat dilakukan dengan Persamaan (4):

$$Skor(X, C_i) = Weight_i \left( \sum_{d_j \in KNN(x)} \left( \left( \sqrt{\sum_{i=1}^n (x_{ik} - y_{ik})^2} \right) \delta(d_j, C_i) \right) \right), \tag{4}$$

Keterangan:

$Weight_i$  = bobot pada kelas  $i$ ,

$d_j \in KNN(x)$  = data *training* ke- $j$  dengan himpunan ketetanggaan terdekat dari data *testing* ke- $x$ ,

$$\delta(d_j, C_i) = \begin{cases} 1, & d_j \in C_i \\ 0, & d_j \notin C_i \end{cases}$$

$C_i$  = kelas ke  $i$

Perhitungan nilai bobot kelas yang telah dilakukan dilanjutkan dengan menghitung nilai skor yang dilihat skor terbesar pada kelas data status pembayaran premi. Hasil dari prediksi menggunakan NWKNN dapat dilihat pada Tabel 5 sebagai berikut.

**Tabel 5. Hasil Prediksi Kelas**

Data	K								Data Asli
	2	3	4	5	6	7	8	9	
102	1	1	1	1	1	1	1	1	1
14	0	0	0	0	0	0	0	0	1
4	1	0	1	1	1	1	1	1	0
29	1	1	1	1	1	1	1	1	1
113	1	1	1	1	1	1	1	1	1
73	1	1	1	0	0	0	0	0	1
58	1	1	0	1	0	1	1	1	0
90	1	0	1	0	1	1	1	1	1
13	0	0	0	0	0	0	0	0	0
116	1	1	1	1	1	1	0	1	1
<b>Jumlah Benar</b>	7	7	8	5	7	6	5	6	

Berdasarkan Tabel 5, semakin banyak prediksi yang sama dengan data asli maka nilai  $K$  tersebut menjadi lebih baik atau optimal untuk digunakan dalam prediksi klasifikasi. Kelas dengan tulisan tebal merupakan klasifikasi benar. Selanjutnya melakukan perhitungan akurasi prediksi klasifikasi pada Persamaan (5) sebagai berikut:

$$akurasi = \frac{\text{jumlah data yang diprediksi benar}}{\text{banyak data dalam satu subset data testing}} \times 100\%, \quad (5)$$

Pengujian dilakukan dengan menggunakan nilai  $K=2,3,4,5,6,7,8,9$  dan  $E=2,3,4,5,6,7,8,9$ . Hasil pengujian dapat dilihat Tabel 6 sebagai berikut.

**Tabel 6. Pengaruh Nilai  $K$ ,  $E$  Terhadap Akurasi Prediksi Akurasi**

$K$	$E$							
	2	3	4	5	6	7	8	9
2	64,45%	65,27%	58,64%	61,64%	67,36%	62,55%	59,55%	56,91%
3	64,45%	65,27%	58,64%	61,64%	68,27%	62,55%	59,55%	56,91%
4	64,45%	65,27%	59,64%	62,64%	68,27%	62,55%	58,55%	55,91%
5	64,45%	65,27%	60,55%	62,64%	68,27%	62,55%	59,55%	55,91%
6	64,45%	65,27%	60,55%	62,64%	67,36%	62,55%	58,55%	55,91%
7	64,45%	65,27%	60,55%	62,64%	67,36%	63,55%	58,55%	55,91%
8	64,45%	65,27%	59,55%	63,64%	66,36%	62,55%	58,55%	55,91%
9	64,45%	65,27%	59,55%	63,64%	66,36%	63,55%	58,55%	55,91%

Dapat dilihat bahwa perubahan pada nilai  $K$  mempengaruhi hasil akurasi. Akurasi yang dihasilkan terlihat tidak stabil, karena akurasi yang didapatkan terlihat naik turun. Pada saat  $E=6$ , nilai akurasi prediksi menunjukkan hasil yang optimal dibandingkan lainnya. Selanjutnya dilakukan pengujian setiap subset untuk mendapatkan rata-rata persentase akurasi sebagaimana pada Tabel 7 dengan nilai  $K=3,4,5$  yang digunakan karena memiliki hasil yang optimal.

**Tabel 7. Pengaruh Nilai  $K$ ,  $E$**

$K$	Prediksi Benar	Akurasi
3	9	75%
4	7	58,33%
5	7	58,33%

#### 4. KESIMPULAN

Berdasarkan hasil dan pembahasan yang diperoleh dapat disimpulkan bahwa algoritma NWKNN yang digunakan pada data status pembayaran premi di PT. Bumiputera diperoleh nilai optimal  $K=3$  dan  $E=6$  dengan nilai akurasi sebesar 75% yang dapat membantu perusahaan untuk memprediksi nasabah dalam pembayaran premi.

#### DAFTAR PUSTAKA

- [1] Prasetyo, E., *Data Mining - Mengolah Data menjadi Informasi Menggunakan Matlab*, ANDI, Yogyakarta: 2014
- [2] Rivaldi, A., Putra, P.A., & Sigit, A., *Klasifikasi Penyimpangan Tumbuh Kembang pada Anak Menggunakan Metode Neighbor Weighted K-Nearest Neighbor (NWKNN)*, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Universitas Brawijaya: 2018

- [3] Siringoringo, R., *Klasifikasi Data Tidak Seimbang Menggunakan Algoritma Smote dan K-Nearest Neighbor*. Jurnal ISD, Universitas Methodist Indonesia: 2018
- [4] Suyanto, *Data Mining untuk Klasifikasi dan Klusterisasi Data*, Infomatika Bandung, Bandung: 2017
- [5] Tan, S., *Neighbor Weighted K-Nearest Neighbor for Unbalanced Text Corpus*, Elsevier Inc, China: 2005
- [6] Betrisandi, *Klasifikasi Nasabah Asuransi Jiwa Menggunakan Algoritma Naive Bayes Berbasis Backward Elimination*, ILKOM Jurnal Ilmiah, UNISAN Gorontalo: 2017
- [7] Banjarsari, M.A., Budiman, H., & Andi, F., *Penerapan K-Optimal pada Algoritma K-NN untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan IP sampai dengan Semester 4*, Kumpulan Jurnal Ilmu Komputer (KLIK), ULM: 2015
- [8] Guntara, D., *Asuransi dan Ketentuan-Ketentuan Hukum yang Mengaturnya*, Jurnal Justisi Ilmu Hukum, UBP karawang: 2016
- [9] Hadi, A.H., Dian, E.R., & Candra, D., *Indetifikasi Penyakit Gagal Ginjal Menggunakan Metode Metode Neighbor Weighted K-Nearest Neighbor (NWKNN)*, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Universitas Brawijaya: 2018
- [10] Indriati, & Achmad, R., *Sentiment Analysis For Review Mobile Applications Using Neighbor Weighted K-Nearest Neighbor (NWKNN)*, Jurnal of Environmental Engineering & Sustainable Technology, SUNWAY University Malaysia: 2016
- [11] Paramananda, N., & Agus, S. P., *Pengaruh Tarif Premi, Kualitas Pelayanan, Daya Tarik Iklan, dan Citra Merek Terhadap Royalitas Nasabah*, Jurnal Manajemen Unud, Unud: 2015
- [12] Santosa, B., *Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Graha Ilmu, Yogyakarta: 2007
- [13] Satria, A., Marji, & Ratnawati, D. E., *Klasifikasi Jenis Kanker Berdasarkan Struktur Protein Menggunakan Metode Neighbor Weighted K-Nearest Neighbor (NWKNN)*. Jurnal Pengembangan Teknologi dan Informatika dan Ilmu Komputer, Universitas Brawijaya: 2019
- [14] Supranto, *Analisis Multivariat Arti & Interpretasi*, Informatika Bandung: Bandung: 2010
- [15] Tan, S. (2005). Neighbor Weighted K-Nearest Neighbor for Unbalanced Text Corpus. *Elsevier Inc.*
- [16] Vuldari, R.T., *Data Mining Teori dan Aplikasi Rapidminer*, Penerbit Gava Media, Yogyakarta: 2017
- [17] Widyastomo, D.L., Indriati, & Rizal, S.S., *Implementasi Fuzzy K-Nearest Neighbor (FK-NN) untuk Pemilihan Keminatan Mahasiswa Teknik Informatika*, Jurnal Pengembangan Teknologi dan Informatika dan Ilmu Komputer, Universitas Brawijaya: 2018
- [18] Yuhdawati, P.S., & Achmad, H., *Analisis Sistem Pengendalian Intern Terhadap Sistem Akuntansi Pengeluaran Kas Atas Pengajuan Pembayaran Klaim Asuransi Jiwa*, Jurnal Administrasi Bisnis (JAB), Universitas Brawijaya: 2017