

Algoritma K-Nearest Neighbour Dalam Mengklasifikasi Penilaian Peserta Didik

¹ Anisa Mardhyath, ²Herlina Latipa Sari, ³Rina Julita

^{1,2,3} Prodi Informatika Universitas Dehasen Bengkulu, Indonesia

¹anisamardhyath@gmail.com, ²herlinalatipasari@unived.ac.id, ³rinajulita@unived.ac.id

Article Info

Article history:

Received, 26/11/2020

Revised, 28/11/2020

Accepted, 1/12/2020

Kata Kunci :

Algoritma K-Nearest Neighbour,
Klasifikasi Peserta Didik

Keywords:

K-Nearest Neighbor Algorithm,
Classification of Students

ABSTRAK

Klasifikasi peserta didik pada PAUD IT Al Hasanah dilakukan untuk mengetahui kompetensi dari masing-masing peserta didik terhadap pelajaran yang diberikan, penggunaan algoritma K-Nearest Neighbour dalam mengklasifikasi penilaian peserta didik dengan menggunakan data penilaian peserta didik berdasarkan dengan pedoman penilaian yang telah ditentukan. Training data untuk melakukan proses klasifikasi (*complete storage*) dengan menggunakan *means* data sampel penilaian Tahun Pelajaran 2017/2018 Semester Ganjil dengan total 20 peserta didik, didapatkan hasil 16 peserta didik termasuk dalam klasifikasi Berkembang Sesuai Harapan (80%) dan 4 peserta didik termasuk dalam klasifikasi Berkembang Sangat Baik (20%).

ABSTRACT

The classification of students in PAUD IT Al Hasanah is carried out to determine the competence of each student regarding the lessons given, the use of the K-Nearest Neighbor algorithm in classifying student assessments using student assessment data based on predetermined assessment guidelines. Training data to carry out the classification process (*complete storage*) using the means of assessment sample data for the 2017/2018 Academic Year Odd Semester with a total of 20 students, the results obtained were 16 students were classified as Developing according to Expectations (80%) and 4 students were included in classification Very Good Developed (20%).

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



Penulis Korespondensi:

Herlina Latipa Sari

Program Studi Informatika,

Universitas Dehasen Bengkulu,

Email: herlinalatipasari@unived.ac.id

1. PENDAHULUAN

Pendidikan anak usia dini adalah masa pembinaan awal anak untuk mendapatkan pendidikan yang memberikan bantuan rangsangan dalam tumbuh kembang anak secara rohani dan jasmani. Salah satu Pendidikan anak usia dini adalah PAUD IT Al Hasanah yang ada di Kota Bengkulu. Pengklasifikasian penilaian peserta didik digunakan sebagai pedoman bagi PAUD IT Al Hasanah dalam mengetahui kompetensi dari masing-masing peserta didik terhadap pelajaran yang diberikan, Pengklasifikasian data peserta didik berdasarkan pedoman penilaian.

Selama ini proses pengklasifikasian peserta didik dilakukan dengan cara mendata semua peserta didik yang mengikuti setiap pelajaran, kemudian guru memberikan nilai pada masing-masing peserta didik. Nilai tersebut akan diklasifikasikan guru untuk mengetahui kompetensi dari masing-masing peserta didik. Permasalahan yang terjadi di PAUD IT Al Hasanah adalah proses pengklasifikasian peserta didik tersebut masih dilakukan secara konvensional dengan menggunakan banyaknya lembar penilaian. Oleh karena itu dibutuhkan suatu aplikasi yang dapat membantu mengelompokkan data penilaian peserta didik berdasarkan kompetensi penilaian yang telah diberikan sesuai dengan pedoman penilaian. Terdapat beberapa algoritma yang dapat digunakan untuk proses pengklasifikasian data [1], salah satunya adalah Algoritma K-

Nearest Neighbor (K-NN) yang memiliki konsistensi yang kuat dan efektif dalam melakukan training data yang besar [2].

Algoritma *K-Nearest Neighbor* (K-NN) merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut [3]. Tujuan dari algoritma adalah mengklasifikasikan objek baru berdasarkan atribut dan *trainingsample*. Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk data *test/uji* yang klasifikasinya belum diketahui [4]. Tujuan dari penelitian ini adalah Penggunaan algoritma k-NN diharapkan dapat memberi informasi yang berguna tentang klasifikasi penilaian peserta didik.

2. STUDI LITERATUR

Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [5]. Algoritma klasifikasi yang banyak digunakan secara luas, yaitu Decision/classification trees, Bayesian classifiers/ Naïve Bayes classifiers, Neural networks, Analisa Statistik, Algoritma Genetika, Rough sets, k-nearest neighbor, Metode Rule Based, Memory based reasoning, dan Support vector machines (SVM) [6]. Klasifikasi data terdiri dari 2 langkah proses [7] yaitu sebagai berikut :

Table 1. Langkah Proses dalam Klasifikasi

No	Langkah Proses
1	Proses Learning (fase training), dimana algoritma klasifikasi dibuat untuk menganalisa data training lalu direpresentasikan dalam bentuk rule klasifikasi
2	Proses klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari rule klasifikasi

Proses klasifikasi didasarkan pada empat komponen [8] : a. Kelas Variabel dependen yang berupa kategorikal yang merepresentasikan ‘label’ yang terdapat pada objek. Contohnya: resiko penyakit jantung, resiko kredit, customer loyalty, jenis gempa. b. Predictor Variabel independen yang direpresentasikan oleh karakteristik (atribut) data. Contohnya: merokok, minum alkohol, tekanan darah, tabungan, aset, gaji. c. Training dataset Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan predictor. d. Testing dataset Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi

Algoritma *Nearest Neighbor* (kadang disebut *K-Nearest Neighbor/K-NN*) merupakan algoritma yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data yang lain. k-nearest neighbor (kNN) termasuk kelompok instance-based learning [9]. Algoritma ini juga merupakan salah satu teknik lazy learning. kNN dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing [10]. Pada algoritma *K-NN*, data berdimensi q, jarak dari data tersebut ke data yang lain dapat dihitung. Nilai jarak inilah yang digunakan sebagai nilai kedekatan/kemiripan antara data uji dengan data latih [11]. Nilai K pada *K-NN* berarti K data terdekat dari data uji.

Nearest Neighbor merupakan teknik klasifikasi yang sederhana, tetapi mempunyai hasil kerja yang cukup bagus [12]. Meskipun begitu, *K-NN* juga mempunyai kelebihan dan kekurangan. Beberapa karakteristik *K-NN* [13] adalah sebagai berikut :

Tabel 2. Karakteristik K-Nearest Neighbour

No	Karakteristik K-Nearest Neighbor
1	<i>K-NN</i> merupakan algoritma yang menggunakan seluruh data latih untuk melakukan proses klasifikasi (<i>complete storage</i>). Hal ini mengakibatkan proses prediksi yang sangat lama untuk data dalam jumlah sangat besar. Pendekatan lain adalah dengan menggunakan <i>means</i> data dari setiap kelas, kemudian menghitung jarak terdekat data uji ke <i>mean</i> data setia kelas tersebut. Hal ini memberi keuntungan kerja yang lebih cepat, tetapi hasilnya kurang memuaskan karena model hanya membentuk hyperplane linear tepat di tengah-tengah di antara 2 kelas yang memisahkan 2 kelas
2	Algoritma <i>K-NN</i> tidak membedakan setiap fitur dengan suatu bobot seperti pada <i>Artificial Neural Network</i> (ANN) yang berusaha menekan fitur yang tidak mempunyai kontribusi terhadap klasifikasi menjadi 0 pada bagian bobot. <i>K-NN</i> tidak memiliki bobot untuk masing-masing fitur.
3	Karena <i>K-NN</i> masuk kategori <i>lazy learning</i> yang menyimpan sebagian atau semua data dan hampir tidak ada proses pelatihan. <i>K-NN</i> sangat cepat dalam proses pelatihan (karena memang tidak ada, tetapi sangat lambat dalam proses prediksi,
4	Hal yang rumit adalah menentukan nilai K yang paling sesuai.

5	Karena <i>K-NN</i> pada prinsipnya memilih tetangga terdekat, parameter jarak juga penting untuk dipertimbangkan sesuai dengan kasus datanya. <i>Euclidean</i> sangat cocok untuk menggunakan jarak terdekat (lurus) antara dua data, tetapi <i>Manhattan</i> sangat teguh (<i>robust</i>) untuk mendeteksi <i>outlier</i> dalam data.
---	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

K-Nearest Neighbor merupakan salah satu metode yang digunakan dalam pengklasifikasian. Prinsip kerja K-Nearest Neighbor (KNN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K-Nearest (neighbor) terdekatnya dalam data pelatihan [1]. Salah satu metode klasifikasi yang sering digunakan adalah metode K-Nearest Neighbor. Penggunaan K-Nearest Neighbor bertujuan untuk mengklasifikasikan objek baru berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek baru tersebut [14]. Teknik algoritma K-Nearest Neighbor mudah untuk diimplementasikan. Dalam hal ini jumlah data atau biasa disebut dengan tetangga terdekat ditentukan oleh user yang dinyatakan oleh k [15]. Langkah-langkah dari algoritma *K-Nearest Neighbors* (KNN) [12], antara lain:

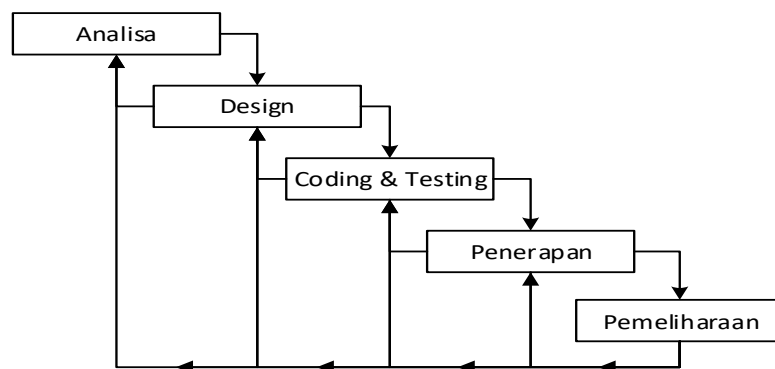
- a) Tentukan parameter K = jumlah banyaknya tetangga terdekat
- b) Hitung jarak antara data baru dan semua data yang ada di data *training*
- c) Untuk menghitung jarak antara data baru dan semua data yang ada menggunakan rumus Euclidean Distance :

$$D(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^p |x_{2j} - x_{1j}|^2}$$

- d) Urutkan jarak tersebut dan tentukan tetangga mana yang terdekat berdasarkan jarak minimum ke K
- e) Tentukan kategori dari tetangga terdekat
- f) Gunakan kategori mayoritas yang sederhana dari tetangga yang terdekat tersebut sebagai nilai prediksi dari data yang baru.

3. METODE PENELITIAN

Metode penelitian yang digunakan penulis adalah metode *Waterfall*. Metode *waterfall* sering dinamakan siklus hidup klasik (*classic life cycle*), dimana hal ini menggambarkan pendekatan yang sistematis dan juga berurutan pada pengembangan perangkat lunak, dimulai dengan spesifikasi kebutuhan pengguna lalu berlanjut melalui tahapan-tahapan perencanaan (*planning*), permodelan (*modeling*), konstruksi (*construction*), serta penyerahan sistem ke para pelanggan/pengguna (*deployment*), yang diakhiri dengan dukungan pada perangkat lunak lengkap yang dihasilkan. Adapun tahapan dari metode Waterfall terlihat pada Gambar 1.



Gambar 1. Metode Waterfall

4. HASIL DAN ANALISIS

a. Implementasi algoritma *K-Nearest Neighbors* (KNN) pada Klasifikasi Penilaian Peserta Didik

Dalam penelitian ini pengklasifikasian data peserta didik dilakukan dengan melihat hasil penilaian pada masing-masing peserta didik berdasarkan aspek penilaian peserta didik sebagai berikut : 1) A1 = Aspek perkembangan nilai agama dan moral, 2) A2 = Aspek perkembangan fisik motoric, 3) A3 = Aspek perkembangan Kognitif, 4) A4 = Aspek perkembangan Bahasa, 5) A5 = Aspek perkembangan sosial emosional, 6) A6 = Aspek perkembangan seni

Pada 6 aspek penilaian tersebut terdapat indikator penilaian (data terlampir), dimana pengisian nilai tersebut sesuai dengan kategori penilaian yang selama ini berjalan yaitu :

BB = Belum Berkembang (1)

- MB = Mulai Berkembang (2)
- BSH = Berkembang Sesuai Harapan (3)
- BSB = Berkembang Sangat Baik (4)

Ke-4 nilai tersebut akan dijadikan sebagai penentuan dalam klasifikasi siswa. Sampel data penilaian peserta didik dari 6 aspek tersebut terlampir. Dari data penilaian peserta didik tersebut akan dicari nilai rata-rata sehingga didapatkan nilai peserta didik pada masing-masing aspek penilaian seperti terlihat pada Tabel1.

Tabel 3. Sampel Data

No	Nama	Nilai Rata-rata						Keterangan
		A1	A2	A3	A4	A5	A6	
1	Azka Risky Aditya	2,909	3,083	3,217	3,454	3,125	3	Berkembang Sesuai Harapan
2	M. Rafa Ghaisan	2,909	3	3,043	3,1818	3,25	3	Berkembang Sesuai Harapan
3	Najatul Lathifah	3	3,833	3,565	3,272	3,416	3,6	Berkembang Sesuai Harapan
4	M. Akram Anugrah	2,545	2,833	2,652	2,909	3	3	Berkembang Sesuai Harapan
5	M. Ali Ukasyah	4	4	4	4	3,916	4	Berkembang Sangat Baik
6	Rina	3,272	2,833	3,173	3,09	3,208	4	?

Adapun Tahapan penyelesaian dari Algoritma K-NN antara lain :

1. Menentukan parameter K = jumlah banyaknya tetangga terdekat Parameter K = 2 tetangga terdekat.
2. Hitung jarak antara data baru dan semua data yang ada di data training
Untuk menghitung jarak antara data baru dan semua data yang ada menggunakan rumus Euclidean Distance :

$$D(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^p |x_{2j} - x_{1j}|^2}$$

Tabel 4. Hitung Euclidean Distance

No.	Nama	Distance
1	d1, d6	$\sqrt{(3,272 - 2,909)^2 + (2,833 - 3,083)^2 + (3,173 - 3,217)^2 + (3,09 - 3,454)^2 + (3,208 - 3,125)^2 + (4 - 3)^2} = 1,155$
2	d2, d6	$\sqrt{(3,272 - 2,909)^2 + (2,833 - 3)^2 + (3,173 - 3,043)^2 + (3,09 - 3,1818)^2 + (3,208 - 3,25)^2 + (4 - 3)^2} = 1,089$
3	d3, d6	$\sqrt{(3,272 - 3)^2 + (2,833 - 3,833)^2 + (3,173 - 3,565)^2 + (3,09 - 3,272)^2 + (3,208 - 3,416)^2 + (4 - 3,6)^2} = 1,209$
4	d4, d6	$\sqrt{(3,272 - 2,545)^2 + (2,833 - 2,833)^2 + (3,173 - 2,652)^2 + (3,09 - 2,909)^2 + (3,208 - 3)^2 + (4 - 3)^2} = 1,369$

5	d5, d6	$\sqrt{\begin{matrix} (3,272 - 4)^2 + (2,833 - 4)^2 + \\ (3,173 - 4)^2 + (3,09 - 4)^2 + \\ (3,208 - 3,916)^2 + (4 - 4)^2 \end{matrix}} = 1,976$
---	--------	-------------------------------------------------------------------------------------------------------------------------------------------------

3. Urutkan jarak tersebut dan tentukan tetangga mana yang terdekat berdasarkan jarak minimum ke-K

Tabel 5. Pengurutan Nilai Distance

No.	Nama	Distance
1	d1, d6	1,155
2	d2, d6	1,089
3	d3, d6	1,209
4	d4, d6	1.369
5	d5, d6	1,976

4. Tentukan kategori dari tetangga terdekat

Pada tahap ini akan diambil tetangga terdekat berdasarkan parameter k yang telah ditentukan yaitu 2, sehingga hasilnya sebagai berikut :

Tabel 6. Parameter K=2

No.	Nama	Distance
1	d1, d6	1,155
2	d2, d6	1,089
3	d3, d6	1,209
4	d4, d6	1.369
5	d5, d6	1,976

Jadi dari data 2 terbaik adalah data nomor 1, 2 yaitu :

Tabel 7. Hasil 2 Data Terbaik

No	Nama	Distance
1	d1, d6	1,155
2	d2, d6	1,089

5. Gunakan kategori mayoritas yang sederhana dari tetangga yang terdekat tersebut sebagai nilai prediksi dari data yang baru

Tabel 8. Kategori Mayoritas

No.	Nama	A1	A2	A3	A4	A5	A6	Keterangan
1	Azka Risky Aditya	2,909	3,083	3,217	3,454	3,125	3	Berkembang Sesuai Harapan
2	M. Rafa Ghaisan	2,909	3	3,043	3,1818	3,25	3	Berkembang Sesuai Harapan
3	Najatul Lathifah	3	3,833	3,565	3,272	3,416	3,6	Berkembang Sesuai Harapan

4	M. Akram Anugrah	2,545	2,833	2,652	2,909	3	3	Berkembang Sesuai Harapan
5	M. Ali Ukasyah	4	4	4	4	3,916	4	Berkembang Sangat Baik
6	Rina	3,272	2,833	3,173	3,09	3,208	4	?

Tabel 9. Nilai 2 terbaik berdasarkan parameter K= 2NN

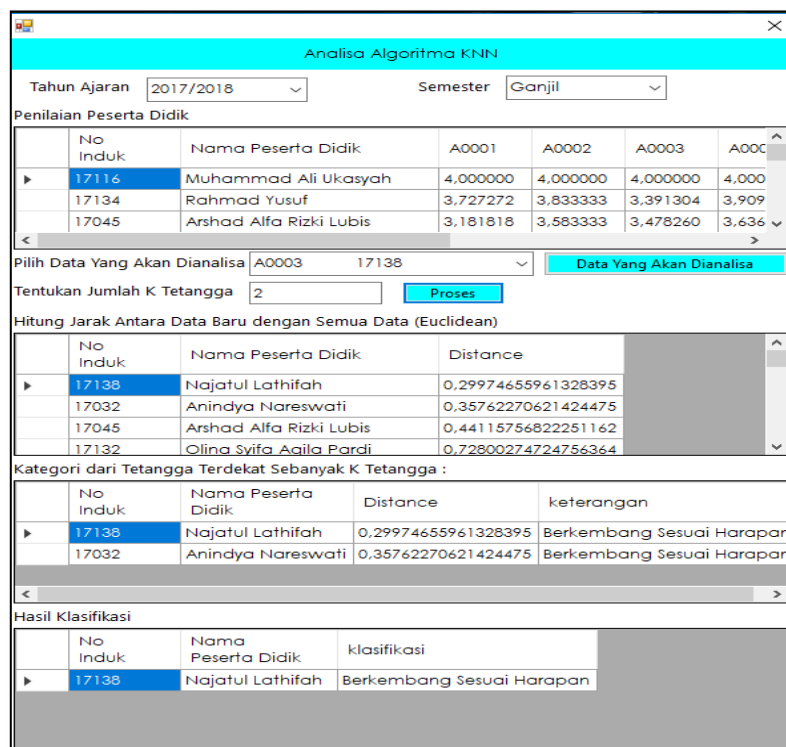
No.	Nama	A1	A2	A3	A4	A5	A6	Keterangan
1	Azka Risky Aditya	2,909	3,083	3,217	3,454	3,125	3	Berkembang Sesuai Harapan
2	M. Rafa Ghaisan	2,909	3	3,043	3,1818	3,25	3	Berkembang Sesuai Harapan

Tabel 10. Hasil Akhir

6	Rina	3,272	2,833	3,173	3,09	3,208	4	Berkembang Sesuai Harapan
---	------	-------	-------	-------	------	-------	---	---------------------------

b. Aplikasi Penilaian Peserta Didik PAUD IT Al Hasanah menggunakan Algoritma K-Nearest Neighbors (KNN)

Aplikasi pengklasifikasian data peserta didik di PAUD IT Al Hasanah merupakan suatu alternatif upaya yang dapat membantu guru dalam mengklasifikasi penilaian peserta didik dimana pada pengujian yang telah dilakukan pada Tahun Pelajaran 2017/2018 Semester Ganjil dengan total 20 peserta didik, penerapan algoritma dalam aplikasi seperti yang terlihat digambar 3.



Gambar 2. Penerapan Algoritma KNN dalam Klasifikasi Penilaian Peserta Didik

5. KESIMPULAN

Aplikasi pengklasifikasian data peserta didik pada studi kasus PAUD IT Al Hasanah menerapkan Algoritma KNN, sehingga hasil akhir klasifikasi yang diperoleh adalah dengan melihat jumlah K tetangga terdekat yang ditentukan dari hasil analisis pada pengujian yang telah dilakukan pada Tahun 2017/2018 Semester Ganjil dengan total 20 peserta didik, didapatkan hasil 16 peserta didik termasuk dalam klasifikasi Berkembang Sesuai Harapan (80%) dan 4 peserta didik termasuk dalam klasifikasi Berkembang Sangat Baik (20%). Aplikasi pengklasifikasian data peserta didik di PAUD IT Al Hasanah sangat membantu dalam menentukan klasifikasi peserta didik berdasarkan 4 jenis kategori BB (Belum Berkembang), MB (Mulai Berkembang), BSH (Berkembang Sesuai Harapan), BSB (Berkembang Sangat Baik) dan juga membantu dalam memberikan alternatif untuk penilaian peserta didik.

REFERENSI

- [1] Zul, M. I., Muslim, I., & Hakim, L. (September 2017). Human Activity Recognition by Using Nearest Neighbor Algorithm from Digital Image. *International Conference on Soft Computing, Intelligent System and Information Technology (ICSIT)* (pp. 58-61). Denpasar, Bali: IEEE Computer Society.
- [2] Anshori, L., Putri, R. R., & Tibyani. (2018). Implementasi Metode K-Nearest Neighbor untuk Rekomendasi Keminatan Studi (Studi Kasus: Jurusan Teknik Informatika Universitas Brawijaya). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 7, 2745-2753.
- [3] Hermawan, F., & Agung, H. (2017). Implementasi Metode K-Nearest Neighbor Pada Aplikasi Data Penjualan PT. Multitek Mitra Sejati. *Jurnal Sains dan Teknologi (Kalbiscentia)*, Volume 4 Nomor 2 Agustus 2017, 103-109.
- [4] Leidiyana, H. (Februari 2017). Penerapan Metode K-Nearest Neighbor Pada Penentuan Grade Dealer Sepeda Motor. *Jurnal Ilmu Pengetahuan dan Teknologi Komputer Vol. 2. No. 2*, 108-112.
- [5] Vedayoko, L. G., Sugiharti, E., & Muslim, M. A. (2017). Expert System Diagnosis of Bowel Disease Using Case Based Reasoning with Nearest Neighbor Algorithm. *Scientific Journal of Informatics Vol. 4, No. 2, November 2017*, 134-142.
- [6] Aamodt, A., & Plaza, E. (1994). Case-Based Reasoning: Foundational Issues Methodological Variations, and System Approaches. *AI Communications Vol. 7 Nr. 1, March 1994*, 39-59.
- [7] Vercellis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*. West Sussex: John Wiley & Sons Ltd.
- [8] Gorunescu, F. (2011). *Data Mining Concepts, Models and Techniques ISBN 978-3-642-19720-8*. Verlag Berlin Heidelberg: Springer.
- [9] Ohsaki, M., Wang, P., Matsuda, K., Katagiri, S., Watanabe, H., & Ralescu, A. (September 2017). Confusion-Matrix-Based Kernel Logistic Regression for Imbalanced Data Classification. *IEEE Transactions on Knowledge and Data Engineering, VOL. 29, NO. 9* (pp. 1806-1819). IEEE.
- [10] Prasetyo, E. (2012). *Data Mining Konsep dan Aplikasi Menggunakan Matlab ISBN 978-979-29-3282-9*. Yogyakarta: Andi.
- [11] Krisandi, N., Helmi, & Prihandono, B. (2013). Algoritma K-Nearest Neighbor Dalam Klasifikasi Data Hasil Produksi Kelapa Sawit Pada Pt. Minamas Kecamatan Parindu. *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster) Volume 02, No.1*, 33-38.
- [12] Pratama, T. C. (2018). Penerapan Metode K-Nearest Neighbour Dalam Menentukan Kelayakan Calon Nasabah Yang Layak Untuk Kredit Mobil (Studi Kasus : Pt. Astra International, Tbk-Toyota). *Jurnal Riset Komputer (JURIKOM)*, Vol. 5 No. 4, Agustus 2018, 402-408.
- [13] Tripathya, A., Agrawal, A., & Rath, S. K. (2015). Classification of Sentimental Reviews Using Machine Learning Techniques. *3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015)* (pp. 821-829). Procedia Computer Science.
- [14] Kuspa, K., & Pratkanis, T. (2013). *Classification of mobile device accelerometer data for unique activity identification unique activity identification*. Semantic Scholar.
- [15] Witten, I., Frank, E., & Hall, M. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco,: Morgan Kaufmann .