

Sistem Rekomendasi Psikotes untuk Penjurusan Siswa SMA menggunakan Metode Modified K-Nearest Neighbor

Muchlas Mughniy¹, Randy Cahya Wihandika², Barlian Henryranu Prasetyo³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹muchlas.muchlas@gmail.com, ²rendicahya@ub.ac.id, ³barlian@ub.ac.id

Abstrak

Penjurusan belajar siswa SMA merupakan pengelompokan minat belajar untuk mempermudah siswa dalam menekuni bidang ilmu dalam pendidikan tingkat lanjut. Namun penentuan potensi akademik siswa melalui bimbingan konseling membutuhkan waktu proses yang lama dengan jumlah peserta didik yang banyak. Sistem rekomendasi penjurusan bertujuan untuk memberikan rekomendasi potensi akademik siswa yang didasarkan pada potensi kemampuan kognitif menggunakan *Intelligenz Struktur Test* (IST). Sistem rekomendasi menggunakan metode *Modified k-Nearest Neighbor* (MKNN) yang mampu mengklasifikasikan potensi siswa berdasarkan kedekatan pada data training yang memiliki parameter sembilan kecerdasan kognitif dan dua kelas penjurusan sehingga didapatkan output sistem berupa rekomendasi kelas penjurusan. Berdasarkan pengujian yang dilakukan pada sistem menggunakan metode *Modified k-Nearest Neighbor* diperoleh rata-rata akurasi maksimum sebesar 67,95% pada 90% dataset, rata-rata akurasi sebesar 63,58% menggunakan *4-fold Cross Validation*, rata-rata *Sensitivity* serta *Specificity* sebesar 23,64% dan 92,34%, perbandingan akurasi MKNN dengan KNN adalah 63,58% banding 57,11%, serta akurasi maksimum sebesar 55,26% pada reduksi hingga menjadi 6 fitur menggunakan PCA. Sehingga sistem rekomendasi penjurusan dengan metode *Modified k-Nearest Neighbor* belum optimal dalam memberikan rekomendasi penjurusan.

Kata kunci: rekomendasi, penjurusan SMA, Modified k-Nearest Neighbor (MKNN)

Abstract

Major selection for high school student intended to facilitate students to focusing on specific field for higher education. However, assigning academic potential for each student through school counselor needs plenty of time. A recommendation system can generate a recommendation for major selection based on cognitive ability by *Intelligenz Struktur Test* (IST). Modified k-Nearest Neighbor is applied to system which classifying academic potential based on neighborhood by training data. Training data consist of nine cognitive intelligences and two majors. So, system will provide a recommendation for major. From the testing process that has been done, has obtain highest averaged accuracy on 90% dataset is 67,95%, averaged accuracy on 4-fold Cross Validation is 63,58%, averaged Sensitivity and Specificity is 23,64% and 92,34%, accuracy comparison between MKNN and KNN is 63,58% and 57,11%, and then highest accuracy for feature reduction using PCA is 55,26% which is reduced to 6 features. According to test result indicate that Modified k-Nearest Neighbor recommendation system not optimal yet to generate a recommendation for major selection.

Keywords: recommendation, major selection for high school students, Modified k-Nearest Neighbor (MKNN)

1. PENDAHULUAN

Penentuan jurusan belajar siswa Sekolah Menengah Atas (SMA) ditetapkan dalam Kurikulum 2013, yang proses pelaksanaan penjurusannya dimulai dari kelas X. Jurusan tersebut merupakan pengelompokan minat belajar siswa yang digolongkan menjadi tiga

jenis, yaitu peminatan Matematika dan Ilmu Pengetahuan Alam (IPA), Ilmu Pengetahuan Sosial (IPS), serta Bahasa dan Budaya. Tujuan dari diadakannya penjurusan di tingkat SMA adalah untuk mempermudah siswa bidang ilmu yang akan ditekuni dalam pendidikan tingkat lanjut, serta sebagai bekal dalam pengarahannya karir kelak. Proses penentuan jurusan telah diatur dalam Peraturan Menteri Pendidikan dan

Kebudayaan Republik Indonesia Nomor 64 tahun 2014 tentang Peminatan pada Pendidikan Menengah serta Buku Pedoman Peminatan Peserta Didik Tahun 2013, dimana proses penetapan penjurusan dapat didasarkan pada nilai akademik siswa, angket minat dan bakat, maupun tes minat bakat dan potensi akademik yang diawasi dan dilaksanakan oleh Bimbingan Konseling (Kebudayaan, 2013). Penggunaan nilai akademik siswa dapat dijadikan tolak ukur yang sederhana untuk menentukan potensi akademik. Namun hal ini tidak semata-mata mendasari potensi siswa, karena akan menimbulkan ketidaksesuaian minat dan bakat siswa yang dapat mempengaruhi proses belajar secara optimal (Nurmalasari, 2014). Selain itu proses yang telah ada juga membutuhkan proses penentuan yang relatif lama dengan jumlah peserta didik yang banyak.

Dengan adanya sistem rekomendasi penjurusan siswa SMA menggunakan *Modified k-Nearest Neighbor* mengingat diperlukannya rekomendasi jurusan secara efisien dengan membandingkan dan mengelompokkan potensi intelegensi peserta didik pada kelompok minat bakat IPA maupun IPS. Sistem akan menggunakan data potensi intelegensi peserta didik dari alat tes psikologi *Intelligenz Struktur Test* (IST) yang memberikan penilaian terhadap kemampuan kognitif diantaranya kecerdasan umum, kecerdasan verbal, kecerdasan numerik dan kecerdasan figural.

Data yang digunakan dalam penelitian ini berupa lembar jawaban psikotes IST sebanyak 132 data yang didapatkan dari Lembaga Psikologi Terapan Fakultas Psikologi Universitas Islam Negeri Malang. Parameter yang digunakan adalah nama siswa, nilai dari 9 subtes IST, dan kelas penjurusan.

2. PENJURUSAN

Peminatan peserta didik SMA adalah upaya untuk menyiapkan manusia yang mampu hidup mandiri, matang, dan terampil di masyarakat. Selain itu juga sebagai bentuk persiapan untuk melanjutkan ke perguruan tinggi sesuai pilihan dan pendalaman mata pelajaran sewaktu SMA. Kelompok peminatan akademik pada satuan pendidikan tingkat menengah atas terbagi menjadi tiga kelompok, diantaranya kelompok peminatan Matematika dan Ilmu Pengetahuan Alam (IPA), peminatan Ilmu Pengetahuan Sosial (IPS), dan peminatan Ilmu Bahasa dan Budaya. Aspek pemilihan dan penetapan peminatan salah

satunya adalah dengan deteksi potensi dengan menggunakan instrumen tes psikologi atau tes minat bakat. Hasil deteksi potensi tersebut dapat dijadikan rekomendasi dalam pertimbangan peminatan peserta didik (Kebudayaan, 2013).

3. TES INTELEGENSI

Tes intelegensi yang umum digunakan untuk mengetahui potensi kecerdasan kognitif siswa adalah *Intelligenz Struktur Test* (IST) yang dikembangkan oleh Rudolf Amthauer di Jerman pada tahun 1953. IST menggambarkan komponen kecerdasan terdiri atas kecerdasan verbal, kecerdasan numerik, kecerdasan figural, dan kecerdasan umum. Dalam IST terdapat sembilan subtes yang secara keseluruhan terdiri dari 176 soal. Subtes tersebut diantaranya (Kumolohadi & Suseno, 2012) :

- a. *Satzerganzung* (SE) – Melengkapi kalimat
- b. *Wortauswahl* (WA) – Mencari kata yang berbeda
- c. *Analogien* (AN) – Mencari hubungan kata
- d. *Gemeinsamkeiten* (GE) – Mencari kata yang mencakup dua pengertian
- e. *Rechen Aufgaben* (RA) – Hitungan sederhana
- f. *Zahlen Reihen* (ZR) – Deret angka
- g. *Form Auswahl* (FA) – Menyusun bentuk
- h. *Würfelaufgaben* (WU) – Kubus
- i. *Merk Aufgaben* (ME) – Mengingat kata

4. MODIFIED K-NEAREST NEIGHBOR

Modified k-Nearest Neighbor (MKNN) merupakan metode klasifikasi yang didasarkan pada kedekatan pada data training. MKNN merupakan peningkatan performa dan kehandalan dari metode *k-Nearest Neighbor* (KNN) dimana tiap data *sample* akan dilakukan validasi untuk mengatasi adanya data *outlier* sehingga pembobotan tiap datanya mampu memberikan hasil yang handal. Dalam metode MKNN, tiap data latih harus dihitung nilai *validity* dimana proses perhitungannya didasarkan pada tetangga terdekat. *Validity* pada tiap data latih akan dijadikan variabel dalam melakukan pembobotan pada proses berikutnya. Pada Persamaan 1 berikut untuk menghitung nilai *validity* (Parvin, et al., 2010):

$$Validity(x) = \frac{1}{H} \sum_{i=1}^H S(lbl(x), lbl(N_i(x))) \quad (1)$$

dimana:

H : Jumlah titik terdekat

$lbl(x)$: Kelas x

$Ni(x)$: Label kelas titik terdekat x

Persamaan 2 digunakan untuk menghitung kesamaan data x dengan data ke- i tetangga terdekat. Persamaan dari fungsi S didefinisikan sebagai berikut (Parvin, et al., 2010):

$$S(a, b) = \begin{cases} 1, & a = b \\ 0, & a \neq b \end{cases} \quad (2)$$

dimana:

a : kelas a pada data latih

b : kelas lain selain a pada data latih

Dalam metode MKNN, pertama *weight* masing-masing tetangga dihitung dengan menggunakan $1 / (d_e + 0,5)$. Kemudian, *validity* dari setiap data pada data latih dikalikan dengan *weight* berdasarkan pada jarak *Euclidean*. Sehingga metode MKNN, didapatkan persamaan *weight voting* tiap tetangga seperti pada Persamaan 3 (Parvin, et al., 2010):

$$W(x) = Validity(x) \times \frac{1}{(d_e + 0,5)} \quad (3)$$

dimana:

$W(x)$: *Weight voting*

$Validity(x)$: nilai *validity*

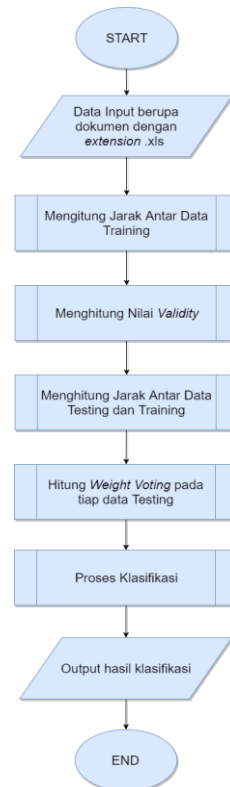
d_e : Jarak *Euclidean* data *training* dengan data *testing*

Gambar 1 merupakan diagram alir metode MKNN. Input sistem berupa data yang tersimpan dalam file dengan *extension .xls* yang berisi sembilan fitur hasil subtes IST dan kelas penjurusan. Proses awal adalah perhitungan jarak tiap data *training* dan perhitungan nilai *validity* untuk masing-masing data *training*. Selanjutnya akan dilakukan perhitungan jarak antar data *training* dan *testing* serta perhitungan *weight voting*. Kemudian proses klasifikasi akan menghasilkan keluaran berupa kelas penjurusan baru.

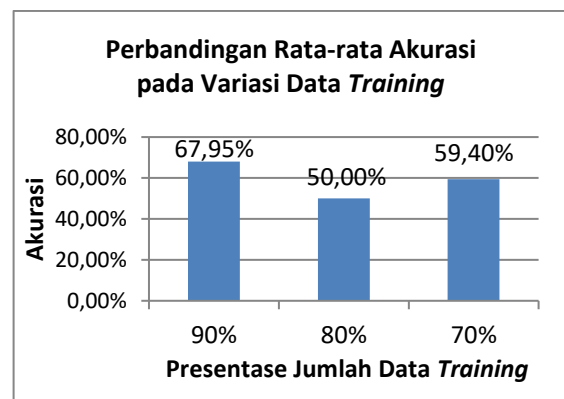
5. HASIL DAN PEMBAHASAN

5.1 Pengujian Pengaruh Nilai k terhadap Akurasi

Pengujian ini bertujuan untuk mengetahui pengaruh nilai k terhadap akurasi yang dihasilkan sistem. Pengujian dilakukan dengan perbandingan data *training* dan *testing* sebesar 90%:10%, 80%:20%, dan 70%:30% dengan nilai k sama dengan 5 hingga 65 yang berinterval 5. Hasil pengujian pengaruh nilai k tertera pada Gambar 2.



Gambar 1 Diagram Alir metode MKNN



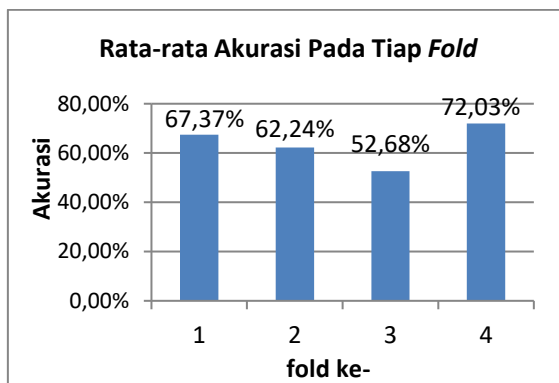
Gambar 2 Perbandingan Rata-rata Akurasi pada Variasi Data Training

Rata-rata akurasi untuk jumlah data *training* sebanyak 90% adalah sebesar 67,95%, pada data *training* sebanyak 80% sebesar 50%, dan pada data *training* sebanyak 70% sebesar 59,40%. Hal ini mengindikasikan bahwa jumlah data *training* semakin banyak maka tinggi kemungkinan sistem untuk mengklasifikasikan dengan tepat.

5.2 Pengujian Akurasi menggunakan k -fold Cross Validation

Pengujian ini dilakukan untuk mengevaluasi tiap *fold* pada *dataset* yang digunakan menjadi data *training* maupun data

testing. Pengujian dilakukan dengan menggunakan *4-fold Cross Validation* dimana dari 132 data akan terbagi menjadi *4-fold* yang terdiri dari masing-masing 33 data. Gambar 3 merupakan hasil pengujian menggunakan *4-fold Cross Validation*.



Gambar 3 Rata-rata Akurasi pada Tiap Fold

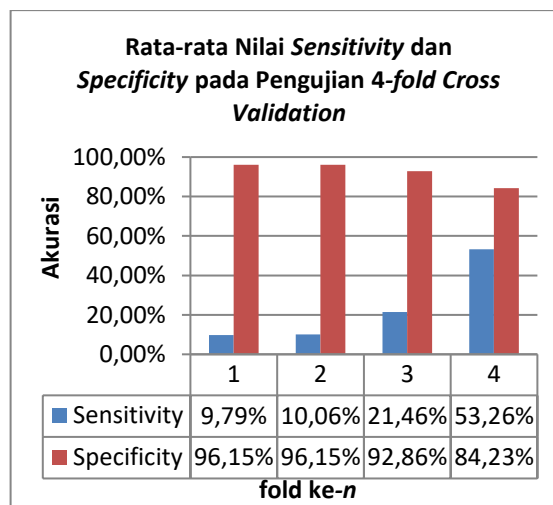
Rata-rata akurasi pada *fold* ke-1 yaitu sebesar 67,37%, pada *fold* ke-2 sebesar 62,24, pada *fold* ke-3 sebesar 52,68% dan pada *fold* ke-4 sebesar 72,03%. Pada *fold* ke-3 terjadi penurunan akurasi yang disebabkan perbedaan jumlah kelas penjurusan pada data training yaitu kelas penjurusan IPA sebanyak 37 data dan kelas penjurusan IPS sebanyak 62. Sedangkan pada *fold* ke-1, ke-2, maupun ke-4 jumlah kelas penjurusan IPA dan IPS mendekati sama. Berdasarkan pengujian ini dapat diindikasikan bahwa tiap *fold* memiliki tingkat akurasi yang mendekati sama sehingga mampu mengklasifikasikan sesuai dengan kedekatan dengan *k*-tetangga.

5.3 Pengujian Sensitivity dan Specificity

Pengujian *Sensitivity* dan *Specificity* bertujuan untuk mengetahui hasil klasifikasi sistem terhadap kelas aktual pada *dataset*. Kelas aktual pada *dataset* terdiri dari kelas penjurusan IPA dan IPS. Hasil pengujian nilai *Sensitivity* dan *Specificity* tertera pada Gambar 4.

Rata-rata *sensitivity* dan *specificity* untuk *fold* ke-1 hingga *fold* ke-4 secara berturut-turut sebesar 9,79% dan 96,15%, 10,06% dan 96,15%, 21,46% dan 92,86%, serta 53,26% dan 84,23%. Secara keseluruhan, nilai *sensitivity* memiliki nilai yang jauh lebih rendah dibandingkan dengan nilai *specificity* sehingga dapat disimpulkan bahwa data dengan kelas aktual IPA banyak yang terklasifikasikan sebagai kelas IPS, sedangkan data dengan kelas aktual IPS telah terklasifikasi dengan baik sebagai kelas

IPS.

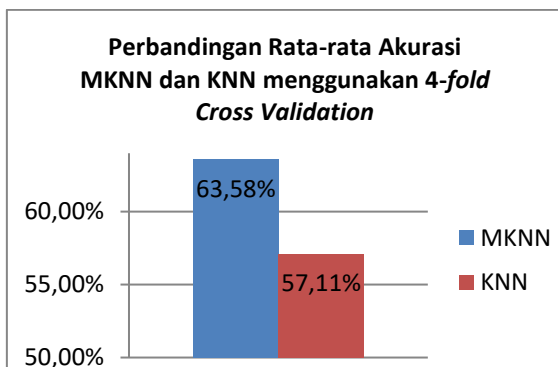


Gambar 4 Rata-rata Nilai Sensitivity dan Specificity

5.4 Pengujian Perbandingan Akurasi MKNN dan KNN

Pengujian perbandingan akurasi MKNN dan KNN bertujuan untuk mengevaluasi performa MKNN yang mampu menangani data *outlier* terhadap performa dari KNN. Pengujian dilakukan dengan menggunakan *4-fold Cross Validation*. Gambar 5 merupakan hasil pengujian perbandingan MKNN dengan KNN.

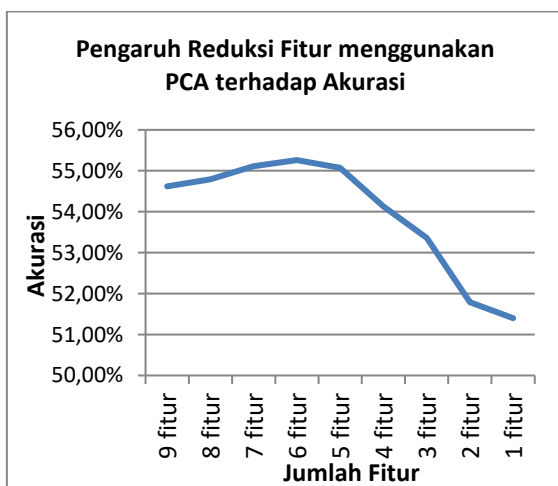
Berdasarkan pengujian perbandingan akurasi MKNN dengan KNN dihasilkan rata-rata akurasi secara keseluruhan sebesar 63,58% untuk metode MKNN dan 57,11% untuk metode KNN. Sehingga dapat diindikasikan bahwa MKNN mampu mengungguli performa KNN dalam penanganan data *outlier* pada *dataset*, dimana data *outlier* akan memiliki nilai *validity* sebesar 0 sehingga pada perhitungan *weight voting*, data *outlier* tersebut akan memiliki nilai 0. Data *outlier* dengan nilai 0 tidak akan mempengaruhi pembobotan pada data *testing* terdekatnya. Sedangkan KNN akan tetap menghitung bobot pada data *outlier*, sehingga akan mempengaruhi penentuan kelas pada data *testing* yang dekat dengan data *outlier* tersebut.



Gambar 5 Perbandingan Rata-rata Akurasi MKNN dengan KNN

5.5 Pengujian Reduksi Fitur Menggunakan *Principal Component Analysis (PCA)*

Pengujian reduksi fitur PCA bertujuan untuk mengevaluasi pengaruh reduksi fitur pada dataset yang ditransformasikan menggunakan PCA. Fitur dari data yang telah ditransformasikan menggunakan PCA kemudian dilakukan reduksi hingga menjadi satu fitur dan dihitung nilai akurasi. Sehingga pada Gambar 6 didapatkan hasil akurasi pada tiap reduksi fitur sebagai berikut.



Gambar 6 Pengaruh Reduksi Fitur menggunakan PCA terhadap Akurasi

Tabel 1 Pengaruh Reduksi fitur terhadap Akurasi

Jumlah Fitur	Akurasi
9 fitur	54,62%
8 fitur	54,80%
7 fitur	55,12%
6 fitur	55,26%
5 fitur	55,07%
4 fitur	54,13%

3 fitur	53,36%
2 fitur	51,79%
1 fitur	51,40%

Berdasarkan pengujian reduksi fitur menggunakan PCA didapatkan nilai akurasi seperti pada Tabel 1. Terjadi peningkatan akurasi menjadi 55,26% pada reduksi hingga menyisakan 6 fitur. Hal ini mengindikasikan bahwa fitur ke-1 hingga fitur ke-6 mengandung informasi yang penting dari hasil transformasi menggunakan PCA. Namun pada reduksi fitur pada jumlah fitur sebanyak enam hingga menyisakan satu fitur terjadi penurunan akurasi hingga menjadi 51,40%. Hal ini mengindikasikan bahwa banyak informasi penting yang hilang dari hasil transformasi PCA, sehingga menyebabkan ketidak-akuratan klasifikasi.

6. KESIMPULAN

Berdasarkan pengujian yang dilakukan pada sistem menggunakan metode *Modified k-Nearest Neighbor* diperoleh rata-rata akurasi maksimum sebesar 67,95% pada 90% dataset, rata-rata akurasi sebesar 63,58% menggunakan *4-fold Cross Validation*, rata-rata *Sensitivity* serta *Specificity* sebesar 23,64% dan 92,34%, perbandingan akurasi MKNN dengan KNN adalah 63,58% banding 57,11%, serta akurasi maksimum sebesar 55,26% pada reduksi hingga menjadi 6 fitur menggunakan PCA. Hal ini mengindikasikan bahwa sistem rekomendasi masih belum optimal dalam memberikan rekomendasi penjurusan. Oleh sebab ini pada pengembangan lebih lanjut dari metode MKNN dapat diimplementasikan perhitungan jarak menggunakan *Hamming distance* maupun *cosine similarity* pada dataset dengan dimensi yang tinggi. Serta pengembangan sistem dapat dilakukan menggunakan metode *Approximate Nearest Neighbor (ANN)* maupun *Locality-Sensitive Hashing (LSH)* untuk jenis data dengan dimensi yang tinggi.

DAFTAR PUSTAKA

- Aggarwal, C. C., 2015. *Data Classification: Algorithms and Applications*. s.l.:CRC Press.
- Andhina, R. K., 2016. *Pemodelan Sistem Pakar untuk Diagnosa Penyakit Tanaman Jagung Menggunakan Metode Modified K-Nearest Neighbour (MKNN)*.

- Arhami, M., 2004. Konsep Dasar Sistem Pakar. Yogyakarta: Penerbit ANDI.
- Arockiasamy, M., 1992. Expert Systems: Applications for Structural, Transportation, and Environmental Engineering. Florida: CRC Press.
- Diana, W., 2011. Penerapan Algoritma Improved K-Nearest Neighbor untuk Pengkategorian Dokumen Teks Berita. Malang: Universitas Brawijaya.
- Gregory, R. J., 2013. Tes Psikologi - Sejarah, Prinsip, dan Aplikasi. Jakarta: Penerbit Erlangga.
- Han, J. & Kamber, M., 2006. Data Mining: Concepts and Techniques. 2nd ed. San Francisco: Morgan Kaufmann Publisher.
- Kebudayaan, K. P. d., 2013. Pedoman Peminatan Peserta Didik. Jakarta: Kementerian Pendidikan dan Kebudayaan.
- Kumolohadi, R. & Suseno, M. N., 2012. Intelligenz Struktur Test Dan Standard Progressive Matrices: (Dari Konsep Intelligensi Yang Berbeda Menghasilkan Tingkat Intelligensi Yang Sama). Jurnal Inovasi dan Kewirausahaan , 1(2), pp. 79-85.
- Kung, S. Y., 2014. Kernel Methods and Machine Learning. Cambridge: Cambridge University Press.
- Kusrini, 2006. Sistem Pakar, Teori dan Aplikasi. Yogyakarta: Penerbit ANDI.
- Kusrini, 2008. Aplikasi Sistem Pakar : Menentukan Faktor Kepastian Pengguna dengan Metode Kuantifikasi Pertanyaan. Yogyakarta: Penerbit ANDI.
- Nurmalasari, E., 2014. Validitas Prediktif Skor Intelligenz Struktur Test (IST) Dan Inventori Motif Berprestasi Terhadap Prestasi Belajar Siswa.
- Olson, L. D. & Dursun, D., 2008. Advanced Data Mining Techniques. Berlin: Springer.
- Parvin, H., Alizadeh, H. & Minaei-Bidgoli, B., 2010. A Modification on K-Nearest Neighbor Classifier. Global Journal of Computer Science and Technology, 10(14), p. 37.
- Turban, E., Aronson, J. E. & Liang, T.-P., 2005. Decision Support Systems and Intelligent Systems (Sistem Pendukung Keputusan dan Sistem Cerdas). 7 ed. Yogyakarta: Penerbit ANDI.
- Zainuddin, S., 2014. Penerapan Algoritma Modified K-Nearest Neighbour (MKNN) Pada Pengklasifikasian Penyakit Tanaman Kedelai.
- Zulfikar & Budiantara, I. N., 2015. Manajemen Riset dengan Pendekatan Komputasi Statistika. Yogyakarta: Deepublish.