

# Pemodelan *Modified K-Nearest Neighbor* Dalam Klasifikasi Jurusan Siswa Di SMAN 6 Pekanbaru

Mustakim<sup>1</sup>, Ulya Ramadhani<sup>2</sup>, Shinta Ayunda Putri<sup>3</sup>

<sup>1,2,3</sup>. Program Studi Sistem Informasi Fakultas Sains dan Teknologi  
Universitas Islam Negeri Sultan Syarif Kasim Riau

Jl.HR. Soebrantas No.155 Simpang Baru, Tampan, Pekanbaru, Riau – Indonesia 28293  
e-mail: mustakim@uin-suska.ac.id, 11653203655@students.uin-suska.ac.id, shintaikhtiar@gmail.com

## Abstrak

Pengembangan sistem informasi pada lembaga pendidikan merupakan salah satu pengembangan terhadap manajemen sekolah dalam menjalankan manajerial pendidikan. Berdasarkan survei yang dilakukan di SMAN 6 Pekanbaru, proses penjurusan Siswa memiliki permasalahan seperti kesulitan yang dialami pihak sekolah dalam menganalisis dan mengevaluasi secara manual saat menentukan jurusan Siswa satu persatu. Hal ini tentunya akan menyita banyak waktu dan tenaga. Pada Data Mining terdapat suatu teknik klasifikasi yang digunakan untuk menggolongkan data sehingga mempermudah dalam mengklasifikasikan jurusan Siswa. Pengklasifikasian jurusan siswa menggunakan algoritma *Modified K-Nearest Neighbor* (MKNN) dengan nilai evaluasi confusion matrix menghasilkan akurasi sebesar 82,29%; 100% nilai presisi; dan 75% nilai recall. Pada penelitian ini, penggunaan information gain dan algoritma K-Means secara berturut dilakukan untuk menyeleksi atribut serta pembagian data latih dan data uji. Hasil pemodelan klasifikasi diimplementasikan pada sistem berbasis web yang kemudian dilakukan perbandingan simulasi parameter dengan hasil akurasi maksimal pada nilai  $k=5$  dengan hasil akurasi optimal 85,4%.

**Kata kunci:** Information Gain, K-Means, MKNN, Jurusan Siswa

## Abstract

The development of information systems in educational institutions is one of the developments of school management in carrying out managerial education. Based on a survey conducted at SMAN 6 Pekanbaru, the student majors process has problems such as difficulties experienced by the school in analyzing and evaluating manually when determining student majors one by one. This of course will take a lot of time and energy. In Data Mining there is a classification technique used to classify data so that it is easier to classify student majors. Classification of student majors using the *Modified K-Nearest Neighbor* (MKNN) algorithm with the evaluation value of the confusion matrix resulting in an accuracy of 82.29%; 100% precision value; and 75% recall value. In this study, the use of information gain and the K-Means algorithm was used to select attributes and to divide training and test data, respectively. The results of the classification modeling are implemented to a web-based system which is then compared with the parameter simulation with the maximum accuracy at  $k=5$  with an optimal accuracy of 85.4%.

**Keywords:** Information Gain, K-Means, MKNN, Student Majority.

## 1. Pendahuluan

Pengembangan terhadap manajemen sekolah dalam menjalankan manajerial pendidikan salah satunya didukung oleh pengembangan sistem informasi pada lembaga pendidikan [1]. Salah satu lembaga pendidikan yang berada di Pekanbaru yaitu SMAN 6 Pekanbaru. SMAN 6 Pekanbaru memiliki dua jurusan didalamnya yaitu Matematika dan Ilmu Pengetahuan Alam (MIPA) dan Ilmu Pengetahuan Sosial (IPS). Penentuan jurusan dilakukan sejak Siswa berada dibangku kelas X dengan mengikuti serangkaian tes yang dilaksanakan. Berdasarkan survei yang dilakukan di SMAN 6 Pekanbaru, penentuan jurusan yang dilakukan selama ini memiliki permasalahan seperti kesulitan yang dialami pihak sekolah dalam menganalisis dan mengevaluasi untuk menentukan jurusan Siswa. Dalam menentukan jurusan Siswa, pihak sekolah perlu mengevaluasi data dari Angket Peminatan Jurusan serta hasil Test Psikotest Siswa. Dikarenakan ada banyaknya data, proses penilaian ini menghabiskan waktu

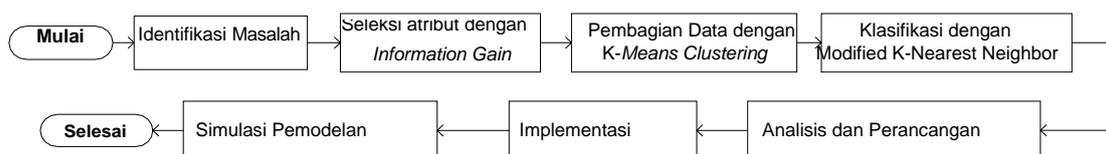
selama 2 bulan. Menurut Hidayah (2014), menentukan jurusan secara manual memiliki banyak kelemahan seperti menghabiskan banyak waktu dan tenaga [2].

Berdasarkan permasalahan tersebut, pada penelitian ini diperlukan sebuah teknik yang berguna untuk mengklasifikasi penentuan jurusan Siswa di SMA Negeri 6 Pekanbaru dengan akurat sehingga dapat meminimalisir kesalahan serta risiko dalam penentuan jurusan. Teknologi data mining merupakan salah satu alat bantu untuk penggalian data pada basis data berukuran besar dengan spesifikasi kerumitan tinggi dan telah banyak digunakan pada lingkungan aplikasi bisnis seperti perbankan, *provider* telekomunikasi, perusahaan pertambangan, perminyakan, dan sebagainya [3]. Dalam data mining terdapat suatu teknik yang digunakan untuk menggolongkan data yaitu teknik klasifikasi [4]. Adapun algoritma yang akan digunakan dalam klasifikasi jurusan Siswa adalah Modified K-Nearest Neighbor (MKNN) yang merupakan penyempurnaan dari K-Nearest Neighbor (KNN). Kelebihan MKNN yaitu terdapat perhitungan nilai *validity* sehingga dapat mengatasi masalah *outlier* dan memiliki tingkat akurasi yang lebih tinggi [5]. Dalam penelitian ini juga menggunakan *information gain* dan K-Means Cluster yang secara berturut digunakan pada tahap seleksi atribut serta dalam pembagian data training dan data testing. Penelitian mengenai teknik pembagian data yang dilakukan oleh Mustakim (2017) menggunakan K-Means dan K-Fold Cross Validation secara berturut menghasilkan akurasi sebesar 93,4% dan 77,8%, sehingga untuk kasus klasifikasi tersebut teknik pembagian data mengimplementasikan K-Means lebih baik daripada menggunakan K-Fold Cross Validation [6]. Selanjutnya penelitian mengenai klasifikasi menggunakan algoritma MKNN yang dilakukan oleh Gazalba (2017) dalam mengklasifikasikan peserta Program Keluarga Harapan didapatkan hasil bahwa K=1 dan cross 2 mempunyai tingkatan akurasi yang tinggi dengan nilai 94.95% serta K=1 yang selanjutnya dijadikan K tetap dalam menghitung MKNN [7].

Berdasarkan pemaparan singkat permasalahan diatas serta didukung dengan penelitian terkait, sehingga pada penelitian ini akan dilakukan analisis terkait klasifikasi penjurusan Siswa menggunakan algoritma MKNN dan hasil pemodelan klasifikasi diterapkan ke dalam sebuah sistem berbasis web ini diharapkan dapat membantu sekolah menentukan jurusan siswa SMAN 6 Pekanbaru.

## 2. Metodologi Penelitian

Adapun tahapan pada penelitian ini dilihat pada Gambar 1.



Gambar 1. Metodologi Penelitian

### 2.1. Modified K-Nearest Neighbor (MKNN)

*Modified K-Nearest Neighbor* merupakan suatu algoritma yang dikembangkan dari algoritma *K-Nearest Neighbor* [8]. Proses algoritma MKNN yaitu dengan menghitung jarak antar data latih dan menghitung jarak data uji dan data latih, selanjutnya mencari bobot terendah berdasarkan nilai K. Bobot yang diperoleh harus divalidasi terlebih dahulu. Selanjutnya menghitung nilai *weight voting* yang akan menghasilkan keluaran berupa bobot pada masing-masing data [9]. Dalam mengatasi kelemahan dari setiap data yang memiliki banyak outlier yaitu dengan mengalikan validitas dengan hasil jarak *euclidean* data training dan data testing.

### 2.2. K-Means Clustering

Algoritma K-Means merupakan algoritma pengelompokan dengan yang bersifat *unsupervised learning* [10]. Algoritma K-Means mengelompokkan data yang memiliki karakteristik yang serupa kedalam beberapa *cluster* dengan mengambil jarak *euclidean* sebagai ukuran kemiripan [11]. Salah satu metode yang digunakan untuk mengevaluasi cluster adalah *Davies-Bouldin Index* (DBI) [12].

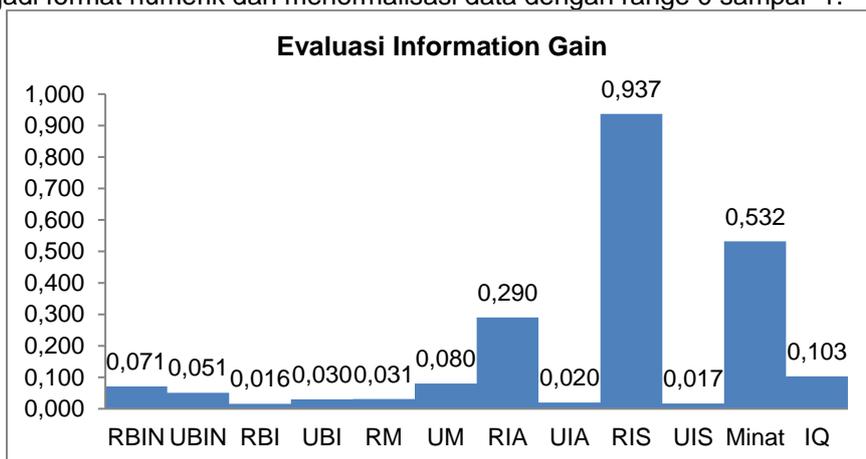
### 2.3. Information Gain (IG)

Information Gain yakni suatu metode seleksi fitur paling relevan dengan menghitung nilai *entropy* berdasarkan peluang kejadian atau atribut tertentu [13]. Hasil dari nilai *entropy* selanjutnya digunakan untuk menentukan atribut mana yang akan dipergunakan dalam proses klasifikasi [14].

### 3. Hasil dan Pembahasan

Pada penelitian ini menggunakan data Siswa kelas sepuluh sebanyak 357 orang yang akan diklasifikasikan ke dalam jurusan MIPA sebanyak 6 kelas dan IPS sebanyak 4 kelas. Atribut yang digunakan sebanyak 12 atribut yaitu, Rerata Raport Bahasa Indonesia (RBIN), Rerata Raport Bahasa Inggris (RBI), Rerata Raport Matematika (RM), Rerata Raport IPS (RPIS), Rerata Raport IPA (RIA), USBN Bahasa Indonesia (UBIN), USBN Bahasa Inggris (UBI), USBN Matematika (UM), USBN IPS (UIS), USBN IPA (UIA), Minat dan Intellectual Quotient (IQ).

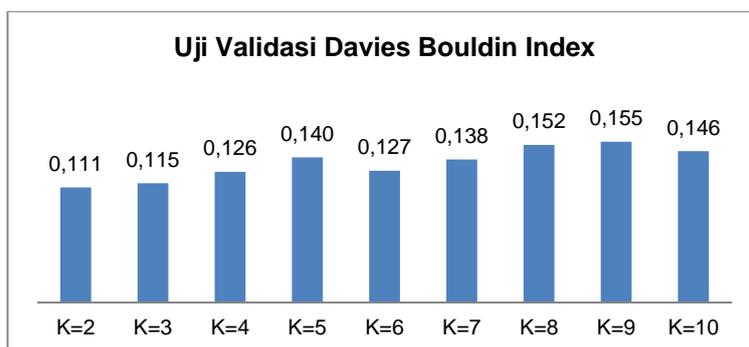
Data yang diperoleh akan melalui tahap *pre-processing* yaitu dengan menghilangkan data yang tidak lengkap, *noise*, hingga penciran data (*outlier*). Selanjutnya dilakukan proses seleksi atribut menggunakan nilai *information gain*. Adapun grafik hasil dari proses seleksi atribut dapat dilihat pada Gambar 2. Pada percobaan ini, dihasilkan nilai gain tertinggi sebesar 0,937 pada atribut RIS, sedangkan nilai *gain* terendah sebesar 0,016 dihasilkan dari atribut RBI. Suatu atribut dapat digunakan pada saat nilai *gain*>0 [15]. Tahap selanjutnya yaitu mengubah data menjadi format numerik dan menormalisasi data dengan range 0 sampai 1.



Gambar 2. Nilai *Information Gain* pada Atribut Penentuan Jurusan

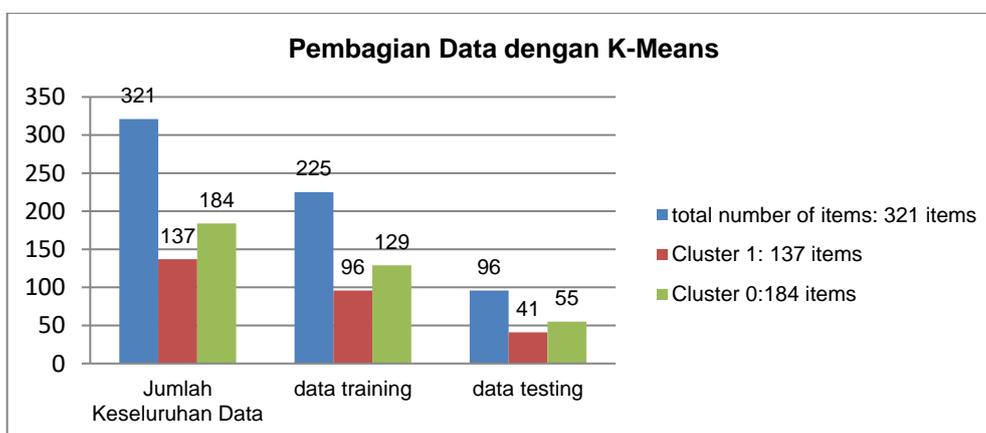
### 3.2. Pembagian Data dengan K-Means Clustering

Hasil data dari tahap *preprocessing* selanjutnya akan dilakukan pembagian data training serta data testing dengan menggunakan algoritma *K-Means Clustering*. Dari tahap *preprocessing* didapatkan data sebanyak 321 *record* yang akan dibagi kedalam beberapa *cluster*. Untuk mendapatkan *cluster* terbaik dibutuhkan percobaan dengan menggunakan nilai *k* yang bervariasi dari rentang *k*=2 sampai *k*=10. Proses evaluasi menggunakan nilai uji validasi DBI untuk mendapatkan hasil nilai *k* terbaik.



Gambar 3. Hasil Uji validasi DBI K-Means Clustering

Dari Gambar 3 diketahui bahwa range nilai uji validasi terendah didapatkan pada saat nilai k=2 dengan nilai sebesar 0,111. Sesuai dengan kaidah, pengelompokan data akan semakin baik jika nilai DBI semakin kecil, sehingga pada penelitian ini akan menggunakan nilai k=2 untuk proses pembagian data training dan data testing. Hasil dari pembagian data training serta data testing menerapkan algoritma K-Means Clustering dapat dilihat pada Gambar 4.



Gambar 4. Hasil Pembagian Data dengan K-Means

### 3.3. Klasifikasi dengan Algoritma Modified K-Nearest Neighbor

Setelah data dibagi menjadi data training dan data testing, selanjutnya yaitu penerapan algoritma MKNN yang ditujukan untuk mendapatkan hasil klasifikasi jurusan Siswa baru. Tabel 1. menampilkan hasil klasifikasi jurusan Siswa baru.

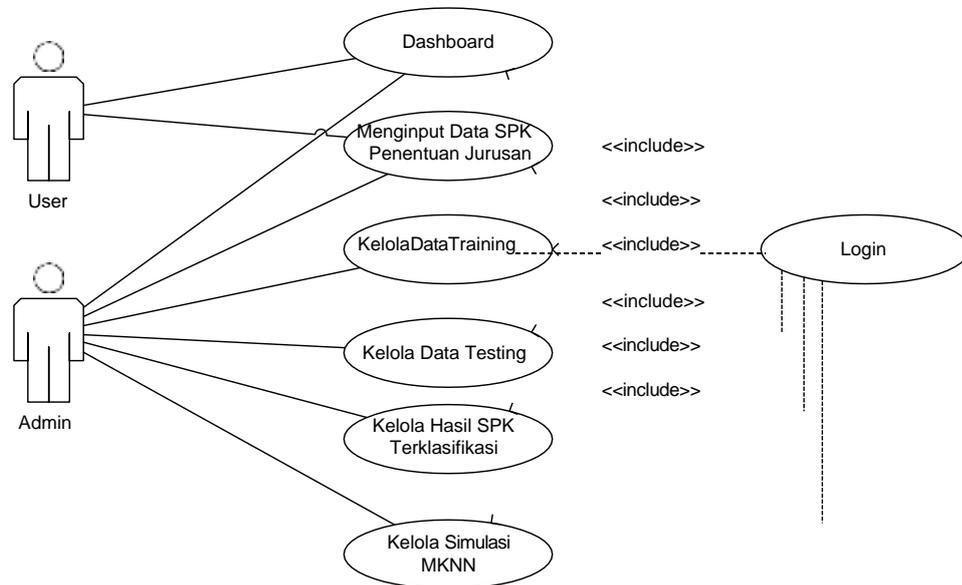
Tabel 1. Hasil Pemodelan Klasifikasi MKNN

No	Nama Siswa	Bahasa Indonesia		Bahasa Inggris		Matematika		IPA		IPS		Minat	IQ	Kelas Aktual	Kelas Prediksi
		RBIN	UBIN	RBI	UBI	RM	UM	RIA	UIA	RIS	UIS				
1	M-107	1.000	0.333	1.000	0.333	0.667	0.667	1.000	0.667	0.500	0.333	1.000	0.500	IPS	IPS
2	M-108	0.500	0.000	0.000	0.000	0.333	0.000	0.000	0.000	0.000	0.000	1.000	0.250	IPS	IPS
3	M-109	0.500	0.667	0.500	0.667	0.667	0.667	0.500	0.667	1.000	0.667	1.000	0.500	IPS	IPS
4	M-111	0.500	0.333	0.500	0.333	0.333	0.333	0.500	0.333	0.500	0.333	1.000	0.250	IPS	IPS
5	M-113	1.000	0.333	1.000	0.333	0.667	0.667	1.000	0.333	1.000	0.333	1.000	0.750	IPS	IPS
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
96	M-321	0.500	0.333	0.500	0.333	0.333	0.333	1.000	0.333	0.500	0.333	0.000	0.500	MIPA	MIPA

Selanjutnya dilakukan evaluasi terhadap pemodelan klasifikasi dengan menggunakan *confusion matrix*. Nilai evaluasi confusion matrix terhadap pemodelan klasifikasi menggunakan MKNN menghasilkan akurasi sebesar 82,29%; 100% nilai presisi; dan 75% nilai evaluasi recall.

### 3.5. Analisis dan Perancangan Sistem

Secara fungsional, analisis kebutuhan sistem diuraikan pada Usecase diagram pada Gambar 5.



Gambar 5. Use Case Diagram pada Sistem yang Dihasilkan

### 3.6. Implementasi Database dan Pengkodean Sistem

Implementasi database dirancang menggunakan database MySQL PhpMyAdmin. Tahap pengkodean sistem yang dilakukan menggunakan bahasa pemrograman php. Form dashboard menampilkan informasi berupa akumulasi data uji, latih, dan data diklasifikasi oleh sistem. Sedangkan pada Gambar 7 ialah laman yang dipakai admin untuk memasukkan data peserta didik baru yang ingin diklasifikasikan.



Gambar 6. Halaman Beranda Admin



Gambar 7. Form Input Data

Pada Gambar 8, tabel yang menampilkan hasil klasifikasi data siswa seluruhnya dengan hasil perkiraan jurusan sesuai pemodelan algoritma MKNN. Sedangkan pada Gambar 9 berupa tampilan laman yang digunakan Admin untuk melakukan percobaan dengan memasukkan nilai parameter  $k$  pada proses klasifikasi jurusan.



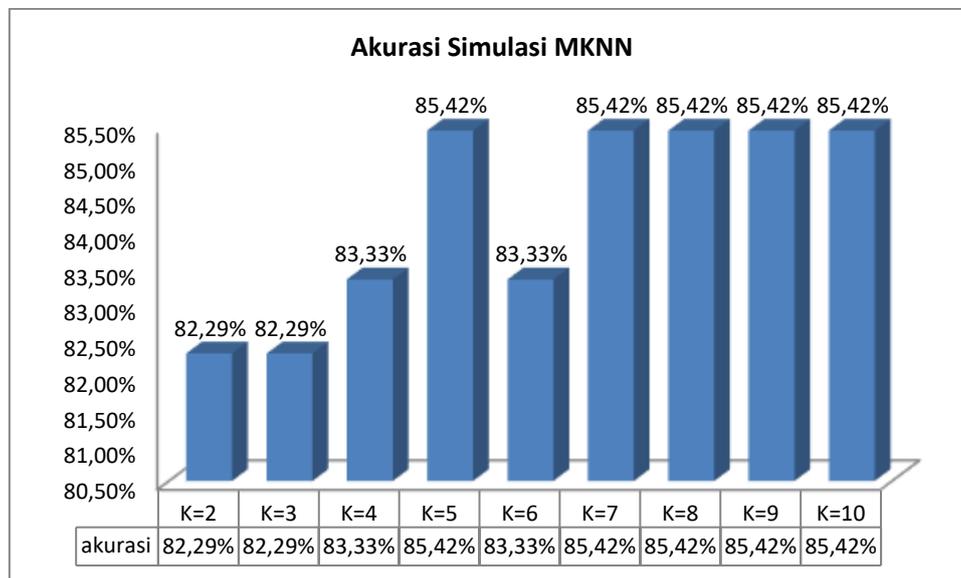
Gambar 8. Halaman Siswa Terklasifikasi



Gambar 9. Halaman Simulasi MKNN

### 3.7. Perbandingan Hasil Simulasi Parameter $K$

Perbandingan simulasi dilakukan guna mengukur akurasi maksimal nilai parameter  $k$  pada pemodelan MKNN. Grafik perbandingan nilai akurasi simulasi dapat dilihat pada gambar 10.



Gambar 10. Akurasi Hasil Simulasi MKNN

## 4. Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa seluruh atribut yang ada dalam penelitian ini dapat dipergunakan. Atribut dengan nilai *gain* tertinggi yaitu Rerata Raport IPS (RIS) dengan nilai 0,937, sedangkan atribut Rerata Raport Bahasa Inggris (RBI) menghasilkan nilai *gain* terendah dengan nilai 0,016. Kemudian, pada proses pembagian data menggunakan K-Means diperoleh hasil Uji Validasi DBI terendah sebesar 0,111 yang terdapat pada percobaan saat nilai  $k=2$ . Adapun pada percobaan klasifikasi penjurusan Siswa menggunakan MKNN didapatkan hasil bahwa nilai akurasi sebanyak 82.29%, presisi sebanyak 100.00%, dan juga recall sebanyak 75.00%. Selanjutnya, dilakukan implementasi hasil pemodelan MKNN ke dalam sistem Klasifikasi Jurusan Siswa yang sudah dirancang bangun dan mendapatkan nilai prediksi jurusan yang sesuai dengan pemodelan MKNN. Percobaan simulasi menggunakan nilai  $k$  yang bervariasi diperoleh hasil nilai akurasi terbaik sebesar 85,42% pada saat  $k=5$ . Berdasarkan pengujian sistem dengan teknik black-box testing dapat disimpulkan bahwa sistem telah terbebas dari kesalahan fungsional dan berhasil untuk dijalankan.

## Daftar Pustaka

- [1] Kadafi AR. Perbandingan Algoritma Klasifikasi untuk Penjurusan Siswa SMA. Jurnal ELTIKOM. 2018; 2(2); 67-77.

- [2] Hidayah N. Klasifikasi Penjurusan Program Studi Sekolah Menengah Atas dengan Algoritma Naïve Bayes Classifier pada SMAN 1 Subah. Universitas Dian Nuswantoro. Semarang. 2013.
- [3] Jananto A. Memprediksi Kinerja Mahasiswa Menggunakan Teknik Data Mining (Studi Kasus Data Akademik Mahasiswa Unisbank). Yogyakarta: Universitas Gajah Mada. 2010.
- [4] Suprawoto T. Klasifikasi Data Mahasiswa Menggunakan Metode K-Means untuk Menunjang Pemilihan Strategi Pemasaran. Jurnal Informatika dan Komputer (JIKO). 2016; 1(1).
- [5] Mutrofin S, Izzah A, Kurniawardhani A, Masrur M. Optimasi Teknik Klasifikasi Modified K-Nearest Neighbor Menggunakan Algoritma Genetika. Jurnal Gamma. 2015; 10(1).
- [6] Mustakim. Effectiveness of K-Means Clustering to Distibute Training Data and Testing Data on K-Nearest Neighbor Classification. Journal of Theoretical and Applied InformationTechnology. 2017; 95(21).
- [7] Gazalba I. Implementasi Algoritma Modified K-Nearest Neighbor untuk Klasifikasi Program Keluarga Harapan Kota Pekanbaru (Doctoral dissertation, Universitas Islam Negeri Sultan Syarif Kasim Riau). 2017.
- [8] Rahmadano ZF. Penerapan Metode Modified K-Nearest Neighbor untuk Klasifikasi Komentar *Fans* dan *Haters* di Instagram (Doctoral Dissertation, Universitas Islam Negeri Sultan Syarif Kasim Riau). 2021.
- [9] Mustakim, Gazalba I, Reza NGI. Comparative Analysis of K-Nearest Neighbor and Modified K-Nearest Neighbor Algorithm for Data Classification. In 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE). IEEE. 2017; 294-298.
- [10] Nabila Z, Isnain AR, Permata, Abidin Z. Analisis Data Mining untuk Clustering Kasus Covid-19 di Provinsi Lampung dengan Algoritma K-Means. Jurnal Teknologi Sistem Informasi (JTS). 2021; 2(2); 100-108.
- [11] Zheng X, Lei Q, Yao R, Gong Y, & Yin Q. Image Segmentation Based on Adaptive K-Means Algorithm. EURASIP Journal on Image and Video Processing. 2018; (1); 1-10.
- [12] Sinaga KP, Yang MS. Unsupervised K-Means Clustering Algorithm. IEEE Access. 2020; 8; 80716-80727.
- [13] Chormunge S, Jena S. Efficient Feature Subset Selection Algorithm for High Dimensional Data. International Journal of Electrical and Computer Engineering (IJECE). 2016; 6; pp.1880-1888.
- [14] Atallah DM, Badawy M, El-Sayed A, Ghoneim MA. Predicting Kidney Transplantation Outcome Based on Hybrid Feature Selection and KNN Classifier. Multimedia Tools and Application. 2019; 78(14); 20383-20407.
- [15] Essra A. Analisis Information Gain Attribute Evaluation untuk Klasifikasi Serangan Intrusi. Journal Information System Development (ISD). 2016; 1(2).