ALGORITMA K-NEAREST NEIGBOR MODEL EUCLIDEAN DISTANCE DALAM KLASIFIKASI KELULUSAN PESERTA BAHASA INDONESIA PENUTUR ASING PADA BALAI BAHASA SUMATERA UTARA

Hot Riris Siburian¹, Efori Buulolo², Hukendik Hutabarat³

1,2,3 Program Studi Teknologi Informatika STMIK Budi Darma, Medan, Indonesia Email: ¹hotririssiburian6@gmail.com, ²buuloloefori21@gmail.com, ³hukendikhutabarat@gmail.com

Abstrak

Balai Bahasa Sumatera Utara juga menyelenggarakan program Bahasa indonesia bagi penutur asing (BIPA) adalah program pembelajaran keterampilan berbahasa Indonesia (berbicara, menulis, membaca, dan mendengarkan) bagi penutur asing. Bahwa selama ini pada kantor Balai Bahasa memiliki kesulitan mengklasifikasikan kelulusan peserta Bahasa Indonesia Penutur Asing (BIPA) yang layak lulus dan yang tidak lulus. Sehingga terjadi komplain dari peserta. Untuk mengatasi masalah diatas maka perlu pengklasifikasian kelulusan berdasarkan data-data kelulusan peserta BIPA sebelumnya. Algoritma K-Nearest Neigbor model Euclidean Distance yang merupakan sebuah metode untuk melakukan klasifikasi/pengelompokan terhadap suatu objek berdasarkan kriteria tertentu. Dengan menggunakan algoritma K-Nearest neigbor dalam pengelompokan peserta BIPA dengan menggunakan berbagai kriteria diharapkan dapat lebih cepat orang Asing untuk berbahasa Indonesia.

Kata Kunci: Bahasa Indonesia Penutur Asing, Algoritma K-Nearest Neigbor

Abstract

The North Sumatra Language Center also organizes an Indonesian language program for foreign speakers (BIPA) is an Indonesian language learning skills program (speaking, writing, reading, and listening) for foreign speakers. That so far the Balai Bahasa office has had difficulty classifying the graduation of Indonesian Foreign Speakers (BIPA) participants who are eligible to pass and who have not passed. So that complaints occur from participants. To overcome the above problems, it is necessary to classify graduation based on previous BIPA participant data. K-Nearest Neigbor Euclidean Distance model which is a method for classifying / grouping an object based on certain criteria. By using the K-Nearest neigbor algorithm in grouping BIPA participants by using various criteria, it is hoped that foreigners can speak Indonesian more quickly.

Keywords: Indonesian Foreign Speakers, K-Nearest Neigbor Algorithm

1. PENDAHULUAN

Kelulusan adalah kegiatan paling akhir dari manejemen seorang peserta didik, Kelulusan juga dapat diartikan dengan pernyataan dari lembaga pendidikan tentang telah diselesaiakannya program pendidikan yang harus diikuti peserta. Bahasa indonesia bagi penutur asing (BIPA) adalah program pembelajaran keterampilan berbahasa Indonesia (berbicara, menulis, membaca, dan mendengarkan) bagi penutur asing. BIPA diselanggarakan dalam upaya untuk mendukung usaha peningkatan fungsi bahasa Indonesia menjadi bahasa Internasional yang sedang digalang oleh Kementerian Pendidikan dan Kebudayaan (Kemdikbud). Bahasa indonesia dikelompokkan kedalam tiga (tiga) tingkatan yaitu kelas pemula, menengah dan atas. Pengajaran bahasa Indonesia bagi penutur asing atau pengajaran BIPA mempunyai peran yang amat penting dan strategis dalam memperkenalkan Indonesia kepada masyarakat Internasional. Hal ini karena pengajaran BIPA di samping merupakan media untuk menyebarluaskan bahasa Indonesia, juga merupakan media untuk menyampaikan berbagai informasi tentang Indonesia.[1]

Data Mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data dengan melakukan penggalian pola-pola dari data dengan tujuan untuk memanipulasi data menjadi informasi yang lebih berharga yang diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dari dalam basis data[2].

Klasifikasi Kelulusan Peserta Bahasa Indenesia Penutur Asing dinilai berdasarkan beberapa faktor kemampuan peserta yaitu mendengarkan, merespon kaidah, membaca, menulis dan berbicara. Bahwa selama ini pada kantor Balai Bahasa memiliki kesulitan mengklasifikasikan kelulusan peserta Bahasa Indonesia Penutur Asing (BIPA) yang layak lulus dan yang tidak lulus. Sehingga terjadi komplain dari peserta.

2. TEORITIS

2.1 Implementasi

Dalam kenyataannya, implementasi merupakan proses untuk melaksanakan ide, program atau seperangkat aktivitas baru dengan harapan orang lain dapat menerima dan melakukan perubahan. Dalam konteks implementasi kurikulum pendekatan-pendekatan yang telah dikemukakan diatas memberikan tekanan pada proses, suatu aktivitas yang digunakan untuk mentransfer ide atau gagasan, program yang dituangkan dalam bentuk kurikulum desain (tertulis) agar dilaksanakan sesuai dengan desain tersebut. Masing-masing pendekatan itu mencerminkan tingkat pelaksanaan yang berbeda[2][3].

ISSN 2597-4645 (media online)

2.2 Data Mining

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (machine learning) untuk menganalisa dan mengekstrak pengetahuan secara otomatis. Data mining berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan diwaktu yang akan datang. Polapola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan lainnya[3].

2.3 Algoritma K-Nearest Neigbor

Dalam Algoritma K-Nearest Neigbor kita mengelompokkan suatu data baru berdasarkan jarak data baru ke beberapa data/tetangga (Neigbor) terdekat, daalam hal ini jumlah data/tetangga terdekat ditentukan oleh user yang dinyatakan dengan K. K-Nearest Neigbor merupakan salah satu model yang digunakan dalam algoritma supervised yang hasil instance diklasifikasikan berdasarkan mayoritas dari kategori pada K-Nearest Neigbor.K-Nearest Neigbor bekera berdasarkan jarak minimum dari data baru ke tetangga terdekat yang sudah diterapkan

Tujuan dari algoritma ini $A = \pi r^2$ adalah mengklasifikasikan objek baru berdasarkan atribut dan training sample. Jauh dekatnya tetangga biasanya dihitung berdasarkan Euclidean Distance yang dipresentasikan sebagai berikut [4]:

$$E(x,y) = \sqrt{\sum_{i=0}^{n} (xi - yi)^2}$$
 (1)

Keterangan:

xi = Sample Data

yi = Data uji atau Testing

n= Deimensi Data

i= Variable Data

Algoritma K-Nearest Neighbor (KNN) yaitu:

- 1. Mentukan parameter K, K merupakan Tetangga terdekat
- 2. Setelah mendapatkan K, kemudian hitung jarak antara data baru dengan semua data training
- 3. Urutkan jarak tersebut dan tetapkan tetangga terdekat berdasarkan jarak minimum ke-K
- 4. Periksa kelas dari tetangga terdekat
- 5. Gunakan mayoritas sederhana dari kelas tetangga terdekat sebagai nilai prediksi data baru.

3. ANALISA DAN PEMBAHASAN

3.1 Analisa Masalah

Analisa sistem sangat tergantung pada teori sistem umum sebagain sebuah landasan konseptual. Terdapat banyak pendekapan untuk analisa sistem dari pada dasarnya semua mempunyai tujuan yang sama, yaitu memahami sistem yang rumit kemudian melakukan moditifikasi dengan berbagai cara. Hasil modifikasi dapat berupa subsistem baru, komponen baru atau serangkain transformasi baru dan lain lain. Tujuannya adalah untuk memperbaiki berbagai fungsi di dalam sistem agar lebih efisien, untuk mengubah sasaran sistem untuk mrngganti output, untuk mencapai tujuan yang sama dengan seperangkat input yang lain atau untuk melakukan beberapa perbaikan serupa.

Data yang digunakan dalam mengelompokkan peserta bahasa penutur asing adalah seluruh data BIPA di Balai Bahasa Sumatera Utara. Selanjutnya akan dilakukan proses untuk menghasilkan data kasus pengelompokan. Dari data yang ada, kolom yang diambil sebagai atribut dalam pembentukan pengelompokan BIPA adalah:

- Seksi I Mendengarkan (Dengan jumlah soal sebanyak 40 soal)
- Seksi II Merespon Kaidah (Dengan jumlah soal sebanyak 25 soal) 2.
- Seksi II Membaca (Dengan jumlah soal sebanyak 40 soal) 3.
- 4. Seksi IV Menulis (Dengan jumlah soal sebanyak 6 soal)
- Seksi V Berbicara (Dengan jumlah soal sebanyak 40 soal)

Tabel 1. Daftar Nama dan Nilai Peserta BIPA di Balai Bahasa Sumatra Utara

No	Nama Peserta	Asal	Mendengarkan	Merespon	Mem	Menulis	Berbic	Kesimpulan
	BIPA	Negara		Kaidah	Baca		ara	
1	Yu Hao Min	China	440	365	485	1	3	Tidak lulus
2	Chong Haobing	China	625	395	380	3	2	Lulus
3	Ye Fang Ting	China	470	530	440	2	2	Lulus
4	Jessie Panazzalo	Australia	395	440	470	2	3	Tidak lulus
5	Jumpei Fukata	Jepang	470	485	515	3	3	Lulus
6	Majeed	Arab Saudi	320	395	410	3	1	Tidak lulus
7	Dhenver	Filipina	470	485	515	2	3	Lulus

	Quiliza Reyesn							
8	Pham T hi Minh Ngol	Vietnam	440	500	536	3	2	Lulus
9	Jamil Rizwan	Afganistan	320	395	360	1	2	Tidak lulus
10	Brian Kiprop Ngetich	Kenya	500	512	575	3	3	Lulus
11	Mc Donald Baluti	Malawi	488	470	515	3	2	Lulus
12	Daomingse	China	605	590	620	3	3	Lulus
13	Brayden	Australia	455	425	515	3	3	Lulus
14	Tiran	Kenya	395	380	320	1	2	Tidak lulus
15	Yuki Fira	Jepang	515	455	485	3	2	Lulus

3.2 Pre-Procesing Data

Pre procesing data merupakan proses yang dilakukan untuk pengerjaan klasifikasi kelulusan peserta bahasa indonesia penutur asing

Contoh kasus

Ada yang mengikuti peserta BIPA yang baru dengan nama Cristian Matheuss dari Negara Inggris, dengan nilai hasil ujiannya adalah mendegarkan memperoleh nilai 400, merespon kaidah memperoleh nilai 450, membaca memperoleh nilai 304, menulis memperoleh nilai 4, berbicara memperoleh nilai 4} yang menjadi pertanyaan apakah yang bernama Cristian Matheuss lulus atau tidak?.

Solusi:

Sesuai dengan algoritma K-NN maka langkah pertama yang harus dilakukan adalah menentukan paramater K. Pada kasus ini parameter K yang digunakan adalah 3. Selanjutkan menghitung jarak terdekat dengan menggunakan model Euclidean Distance. Perhitungannya sebagai berikut:

model Euclidean Distance. Perhitungannya sebagai berikut:

1.
$$C = \sqrt{(a_1 - a_a)^2 + (b_1 - b_b)^2 + (c_1 - c_c)^2 + (d_1 - d_d)^2 + (e_1 - e_e)^2}$$

$$= \sqrt{(440 - 400)^2 + (365 - 450)^2 + (485 - 304)^2 + (1 - 4)^2 + (3 - 4)^2}$$

$$= \sqrt{(1,600 + 7,225 + 32,761 + 9 + 1)}$$

$$= \sqrt{41,596}$$

$$= 203,9509$$
2. $= \sqrt{(625 - 400)^2 + (395 - 450)^2 + (380 - 304)^2 + (3 - 4)^2 + (2 - 4)^2}$

$$= \sqrt{(50,625 + 3,025 + 18,496 + 1 + 4)}$$

$$= \sqrt{72,151}$$

$$= 268,6093$$
3. $= \sqrt{(470 - 400)^2 + (530 - 450)^2 + (440 - 304)^2 + (2 - 4)^2 + (2 - 4)^2}$

$$= \sqrt{(4,900 + 6,400 + 18,496 + 4 + 4)}$$

$$= \sqrt{29,804}$$

$$= 172,6383$$
4. $= \sqrt{(395 - 400)^2 + (440 - 450)^2 + (470 - 304)^2 + (2 - 4)^2 + (3 - 4)^2}$

$$= \sqrt{(25 + 100 + 27,556 + 4 + 1)}$$

$$= \sqrt{27,686}$$

$$= 1166,3911$$
5. $= \sqrt{(470 - 400)^2 + (485 - 485)^2 + (515 - 304)^2 + (3 - 4)^2 + (3 - 4)^2}$

$$= \sqrt{(4,900 + 225 + 44,521 + 1 + 1)}$$

$$= \sqrt{50,648}$$

$$= 225,051$$

Tabel 2. Hasil perhitungan dengan model Euclidean Distance

 I	
No	Hasil Perhitungan
1	203,9509
2	268,6093
3	172,6383
4	1166,3911
5	225,051
6	143.7741

Volume 3, Nomor 1, Oktober 2019 DOI: 10.30865/komik.v3i1.1585

7	225,0577
8	240,6844
9	112,1338
10	295,4437
11	229,4994
12	401,8494
13	219,4903
14	72,0694
15	214,5134

Selanjutnya adalah pengurutan data hasil perhitungan menggunakan K-NN sebagai berikut:

Tabel 3. Pengurutan Data Hasil Perhitungan

Ranking	Nomor urut	Hasil Pengurutan Data		
1	14	72.0694		
2	9	112,1338		
3	6	143,7741		
4	4	1166,3911		
5	3	172,6383		
6	1	203,9509		
7	15	214,5134		
8	13	219,4903		
9	5	225,051		
10	7	225,0577		
11	11	229,4994		
12	8	240,6844		
13	2	268,6093		
14	10	295,4437		
15	12	401,8494		

Berdasarkan uraian diatas data nomor 15 atas nama Tiran asal Negara Kenya Pada Tabel 4.1 Dinyatakan Tidak Lulus, Data nomor 9 atas nama Jamil Rizwan asal Negara Afganistan pada Tabel 4.1 Dinyatakan Tidak Lulus, Data nomor 6 atas nama Majeed asal Negara Arab Saudi Pada Tabe 4.1 Dinyatakan Tidak Lulus. Berdasarkan parameter di atas maka dapat diklasifikasikan atas nama Cristian Matheuss dari Negara Inggris dinyatakan Tidak Lulus.

4. IMPLEMENTASI

Implementasi merupakan suatu tindakan untuk mewujudkan terlaksananya rencana dalam mencapai tujuan tertentu.Implementasi system meliputi spesifikasi kebutuhan perangkat keras (hardware), perangkat lunak (software) dan hasil pengujian. Perangkat keras merupakan salah satu komponen dari sebuah komputer yang sifat alatnya bias dilihat dan dapat disentuh oleh manusia secara langsung atau berbentuk nyata, yang berfungsi untuk mendukung proses komputerisasi. Perangkat lunak merupakan program-program komputer yang berguna untuk menjalankan suatu pekerjaan sesuai dengan dikehendaki. Program tersebut ditulis dengan bahasa khusus yang dimengerti oleh komputer.

Tampilan Menu Utama

Tampilan ini merupakan tampilan dari Form menu utama dimana berfungsi untuk melakukan pemanggilan

terhadap form-form yang lain.[5]



Gambar 1. Tampilan Menu Utama

Tampilan Form Input

Dibawah ini adalah tampilan Form input sebagai tempat menyimpan data. Berikut tampilan jika di klik Simpan maka data akan tersimpan, jika di klik Edit maka data akan di edit, jika di klik Hapus maka data akan terhapus, jika di klik Batal maka data batal disimpan dan jika di klik Keluar maka kembali ke menu utama.

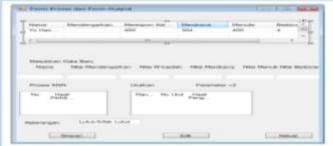
Volume 3, Nomor 1, Oktober 2019 DOI: 10.30865/komik.v3i1.1585

Perry Paper Character Char

Gambar 2. Tampilan Form Input

3. Tampilan Form Proses

Tampilan ini adalah proses dari klasifikasi kelulusan peserta BIPA



Gambar 3. Tampilan Form proses

5. KESIMPULAN

Berdasarkan uraian pada bab-bab sebelumnya, maka penulis mengambil keputusan sebagai berikut.

- 1. Dengan menggunakan Algoritma K- Nearest Neigbor Model Euclidean Distance dapat menjadi salah satu pemecah masalah dalam pengkasifikasian kelulusan peserta bahasa Indonesia pada Balai Bahasa Sumatera Utara
- 2. Dengan menggunakan algoritma K-Nearest Neigbor dapat memperoleh hasil pengklasifikasian kelulusan
- 3. Sistem yang digunakan untuk merancang hasil pengklasifikasian kelulusan peserta BIPA menggunakan *tools Visual Basic Net 2008*

6. REFERENCES

- [1] www.badanbahasa.kemdikbud.go.id/ukbi/v2, "Badan Bahasa."
- [2] F. A. Hermawati, Data Mining. Yogyakarta: CV.ANDI OFFSET, 2013.
- [3] E. Prasetyo, *Data Mining Mengolah data menjadi informasi menggunakan matlab*, 1st ed. Yogyakarta: CV.ANDI OFFSET, 2014.
- [4] A. Z. F. Nursetia wati, Abdul Syukur, P. T. I. U. D. N. Semarang, and E. Nursetiawati.imran@gmail.com, "PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN K-NEAREST NEIGHBOR BERBASIS PARTICLE SWARM OPTIMIZATION," vol. 6, no. 2, p. 103, 1393.
- [5] Hendrayudi, DASAR-DASAR PEMROGRAMAN MICROSOFT VISUAL BASIC 2008. Bandung: PT.SARANA TUTORIAL NURANI SEJAHTERA, 2010