

# Implementasi Algoritma K-Means Clustering pada Permasalahan Seleksi Mahasiswa Berprestasi menggunakan Visual Basic

Luthfiana Erlistya<sup>1</sup>, Patricia Saroha Sumarna<sup>2</sup>, Ferfigha Alsaka<sup>3</sup>

Universitas Airlangga<sup>1</sup>

Universitas Airlangga<sup>2</sup>

Universitas Airlangga<sup>3</sup>

erlistyaa@gmail.com

**Abstrak**— Predikat mahasiswa berprestasi merupakan predikat yang diunggulkan pada jenjang bangku perkuliahan. Hal ini yang dirasakan oleh pihak Fakultas Sains dan Teknologi Universitas Airlangga dengan kondisi mahasiswanya yang sangat antusias untuk berlomba-lomba meningkatkan kualitas individu masing-masing agar mampu mendapatkan predikat tersebut. Namun, terjadinya pandemi Covid-19 yang sudah menghancurkan segala aspek kehidupan pada setiap negara termasuk negara Indonesia, menjadikan kegiatan pada proses pendidikan mengalami perubahan dan permasalahan yang cukup drastis. Pada paper kali ini kami bertujuan untuk membantu permasalahan proses seleksi mahasiswa berprestasi menggunakan cabang ilmu *Data Mining* dengan Algoritma *K-Means clustering* agar mempermudah serta mengelompokkan pengambilan keputusan, kemudian data tersebut disimpan pada *Microsoft Access* sebagai *database* dan menampilkannya dengan *Visual Basic* untuk mempercepat proses seleksi, diharapkan juga proses seleksi ini bersifat transparan dengan adanya bantuan dari program tersebut. Sumber data diambil dari nilai Indeks Prestasi Kumulatif (IPK) dan nilai TOEFL pada beberapa mahasiswa Fakultas Sains dan Teknologi Universitas Airlangga yang telah mengisi *g-form* atau kuesioner. Hasil akhir pada program ini akan membantu pihak kampus dalam menganalisa para calon peserta seleksi mahasiswa berprestasi sesuai syarat dan kriteria yang telah tersedia.

**Kata kunci:** *Data Mining, K-Means Clustering, Microsoft Access, Visual Basic*

## I. PENDAHULUAN

Pandemi Covid-19 telah secara “ganas” meluluhlantakkan segala bidang yang ada pada seluruh negara di dunia. Jika dipandang dari bidang pendidikan, tentunya pendidikan di Indonesia mengalami perubahan yang sangat amat drastis. Terbiasa selalu menggunakan cara mengajar tatap muka, namun kini dihadapkan dengan realita yang mengharuskan mengajar dengan cara *online* atau secara virtual. Tidak hanya dari segi pengajaran, dari segi kegiatan non akademis seperti lomba-lomba pendidikan, penampilan minat bakat, olimpiade, seminar, dan masih banyak lagi kegiatan dari segi pendidikan yang terpaksa harus dilakukan secara virtual. Hal ini juga yang menjadikan masyarakat secara cepat bahu-membahu melakukan inovasi untuk mempermudah kegiatan yang sekarang dikenal dengan kegiatan di era *new normal* atau era kebiasaan baru. Pada era kebiasaan baru, tentunya kegiatan pendidikan hingga saat ini masih tetap dilakukan secara virtual, termasuk proses seleksi pemilihan mahasiswa berprestasi atau dikenal dengan pilmapres pada setiap fakultas di Universitas Airlangga.

Pilmapres atau Pemilihan Mahasiswa Berprestasi merupakan kompetisi mahasiswa yang diselenggarakan oleh Direktorat Jenderal Pembelajaran dan Kemahasiswaan yang ada di bawah naungan Kementerian Riset, Teknologi, dan Pendidikan Tinggi setiap tahunnya. Tujuannya untuk memberikan apresiasi kepada mahasiswa terbaik yang siap menjadi agen perubahan untuk membangun Indonesia yang lebih baik. Sedangkan mahasiswa berprestasi merupakan mahasiswa yang mencapai prestasi tinggi dalam bidang akademik maupun non akademik.

Untuk tetap menunjukkan eksistensi dan kualitas dari Fakultas Sains dan Teknologi Universitas Airlangga mengadakan pemilihan mahasiswa berprestasi secara virtual pada tahun 2020. Namun, dari situ kami melihat masih banyak kekurangan yang perlu diperbaiki dan ditingkatkan, sehingga penulis terinspirasi untuk melakukan sebuah pembuatan program untuk mempermudah tahapan seleksi mahasiswa berprestasi pada fakultas tersebut. Program ini menggunakan terapan

pengetahuan dari bidang ilmu *data mining* menggunakan algoritma *K-Means Clustering* yang akan ditampilkan menggunakan *Visual Basic*.

Data mining adalah analisa terhadap data untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut [1]. Sedangkan pada pengertian lain data mining adalah metoda yang digunakan untuk mengekstraksi informasi prediktif tersembunyi pada database, ini adalah teknologi yang sangat potensial bagi perusahaan yang sangat potensial bagi perusahaan dalam memberdayakan data warehouse [2]. Secara garis besar, data mining dapat dikelompokkan menjadi 2 kategori utama, yaitu deskriptive mining dan predictive.

Pada dasarnya *clustering* merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain [3]. *Clustering* merupakan teknik untuk mencari dan mengelompokkan objek atau data yang berkarakteristik sama kedalam sebuah kelas atau cluster, dimana clustering ini bersifat tanpa arahan (*unsupervised*) yang dapat dipelajari sendiri serta tidak memerlukan target output.

*K-means clustering* merupakan salah satu metode data *clustering* non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan cluster/ kelompok yang lain sehingga data yang berada dalam satu cluster/kelompok memiliki tingkat variasi yang kecil [3].

Visual Basic adalah salah satu aplikasi pemrograman under Windows yang berbasis pada visual atau grafis. Aplikasi ini dikeluarkan oleh Microsoft Cooperation yang juga pemilik dari sistem operasi Microsoft Windows. Visual Basic juga dapat diartikan sebagai sebuah bahasa pemrograman yang menawarkan *Integrated Development Environment (IDE)* visual untuk membuat program perangkat lunak berbasis sistem operasi *Microsoft Windows* dengan menggunakan model pemrograman (COM). [4].

Pada permasalahan pemilihan mahasiswa berprestasi secara virtual ini, akan dijelaskan topik tentang bagaimana proses seleksi mahasiswa berprestasi berdasarkan nilai IPK dan TOEFL menggunakan algoritma *K-Means clustering* dengan *visual basic* dan bagaimana hasil dari implementasi algoritma *K-Means clustering* pada studi kasus seleksi mahasiswa berprestasi. Secara garis besar, penelitian ini bertujuan untuk menjelaskan cara kerja seleksi mahasiswa berprestasi berdasarkan nilai IPK dan TOEFL menggunakan algoritma *K-Means clustering* dengan *visual basic* dan mengetahui hasil kelulusan pada seleksi mahasiswa berprestasi berdasarkan nilai IPK dan TOEFL menggunakan algoritma *K-Means clustering* dengan *visual basic*.

Paper ini memiliki beberapa manfaat, yang pertama yaitu untuk mengetahui pengertian, fungsi, cara kerja, dan penyelesaian suatu masalah menggunakan algoritma *K-Means clustering*, kemudian untuk memberikan informasi dan implementasi mengenai program *Microsoft Access* dan *Visual Basic* pada penyelesaian permasalahan sehari-hari, dan yang terakhir untuk mempermudah dan mempercepat kinerja civitas akademika kampus untuk melangsungkan kegiatan proses seleksi pemilihan mahasiswa berprestasi, khususnya pada Fakultas Sains dan Teknologi Universitas Airlangga, dan seluruh kampus yang ada di Indonesia.

## II. METODE PENELITIAN

Adapun langkah-langkah yang digunakan dalam penelitian ini adalah sebagai berikut:

### A. Studi Pustaka

Mempelajari mengenai algoritma *K-Means Clustering* dan penggunaan *Visual Basic* serta *Microsoft Access* untuk merancang aplikasi.

### B. Subjek Penelitian

Subjek penelitian ini adalah mahasiswa Fakultas Sains dan Teknologi Universitas Airlangga. Subjek ini dipilih berdasarkan hal yang ingin dikelompokkan yaitu mahasiswa berprestasi dengan parameter yang diambil adalah Indeks Perhitungan Kumulatif (IPK) dan skor TOEFL (*Test Of English as a Foreign Language*).

### C. Instrumen Penelitian

Instrumen Penelitian adalah alat atau fasilitas yang digunakan peneliti dalam mengumpulkan data agar pekerjaannya lebih mudah dan hasilnya lebih baik, dalam arti lebih cermat, lengkap, dan sistematis sehingga lebih mudah diolah [5]. Pada penelitian ini digunakan *google form* sebagai alat untuk mengumpulkan data. *Google form* adalah layanan dari *Google* yang memungkinkan

penggunanya untuk membuat survey, dengan fitur formulir online yang dibuat sesuai dengan kebutuhan sehingga pengguna bisa mendapatkan jawaban secara langsung dari responden yang mengikuti survey.

*D. Sumber Data*

Sumber data penelitian dibedakan menjadi dua, yaitu data primer dan data sekunder. Data primer adalah sumber data yang langsung memberikan data kepada pengumpul data sedangkan data sekunder adalah sumber data yang tidak langsung memberikan data kepada pengumpul data, misalnya lewat orang lain atau dokumen [6]. Data yang digunakan dalam penelitian ini adalah data primer dan sekunder. Data primer diperoleh dari menyebar *google form* ke mahasiswa Fakultas Sains dan Teknologi. Data sekunder dalam penelitian ini berupa data-data pendukung lainnya, baik berupa buku-buku, *software*, maupun internet.

*E. Teknik Pengumpulan Data*

Teknik pengumpulan data merupakan cara untuk memperoleh suatu data pada saat melakukan kegiatan penelitian. Pada penelitian, pengumpulan data dilakukan dengan memberikan kuisioner kepada mahasiswa Fakultas Sains dan Teknologi yang berupa *google form*. Data diambil dari tanggal 10 Maret 2021 sampai 19 Maret 2021 serta diperoleh 25 data dengan rincian sebagai berikut.

TABEL 1. DATA IPK DAN SKOR TOEFL MAHASISWA FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS AIRLANGGA

ID	Nama Responden	Jurusan	Skor Test TOEFL-PBT ( <i>paper-based test</i> )	IPK
D01	Asta	Matematika	550	3.45
D02	Muhammad Fajrul	Matematika	450	2.81
D03	Nur Farita Zakiyah Irfiani	Matematika	443	3.49
D04	Serly Elvita	Matematika	395	3.55
D05	Eka Putri Armadani	Matematika	447	3.25
D06	Aufa Wira Andhika	Matematika	587	3.67
D07	Yosefa Febrina	Matematika	460	3.62
D08	Mohammad Rafii Addarimi	Matematika	467	3.09
D09	Bagas Erryanto	Matematika	405	2.94
D10	Nur iftitahus Sa'adah Sydana	Matematika	443	3.1
D11	Ahmad Syarif N.A.	Matematika	580	3.68
D12	Phelia	Kimia	577	3.83
D13	Rizka Ismi Prastiwi	Matematika	390	3.29
D14	Nidya Putri Andini	Matematika	580	3.32
D15	Difandi Wahyu Hibatur	Matematika	450	3.53
D16	Della Nur Fitri	Matematika	468	3.23
D17	Fidayanti Ayu Hariyono	Matematika	470	3.17
D18	Fasya	Matematika	560	3.34
D19	Anak Agung Ayu Dhea Crystalia Putri	Matematika	600	3.58
D20	Sava Tasya Idelia	Matematika	387	2.89
D21	Erly	Statistika	480	3.6

D22	Luthfiana Erlistya	Matematika	355	3.14
D23	Patricia Saroha Sumarna	Matematika	452	3.52
D24	Ferfigha Alsaka	Matematika	550	3.89

#### F. Teknik Pengolahan Data

Teknik pengolahan data dalam penelitian ini menggunakan perhitungan komputasi *K-Means Clustering* yang diterapkan pada *Visual Basic*. Metode *K-Means* adalah salah satu metode dalam fungsi *clustering* atau pengelompokan. *Clustering* mengacu pada pengelompokan data, observasi atau kasus berdasar kemiripan objek yang diteliti. Sebuah *cluster* adalah suatu kumpulan data yang mirip dengan lainnya atau ketidakmiripan data pada kelompok lain [7]. *Clustering* adalah membagi objek data (bentuk, entitas, contoh, ketaatan, unit) ke dalam beberapa jumlah kelompok (grup, bagian atau kategori). Tujuan proses *clustering* adalah meminimalkan terjadinya *objective function* yang diset dalam proses *clustering*, yang pada umumnya digunakan untuk meminimalisasikan variasi dalam suatu *cluster* dan memaksimalkan variasi antar *cluster* [7].

Algoritma dasar *clustering* data menggunakan metode *K-Means* dapat dilakukan dengan cara [8]:

- Tentukan jumlah k *cluster*.
- Pilih titik secara random sebanyak K buah, dimana titik ini akan menjadi pusat (*centroid*) dari masing-masing kelompok (*clusters*).
- Alokasikan setiap data atau objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak antar kedua objek tersebut. Jarak paling dekat antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk ke dalam *cluster* yang mana. Rumus jarak *Euclidian* :

$$D(a, b) = \sqrt{\sum_{k=1}^n (a_k - b_k)^2}$$

dimana n adalah jumlah dimensi (atribut),  $a_k$  dan  $b_k$  adalah atribut ke-K dari objek data p dan q.

- Hitung kembali pusat *cluster* dengan anggota *cluster* yang sekarang. Pusat *cluster* adalah rata-rata semua data atau objek dalam *cluster*.
- Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* sudah tidak berubah lagi, maka proses peng-*cluster*-an selesai.
- Kembali ke langkah 3 sampai pusat *cluster* tidak berubah lagi

Penerapan *Visual Basic* untuk membuat aplikasi *K-Means Clustering* yang nantinya dapat digunakan oleh administrator, dimana admin dapat melakukan perhitungan *K-Means* dan mengubah *database* yang dimiliki sebelumnya. Sedangkan dari segi mahasiswa dapat mengetahui hasil *clustering* dari program yang menyatakan mahasiswa tersebut lolos atau tidak lolos sebagai mawapres dan dilanjutkan untuk mengikuti seleksi pada tahap selanjutnya. Langkah yang perlu dipahami dari proses kerja penelitian ini yaitu pada tahap pertama yang dilakukan adalah menghubungkan *database* (*Microsoft Acces*) dengan *Visual Basic*, pada tahap ini dibuatkan *module* untuk *syntax* yang akan menghubungkan *database* dengan *Visual Basic*. Setelah terhubung, dibuat desain aplikasi dengan membuat *form* untuk setiap tampilan. *Database* yang ada nanti akan dikelompokkan menggunakan *K-Means Clustering* pada *form calculate*. Berikutnya dimasukkanlah *syntax* untuk menghubungkan antar *form*, kemudian perhitungan untuk melakukan *K-Means Clustering*.

### III. HASIL DAN PEMBAHASAN

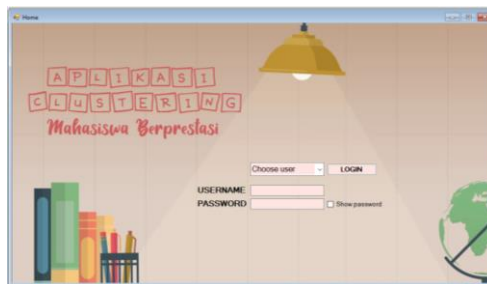
Berdasarkan penelitian yang dilakukan, dapat disimpulkan bahwa algoritma *K-Means* bisa digunakan untuk mengelompokkan permasalahan seleksi mahasiswa berprestasi berdasarkan nilai IPK dan nilai TOEFL. Dari data yang dimiliki, didapatkan 2 kelompok hasil dari pengolahan data, yaitu :

- *Cluster* pertama memiliki kode C1 yang menyatakan mahasiswa tersebut tidak lolos pada seleksi tersebut
- *Cluster kedua* memiliki kode C2 yang menyatakan mahasiswa tersebut lolos pada seleksi tersebut, sehingga dapat melanjutkan ke tahap wawancara

Didapatkan hasil centroid atau pusat *cluster* beraneka ragam hasil, ini dikarenakan titik awal centroid yang dapat dipilih secara bebas. Namun, untuk memaksimalkan peluang keberhasilan algoritma, kami mencoba untuk mengolah data pada *button* "calculate" kemudian kami menemukan suatu kesimpulan yaitu terdapat syarat dalam memilih nilai C, yang mana  $C1 < C2$ . Sehingga saat nilai C1 lebih kecil daripada C2, ini berarti nilai C1 merupakan nilai yang rendah, sehingga merujuk ke pendefinisian tidak lolos. Sedangkan untuk C2 lebih besar dari pada C1, akan merujuk pada pendefinisian C2 merupakan nilai yang lebih tinggi, maka merujuk ke pendefinisian lolos.

#### A. Simulasi Program

Pada bagian ini akan dijelaskan mengenai langkah-langkah menjalankan program serta gambar. Jika aplikasi dibuka atau dijalankan, akan muncul tampilan awal atau *home* seperti berikut.



GAMBAR 1. FORM HOME

Kemudian *user*/pengguna bisa memilih opsi apakah ingin masuk sebagai admin atau mahasiswa. Jika masuk sebagai admin maka pastikan username dan password yang dimasukkan benar. Jika ingin memastikan apakah password yang dimasukkan benar maka Anda bisa mencentang show password sehingga password terlihat. Berikut tampilan jika masuk sebagai mahasiswa dan admin secara berturut-turut.

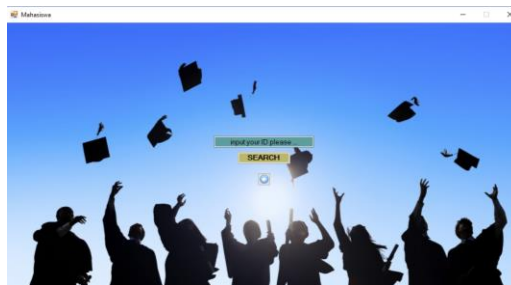


GAMBAR 2. FORM MAHASISWA



GAMBAR 3. FORM ADMIN

Pada form mahasiswa terdapat *click button* show your result dan about “mahasiswa berprestasi”. Pada form admin terdapat *click button* database dan calculate.



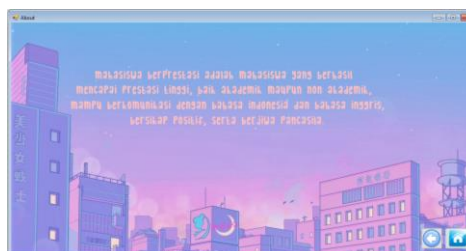
GAMBAR 4. FORM MAHASISWA → SHOW YOUR RESULT

Anda diminta untuk memasukkan ID pendaftaran (contoh D01). Kemudian untuk melihat hasil apakah Anda lolos sebagai mahasiswa berprestasi atau tidak maka pilih search. Berikut salah satu contoh tampilan ID mahasiswa.



GAMBAR 5. FORM MAHASISWA → SHOW YOUR RESULT → SEARCH

Ditampilkan pengumuman hasil yang menyatakan lolos atau tidak lolos sesuai dengan ID yang dimasukkan pada tahap sebelumnya.



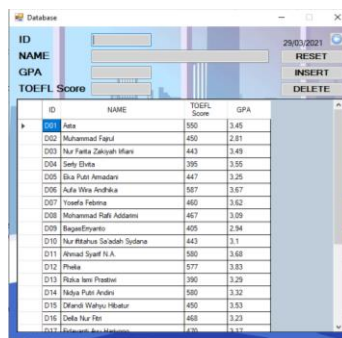
GAMBAR 6. FORM ABOUT

Tampilan di atas merupakan tampilan jika Anda memilih *button* About “Mahasiswa Berprestasi”. Dijelaskan singkat mengenai apa itu mahasiswa berprestasi.



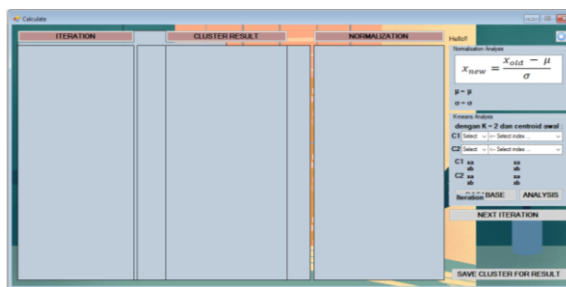
GAMBAR 7. FORM ADMIN

Sedangkan bila Anda masuk sebagai admin maka akan muncul *button* database dan calculate. Jika Anda ingin melihat, mengubah/edit, menghapus, atau menambahkan *database* maka Anda dapat memilih *button database*, namun jika Anda ingin menjalankan algoritma *K-Means Clustering* maka Anda pilih *Calculate*.



GAMBAR 8. FORM DATABASE

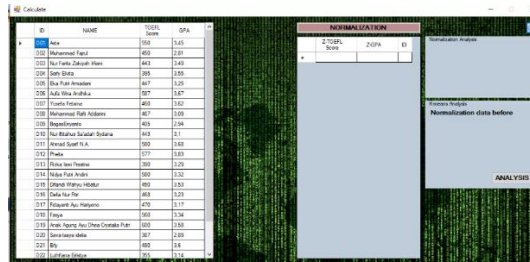
Anda dapat melihat, mengedit, menghapus, dan menambahkan *database* sesuai tampilan *button* yang ada. Saat akan menambahkan data, maka perlu mengisi kolom ID, Nama, IPK, dan Skor TOEFL kemudian pilih *insert*, maka data akan otomatis ditambahkan pada *database*.



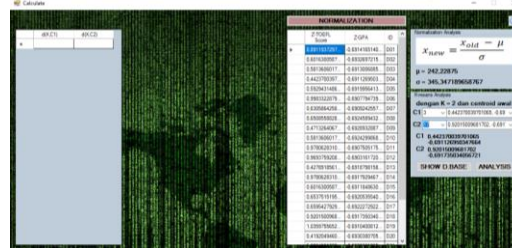
GAMBAR 9. FORM CALCULATE

Untuk memulai kalkulasi pilih tombol normalization, kemudian pilih centroid pada *listbox*. Selanjutnya pilih *button analysis* dan pilih *button iteration* dan akan muncul tabel cluster result. Lalu *click* next iteration sampai nilai pusat tidak berubah, dan Anda dapat menyimpan hasil kalkulasi pada Save Cluster For Result. Pada bagian ini, normalization berguna untuk menghindari terjadinya ketimpangan data. Dikarenakan perbedaan perbandingan skala antara variable IPK dan TOEFL, maka perlu adanya normalization agar nilai data menjadi tidak terlalu jauh jaraknya atau tidak terlalu timpang. Proses ini sering juga disebut *preprocessing*, sehingga akan dimiliki data yang lebih baik dan siap untuk diolah pada proses selanjutnya

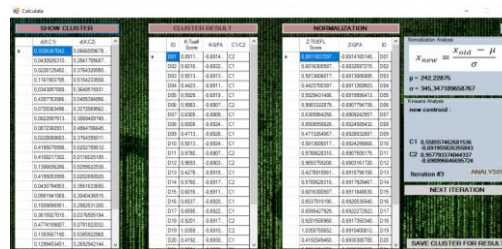




GAMBAR 10. TAMPILAN AWAL



GAMBAR 11. ITERASI AWAL (ITERASI KE-0)



GAMBAR 12. ITERASI KETIGA

Iterasi diakhiri pada iterasi ketiga dengan ciri-ciri pada kotak “K-Means Analysis” nilai C1 dan C2 sudah tidak berubah atau saat di *click* pada *button* “next iteration” nilai tersebut akan muncul dengan hasil yang sama. Selain itu terdapat tombol *back* (silang) yang terletak di pojok kanan atas, untuk kembali ke halaman sebelumnya, dan untuk *form* admin terdapat tombol *log out* untuk keluar dari akun yang telah masuk atau *log in*. Tidak ditampilkannya proses iterasi 1 dan 2 untuk meringkas pembahasan.

#### IV. SIMPULAN DAN SARAN

Pada hasil akhir penelitian yang dilakukan dari paper ini, terdapat beberapa kesimpulan yang diambil. Pertama, algoritma *K-Means clustering* mampu menentukan hasil seleksi pilmapres berdasarkan kelompok yang ada. Pada *database* dengan *centroid* awal antara data ke-3 dan data ke-17, menghasilkan 8 Mahasiswa yang lolos pada pilmapres, dan 16 Mahasiswa tidak lolos seleksi pilmapres. Namun saat *centroid* awal diganti, ternyata hasil keputusan / *output* bagi seleksi mahasiswa tetap sama. Sehingga ini berarti, algoritma *K-Means clustering* cukup akurat pada studi kasus kali ini. Kedua, Semakin banyak *data base* yang dimiliki, maka semakin banyak *clustering centroid* yang terbentuk diproses metode *K-Means*.

Selain mendapatkan kesimpulan, penulis juga menyampaikan saran kepada pembaca karena tentunya penulis memiliki kekurangan yang mana kekurangan itu diharapkan dapat diperbaiki oleh pembaca. Pertama, diharapkan pada penelitian selanjutnya dapat dilakukan penelitian berbasis *web* dan menggunakan *data base* berbasis internet, sehingga dapat diakses melalui laman *web* online sehingga dapat dijangkau dari segala wilayah domisili seluruh mahasiswa yang mengikuti pilmapres tersebut. Kemudian juga diharapkan untuk mempermudah proses pengumpulan data, penulis dapat mengakses *cyber campus* sebagai admin atau layanan database web kampus mahasiswa, agar pengumpulan data tidak manual namun otomatis diambil nilai IPK dan hasil TOEFL yang ada pada *cyber campus*.



DAFTAR PUSTAKA

- [1] Mudzakir, Bayu Dwiputra. (2018). Pengelompokan Data Penjualan Produk Pada PT Advanta Seeds Indonesia Menggunakan Metode K-Means. Malang. JATI (Jurnal Mahasiswa Teknik Informatika).
- [2] Darmi, Yulia., Setiawan, Agus. (2016). Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk. Bengkulu. Jurnal Media Infotama Vol.12 No.2.
- [3] Sari, Venny Novita., Yupianti, Maharani, Dewi., (2018). Penerapan Metode K-Means Clustering Dalam Menentukan Predikat Kelulusan Mahasiswa Untuk Menganalisa Kualitas Lulusan. Bengkulu. JURTEKSI (Jurnal Teknologi dan Sistem Informasi) Vol. IV No.2, Jun 2018, hlm 133-140.
- [4] Wikipedia.com. (2021, 26 Januari). Visual Basic. Diakses pada 17 Maret 2021, dari [https://id.wikipedia.org/wiki/Visual\\_Basic](https://id.wikipedia.org/wiki/Visual_Basic).
- [5] Arikunto. 2006. Prosedur Penelitian Suatu Pendekatan Praktik. Edisi Revisi VI. Jakarta. Rineka Cipta.
- [6] Sugiyono (2015). Metode Penelitian Kombinasi (Mix Methods). Bandung: Alfabeta.
- [7] Rahmat, Brilian C.T.I, Gafar, Agum Agditama, Fajriani, Nurul. Dkk., (2017). Implementasi K-Means Clustering Pada Rapidminer Untuk Analisis Daerah Rawan Kecelakaan. Kendari.
- [8] Ellymunig.wordpress.com. (2014, 24 Maret). Metode K-Means. Diakses pada 25 Maret 2021, dari <https://ellymunig.wordpress.com/2014/03/24/metode-k-means/>
- [9] Hastuti, N. F., Saptono, R., & Suryani, E. (2012). Pemanfaatan Metode K-Means Clustering Dalam Penentuan Penerima Beasiswa. Jurnal Informatika.
- [10] Yurnas, D. O. (2015). Pengelompokan Penerima Bantuan Kesejahteraan Masyarakat Dengan Pendekatan Data Mining Terintegrasi Sistem Pendukung Keputusan. SENATKOM Vol.1, 15- 16.
- [11] Setiawan, R. (2016). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru. Jurnal Lentera ICT Vol.3 No.1, 78-79
- [12] Nisbet, R., Elder IV, J., & Liner, G. (2009). Handbook of Statistical Analysis and Data Mining Applications. Elsevier Inc.
- [13] Velmurugan, T. (2012). Efficiency of k-Means and K-Medoids Algorithms for Clustering Arbitrary Data Points. Int.J. Computer Technology & Applications, 1759.
- [14] Oyelade, Oladipupo, Obagbuwa, "Application of K-Means clustering algorithm for prediction of students academic performance", International Journal of Computer Science and Information Security, Volume 7, 2010.