

Penerapan Algoritma Klusterisasi dan Klasifikasi pada Tingkat Kepentingan Sistem Pembelajaran di Universitas Terbuka

Nadiyah Hidayati¹, Ade Irma Rizmayanti², Chintamia Bunga Sari Dewi³, Rhini Fatmasari⁴, Windu Gata⁵

^{1,2,3,5}STMIK Nusa Mandiri

e-mail: ¹114002368@nusamandiri.ac.id, ²14002347@nusamandiri.ac.id,
³chintamia.cbd@nusamandiri.ac.id, ⁵windu@nusamandiri.ac.id

⁴Universitas Terbuka

e-mail: riens@ecampus.ut.ac.id

Abstrak

Universitas Terbuka merupakan Perguruan Tinggi Negeri (PTN) ke-45 di Indonesia yang menerapkan sistem belajar terbuka dan jarak jauh, keberhasilan pembelajaran lebih ditentukan oleh adanya jiwa kemandirian dan motivasi tinggi dari mahasiswa. Untuk mengetahui keberhasilan sistem pembelajaran yang diberikan, dilakukan survei menggunakan kuesioner yang dibagikan kepada mahasiswa untuk mengetahui penilaian dari masing-masing mahasiswa. Tujuan dari penelitian ini adalah untuk mengkluster dan mengklasifikasi data hasil kuesioner tingkat kepentingan sistem pembelajaran Universitas Terbuka dengan menggunakan *software RapidMiner* 9.0.0.3. Metode klusterisasi yang digunakan adalah algoritma *k-medoids*, sedangkan metode yang digunakan untuk klasifikasi adalah algoritma *Naïve Bayes*, *k-NN*, dan *C4.5*. Dari pengolahan data tersebut didapatkan hasil 2 kluster dengan pembagian data sebanyak 273 pada kluster 0 dan kluster 1 sebanyak 97 data. Pada proses klasifikasi, algoritma *Naïve Bayes* memperoleh nilai akurasi paling tinggi dibandingkan dengan algoritma yang lain dengan nilai akurasi sebesar 72,70% dengan nilai *AUC* sebesar 0,499. Sedangkan algoritma *k-NN* memperoleh nilai akurasi sebesar 71,62% dengan nilai *AUC* sebesar 0,438 dan algoritma *C4.5* memperoleh nilai akurasi sebesar 68,92% dengan nilai *AUC* sebesar 0,450.

Kata Kunci: Klusterisasi, Klasifikasi, *RapidMiner*, Tingkat Kepentingan

Abstract

Universitas Terbuka is the 45th State University (PTN) in Indonesia which applies an open and distance learning system, the success of learning is more determined by the existence of a spirit of independence and high motivation from students. To find out the success of the learning system provided, a survey was conducted using a questionnaire that was distributed to students to determine the assessment of each student. The purpose of this study is to cluster and classify the results of questionnaire data on the importance of the Universitas Terbuka learning system using the RapidMiner 9.0.0.3 software. The clustering method used is the k-medoids algorithm, while the method used for classification is the Naïve Bayes algorithm, k-NN, and C4.5. From the data processing, 2 cluster results were obtained with data distribution of 273 in cluster 0 and cluster 1 for 97 data. In the classification process, the Naïve Bayes algorithm obtained the highest accuracy value compared to other algorithms with an accuracy value of 72.70% with an AUC value of 0.499. While the k-NN algorithm obtained an accuracy value of 71.62% with an AUC value of 0.438 and the C4.5 algorithm obtained an accuracy value of 68.92% with an AUC value of 0.450.

Keywords: Clustering, Classification, *RapidMiner*, Level of importance

1. Pendahuluan

Universitas Terbuka merupakan Perguruan Tinggi Negeri (PTN) ke-45 di Indonesia yang menerapkan sistem belajar

terbuka dan jarak jauh. Sistem belajar ini terbukti efektif untuk meningkatkan daya jangkau dan pemerataan kesempatan pendidikan tinggi yang berkualitas bagi

semua warga negara Indonesia, termasuk mereka yang tinggal di daerah-daerah terpencil, baik di seluruh nusantara maupun di berbagai belahan dunia. Sebagai institusi pendidikan jarak jauh, salah satu karakteristik dari Universitas Terbuka (UT) adalah sistem belajar secara mandiri sehingga adanya keterpisahan secara fisik antara mahasiswa dengan pengajar/dosen. Pendidikan Jarak Jauh (PJJ) mengutamakan inisiatif dan keaktifan siswa dalam proses belajar (*student centered*). Siswa yang memiliki kemandirian yang tinggi dalam belajar digambarkan sebagai orang yang mampu mengontrol proses belajar (Irmawaty, Iswanto, & Nupikso, 2018)

Penerapan teknologi informasi dalam dunia pendidikan juga dapat menghasilkan data yang berlimpah mengenai mahasiswa dan proses pembelajaran yang dihasilkan. Penumpukan data mahasiswa secara terus menerus akan memperlambat pencarian informasi terhadap data tersebut. Untuk mengatasi permasalahan tersebut maka perlu dibuatkan sistem yang memudahkan untuk pengelompokan data tersebut menggunakan teknik *data mining* dengan metode Klusterisasi.

Data mining merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar (Haryati, Sudarsono, & Suryana, 2015). *Data mining* melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan satu metode yang melakukan pengelompokan data dengan sistem partisi (Ainun, Isti, & Fachri, 2020). Metode *data mining* yang digunakan dalam penelitian ini adalah *clustering* dan *classification*, metode *clustering* mengidentifikasi objek yang memiliki kesamaan karakteristik tertentu, dan kemudian menggunakan karakteristik tersebut sebagai "vektor karakteristik" atau "centroid" (Yunita, 2018). *Clustering* memiliki peran yang penting dalam *data mining*, dimana teknik ini akan membagi data kedalam beberapa *cluster* sesuai dengan kemiripannya.

Pada *Clustering* terdapat beberapa algoritma pengelompokan untuk mengelompokan data secara mudah. Salah

satu algoritma pengelompokan adalah *K-medoids*. *K-Medoids* atau bisa disebut *Partitioning Around Medoids* (PAM) adalah salah satu dari algoritma *clustering* yang mirip dengan algoritma *K-Means*. Perbedaan dari kedua algoritma ini yaitu algoritma *K-Medoids* atau PAM menggunakan objek sebagai perwakilan (*medoid*) sebagai pusat *cluster* untuk setiap *cluster*, sedangkan *K-Means* menggunakan nilai rata-rata (*mean*) sebagai pusat *cluster*. Kelebihan dari Algoritma *K-Medoids* lebih mudah untuk mengatasi kelemahan yang sering terjadi pada algoritma *K-Means* seperti sensitive terhadap noise dan *outlier*, dimana objek dengan nilai yang besar yang memungkinkan menyimpang pada dari distribusi data. Dan hasil proses *clustering* dari algoritma *K-Medoids* tidak bergantung pada urutan masuk dari dataset (Riadi, Azhar, & Wicaksono, 2020).

Sedangkan metode *classification* adalah proses menempatkan objek tertentu (konsep) dalam satu set kategori, berdasarkan masing-masing objek (konsep) *property* (Septiani, 2017). Klasifikasi merupakan proses menemukan sebuah model dan fungsi yang mendeskripsikan dan membedakan data ke dalam kelas-kelas. Proses klasifikasi didasarkan pada empat komponen mendasar yaitu kelas, prediktor, pelatihan *dataset*, dan pengujian *dataset* (Nawangsih & Amirudin, 2019). Pada *classification* terdapat model yang paling populer diantaranya, *Decision/Classification Trees*, *Bayesian Classifiers/Naïve Bayes Classifiers*, *Neural Networks*, *Statistical Analysis*, *Genetic Algorithms*, *Rough Sets*, *K-Nearest Neighbor Classifier*, *Rule based Methods*, *Memory Based Reasoning*, *Support Vector Machines* (Rahmadi, Kaurie, & Susanti, 2020).

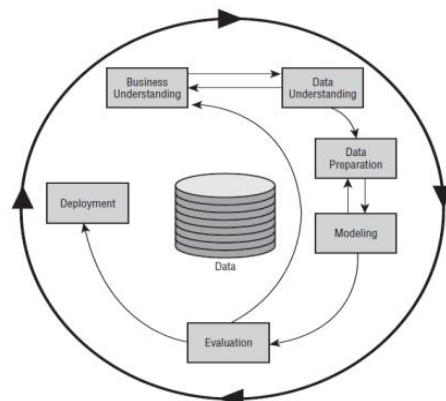
Maka dari itu peneliti menyelesaikan permasalahan mengenai pengelompokan data tingkat kepentingan sistem pembelajaran UT dengan menggunakan metode data mining *k-medoids clustering* dan komparasi klasifikasi algoritma *Naïve Bayes*, *k-NN*, dan *C4.5*.

Beberapa penelitian terkait penerapan data mining menggunakan metode *k-medoids clustering* yang pernah dilakukan diantaranya, (Simamora, Furqon, & Priyambadha, 2017) melakukan pengelompokan data kejadian tsunami yang disebabkan oleh gempa bumi.

(Defiyanti, Jajuli, & Rohmawati, 2017) meneliti pengelompokan mahasiswa yang berhak, dipertimbangkan atau tidak berhak menerima beasiswa Bantuan Belajar Mahasiswa (BBM). (Amalia, Narasati, & Faqih, 2019) meneliti tentang minuman yang sedang menjadi tren dalam masyarakat karena melihat minat yang besar di kalangan masyarakat. (Astria et al., 2020) melakukan pengklasteran wilayah pendistribusian listrik di Indonesia yang terdiri dari 34 provinsi. (Arianto, 2019) melakukan pengelompokan data untuk membantu pemerintah desa dalam pengelompokan penduduk kurang mampu sehingga bantuan dari pemerintah desa dapat tersalurkan dengan tepat. (Wira, Budiarto, & Wiguna, 2019) meneliti pola pemilihan program studi mahasiswa baru tahun 2018 di Universitas Kanjuruhan Malang. (Pramesti, Furqon, & Dewi, 2017) melakukan pengelompokan wilayah-wilayah dengan potensi kebakaran hutan/lahan dengan memanfaatkan data titik panas. (Damanik et al., 2019) melakukan pengelompokan desa yang memiliki fasilitas sekolah berdasarkan provinsi dan tingkat pendidikan. (Sundari et al., 2019) meneliti pengelompokan data imunisasi campak balita di Indonesia. (Ningsih et al., 2019) melakukan pengklasteran penduduk buta huruf menurut provinsi.

Dalam penelitian-penelitian sebelumnya mengenai penerapan data mining menggunakan metode *k-medoids clustering* belum ada yang menggunakan data tingkat kepentingan sistem pembelajaran yang didapatkan dari kuisioner mahasiswa UT. Data tersebut berisi mengenai identitas diri seperti jenis kelamin, usia, latar belakang pendidikan, UPBJJ-UT (Unit Program Belajar Jarak Jauh-Universitas Terbuka), program studi dan tahun registrasi, serta berisi penilaian sistem pembelajaran tersebut.

Penelitian ini menggunakan metode *Cross Industry Standard Process for Data Mining* (CRISP-DM) yaitu standar proses pengolahan *data mining* yang secara umum dapat dilihat pada Gambar 1. berikut ini:



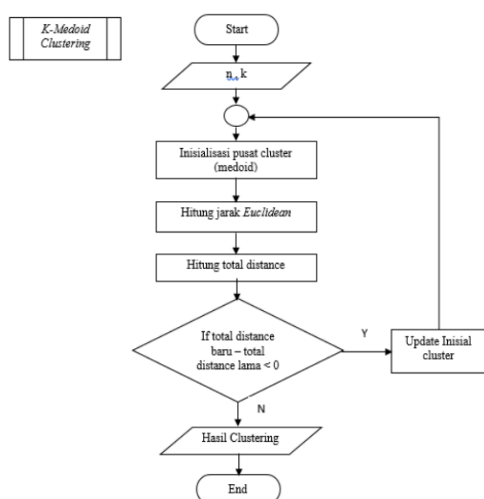
Gambar 1. Tahapan Penelitian CRISP-DM
Sumber: (Defiyanti et al., 2017)

Berikut ini merupakan penjelasan dari metode CRISP-DM (Feblian & Daihani, 2017):

- a. *Business Understanding*
Business Understanding merupakan tahap awal dalam metode CRISP-DM. Dalam *business understanding* dilakukan penentuan tujuan dari penelitian yang dilaksanakan dan menyiapkan strategi untuk mencapai tujuan yang telah ditetapkan.
- b. *Data Understanding*
Data Understanding merupakan tahapan pengumpulan data, mempelajari data, dan mengidentifikasi masalah yang berkaitan dengan data.
- c. *Data Preparation*
Pada tahap ini struktur basis data akan dipersiapkan sehingga mempermudah proses *mining*.
- d. *Modelling*
Modelling adalah tahap menentukan teknik data mining yang digunakan, menentukan *tools data mining*, algoritma *data mining*, dan menentukan parameter dengan nilai yang optimal.
- e. *Evaluation*
Evaluation adalah tahap interpretasi terhadap hasil *data mining* yang ditunjukkan dalam proses pemodelan.
- f. *Deployment*
Deployment adalah tahap penyusunan laporan atau presentasi dari pengetahuan yang didapat dari evaluasi pada proses *data mining*.

Penelitian ini bertujuan untuk menentukan tingkat kepentingan sistem pembelajaran di Universitas Terbuka dengan menerapkan metode *k-medoids clustering*. Dengan menggunakan metode

ini, data-data yang telah didapatkan dapat dikelompokkan kedalam beberapa *cluster* berdasarkan kemiripan dari data-data tersebut, sehingga data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster* dan yang memiliki karakteristik yang berbeda dikelompokkan dalam *cluster* yang lain yang memiliki karakteristik yang sama.



Gambar 2. Flowchart Algoritma *K-Medoids Clustering*

Sumber: (Pramesti et al., 2017)

Langkah-langkah melakukan *clustering* dengan metode *K-Medoids* adalah sebagai berikut (Hardiyanti, Tambunan, & Saragih, 2019):

1. Inisialisasi pusat *cluster* sebanyak *k* (jumlah *cluster*)
2. Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidian Distance* dengan persamaan:

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Dimana:

x_i : objek *x* ke-*i*

y_i : daya *y* ke-*i*

n: banyaknya objek

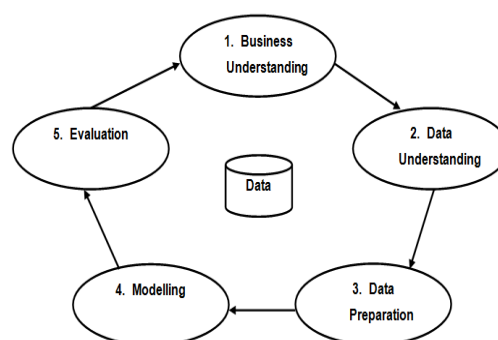
3. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat *medoid* baru.
4. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoids* baru.
5. Hitung total simpangan (*S*) dengan menghitung nilai total *distance* baru – total *distance* lama. Jika $S < 0$, maka tukar objek dengan data *cluster* untuk

membentuk sekumpulan *k* objek baru sebagai *medoids*.

6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan *medoids*, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

2. Metode Penelitian

Dalam penelitian ini peneliti menggunakan data mining dengan tahapan CRISP-DM, namun dalam penelitian ini hanya menggunakan lima tahapan. Berikut tahapan yang digunakan dalam penelitian ini:



Gambar 3. Tahapan Penelitian CRISP-DM

2.1. Business Understanding

Pada tahap ini berfokus pada pemahaman mengenai tujuan dari proyek dan kebutuhan secara persepektif bisnis, kemudian mengubah pengetahuan ini menjadi definisi masalah penambangan data dan rencana awal yang dirancang untuk mencapai tujuan (Kurniawati, Indrajit, & Fauzi, 2017). Tujuan utama dari penelitian ini adalah mengkluster sekaligus mengklasifikasi data tingkat kepentingan sistem pembelajaran Universitas Terbuka. Pada tahap ini juga dilakukan kegiatan untuk mempersiapkan strategi awal untuk mencapai tujuan tersebut.

2.2. Data Understanding

Pada tahap ini dilakukan pengumpulan, pengidentifikasi, dan pemahaman data yang akan digunakan pada penelitian ini. *Dataset* yang digunakan dalam penelitian ini yaitu data tingkat kepentingan sistem pembelajaran Universitas Terbuka yang didapatkan dari hasil kuisioner yang diisi mahasiswa.

2.3. Data Preparation

Pada tahap ini akan dilakukan beberapa penyeleksian untuk menghasilkan data yang dibutuhkan,

tahapannya yaitu (Iriadi & Nuraeni, 2016):

- a. *Data Cleaning* untuk membersihkan nilai yang kosong atau *tuple* yang kosong.
- b. *Data Integration* yang berfungsi menyatukan tempat penyimpanan yang berbeda kedalam satu data.
- c. *Data Reduction*, jumlah atribut yang digunakan terlalu banyak dan tidak semua atribut menjadi syarat atas atribut penentu.

Gambar 4. Data Tingkat Kepentingan

2.4. Modeling

Pada tahap *modeling* dilakukan proses klusterisasi dengan model yang diusulkan dalam penelitian ini adalah *k-medoids* dengan *software RapidMiner 9.0* untuk pengelompokan tingkat kepentingan sistem pembelajaran jarak jauh pada UT. Selanjutnya dilakukan klasifikasi dengan algoritma *Naïve Bayes*, *k-NN*, dan *C4.5*.

2.5. Evaluation

Pada tahap *evaluation* dilakukan proses klasifikasi dengan beberapa algoritma yaitu algoritma *Naïve Bayes*, *k-NN*, dan *C4.5* untuk melihat hasil *accuracy*, *AUC* dan *Curva ROC* pada *RapidMiner*.

3. Hasil dan Pembahasan

3.1. Tahap Pemahaman Bisnis (Business Understanding)

Pada tahap ini dilakukan pemahaman dari objek penelitian yang dituju, pemahaman ini mencakup tujuan objek penelitian yaitu pengklasteran dan pengklasifikasi terhadap data tingkat kepentingan sistem pembelajaran UT. Pada penelitian ini untuk dapat menghasilkan pengelompokan data dan nilai akurasi dilakukan penerapan pengklasteran dengan *K-Medoids clustering* dan beberapa model algoritma klasifikasi diantaranya, algoritma *Naïve Bayes*, *k-NN*, dan *C4.5*.

3.2. Tahap Pemahaman Data (Data Understanding)

Data Understanding merupakan dataset yang diambil dari hasil kuisisioner yang diisi mahasiswa UT tentang tingkat kepentingan sistem pembelajaran di UT. Berikut merupakan *data understanding* pada penelitian ini dapat dilihat pada Gambar 4. dibawah ini:

Pada Gambar 4. Data yang digunakan yakni identitas diri mahasiswa seperti usia, jenis kelamin, latar belakang pendidikan, asal kampus, program studi, dan tahun registrasi masuk serta hasil kuisisioner yang telah diisi mahasiswa UT. Tahap ini dimulai dengan pengumpulan data yang kemudian akan dilanjutkan dengan proses untuk mendapatkan pemahaman yang mendalam tentang data, mengidentifikasi masalah kualitas data, atau untuk mendeteksi adanya bagian yang menarik dari data yang dapat digunakan untuk hipotesa untuk informasi yang tersembunyi.

3.3. Tahap Persiapan Data (Data Preparation)

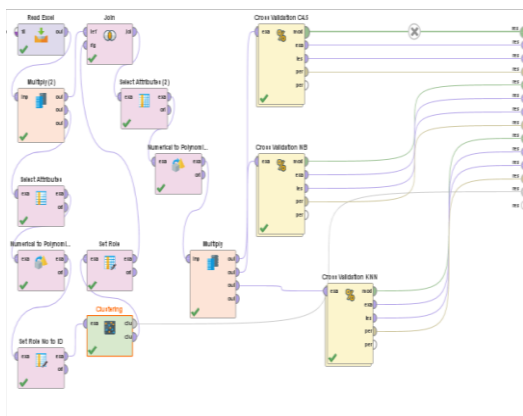
Pada tahap ini dilakukan proses pembersihan dan transformasi data yang awalnya masih berupa data campuran berisi teks dikonversi menjadi data *numeric*. Hal ini dimaksudkan agar data tersebut lebih mudah dibaca dan dimengerti. Berikut merupakan *data preparation* pada penelitian ini dapat dilihat pada Gambar 5. dibawah ini:

Gambar 5. Data Preparation Hasil Kuisisioner Tingkat Kepentingan

3.4. Tahap Pemodelan (Modelling)

Tahap pemodelan dilakukan dengan menyiapkan *dataset* yang akan digunakan dalam penelitian ini yaitu data tingkat kepentingan sistem pembelajaran

UT dengan menggunakan format *.xls seperti pada gambar berikut:

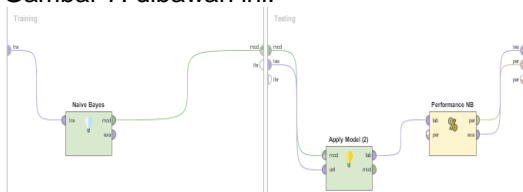


Gambar 6. Tahapan pemodelan data pada Rapidminer 9.0.3

Gambar 6. menunjukkan tahap pemodelan pada penelitian ini dengan melakukan proses klusterisasi terlebih dahulu, kemudian dilakukan proses klasifikasi dengan tiga algoritma yang digunakan yaitu algoritma *Naïve Bayes*, *k-NN*, dan *C4.5*. Penjelasan dari ketiga model algoritma tersebut sebagai berikut:

a. Model Algoritma *Naïve Bayes*

Penerapan model algoritma *Naïve Bayes* pada penelitian ini dapat dilihat pada Gambar 7. dibawah ini:

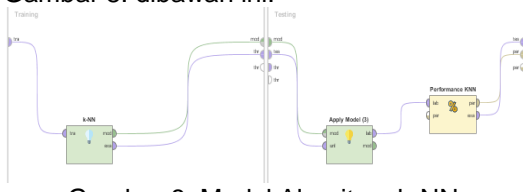


Gambar 7. Model Algoritma *Naïve Bayes*

Berdasarkan gambar diatas, tahapan klasifikasi *Naïve Bayes* dibagi menjadi dua subproses yaitu subproses *training* diisi dengan model *Naïve Bayes* dan subproses *testing* diisi dengan *apply model* dan *performance Naïve Bayes*

b. Model Algoritma *k-NN*

Penerapan model algoritma *k-NN* pada penelitian ini dapat dilihat pada Gambar 8. dibawah ini:



Gambar 8. Model Algoritma *k-NN*

Berdasarkan gambar diatas, tahapan klasifikasi *k-NN* dibagi menjadi dua subproses yaitu subproses *training* diisi dengan model *k-NN* dan subproses *testing* diisi dengan *apply model* dan *performance k-NN*

c. Model Algoritma *C4.5*

Penerapan model algoritma *C4.5* pada penelitian ini dapat dilihat pada Gambar 9. dibawah ini:

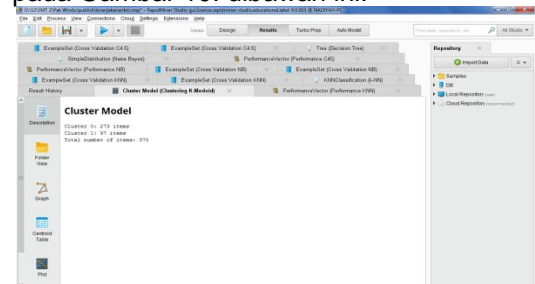


Gambar 9. Model Algoritma *k-NN*

Berdasarkan gambar diatas, tahapan klasifikasi *C4.5* dibagi menjadi dua subproses yaitu subproses *training* diisi dengan model *C4.5* dan subproses *testing* diisi dengan *apply model* dan *performance C4.5*

Hasil *K-Medoids Clustering*

Pada proses *k-medoids clustering* menentukan jumlah *k* kluster terlebih dahulu, yaitu sebanyak 2 kluster dengan menggunakan data sebanyak 370 data. Hasil *k-medoids clustering* dapat dilihat pada Gambar 10. dibawah ini:



Gambar 10. Hasil *K-Medoids Clustering*

Gambar 10. menunjukkan bahwa pada proses *clustering K-Medoids* dengan 2 kluster menghasilkan data sebanyak 273 pada kluster 0 dan kluster 1 sebanyak 97 data.

3.5. Tahap Evaluasi (*Evaluation*)

Klasifikasi yang dilakukan terhadap data tingkat kepentingan sistem pembelajaran UT menggunakan beberapa model algoritma diantaranya, algoritma *Naïve Bayes*, *k-NN*, dan *C4.5*. Dalam hal ini proses evaluasi akan menghasilkan

confusion matrix nilai accuracy dan nilai AUC (Area Under Curve) sebagai berikut:

a. Hasil Akurasi Algoritma Naïve Bayes

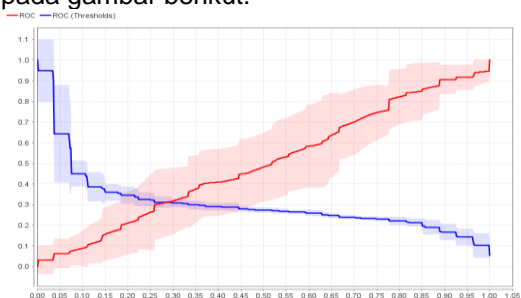
Hasil akurasi dengan model algoritma Naïve Bayes dapat dilihat pada Gambar 11. dibawah ini:

accuracy: 72.70% +/- 3.30% (micro average: 72.70%)

	true cluster_0	true cluster_1	class precision
pred. cluster_0	264	92	74.16%
pred. cluster_1	9	5	35.71%
class recall	96.70%	5.15%	

Gambar 11. Nilai Akurasi Algoritma Naïve Bayes

Berdasarkan gambar diatas, dapat diketahui hasil akurasi yang diperoleh sebesar 72,70%. Sedangkan nilai AUC yang dihasilkan sebesar 0,499 dimana dalam grafik ROC yang dapat diketahui pada gambar berikut:



Gambar 12. AUC Algoritma Naïve Bayes

b. Hasil Akurasi Algoritma k-NN

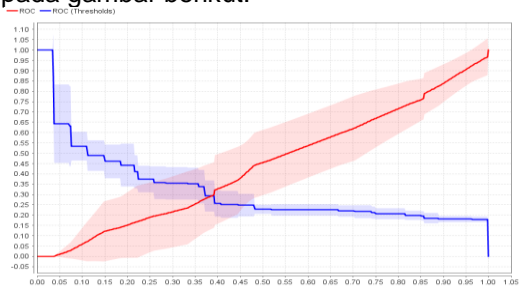
Hasil akurasi dengan model algoritma k-NN dapat dilihat pada Gambar 13. dibawah ini:

accuracy: 71.62% +/- 2.18% (micro average: 71.62%)

	true cluster_0	true cluster_1	class precision
pred. cluster_0	265	97	73.20%
pred. cluster_1	8	0	0.00%
class recall	97.07%	0.00%	

Gambar 13. Nilai Akurasi Algoritma k-NN

Berdasarkan gambar diatas, dapat diketahui hasil akurasi yang diperoleh sebesar 71,62%. Sedangkan nilai AUC yang dihasilkan sebesar 0,438 dimana dalam grafik ROC yang dapat diketahui pada gambar berikut:



Gambar 14. AUC Algoritma k-NN

c. Hasil Akurasi Algoritma C4.5

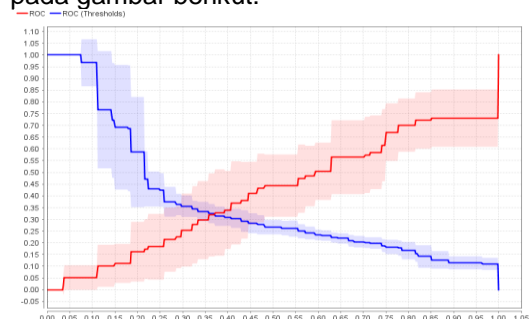
Hasil akurasi dengan model algoritma k-NN dapat dilihat pada Gambar 15. dibawah ini:

accuracy: 68.92% +/- 4.05% (micro average: 68.92%)

	true cluster_0	true cluster_1	class precision
pred. cluster_0	249	91	73.24%
pred. cluster_1	24	6	20.00%
class recall	91.21%	6.19%	

Gambar 15. Nilai Akurasi Algoritma C4.5

Berdasarkan gambar diatas, dapat diketahui hasil akurasi yang diperoleh sebesar 68,92%. Sedangkan nilai AUC yang dihasilkan sebesar 0,450 dimana dalam grafik ROC yang dapat diketahui pada gambar berikut:



Gambar 16. AUC Algoritma C4.5

Berdasarkan hasil yang didapatkan dari ketiga model tersebut, untuk melihat hasil nilai tertinggi dapat dilihat pada tabel berikut:

Tabel 1. Komparasi Hasil Algoritma Klasifikasi

	Naïve Bayes	k-NN	C4.5
Accuracy	72,70%	71,62%	68,92%
AUC	0,499	0,438	0.450

Dari Tabel 1. menunjukkan bahwa model algoritma klasifikasi yang mempunyai nilai tertinggi adalah algoritma Naïve Bayes dengan nilai accuracy sebesar 72,70% dan nilai AUC sebesar 0,499.

4. Kesimpulan

Berdasarkan pembahasan dan pengujian yang dilakukan, dapat disimpulkan bahwa hasil analisa menggunakan algoritma K-Medoids dengan melakukan pembagian cluster sebanyak 2 cluster sangat membantu dalam pengelompokan data tingkat kepentingan sistem pembelajaran di UT. Hasil analisa menggunakan beberapa algoritma

klasifikasi memperoleh nilai akurasi yang paling tinggi ditunjukkan pada algoritma *Naïve Bayes* sebesar 72,70% dengan nilai *AUC* sebesar 0,499, namun nilai akurasi dan *AUC* tersebut masih dikategorikan rendah.

Dengan demikian saran untuk penelitian selanjutnya agar dapat meningkatkan nilai akurasi dan *AUC* dengan menggunakan data yang lebih banyak atau menggunakan metode *data mining* klasifikasi lainnya untuk melakukan perbandingan. Selain itu penelitian ini juga dapat dikembangkan ke tahap *deployment* sesuai dengan CRISP-DM secara lengkap.

Referensi

- Ainun, E., Isti, W., & Fachri, S. (2020). IMPLEMENTASI ALGORITMA K - MEANS CLUSTERING TINGKAT KEPENTINGAN TAGIHAN RUMAH SAKIT DI PT PERTAMINA (PERSERO).
- Amalia, D. R., Narasati, R., & Faqih, A. (2019). Perbandingan Hasil Klasifikasi Rasa Minuman Thai Tea yang Paling Digemari Menggunakan K-means dan K-medoids, 401–407.
- Arianto, J. (2019). PENERAPAN DATA MINING UNTUK PENGELOMPOKAN PENDUDUK KURANG MAMPU DESA SAMBIREJO TIMUR DENGAN ALGORITMA K-MEDOIDS (STUDI KASUS KANTOR KEPALA DESA SAMBIREJO TIMUR), 3, 569–573. <https://doi.org/10.30865/komik.v3i1.1660>
- Astria, C., Hartama, D., Windarto, A. P., & Sudahri, I. (2020). Pengembangan Metode Datamining K-Medoid Pada Kasus Distribusi Listrik di Indonesia, 276–281.
- Damanik, I. I. P., Solikhun, Saragih, I. S., Parlina, I., Suhendro, D., & Wanto, A. (2019). Algoritma K-Medoids untuk Mengelompokkan Desa yang Memiliki Fasilitas Sekolah di Indonesia, (September), 520–527.
- Defiyanti, S., Jajuli, M., & Rohmawati, N. W. (2017). Optimalisasi K - Medoid Dalam Pengklasteran Mahasiswa Pelamar Beasiswa Dengan Cubic Clustering Criterion, 3(1), 211–218.
- Feblian, D., & Daihani, D. U. (2017). IMPLEMENTASI MODEL CRISP-DM UNTUK MENENTUKAN SALES PIPELINE PADA PT X, 1–12.
- Hardiyanti, F., Tambunan, H. S., & Saragih, I. S. (2019). PENERAPAN METODE K-MEDOIDS CLUSTERING PADA PENANGANAN KASUS DIARE DI INDONESIA, 3(2012), 598–603. <https://doi.org/10.30865/komik.v3i1.1666>
- Haryati, S., Sudarsono, A., & Suryana, E. (2015). IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA C4 . 5, 11(2), 130–138.
- Iriadi, N., & Nuraeni, N. (2016). KAJIAN PENERAPAN METODE KLASIFIKASI DATA MINING ALGORITMA C4.5 UNTUK PREDIKSI KELAYAKAN KREDIT PADA BANK MAYAPADA JAKARTA, 132–137.
- Irmawaty, Iswanto, Y., & Nupikso, G. (2018). EFEKTIVITAS PEMBELIAN BUKU MATERI POKOK (MODUL) UT MELALUI TOKO BUKU ONLINE (TBO) ANTARA HARAPAN DAN, 32–38.
- Kurniawati, I., Indrajit, R. E., & Fauzi, M. (2017). Peran Bussines Intelligence Dalam Menentukan Strategi Promosi Penerimaan Mahasiswa Baru, 1(2), 70–79.
- Nawangsih, I., & Amirudin. (2019). PENERAPAN ALGORITMA C4.5 UNTUK PREDIKSI AKURASI KEPUASAN PELANGGAN PO. SINAR JAYA, 8(1), 2009–2012.
- Ningsih, S. R., Damanik, I. S., Windarto, A. P., Tambunan, S. H., & Wanto, A. (2019). Analisis K-Medoids Dalam Pengelompokkan Penduduk Buta Huruf Menurut Provinsi, (September), 721–730.
- Pramesti, D. F., Furqon, M. T., & Dewi, C. (2017). Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan / Lahan Berdasarkan Persebaran Titik Panas (Hotspot), 1(9), 723–732.
- Rahmadi, M., Kaurie, F., & Susanti, T. (2020). Uji Akurasi Dataset Pasien Pasca Operasi Menggunakan Algoritma Naïve Bayes Menggunakan Weka Tools, 7(1), 134–139. <https://doi.org/10.30865/jurikom.v7i1.1761>
- Riadi, M., Azhar, Y., & Wicaksono, G. W. (2020). Implementasi Algoritma C5 . 0 Dan K-Medoids Untuk Klasterisasi, 2(4), 511–523.

-
- Septiani, W. D. (2017). KOMPARASI METODE KLASIFIKASI DATA MINING ALGORITMA C4.5 DAN NAIVE BAYES UNTUK PREDIKSI PENYAKIT HEPATITIS, *13*(1), 76–84.
- Simamora, D. A. S., Furqon, M. T., & Priyambadha, B. (2017). Clustering Data Kejadian Tsunami Yang Disebabkan Oleh Gempa Bumi Dengan Menggunakan Algoritma K-Medoids, *1*(8), 635–640.
- Sundari, S., Damanik, I. S., Windarto, A. P., Tambunan, H. S., Jalaluddin, & Wanto, A. (2019). Analisis K-Medoids Clustering Dalam Pengelompokan Data Imunisasi Campak Balita di Indonesia, (September), 687–696.
- Wira, B., Budianto, A. E., & Wiguna, A. S. (2019). IMPLEMENTASI METODE K-MEDOIDS CLUSTERING UNTUK MENGETAHUI POLA PEMILIHAN PROGRAM STUDI MAHASIWA BARU TAHUN 2018 DI UNIVERSITAS KANJURUHAN MALANG, *1*(3), 54–69.
- Yunita, F. (2018). Yunita, Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Islam Indragiri) *1* 238, 7(September), 238–249.