PREDIKSI KELULUSAN TEPAT WAKTU BERDASARKAN DATA REGISTRASI MAHASISWA DENGAN ALGORITMA *FUZZY C-MEANS* DAN *K-NEAREST NEIGHBORS*

SKRIPSI

UIN SUNAN AMPEL
S U R A B A Y A

DISUSUN OLEH:

SILVANA PUSPA NABILA

NIM: H76216047

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN AMPEL

SURABAYA

2020

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

NAMA : SILVANA PUSPA NABILA

NIM : H76216047

JUDUL : PREDIKSI KELULUSAN TEPAT WAKTU BERDASARKAN DATA REGISTRASI MAHASISWA DENGAN ALGORITMA *FUZZY C-MEANS* DAN *K-NEAREST NEIGHBOR*

Ini telah diperiksa dan disetujui untuk diujikan

Surabaya, 10 Agustus 2020

Menyetujui,

Dosen Pembimbing I	Dosen Pembimbing II
Nurissaidah Ulmnuha, M.Kom	Ahmad Yusuf, M.Kom
NIP: 199011022014032004	NIP: 199001202014031003

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi Silvana Puspa Nabila ini telah dipertahankan
di depan tim penguji skripsi
di Surabaya, 11 Agustus 2020

Mengesahkan,

Dewan Penguji

Penguji I	Penguji II
Nurissaidah Ulinnuha, M.Kom	Ahmad Yusuf M. Kom
NIP: 199011022014032004	NIP : 199001202014031003

Penguji III	Penguji IV
Mujib Ridwan, M.T	Khalid M.Kom
NIP: 198604272014031004	NIP: 197906092014031002

Mengetahui,

Plt. Dekan Fakultas Sains dan Teknologi

UIN Sunan Ampel Surabaya

Dr. Evi Fatimatur Rusydiyah, M.Ag.

NIP. 197312272005012003

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini.

Nama : Silvana Puspa Nabila

NIM : H76216047

Program Studi : SISTEM INFORMASI

Angkatan : 2016

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul: "PREDIKSI KELULUSAN TEPAT WAKTU BERDASARKAN DATA REGISTRASI MAHASISWA DENGAN ALGORITMA *FUZZY C-MEANS* DAN *K-NEAREST NEIGHBORS*". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 12 Agustus 2020

Yang menyatakan,

METERAI TEMPEL

6000 ENAM RIBU RUPIAH

(Silvana Puspa Nabila)

NIM: H76216076

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN
Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : SILVANA PUSPA NABILA

NIM : H76216047

Fakultas/Jurusan : FAKULTAS SAINS DAN TEKNOLOGI / SISTEM INFORMASI

E-mail address : silvana.puspa01@gmail.com / H76216047@uinsby.ac.id

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

■ Sekripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (.....)

yang berjudul :

PREDIKSI KELULUSAN TEPAT WAKTU BERDASARKAN DATA REGISTRASI MAHASISWA DENGAN ALGORITMA *FUZZY C-MEANS* DAN *K-NEAREST NEIGHBORS*

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Ekslusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara ***fulltext*** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 19 Agustus 2020

Penulis

(Silvana Puspa Nabila)
nama terang dan tanda tangan

ABSTRAK

PREDIKSI KELULUSAN TEPAT WAKTU BERDASARKAN DATA REGISTRASI MAHASISWA DENGAN ALGORITMA *FUZZY C-MEANS* DAN *K-NEAREST NEIGHBORS*

Oleh:

Silvana Puspa Nabila

Penerimaan mahasiswa baru di UINSA setiap tahunnya terus meningkat, namun ada sebuah permasalahan yakni tidak semua jumlah mahasiswa dapat lulus secara tepat waktu sesuai dengan kurikulum yang telah di tempuh. Mahasiswa yang berpotensi lulus tidak tepat waktu dapat di minimalisir apabila dapat mengetahui karakteristiknya. Salah satu caranya yakni dengan membuat model prediksi berbasis *machine learning* guna membantu memperbaiki konsep registrasi pada mahasiswa baru dengan metode FCM-KNN yang secara garis besar merupakan penggabungan antara dua metode yaitu tahapan pengelompokan data dengan menggunakan FCM dan tahap kategorisasi data menggunakan KNN. Pada penelitian ini digunakan data registrasi mahasiswa UINSA, kemudian pada pengujian skor digunakan *confusion matrix* dan *10-fold cross validation.* Hasilnya menunjukkan bahwa model prediksi pada metode FCM-KNN dengan pengujian 10-*fold cross validation* dengan skenario k-1 mempunyai rata rata akurasi 71%. Kesimpulannya adalah seiring bertambahnya nilai K, maka nilai akurasi juga berubah-ubah.

***Kata Kunci*—**Prediksi, FCM-KNN, *Machine learning,* Kelulusan tepat waktu, *Confusion Matrix, K-Fold.*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

ABSTRACT

PREDIKSI KELULUSAN TEPAT WAKTU BERDASARKAN DATA REGISTRASI MAHASISWA DENGAN ALGORITMA *FUZZY C-MEANS* DAN *K-NEAREST NEIGHBORS*

Oleh:
Silvana Puspa Nabila

The enrollment of new students at UINSA continues to grow every year, but the problem is that not all students can graduate on time according to the curriculum that has been taken. The potential of late graduation can be minimized if the students can find out their characteristics. One of the ways to achieve it to create a machine learning-based prediction model to improve the process of registration for new students with the FCM-KNN method, which in general is a combination of two methods for data grouping stage using FCM and the data categorization stage using the KNN. This research used the UINSA student registration data and the score testing used confusion matrix and k-fold cross validation. The results showed that the prediction model in the FCM-KNN method with 10-fold cross validation testing with the k-1 scenario has an average accuracy score of 71%. The conclusion is that as the K value increases, the accuracy score is also changed.

Keywords—Prediction, FCM-KNN, Machine learning, Graduate on time, Confusion Matrix, K-Fold.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR ISI

ABSTRAK ..... **vi**
ABSTRACT ..... **vii**
DAFTAR ISI ..... **viii**
DAFTAR TABEL ..... **x**
DAFTAR GAMBAR ..... **xi**
BAB 1 PENDAHULUAN **2**
1.1 Latar Belakang 2
1.2 Perumusan Masalah 4
1.3 Batasan Masalah 5
1.4 Tujuan Penelitian 5
1.5 Manfaat Penelitian 5
BAB II TINJAUAN PUSTAKA **6**
2.1 Tinjauan Pustaka Terdahulu 6
2.2 Dasar Teori 9
2.2.1 Data Mining 9
2.2.2 Metode Pembelajaran Model 9
2.2.3 Klastering 10
2.2.4 Klasifikasi 10
2.2.5 Distance 10
2.2.6 K-Nearest Neighbor 11
2.2.7 *Fuzzy C-Means* 12
2.2.8 FCM-KNN 15
2.2.9 *Confusion Matrix* 15
2.2.10 K-*Fold CrosssValidation* 16
2.2.11 Python 17
2.3 Integrasi Keilmuan 17
BAB III METODOLOGI PENELITIAN **20**
3.1 Desain Penelitian 20

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

3.1.1 Perumusan Masalah 20
3.1.2 Studi Pustaka 21
3.1.3 Pengumpulan Data 21
3.1.4 Pengolahan Data 23
3.1.5 Pelatihan Model 23
3.1.6 Testing 25
BAB IV HASIL DAN PEMBAHASAN 26
4.1 Pengumpulan Data 26
4.2 Pengolahan Data 27
4.2.1 Pre-Processing 27
4.2.2 Perhitungan FCM 31
4.2.3 Perhitungan KNN 42
4.3 Pelatihan Model 43
BAB V PENUTUP 47
5.1 Kesimpulan 47
5.2 Saran 47
DAFTAR PUSTAKA 48

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR TABEL

Tabel 2. 1 Tinjauan Pustaka Terdahulu 6
Tabel 2. 2 Label Confusion Matrix 15
Tabel 2. 3 Perhitungan Rumus Confusion Matrix 16

Tabel 3. 1 Atribut yang ada pada penelitian terdahulu 21
Tabel 3. 2 Deskripsi Atribut Data 22

Tabel 4. 1 Cuplikan data awal 26
Tabel 4. 2 Detail tahapan pre-processing pada setiap atribut 27
Tabel 4. 3 Data validation 29
Tabel 4. 4 Sampel dataset sebelum dilakukan pre processing 30
Tabel 4. 5 Sampel Dataset Sesudah Dilakukan Pre processing 30
Tabel 4. 6 Perbandingan data mahasiswa yang memiliki status lulus dan lulus tidak tepat waktu 31
Tabel 4. 7 Parameter FCM 33
Tabel 4. 8 Matriks pengacakan derajat keanggotaan 34
Tabel 4. 9 Perhitungan pusat cluster 1 pada iterasi ke 1 34
Tabel 4. 10 Perhitungan pusat cluster 2 pada iterasi ke 1 35
Tabel 4. 11 Hasil pusat cluster iterasi 1 35
Tabel 4. 12 Perhitungan fungsi objektif iterasi 1 36
Tabel 4. 13 Perhitungan perubahan matriks U iterasi 1 ke iterasi selanjutnya 36
Tabel 4. 14 Cek kondisi berhenti iterasi 1 37
Tabel 4. 15 Perhitungan iterasi cluster 1 iterasi ke 19 37
Tabel 4. 16 Perhitungan iterasi cluster 2 iterasi ke 19 38
Tabel 4. 17 Hasil pusat cluster iterasi 19 39
Tabel 4. 18 Perhitungan fungsi objektif iterasi 1 39
Tabel 4. 19 Perhitungan perubahan matriks U iterasi 19 ke iterasi selanjutnya 39
Tabel 4. 20 Cek kondisi berhenti iterasi 19 40
Tabel 4. 21 Hasil derajat keanggotaan pada iterasi ke 19 40
Tabel 4. 22 Pemilihan data masuk ke dalam cluster 1 atau 2 41
Tabel 4. 23 Data test 42
Tabel 4. 24 Hasil distance pemilihan centroid terdekat 42
Tabel 4. 25 Hasil jarak data cluster 1 43
Tabel 4. 26 Hasil data terkecil dan data sebenarnya 43
Tabel 4. 27 Skenario pembagian 2500 data. 44
Tabel 4. 28 Hasil pengujian akurasi nilai cluster dan K tanpa cross validation (train 60% dan test 40%) 44
Tabel 4. 29 Hasil pengujian akurasi nilai cluster dan K menggunakan cross validation. 45
Tabel 4. 30 Confusion matrix model FCM-KNN dengan skenario k=7 46

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR GAMBAR

Gambar 2. 1 Skenario Pengujian 10-Fold Cross Validation 17

Gambar 3. 1 Flowchart Model Penelitian 20
Gambar 3. 2 Flowchart Pelatihan dan Pengujian Model..... 24

Gambar 4. 1 Perbandingan grafik jumlah mahasiswa pada periode masuk 27
Gambar 4. 2 Perbandingan grafik jumlah mahasiswa lulus dan tidak lulus tepat waktu 31

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB 1
PENDAHULUAN

1.1 Latar Belakang

Pada setiap tahunnya penerimaan mahasiswa baru di tingkat perguruan tinggi sudah menjadi agenda rutin. Sejak tahun 2014 Universitas Islam Negeri Sunan Ampel Surabaya (UINSA) telah menggunakan sistem penerimaan mahasiswa baru (https://pmb.uinsby.ac.id/) dan sistem informasi akademik (https://ctrl.uinsby.ac.id/)

Penerimaan mahasiswa di UINSA sendiri, mempunyai jumlah kuota penerimaan yang terus meningkat di setiap tahunnya. Berdasarkan data yang didapatkan dari sistem informasi akademik bahwa jumlah mahasiswa baru yang diterima pada tahun 2014 menerima sebanyak 3,051 mahasiswa dan pada tahun 2015 mengalami peningkatan sebanyak 3,544 mahasiswa. Tetapi ada sebuah permasalahan yaitu tidak semua mahasiswa dapat lulus secara tepat waktu sesuai dengan kurikulum yang telah di tempuh.

Kelulusan saat ini menjadi salah satu hal yang berpengaruh terhadap akreditasi perguruan tinggi. Maka dari itu, dengan adanya nilai edukasi yang ideal dapat menjadikan elemen penilaian yang penting untuk kelulusan mahasiswa pada perguruan tinggi (*BANPT*, 2009). Selain itu, pihak perguruan tinggi pun juga akan diuntungkan jika mahasiswanya dapat lulus tepat waktu karena akan mempengaruhi penilaian akreditasi perguruan tinggi.

Dikatakan lulus tepat waktu apabila mahasiswa tersebut dapat menyelesaikan masa studinya kurang dari atau sama dengan empat tahun. Sedangkan dikatakan tidak lulus tepat waktu apabila menyelesaikan perkuliahan lebih dari empat tahun. Hal ini telah dijelaskan pada Undang-Undang tentang Pendidikan Tinggi Pasal 13 Nomor 12 Tahun 2012, yang membahas mengenai kemampuan mahasiswa dalam menyelesaikan program pendidikan sesuai dengan kecepatan belajar masing-masing individu dan tidak melebihi batas maksimum yang telah ditentukan oleh perguruan tinggi.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Pada saat ini, permasalahan yang ada yaitu pihak UINSA masih menggunakan seleksi manual serta belum bisa mengetahui karakteristik mahasiswa yang memiliki kemungkinan lulus tepat waktu lebih tinggi. Hal ini cukup penting untuk bisa membantu UINSA untuk melakukan pengambilan keputusan dalam menyeleksi mahasiswa baru. Oleh karena itu, perlu dilakukannya penelitian ini agar data tersebut bisa bermanfaat yakni dengan mengolah data menggunakan *data mining.* Definisi dari *data mining* itu sendiri merupakan sebuah teknik dalam ilmu statistik, matematika, kecerdasan buatan, dan *machine learning*, gunanya untuk mengektrasi serta mengidentifikasi suatu informasi yang berguna dari suatu pengetahuan mengenai dengan berbagai *database* (Turban dkk., 2005).

Untuk meminimalisir proses pengelompokan data, maka diperlukan beberapa metode dari *data mining* yaitu metode klastering dan klasifikasi. Metode klastering digunakan dalam mengelompokkan suatu data ke dalam *cluster* atau biasa disebut kelompok. Pada setiap *cluster*, data akan dibuat semirip mungkin sedangkan jarak antar *cluster* dibuat sejauh mungkin (Falahi, 2019), sedangkan klasifikasi merupakan sebuah proses untuk menentukan data tersebut masuk ke dalam salah satu kategori yang telah ditentukan sebelumnya (Wijaya & Santoso, 2016).

Nilai merupakan parameter penting dalam kesuksesan mahasiswa. Maka dari itu model prediksi ini sangat penting untuk membantu penyelenggara pendidikan perguruan tinggi guna melakukan sebuah tindakan untuk masa yang akan datang. Penelitian ini dilakukan di UINSA dengan menggunakan data mahasiswa angkatan 2014-2015 sebagai data testing dan data training. Untuk itu, penelitian ini dapat membantu pihak universitas dalam mengevaluasi dan memperbaiki bagaimana sistem pembelajaran yang telah berlangsung di dalamnya sehingga dapat menghasilkan lulusan yang baik dan berkualitas.

Pada penelitian sebelumnya mengenai prediksi menggunakan data registrasi mahasiswa juga dilakukan oleh (Yusuf dkk., 2012), penelitian tersebut juga menggunakan nilai mahasiswa sebagai parameternya. Dari data nilai mahasiswa tersebut akan di cluster menjadi beberapa kelompok yang telah sesuai dengan parameter jumlah cluster (k) yang sudah ditetapkan, dari hasil clustering tersebut maka

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

akan keluar hasil k-cluster data nilai mahasiswa. Hasil dari metode ini cukup baik terlihat dari nilai kesalahan RMSE sekitar 0.05- 0.08 dari dataset yang digunakan.

Selain itu penelitian terdahulu terkait dengan metode dibahas oleh (Afrianto & Kurniawati, 2013), ia menyatakan bahwa prinsip dari metode FCM-KNN yaitu dengan mengelompokkan data yang memiliki kesamaan terlebih dahulu sebelum pemberian label. Pada tahap pemberian label digunakan algoritma FCM. Menurut jurnal tersebut, dengan adanya tahap pengelompokan data terlebih dahulu diharapkan mampu meningkatkan performa kategorisasi data serta mampu menghemat waktu komputasi pada saat memilih k-tetangga terdekat. Karena dengan hal tersebut jarak antara k-tetangga terdekat cukup dicari pada lingkup anggota kelompok data dalam satu cluster dan tidak perlu membandingkan dengan data keseluruhan.

Tentunya penelitian ini bukanlah yang pertama, sudah banyak jurnal yang membahas mengenai topik prediksi kelulusan tepat waktu ini, tetapi jarang sekali yang menggunakan variabel data registrasi mahasiswa. Oleh karena itu, sudah dijelaskan sebelumnya bahwa penelitian ini akan menggunakan algoritma FCM-KNN dengan data registrasi mahasiswa yang diharapkan dapat mengetahui hasil akurasi untuk prediksi. Berdasarkan latar belakang yang sudah dipaparkan di atas, maka dilakukanlah penelitian yang berjudul **"PREDIKSI KELULUSAN TEPAT WAKTU BERDASARKAN DATA REGISTRASI MAHASISWA DENGAN ALGORITMA *FUZZY C-MEANS* DAN *K-NEAREST NEIGHBORS*"**. Diharapkan hasil dari penelitian ini mampu memberikan kontribusi praktis kepada UINSA terkait hasil prediksi kelulusan dan mampu memberikan kontribusi teoritis berupa akurasi algoritma FCM-KNN dalam melakukan prediksi kelulusan.

1.2 Perumusan Masalah

Rumusan masalah dari penelitian ini, adalah:

1. Bagaimana hasil dari membangun model algoritma FCM-KNN untuk prediksi kelulusan tepat waktu berdasarkan data registrasi mahasiswa pada studi kasus di UINSA?

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2. Bagaimana performa akurasi metode FCM-KNN dalam menyelesaikan permasalahan prediksi kelulusan tepat waktu berdasarkan data registrasi mahasiswa di UINSA?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini, adalah:

1. Ruang lingkup untuk penelitian ini dibatasi dengan data mahasiswa di UINSA.
2. Data yang digunakan dalam penelitian merupakan data mahasiswa UINSA yang terdaftar pada tahun 2014 – 2015.
3. Hasil prediksi diharapkan mampu untuk menjadi dasar acuan untuk pengambilan keputusan penerimaan mahasiswa.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini, adalah:

1. Untuk membangun model menggunakan metode FCM-KNN dalam prediksi kelulusan tepat waktu berdasarkan data registrasi mahasiswa di UINSA.
2. Untuk mengetahui performa metode FCM-KNN dalam menyelesaikan permasalahan prediksi kelulusan tepat waktu mahasiswa di UINSA.

1.5 Manfaat Penelitian

Manfaat dalam penelitian ini sangat luas, di antaranya yaitu:

1. Secara Akademis
   a. Mengimplementasikan ilmu pengetahuan yang didapat selama masa perkuliahan.
   b. Mendapatkan pengetahuan dan sumber referensi baru dalam menerapkan metode FCM dan KNN.
2. Secara Aplikatif
   a. Membantu pihak perguruan tinggi dalam melakukan prediksi kelulusan tepat waktu berdasarkan data registrasi mahasiswa.
   b. Membantu civitas akademik dalam memperbaiki konsep registrasi pada mahasiswa baru.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB II
TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka Terdahulu

Tentu saja penelitian ini bukan pertama kalinya dibuat, maka guna memberikan pemahaman antara penelitian terdahulu dengan penelitian yang akan dilakukan, maka akan dijelaskan pada Tabel 2.1 dibawah ini:

Tabel 2. 1 Tinjauan Pustaka Terdahulu

No.	Judul	Metode dan Hasil	Korelasi dengan Penelitian
1.	*Rekomendasi Multilabel Otomatis Pada Artikel Dengan Algoritma FCM-KNN* (Zehansyah dkk., 2019)	Metode yang digunakan yakni FCM-KNN *multilabel* otomatis untuk klasifikasi dokumen. Pada proses pengujian, hasil terbaik dari 5 percobaan *k* tetangga mulai dari *k=1* hingga *k=10.* didapatkan evaluasi F1 pada saat *k=1* dengan hasil evaluasi F1 sebesar 93,33% dan BEP 93,75%. Dari segi akurasi, 18 dokumen dapat terklasifikasi dengan benar dan 2 tidak terklasifikasi.	Menggunakan metode FCM-KNN untuk klasifikasi dokumen *online.*
2.	*Categorizing Document by FCM-KNN Approach* (Priandini dkk., 2017)	Penelitian ini mencoba mencari cluster terbaik dan menghasilkan empat cluster dan hasil evaluasi dari menggabungkan kedua	

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

		metode menunjukkan precision recall sebesar 0,7. Data yang digunakan menggunakan jurnal online berbahasa inggris yakni *sciencedirect.com*	
3.	*Kategori Dokumen Teks Secara Multi Label Menggunakan FCM dan KNN Pada Artikel Berbahasa Indonesia* (Afrianto & Kurniawati, 2013)	Metode FCM-KNN memberikan performa lebih optimal dalam klasifikasi dokumen multi label dibandingkan dengan metode MLKNN, yakni metode pengelompokan multi label lainnya. Data yang digunakan yakni menggunakan 175 berita dokumen online ber bahasa Indonesia.	
4.	*Penerapan K-Optimal Pada Algoritma KNN Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer FMIPA UNLAM Berdasarkan IPS Sampai Dengan Semester 4* (Banjarsari dkk., 2015)	Pada penelitian ini menggunakan data mining untuk memprediksi kelulusan mahasiswa berdasarkan variabel nilai IP dari semester 1 sampai semester 4. Hasilnya menemukan Nilai K-Optimal berdasarkan IP sampai dengan semester 4 adalah K=5. Dan proses K-Fold Cross Validation didapatkan tingkat akurasi untuk K=5 sebesar 80,00%	Menggunakan Data Mining untuk membantu memprediksi potensi kelulusan tepat waktu mahasiswa

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

5.	*Penerapan Metode Clustering Untuk Pengelompokan Mahasiswa Potensial Drop Out Menggunakan Algoritma K-Means ++* (Falahi, 2019)	Dengan menggunakan metode data mining pada pengelompokan mahasiswa yang berpotensial drop out. Mampu mengetahui variabel yang mempengaruhi potensi drop outnya mahasiswa pada perguruan tinggi.	
6.	*Klasifikasi dan Klastering Mahasiswa Informatika Universitas Muhammadiyah Surakarta* (Nugroho, 2015)	Pada penelitian ini menggunakan dua metode data mining untuk memperoleh masa studi yang efektif. Dan variabel yang digunakan yakni nilai IPK mahasiswa, masa studi, rata-rata jumlah SKS. Hasilnya yakni proses klastering menghasilkan 5 klaster (*cluster* 4 – *cluster* 0) dan variabel yang berpengaruh pada penelitian ini rerata SKS	

Dari pembahasan pada tabel 2.1 maka bisa disimpulkan bahwa metode penggabungan FCM-KNN sudah pernah digunakan pada penelitian terdahulu, namun studi kasusnya berbeda-beda. Berdasarkan penelitian sebelumnya, dikatakan metode FCM-KNN memiliki keakuratan dalam cluster dan dapat menghemat waktu komputasi dalam mencari nilai *k* tetangga terdekat (Zehansyah dkk., 2019). Oleh karena itu,

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

metode FCM-KNN ini akan di implementasikan untuk penelitian "Prediksi Kelulusan Tepat Waktu Berdasarkan Data Registrasi Mahasiswa". Diharapkan pada tahap penelitian ini untuk mendapatkan hasil yang optimal.

2.2 Dasar Teori

2.2.1 Data Mining

Data mining bukanlah suatu hal baru di bidang ilmu komputer dan dunia data. Maka dari itu, tujuan *data mining* yakni untuk memperbaiki suatu teknik atau cara tradisional sehingga dapat menangani dimensi data yang tinggi, jumlah data yang sangat besar dan data yang berbeda sifat dan heterogen (Maulana & Fajrin, 2018).

Penjelasan lainnya juga mendefinisikan bahwa *data mining* adalah proses menganalis sebuah pola data atau korelasi dari ratusan *field database* yang besar dengan cara yang berbeda dan menjadikan sebuah informasi menjadi penting yang dapat dipakai untuk meningkatkan sebuah keuntungan (Fansuri, 2012). Dalam pengelompokan *data mining* akan dibagi menjadi enam kelompok yaitu Deskripsi, Estimasi, Prediksi, Klasifikasi, Klaster dan Asosiasi (Maulana & Fajrin, 2018). Sedangkan untuk tahap pembentukan model *data mining* dibagi menjadi dua yakni data *testing* dan data *training.* Pada data *training,* data yang telah diketahui kelompoknya akan melatih model dengan menggunakan teknik algoritma tertentu. Setelah tahap sebelumnya selesai maka selanjutnya tahap *testing* atau tahap uji coba berguna untuk menguji nilai akurasi daripada model tersebut (Aradea dkk., 2011)

2.2.2 Metode Pembelajaran Model

a. Pembelajaran terawasi (*Supervised Learning*)

   Pada teknik pembelajaran ini, proses awalnya memerlukan beberapa data *training* dan data yang telah ditentukan kelas dan kelompok sebelumnya atau yang biasa disebut label dan keluarannya telah diketahui (Fansuri, 2012). Yang termasuk *supervised learning* diantaranya klasifikasi, regresi, dan prediksi.

b. Pembelajaran tak terawasi (*Unsupervised Learning*)

   Sedangkan pada teknik pembelajaran ini, sebuah struktur kelompok data berdasarkan suatu pola tertentu yang belum ditentukan sebelumnya dan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

belum terdapat output yang diperoleh (Darujati & Gumelar, 2012). Perbedaan dari pembelajaran *supervised learning* yaitu data *training* tidak perlu diberikan label.

2.2.3 Klastering

Klastering sendiri merupakan suatu proses pembentukan kelompok data (*cluster*) dari berbagai himpunan data yang belum diketahui kelas sebelumnya dan belum diketahui data tersebut masuk ke dalam *cluster* yang mana. Metode ini akan membagi data dalam beberapa bagian sehingga data yang dihasilkan nantinya mempunyai karakteristik yang sama di dalam satu *cluster*.

Tujuan dari klastering adalah untuk meminimalisasi variasi ke dalam satu *cluster* sehingga didapatkan hasil variasi maksimal antar *cluster*. Klastering termasuk ke dalam teknik pembelajaran *unsupervised learning*, dimana tidak memerlukan *fase learning* serta tidak perlu menggunakan tahap pelabelan pada tiap kelompok (Priyatman dkk., 2019).

2.2.4 Klasifikasi

Klasifikasi adalah sebuah proses awal dari pengelompokan data. Proses ini sangat penting dalam sistem informasi khususnya dalam *data mining* untuk mengetahui pengetahuan bisnis (Natalius, 2011). Adapun beberapa algoritma yang termasuk klasifikasi adalah *Naive Baiyes*, *k-Nearest Neighbor* (KNN), *Decision Tree*, *Artificial Neural Network* (ANN).

Tujuan dari klasifikasi adalah memperkirakan kelas dari suatu data yang labelnya tidak diketahui. Klasifikasi termasuk ke dalam *supervised learning* tujuannya untuk menemukan sebuah pola baru dari suatu data dan dapat menghubungkan data yang sudah ada sebelumnya dengan data baru (Krisandi, dkk., 2013).

2.2.5 Distance

Jarak atau *distance* pada sebuah data dilakukan untuk mengukur sebuah kemiripan suatu objek terhadap objek acuannya. Sebelum dilakukan pengelompokkan data untuk di identifikasi, maka terlebih dahulu ditentukan ukuran jarak kedekatannya

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

antar elemen data, salah satu rumus jarak yang sering digunakan pada umumnya yakni menggunakan jarak euclidean *(euclidean distance)* (Nugraheny, 2017)

2.2.6 K-Nearest Neighbor

Algoritma KNN merupakan suatu algoritma klasifikasi dan termasuk ke dalam algoritma *supervised learning.* Algoritma ini menggunakan klasifikasi terhadap suatu objek berdasarkan data jarak tetangga (*neighbor)* terdekatnya(Mariana dkk., 2015).

Kemiripan data dapat lebih dari satu, maka dari itu K-NN dapat mengklasifikasi sejumlah data *k* yang mirip juga data yang memiliki banyak kemiripan. Jarak yang digunakan pada KNN disebut dengan Jarak *Euclidien.* Selain sederhana dalam mencari jarak terdekat antar data, Kelebihan algoritma ini dapat menggeneralisasi himpunan data training yang relatif kecil (Satriya & Santoso, 2018), Adapun langkah-langkah untuk menghitung metode KNN adalah:

a. Penentuan awal parameter *k*
b. Tentukan jarak antar data yang akan dievaluasi
c. Urutkan jarak yang telah didapat
d. Tentukan jarak terdekat sampai dengan urutan nilai *k*
e. Pasangkan dengan kelas yang sesuai
f. Cari jumlah kelas yang mayoritas dan dapat diprediksi merupakan kategori objek

   Berikut rumus menghitung jarak *euclidien* dijelas pada Persamaan 2.1 (Satriya & Santoso, 2018):

$$d_i \sqrt{\sum_{i=1}^{p} (x_1 - y_1)^2} \qquad \text{.....(2.1)}$$

Keterangan:

x_i = Data uji

y_i = Data testing

i = Variabel data

d = Jarak

p = Dimensi Data

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.2.7 *Fuzzy C-Means*

Algoritma FCM merupakan algoritma klastering yang merupakan teknik untuk menentukan suatu *cluster* yang optimal. Algoritma ini menganut konsep pendekatan *fuzzy* sehingga data dapat menjadi anggota dari semua *cluster*. Nilai matriks keanggotaan terbentuk dengan cara menentukan tingkat keanggotaan dari masing-masing *cluster* yang ada, yaitu bernilai 0 sampai 1 (Afrianto & Kurniawati, 2013).

Algoritma ini memiliki kelebihan dengan cara penempatan nilai *cluster* yang tepat daripada metode lainnya. Yaitu dengan memperbaiki pusat *cluster* secara berulang, maka dapat dilihat bahwa pusat awal *cluster* bergerak pada lokasi yang tepat (Agustini, 2017). Adapun langkah dalam menghitung algoritma ini adalah sebagai berikut (Agustini, 2017)

1. Tentukan jumlah data yang akan di *cluster* berupa matriks yang berukuran n x m (n= jumlah sampel, m= atribut tiap data).

$$\begin{bmatrix} X_{11} & \cdots & X_{1j} \\ \vdots & \ddots & \vdots \\ X_{k1} & \cdots & X_{kj} \end{bmatrix} \quad \text{.....(2.2)}$$

   Keterangan:

   X_{ij} = data sampel ke- i (i=1,2,3,....n) atribut ke – j (j=1,2,3....m)

   k : objek

   j : fitur

2. Tentukan inisialisasi
   - Jumlah cluster (c)
   - Pangkat (w)
   - Maksimum iterasi (MaxIter)
   - Nilai eror terkecil yang diharapkan(ξ)
   - Fungsi objektif awal (P_0)
   - Iterasi awal (t)
3. Inisialisasi nilai centroid secara random μik, i =1,2,3,....,n; k=1,2,3....c; Sebagai nilai matriks partisi awal U. Menghitung jumlah setiap kolom

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$$Q_1 \sum_{k=1}^{c} \mu_{ik} \quad \text{.....(2.3)}$$

dengan j=1,2,3....n

menghitung:

$$\mu_{ik}(baru) = \frac{\mu_{ik}(lama)}{Q_i}$$

Keterangan:

μ_{ik}: Nilai derajat keanggotaan pada *cluster* ke- *i* dan data ke- *k*

c *:* cluster

Q_1 : data ke 1

4. Menghitung pusat cluster ke – k: V_{kj}, dengan k=1,2,3,....c; dan j=1,2,3.....m

$$V_{kj} \frac{\sum_{i=1}^{n} \left((\mu_{ik}) w_{*x_{ij}} \right)}{\sum_{i=1}^{n} (\mu_{ik})\ w} \quad \text{.....(2.4)}$$

Keterangan:

V_{kj} : nilai *i* representasi dari pusat *cluster, j* representasi dari fitur

μ_{ik}: Nilai derajat keanggotaan pada *cluster* ke- *i* dan data ke- *k*

n : banyaknya objek

m: nilai pembobotan (pembobotan > 1)

X_{kj} *: k* representasi dari data dan *j* representasi dari bobot matriks setiap fitur

5. Nilai fungsi objektif awal= 0

$$P_n = \sum_{k=1}^{n} \sum_{i=1}^{c} (\mu_{ik})^w (d_{ik})^2 \quad \text{.....(2.5)}$$

Keterangan:

Pn : nilai fungsi objektif

μ_{ik}: nilai *i* representasi dari *cluster* dan *k* representasi dari nilai data terhadap nilai derajat keanggotaan.

d_{ik} : nilai *k* representasi dari data dan nilai *i* representasi dari cluster terhadap nilai jarak.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

6. Menghitung fungsi objektif pada iterasi ke- t:

$$P_n = \sum_{k=1}^{n} \sum_{i=1}^{c} \left(\left[\sum_{j=1}^{m} (X_{kj} - V_{ij})^2 \right] (\mu_{ik})^w \right) \quad \text{.....(2.6)}$$

Keterangan:

Pn : nilai fungsi objektif

μ_{ik}: nilai *i* representasi dari *cluster* dan *k* representasi dari nilai data terhadap nilai derajat keanggotaan.

X_{kj} *:* *k* representasi dari data dan *j* representasi dari bobot matriks setiap fitur

V_{ij} : nilai *i* representasi dari pusat *cluster, j* representasi dari fitur

7. Update perubahan derajat keanggotaan:

$$\mu ik(baru) = \left[\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m-1}} \right]^{-1} \quad \text{.....(2.7)}$$

Keterangan:

μ_{ik} (baru) : nilai *i* representasi dari *cluster* dan *k* representasi dari nilai data terhadap nilai derajat keanggotaan

d_{ik} *:* nilai *i* representasi dari *cluster* dan nilai *k* representasi dari data terhadap nilai jarak

d_{jk} *:* nilai *j* representasi dari *cluster* lainnya dan nilai *k* representasi dari data terhadap nilai jarak

m : nilai data

8. Cek kondisi berhenti:

Jika : $(|Pt - Pt - 1| < \xi)$ atau

$(t > MaxIter)$ maka berhenti .....(2.8)

Jika tidak; t=t+1, ulangi langkah ke 4

Keterangan:

P_n : Fungsi objektif

ε : nilai eror

9. Jika terdapat nilai derajat keanggotaan paling besar, maka menandakan data tersebut masuk kedalam anggota *cluster* tersebut.

10. Pemilihan centroid terdekat

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.2.8 FCM-KNN

Konsep metode ini yaitu dengan menggabungkan dua tahapan metode. Yakni tahapan awal pengelompokan data menggunakan FCM dan klasifikasi untuk tahap kategorisasi data menggunakan KNN. Adapun tahapan dalam proses ini adalah (Zehansyah dkk., 2019).

1. Setelah melalui tahap pemilihan centroid terdekat maka tahap selanjutnya yakni output FCM berupa centroid akan dimasukan ke dalam KNN dengan menghitung nilai jarak kedekatan tetangga terdekat. *euclidean distance* untuk mengukur jarak objek dengan pusat *cluster*

$$d_i \sqrt{\sum_{i=1}^{p} (x_1 - y_1)^2} \quad \text{.....(2.9)}$$

   Keterangan:

   x_i = Data uji

   y_i = Data testing

   i = Variabel data

   d = Jarak

   p = Dimensi Data

2. Sesudah jarak ditemukan, maka akan dilakukan penentuan kelas pada data testing dengan menggunakan konsep modus.

2.2.9 *Confusion Matrix*

Confusion matrix adalah alat pengukur yang memiliki fungsi untuk mengevaluasi model suatu klasifikasi yang berbentuk tabel matriks. Memiliki 2 kelas negatif dan positif serta terdapat 4 kolom. Pada kolom tersebut, masing masing memiliki nilai *True-Positive* (TP), *True-Negative* (TN), *False-Positive* (FP) *dan False-Negative* (FN) (Han, 2011). Berikut merupakan bentuk tabel *confusion matrix* serta deskripsinya akan dijelaskan pada Tabel 2.2.

Tabel 2. 2 Label Confusion Matrix

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Confusion Matriks		**Nilai Sebenarnya**	
		True	False
Nilai Prediksi	True	TP (Corect Result)	FP (Unexpected Result)
	False	FN (Missing Result)	TN (Corect Absence of Result)

Berikut definisi dari tabel *confusion matrix* diatas:

1. TP (*True Positive*) → digunakan untuk jumlah data sebenarnya positif dan jumlah nilai prediksi positif.
2. FP (*False Positive*) → digunakan untuk jumlah data sebenarnya positif dan jumlah nilai prediksi negatif.
3. TN (*True Negative*) → digunakan untuk jumlah data sebenarnya negatif dan jumlah nilai prediksi negatif.
4. FN (*False Negative*) → digunakan untuk jumlah data sebenarnya negatif dan jumlah nilai prediksi positif.

Dari penjelasan tabel matriks diatas, pada tahap evaluasi dapat menggunakan cara hitung skor akurasi. Tahap akurasi digunakan untuk mengukur seberapa akurat sistem dalam mengklasifikasi data (Putra, 2019). Berikut rumus perhitungan akurasi.

Tabel 2. 3 Perhitungan Rumus *Confusion Matrix*

Akurasi $= \frac{TP + TN}{TP + TN + FN + FP}$ .....(2.10)

Keterangan:

- Akurasi adalah persentase nilai total prediksi yang benar diidentifikasi dari semua data.

2.2.10 K-*Fold Cross Validation*

K-*fold cross validation* merupakan metode pengujian data terhadap algoritma dengan cara membagi sejumlah data n-*fold* yang diinginkan kemudian data tersebut akan dibagi dalam n sebuah partisi dengan jumlah yang sama *A,B,C,D,* selanjutnya

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

dilakukan proses testing sebanyak n kali. Dalam iterasi ke-$_n$ partisi *A* akan dijadikan data testing lalu sisanya akan menjadi data training.

Tahap selanjutnya, dengan melakukan perulangan pengacakan nilai atribut masuk untuk didapatkan hasil uji beberapa atribut input (Pitria, 2014) Pengujian ini bertujuan untuk mengetahui seberapa besar akurasi metode FCM-KNN yang diterapkan pada penelitian ini. Karena dataset pada penelitian ini cukup besar maka untuk pengujian terbaik menurut (García dkk., 2015) dalam bukunya yang berjudul "*Data Preprocessing in Data Mining*" dianjurkan untuk menggunakan 10-*fold cross validation* dalam membuat model. Pada Gambar 2.1 dijelaskan ilustrasi evaluasi menggunakan 10-*fold cross validation* (Ren dkk., 2019).

Tabel 2. 4 Skenario Pengujian 10-*Fold Cross Validation*

	n-data									
	Percobaan A	Percobaan B	Percobaan C	Percobaan D	Percobaan E	Percobaan F	Percobaan G	Percobaan H	Percobaan I	Percobaan J
Percobaan 1	A	B	C	D	E	F	G	H	I	J
Percobaan 2	A	B	C	D	E	F	G	H	I	J
Percobaan 3	A	B	C	D	E	F	G	H	I	J
Percobaan 4	A	B	C	D	E	F	G	H	I	J
Percobaan 5	A	B	C	D	E	F	G	H	I	J
Percobaan 6	A	B	C	D	E	F	G	H	I	J
Percobaan 7	A	B	C	D	E	F	G	H	I	J
Percobaan 8	A	B	C	D	E	F	G	H	I	J
Percobaan 9	A	B	C	D	E	F	G	H	I	J
Percobaan 10	A	B	C	D	E	F	G	H	I	J
Keterangan :										
	Data Training									
	Data Testing									

2.2.11 Python

Python merupakan sebuah bahasa pemrograman paling populer yang berorientasi pada objek yang dinamis dan bisa digunakan di banyak pengembangan perangkat lunak lainnya. Python juga menyediakan banyak integrasi bahasa pemrograman lain serta *library* yang dapat dikembangkan dengan mudah (General Python FAQ, 2019)

2.3 Integrasi Keilmuan

Untuk mengetahui integrasi keislaman dalam penelitian ini, maka dilakukan studi pustaka dalam sudut pandang islam oleh pakar. Berdasarkan hasil wawancara yang dilakukan oleh pakar sekaligus dosen di Fakultas Ushuluddin yang bernama Fejrian Yazdajird Iwanebel,S.Th.I., M.Hum. Mengatakan ada korelasi penelitian ini dengan ayat Al-Qur'an dan hadist,

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Salah satunya yakni hadist tentang nikmat. Nikmat yang diberikan Allah kepada umatnya sangatlah banyak, salah satunya yaitu dengan diberikan nikmat sehat dan waktu yang senggang. Sikap malas dan menyiakan waktu termasuk hal yang tidak disenangi Allah. Nabi shallallahu'alaihi wa sallam bersabda

نِعْمَتَانِ مَغْبُونٌ فِيهِمَا كَثِيرٌ مِنَ النَّاسِ ، الصِّحَّةُ وَالْفَرَاغُ

Artinya: "Ada dua nikmat yang banyak manusia tertipu olehnya, yaitu nikmat sehat dan waktu yang senggang" (HR. Bukhori No. 6412)

Hadist yang telah disebutkan diatas menjelaskan bahwa sesungguhnya manusia telah diberikan banyak nikmat oleh Allah, namun manusia sering kali lalai dengan kenikmatan itu. Korelasi hadist tersebut dengan penelitian yang diangkat yaitu hendaknya kita sebagai mahasiswa jangan terlalu tertipu dengan waktu senggang dan sebaiknya menggunakan waktu seefektif mungkin sehingga dapat mengoptimalkan masa perkuliahan agar tidak terjadi penambahan jumlah semester berikutnya.

Selain nikmat yang telah diberikan, Allah juga menyinggung manusia untuk fokus dan bersungguh-sungguh dalam menjalankan suatu urusan, yaitu dalam QS. Al-Insyirah ayat 7 yang dijelaskan sebagai berikut:

فَإِذَا فَرَغْتَ فَانْصَبْ

Artinya: "maka apabila kamu telah selesai (dari suatu urusan), kerjakanlah dengan sungguh-sungguh (urusan) yang lain."

Al-Insyirah merupakan surah ke 94 yang termasuk golongan surah Makkiyah karena diturunkan di Kota Makkah dan memiliki 8 ayat. Salah satu potongan ayat diatas menjelaskan bahwa sesungguhnya Allah memerintahkan seorang hamba harus bersungguh-sungguh dalam berdo'a dan melakukan suatu pekerjaan.

Tafsir hadist dan ayat diatas apabila diintegrasikan dengan penelitian ini adalah Alllah Subhanahu Wa Ta'ala telah memberikan banyak nikmat kepada kita salah

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

satunya yakni nikmat sehat dan waktu yang senggang sehingga dengan begitu kita sebagai mahasiswa yang masih semangat dalam menuntut ilmu janganlah bermalas-malas an karena Allah sangat tidak menyukai sikap tersebut, serta mahasiswa harus bersungguh-sungguh dalam melakukan perkuliahan dan berusaha semaksimal mungkin untuk bisa mendapatkan hasil yang maksimal agar dapat lulus secara tepat waktu.

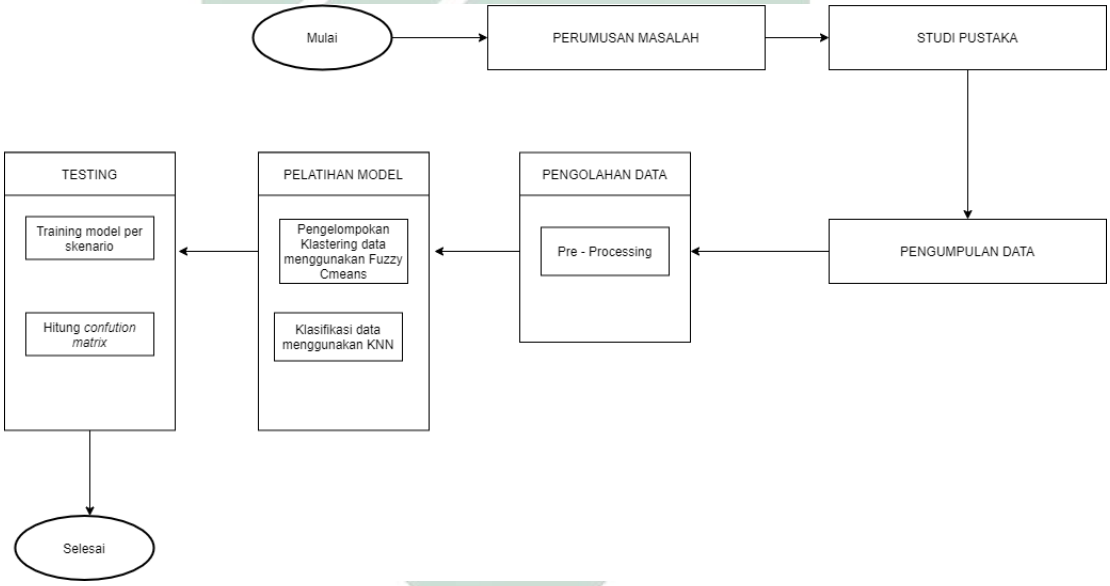
Pada surah ini kita dapat mengambil ibrah (pelajaran) yaitu setiap orang pasti mempunyai kesulitan dalam pekerjaannya, tapi ingat setiap kesulitan sepaket dengan kemudahan. Disini Al-Qur'an memberikan pelajaran kepada manusia untuk meneladani sikap Rasulullah untuk selalu memiliki sikap optimisme dan bersungguh-sungguh dalam melaksanakan suatu pekerjaan.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB III
METODOLOGI PENELITIAN

3.1 Desain Penelitian

Pada tahapan ini akan dipresentasikan dengan bentuk diagram alur yang bertujuan untuk mempermudah dalam penyampaian informasi. Langkah-langkah yang hendak dilakukan pada penelitian prediksi kelulusan tepat waktu berdasarkan data registrasi mahasiswa untuk dijadikan dasar kelulusan menggunakan metode FCM-KNN. Berikut diagram alur dari penelitian skripsi yang disajikan pada Gambar 3.1.



Gambar 3. 1 *Flowchart* Model Penelitian

Berikut merupakan pembahasan dari proses alur metodologi yang telah digambarkan:

3.1.1 Perumusan Masalah

Pada tahapan ini, rumusan masalah yang menjadi latar belakang dibuatnya penelitian tentang prediksi kelulusan tepat waktu berdasarkan data registrasi mahasiswa, merupakan salah satu upaya untuk dapat membantu civitas akademik perguruan tinggi dalam memperbaiki konsep registrasi pada mahasiswa baru**.** Dengan menggunakan teknik *data mining* dalam mengolah data mahasiswa, memungkinkan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

sistem dapat memprediksi apakah mahasiswa tersebut dapat lulus tepat waktu atau tidak.

3.1.2 Studi Pustaka

Pada tahap ini, penulis melakukan studi pustaka dengan melakukan pemahaman lebih lanjut terhadap teknik *data mining*. Dari hasil studi pustaka didapatkan beberapa metode yang populer. Hasilnya, beberapa penelitian mengatakan bahwa metode FCM - KNN dapat menghemat waktu komputasi (Zehansyah dkk., 2019).

3.1.3 Pengumpulan Data

Data yang digunakan untuk penelitian ini adalah data mahasiswa UINSA yang terdaftar pada tahun 2014-2015 yang diperoleh dari sistem informasi akademik. Dengan melakukan pengolahan data dan perhitungan terhadap data sekunder menggunakan teknik *data mining*, diharapkan dapat mengumpulkan data secara rinci dan jelas.

Dari data Tabel 2.1, maka dapat dipilih beberapa variabel ke dalam daftar perbandingan atribut yang digunakan untuk prediksi penentuan kelulusan tepat waktu berdasarkan data registrasi mahasiswa berdasarkan jurnal yang relevan yang akan ditampilkan seperti pada Tabel 3.1:

Tabel 3. 1 Atribut pada penelitian terdahulu

Penelitian terdahulu						
Nama Peneliti	(Zehansyah dkk., 2019)	(Priandini dkk., 2017)	(Afrianto & Kurniawati, 2013)	(Banjarsari dkk., 2015)	(Falahi, 2019)	(Nugroho, 2015)
Atribut yang digunakan :	Artikel online "Kompasiana"	Data jurnal *science direct*	Dokumen berita bahasa Indonesia	NIM	IP	Masa studi mahasiswa
				IP (sem 1-4)	Total SKS	IPK

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

					Lama semester	Jurusan sekolah
					Jalur masuk	Jenis kelamin
					Asal sekolah	Asal sekolah
					Penghasilan ortu	Rata-rata SKS
						Pernah menjadi asisten

Dari pemetaan pada Tabel 3.1, maka dipilih beberapa fitur yang akan digunakan untuk membuat *training model* berdasarkan ketersediaan data yang telah diperoleh. Adapun data yang telah didapatkan dari sistem akademik terdeskripsikan pada Tabel 3.2.

Tabel 3. 2 Deskripsi Atribut Data

No.	Nama Data	Tipe Data	Deskripsi
1.	Asal SMU	Nominal	Digunakan untuk klasifikasi segmentasi asal sekolah mahasiswa. Contohnya “SMAN 2 Surabaya” “SMAN 12 Surabaya”
2.	Pernah Ponpes	Nominal	Digunakan untuk klasifikasi segmentasi apakah mahasiswa pernah menjadi santri sebelumnya. Contoh: “pernah”, “tidak pernah”
3.	Kode Unit	Nominal	Digunakan untuk klasifikasi dalam perihal segmentasi prodi. Contoh: “BKI”,“SI”,”ILKOM”

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4.	Jalur Penerimaan	Nominal	Digunakan untuk klasifikasi segmentasi perihal jalur masuk di UINSA. Contoh: “SPANPTKIN”, “UMPTKIN”
5.	Nilai NEM	Rasio	Digunakan untuk klasifikasi segmentasi nilai ujian masuk mahasiswa. Contoh: “85” “75”
6.	Label Kelulusan	Nominal	Digunakan untuk klasifikasi segmentasi apakah mahasiswa dapat lulus secara tepat waktu atau tidak Contoh: “lulus tepat waktu” “tidak tepat waktu”

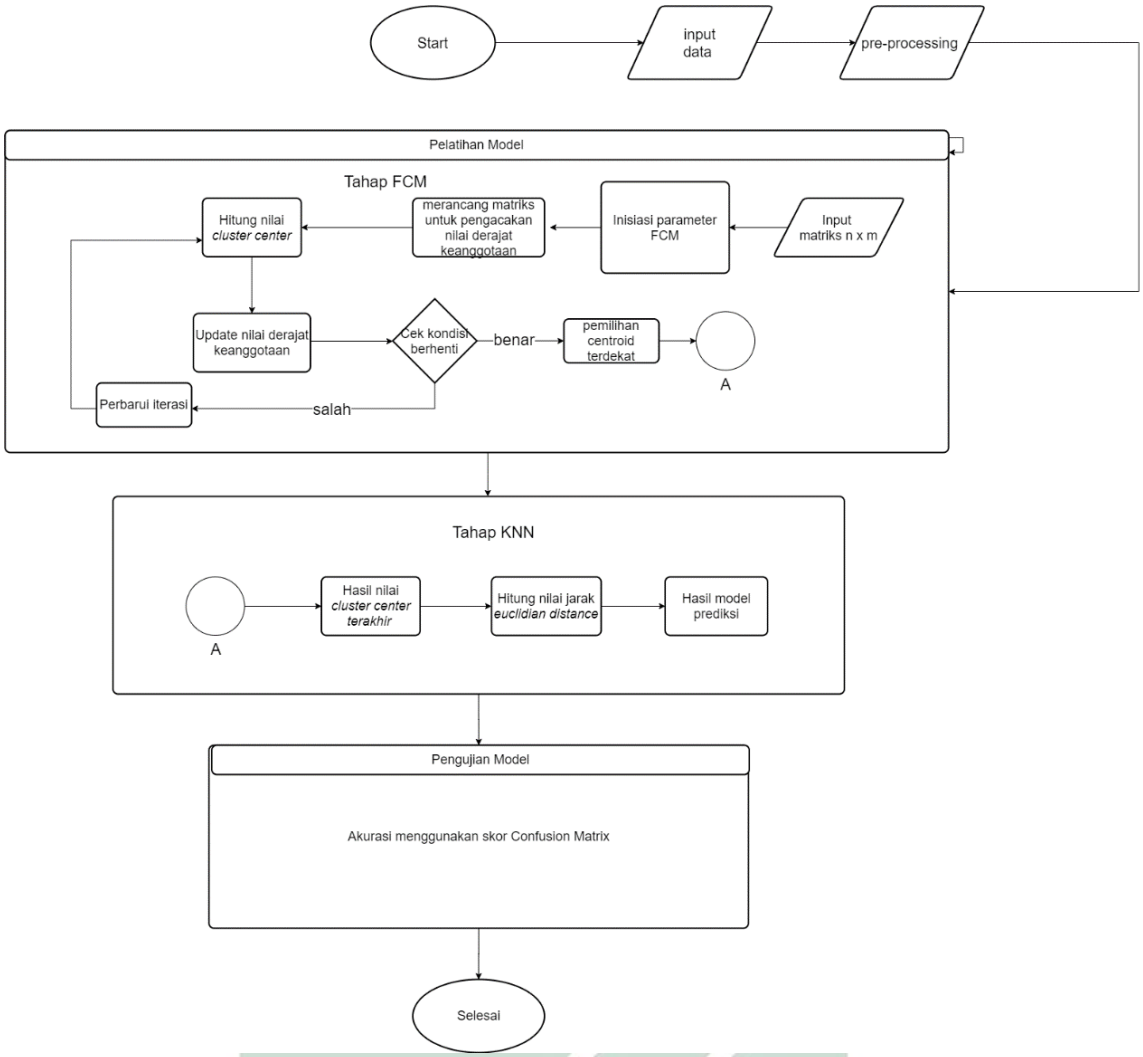
3.1.4 Pengolahan Data

Pada tahap ini, akan dilakukan *pre-processing* data karena masih banyak data yang *noisy* atau masih terdapat data yang menyimpang dari yang diharapkan antara penggunaan nama atau kode, karena data mentah banyak yang memiliki tipe data kategorik maka akan dirubah ke tipe data numerik. Maka, di tahap *pre-processing* ini dilakukan data cleaning untuk menghilangkan nilai data yang salah, normalisasi data, dan menggabungkan data dari beberapa sumber untuk kebutuhan sistem.

3.1.5 Pelatihan Model

Tahap selanjutnya yakni diperlukan tahap pelatihan model yang akan dijelaskan pada *flowchart* dibawah ini.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id



Gambar 3. 2 *Flowchart* Pelatihan dan Pengujian Model

Langkah-langkah dalam *flowchart* akan dijelaskan dibawah ini:

a. Merancang matriks pada data.
b. Tetapkan jumlah cluster, pangkat, nilai max iterasi, parameter nilai error terkecil dan fungsi objektif awal
c. Menginisialisasi nilai centroid secara random.
d. Hitung nilai pusat *cluster*. Menggunakan rumus persamaan 2.4
e. Memperbaiki atau *update* nilai derajat keanggotaan menggunakan rumus persamaan 2.7

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

f. Cek kondisi proses pemberhentian iterasi, jika salah ulangi tahap “e” dan jika benar lanjut tahap selanjutnya.
g. Jika terdapat nilai derajat keanggotaan paling besar diantara lainnya, maka menandakan data akan masuk kedalam anggota *cluster* tersebut.
h. Pemilihan centroid terdekat.
i. Penetapan nilai K lalu menghitung K tetangga terdekat terhadap anggota cluster terpilih dilihat dari centroid terdekat, perhitungan nilai antar jarak pusat cluster menggunakan *euclidean distance* menggunakan rumus Persamaan 2.9.
j. Penentuan kelas pada data testing dengan konsep modus pada data kelas dari k tetangga.

3.1.6 Testing

Pada tahap ini akan dilakukan skenario pengujian dimana akan menggunakan bahasa pemrograman *python*. Selanjutnya data yang telah melalui tahap pre-processing digunakan untuk mengevaluasi kinerja sebuah model atau algoritma untuk menguji keberhasilan perhitungan algoritma pada tahap pelatihan, yakni dengan proses testing data yang telah memiliki label.

Proses testing model dilakukan dengan beberapa skenario pembagian data. Pada tahap ini, digunakan *10 fold Cross Validation.* dengan cara data dibagi menjadi 10 set data untuk dilakukan evaluasi model algoritma. Dari beberapa skenario tersebut bertujuan agar dapat dipilih model dan skenario yang optimal. Sehingga pada proses pelatihan model selanjutnya, hanya dibutuhkan 1 kali pelatihan saja. Ilustrasi skenario pengujian menggunakan *10 fold Cross Validation* telah dijelaskan pada Gambar 2.1 diatas.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB IV
HASIL DAN PEMBAHASAN

Berdasarkan dua rumusan masalah pada bab 1.2, terkait Prediksi Kelulusan Tepat Waktu Berdasarkan Data Registrasi Mahasiswa Dengan Algoritma *Fuzzy C-Means* dan *K-Nearest Neighbors.* Adapun prediksi kelulusan tepat waktu yang dimaksud merupakan hasil model prediksi yang berupa 'lulus tepat waktu' dan 'tidak lulus tepat waktu' yang nantinya akan menjadi dasar acuan untuk pihak akademik dalam menentukan seleksi peserta didik baru. Serta mengetahui performa dari akurasi algoritma yang telah di implementasikan.

4.1 Pengumpulan Data

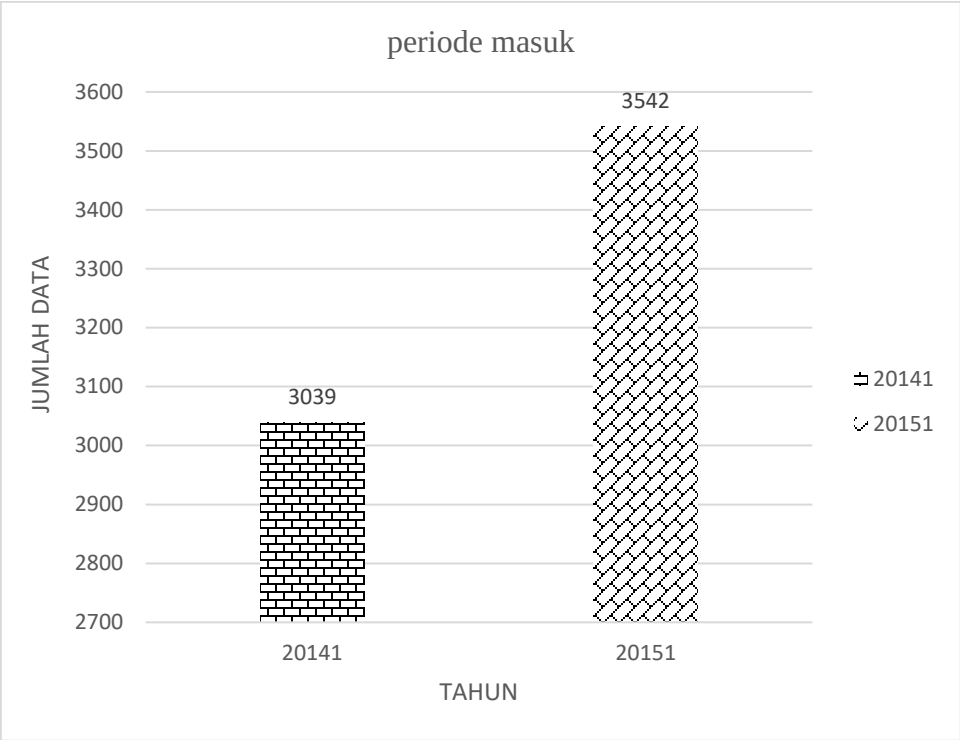
Data mentah awal memiliki **6581** data yang diperoleh dari sistem informasi akademik dan menggunakan data mahasiswa yang terdaftar pada tahun masuk 2014-2015 dan periode wisuda tahun 2017-2018. Berikut merupakan cuplikan data mentah seperti pada Tabel 4.1.

Tabel 4. 1 Cuplikan data awal

Nim	Nama	Sex	…	Pernah ponpes	Asal_smu	Nem_smu
B76xxxx	Windi Setyawan	L		0	Smkn 1 Pungging	-
J71xxxx	Khoirul Mulyanto	L		0	SMK Taruna Surabaya	32,00
B93xxxx	Rofiul Umah	P	…	1	MAN Nganjuk	43,50
A73xxxx	M Khidlir Ali	P		1	MA Fadlillah	47,30
B91xxxx	Meirina Rachmawati	P		0	MAN Sidoarjo	-

Dari cuplikan data mentah diatas, berikut grafik perbandingan data menurut periode masuknya.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id



Gambar 4. 1 Perbandingan grafik jumlah mahasiswa pada periode masuk

4.2 Pengolahan Data

Dalam sub bab ini menjelaskan tentang pengolahan data yang meliputi tahap *pre-processing* data, perhitungan pada algoritma FCM dan KNN. Untuk pembahasan rinci masing-masing tahapan akan dijelaskan sebagaimana dalam sub bab berikut ini.

4.2.1 Pre-Processing

Pre-processing data merupakan suatu tahap pengolahan data yang dilakukan dalam mengolah suatu data mentah menjadi data yang mempunyai kualitas. Dilakukan *pre-processing* agar sebuah data yang akan diolah terhindar dari data yang mengandung error dan outliers (*noisy*), serta data yang tidak konsisten.

Tabel 4. 2 Detail tahapan pre-processing pada setiap atribut

Atribut	Perlakuan	Ekspektasi Hasil
jalur_penerimaan	- Menghapus data yang tidak valid	Snmptn = 0, Sbmptn = 1 , Umptain = 2, Ptain = 3, Mandiri = 4, Umum= 5, Test = 6,

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

	- Mengubah variabel kategorikal menjadi numerik	Kemitraan = 7, Proyek = 8, Prestasi = 9, Mhs Asing = 10.
kode_prodi	- Menghapus data yang tidak valid - Mengubah variabel kategorikal menjadi numerik	PSI=0, BKI=1, SINF=2, SI = 3, KOM = 4, PAI= 5 AS = 6, PGMI= 7, ES = 8, SKI = 9, PMI = 10, PMT = 11, KPI = 12, SJ =13, BSA = 14, PBA= 15, TH = 16, HDS= 17, MD = 18, PA = 19, AT = 20, PBI= 21, PGRA = 22, AKUN = 23, IE= 24, SOS = 25, ART = 26, MNJ = 27, BIO = 28, HI = 29, MTK = 30, AQ = 31, HTN= 32, TL = 33, KI= 34, IK= 35, IP= 36, IF 37, M dan PM = 38 AF = 39, ZW= 40
pernah_ponpes	- Tidak ada perlakuan	0= tidak pernah ponpes 1= pernah ponpes
asal_smu	- Menghapus data yang tidak valid - Mengubah variabel kategorikal menjadi numerik	Sman = 0, Sma = 1, Smkn = 2, Smk = 3, Man = 4, Ma = 5
nem_smu	- Menyamakan rentang nilai dari asal sekolah 1. Asal sekolah "SMA,SMAN,MA,MAN" nilai nem akan dibagi 6 karena jumlah mata pelajaran saat diujikan ada 6 jenis. 2. Asal sekolah "SMK" dan "SMKN" nilai nem akan dibagi 4 karena jumlah	Nilai nem SMA 458.9 → 76.48 Nilai nem SMK 314 → 78.5

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

	mata pelajaran saat diujikan ada 4 jenis.	
Label kelulusan	- Mengubah variabel kategorik ke menjadi	Lulus = 0 Tidak lulus tepat waktu = 1

Dari hasil tahapan Tabel 4.2, maka dapat disimpulkan data atribut pada *dataset* menunjukkan bahwa terdapat nilai atribut yang bersifat valid dan tidak valid. Maka, dapat diilustrasikan seperti Tabel 4.3.

Tabel 4. 3 Data validation

Atribut	Contoh nilai yang valid	Contoh nilai tidak valid
jalur_penerimaan	Snmptn, Sbmptn , Umptain, Ptain , Mandiri, Umum, Test, Kemitraan, Proyek, Prestasi, Mhs Asing	None, kosong
kode_prodi	SINF,PMI, IP, KOM,..	-
pernah_ponpes	0, 1	-
asal_smu	“SMAN 2 Lamongan” ,“SMKN 1 Pungging”, “MAN Sidoarjo”,...	-
nem_smu	32.00 , 48.05 , 55.21, ...	434, 359, 515, ...
Label kelulusan	0,1	-

Dari atribut diatas, dapat diambil beberapa sampel irisan *dataset*, maka diperoleh data seperti Tabel 4.4.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4. 4 Sampel dataset sebelum dilakukan pre processing

jalur_penerimaan	kode_prodi	pernah_ponpes	asal_smu	nem_smu
SBMPTN	PSI	0	SMK Taruna Surabaya	32,00
PTAIN	PAI	0	MAN Surabaya	48,05
SBMPTN	PGMI	0	MAN Mojosari	46,60
SNMPTN	KPI	1	SMA Al-Azhar Menganti	43,30
TES	KPI	0	MAN Lamongan	43,75

Proses pengolahan data dilakukan dengan menggunakan bahasa python, sehingga apabila variabel (X) dan label (Y) memiliki tipe data yang berbeda maka harus dilakukan persamaan tipe data dengan cara label encoder. Label encoder merupakan suatu cara untuk mengkonversi tipe data *string* menjadi *integer*. Setelah dilakukan tahap *pre processing* maka didapatkan irisan *dataset* seperti pada Tabel 4.5.

Tabel 4. 5 Sampel Dataset Sesudah Dilakukan Pre processing

jalur_penerimaan	kode_prodi	pernah_ponpes	asal_smu	nem_smu
1	0	0	3	32,00
3	5	0	4	48,05
1	7	0	4	46,60
0	12	1	1	43,30
6	12	0	4	43,75

Data mentah yang berjumlah 6581 data, kemudian dilakukan seleksi menjadi 2504 data dengan cara menyeleksi data berdasarkan ada atau tidaknya variabel yang dibutuhkan yaitu variabel jalur penerimaan, kode prodi, pernah ponpes, asal smu dan nilai nem smu. Pada tabel dibawah ini dijelaskan komposisi data yang telah diolah. Dari tabel tersebut bisa dilihat perbandingan jumlah data antara mahasiswa yang

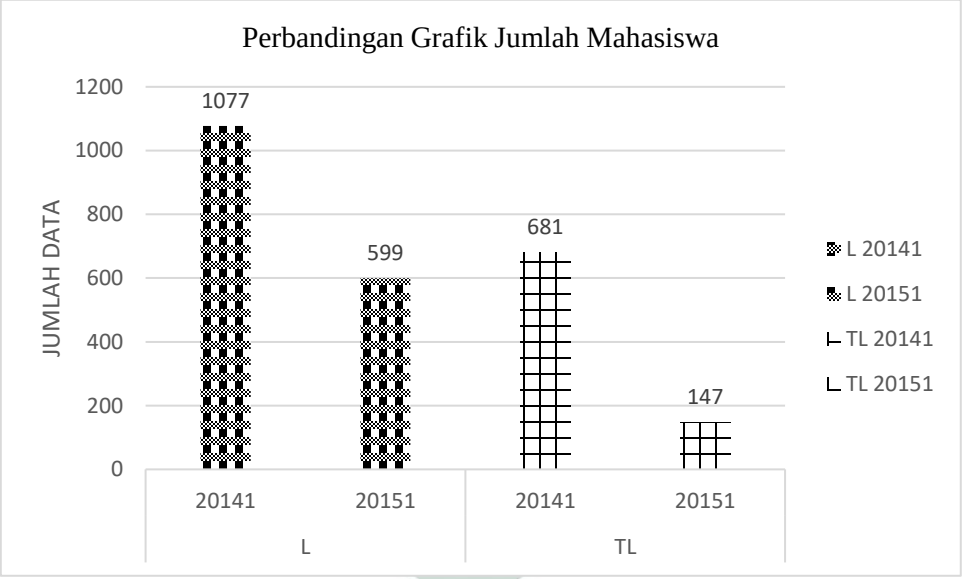
digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

memiliki label lulus dan mahasiswa yang memiliki label tidak lulus tepat waktu. Perbandingan jumlah data yang tidak seimbang dapat dilihat pada Tabel 4.6

Tabel 4. 6 Perbandingan data mahasiswa yang memiliki status lulus dan lulus tidak tepat waktu

Tahun	Status label	
	Lulus	Lulus Tidak tepat waktu
2014	1077	681
2015	599	147
	Total data = 2504	

Untuk lebih jelasnya akan diberi gambaran grafik pada Gambar 4.2.



Gambar 4. 2 Perbandingan grafik jumlah mahasiswa lulus dan tidak lulus tepat waktu

4.2.2 Perhitungan FCM

Setelah dilakukan *pre-processing* data, maka tahap selanjutnya yaitu perhitungan FCM. Dibawah ini merupakan gambar pseudocode dari algoritma FCM.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Import library : `from fcmeans import FCM`

Input cluster FCM dan input data train :

```
def get_FCM_centers_and_labels(x_train):
  fcm = FCM(n_clusters=2)
  fcm.fit(x_train)
```

Hasil FCM → centroid dan label :

```
fcm_centers = fcm.centers
  fcm_labels  = fcm.u.argmax(axis=1)
  return [fcm_centers, fcm_labels]
```

Inisiasi parameter:

```
n_clust = 2 #cluster
  n_n = 5 #jumlah nilai k
  fcm_centers, fcm_labels =get_FCM_centers_and_labels(x_train)
```

mencari jarak terpendek dari data testing ke centroid

```
min = 1000000
    min_index = -1
    for i in range(n_clust):
      dist = distance.euclidean(baris_yang_diperiksa,
fcm_centers.iloc[i, :])
      if min > dist:
        min = dist
        min_index = i #menyimpan centroid terdekat
```

Gambar 4. 3 Pseudocode FCM Python

Berikut merupakan perhitungan manual yang diambil beberapa data untuk dihitung.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4.2.2.1 Bentuk Matriks n x m

$$\begin{bmatrix} 3 & 21 & 1 & 4 & 49.00 \\ 0 & 30 & 0 & 1 & 62.00 \\ 3 & 11 & 0 & 5 & 47.35 \\ 6 & 34 & 1 & 1 & 51.75 \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

4.2.2.2 Inisiasi Parameter FCM

Setelah terbentuk matriks, maka tahap selanjutnya yakni menginisialisasi parameter untuk algoritma FCM. Pada perhitungan manual, disini penetapan parameter yang telah di uji sebelumnya seperti jumlah cluster, pangkat, max iter, error terkecil. Hasil pengujian membuktikan bahwa nilai parameter yang disebutkan sudah memenuhi hasil optimal dalam penempatan cluster dibanding menggunakan cluster 3,4 atau 5. Serta didukung oleh jurnal terkait yang dibahas oleh (Ahmadi & Hartini, 2013) yang membahas tentang nilai pangkat berpengaruh pada jumlah cluster, dan semakin banyak maximum iterasi dan semakin kecil nilai error cluster akan berada pada titik yang tepat, untuk itu parameter FCM akan dijelaskan pada Tabel 4.7.

Tabel 4. 7 Parameter FCM

Cluster	2
Pangkat	2
Max_iter	100
Error terkecil	0,001
Fungsi Objectif (Po)	0

Keterangan:

Cluster = merupakan jumlah cluster yang akan dibentuk sesuai dengan kebutuhan suatu data

Pangkat = Nilai eksponen

MaxIter = Batas *looping* atau perulangan. Disini looping akan berhenti jika nilai maks iterasi terpenuhi.

Eror terkecil = Batasan nilai yang membuat looping akan berakhir jika mendapatkan nilai eror yang diharapkan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Fungsi objektif = Fungsi yang dioptimumkan (maksimum atau minimum). Disini nilai 0 berarti untuk mendapatkan nilai minimum didapatkan dengan rumus persamaan 2.5 dan 2.6.

4.2.2.3 Merancang matriks pengacakan derajat keanggotan

Tabel 4. 8 Matriks pengacakan derajat keanggotaan

C_1	C_2	Total
0.32	0.68	1
0.46	0.54	1
0.375	0.625	1
0.24	0.76	1
0.005	0.995	1

Keterangan:

C_1 = Cluster ke 1.

C_2 = Cluster ke 2.

4.2.2.4 Perhitungan Pusat Cluster ke- 1 Iterasi – 1

Tabel 4. 9 Perhitungan pusat cluster 1 pada iterasi ke 1

Derajat Keanggotaan Pada Cluster Ke-1	Data yang diCluster			$(\mu_{i1})^2$	$(\mu_{i1})^2 * X_{i1}$	…	$(\mu_{i1})^2 * X_{i5}$
	X_{i1}	…	X_{i5}				
0.320	1	…	32.00	0.102	0.102	…	3.277
0.460	3	…	48.05	0.212	0.635	…	10.167
0.375	1	…	46.60	0.141	0.141	…	6.553
0.240	0	…	43.30	0.058	0.000	…	2.494
0.005	6	…	43.75	0.000	0.000	…	0.001
			$\sum$	0.512	0.878		22.492
$\sum [(\mu_{i1}^2) * X_{ij}] / \sum (\mu_{i1}^2)$					1.714		43.909

Keterangan:

X_{i1}– X_{i5} = Data variabel X pada cluster ke 1 iterasi 1

$(\mu_{i1})^2$ = Hasil Nilai derajat keanggotaan dipangkat dua

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$(\mu_{i1})^2$*X_{i1} -$(\mu_{i1})^2$*X_{i5} = Hasil Nilai derajat keanggotaan dipangkat dua data ke 1 sampai ke 5.

4.2.2.5 Perhitungan Pusat Cluster ke-2 Iterasi – 1

Tabel 4. 10 Perhitungan pusat cluster 2 pada iterasi ke 1

Derajat Keanggotaan Pada Cluster Ke-2	Data yang diCluster			$(\mu_{i1})^2$	$(\mu_{i1})^2$*X_{i1}	...	$(\mu_{i1})^2$*X_{i5}
	X_{i1}	…	X_{i5}				
0.680	1	…	32.00	0.462	0.462	…	14.797
0.540	3	…	48.05	0.292	0.875	…	14.011
0.625	1	…	46.60	0.391	0.391	…	18.203
0.760	0	…	43.30	0.578	0.000	…	25.010
0.995	6	…	43.75	0.990	5.940	…	43.314
			∑	2.712	7.668		115.335
∑ [(μ_{i1}^2)*X_{ij}] / ∑ (μ_{i1}^2)					2.827		42.524

Keterangan:

X_{i1}– X_{i5} = Data variabel X pada cluster ke 2 iterasi 1

$(\mu_{i1})^2$ = Hasil Nilai derajat keanggotaan dipangkat dua

$(\mu_{i1})^2$*X_{i1} - $(\mu_{i1})^2$*X_{i5} = Hasil Nilai derajat keanggotaan dipangkat dua data ke 1 sampai ke 5.

4.2.2.6 Menghitung Pusat Cluster

Tabel 4. 11 Hasil pusat cluster iterasi 1

		HASIL PUSAT CLUSTER (V)				
V=	C_1	1.714	5.337	0.112	3.463	43.909
	C_2	2.827	8.481	0.213	3.191	42.524

Keterangan:

V = Hasil pusat cluster

C_1 = Hasil pusat cluster ke 1 iterasi 1 dari jumlah data $(\mu_{i1})^2$*X_{i1} sampai dengan $(\mu_{i1})^2$*X_{i5}

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

C_2 = Hasil pusat cluster ke 2 iterasi 1 dari jumlah data $(\mu_{i1})^2 * X_{i1}$ sampai dengan $(\mu_{i1})^2 * X_{i5}$

4.2.2.7 Penghitungan Fungsi Objektif Iterasi ke - 1

Tabel 4. 12 Perhitungan fungsi objektif iterasi 1

PENGHITUNGAN FUNGSI OBJEKTIF				
Derajat Keanggotaan data ke-i		$\left[\sum_{j=1}^{m}(X_{kj}-V_{1j})^2\right](\mu_{i1})^2$	$\left[\sum_{j=1}^{m}(X_{kj}-V_{2j})^2\right](\mu_{i2})^2$	L_1+L_2
$\mu_{i1}{}^2$	$\mu_{i2}{}^2$	L_1	L_2	
0.102	0.462	17.515	86.055	103.570
0.212	0.292	4.066	12.653	16.719
0.141	0.391	1.521	8.926	10.447
0.058	0.578	3.143	15.245	18.388
0.000	0.990	0.002	24.405	24.407
			Fungsi Objektif	173.530

Keterangan:

L_1 = Hasil pengurangan antara data variabel X (X_{i1} – X_{i5}) dan hasil pusat cluster (C_1) dan kemudian dikali dengan derajat keanggotaan di pangkat 2.

L_2 = Hasil pengurangan antara data variabel X (X_{i1} – X_{i5}) dan hasil pusat cluster (C_2) dan kemudian dikali dengan derajat keanggotaan di pangkat 2.

4.2.2.8 Menghitung Perubahan Matriks Partisi $\sum_{i=1}^{2}\left[\sum_{j=1}^{m}(X_{kj}-V_{ij})^2\right]^{-1}$

Tabel 4. 13 Perhitungan perubahan matriks U iterasi 1 ke iterasi selanjutnya

Matriks Partisi U				
$\sum_{j=1}^{m}(X_{kj}-V_{1j})^2$	$\sum_{j=1}^{m}(X_{kj}-V_{2j})^2$	$\sum_{i=1}^{2}\sum_{j=1}^{m}(X_{kj}-V_{2j})^2$	μ_{i1}	μ_{i2}
L_1	L_2	$L_T = L_1+L_2$	L_1/L_T	L_2/L_T
0.006	0.005	0.011	0.521	0.479
0.052	0.023	0.075	0.693	0.307
0.092	0.044	0.136	0.679	0.321
0.018	0.038	0.056	0.326	0.674
0.016	0.041	0.056	0.281	0.719

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Keterangan:

L_1 = Hasil perubahan matriks U baru dari pengurangan antara data variabel X (X_{i1} – X_{i5}) dan hasil pusat cluster (C_1) di pangkat 2.

L_2 = Hasil perubahan matriks U baru dari pengurangan antara data variabel X (X_{i1} – X_{i5}) dan hasil pusat cluster (C_2) di pangkat 2.

4.2.2.9 Cek Kondisi Berhenti Iterasi ke - 1

Jika nilai fungsi objektif lebih besar dari nilai Max iter yang telah ditetapkan pada tahap sebelumnya maka iterasi berhenti, tetapi jika tidak maka mengulang ke tahap 4.2.2.5

Tabel 4. 14 Cek kondisi berhenti iterasi 1

P_1	173.530
P_0	0
$\|P_1 - P_0\|$	173.5299

Keterangan :

P_0= Fungsi objektif ke 0

P_1= Fungsi objektif ke 1

Dari hasil perhitungan iterasi ke- 1. Hasil belum ditemukan, selisih fungsi objective masih lebih besar dari epsilon (error terkecil yang diharapkan), iterasi dilanjutkan dan iterasi berhenti pada iterasi ke- 19.

4.2.2.10 Perhitungan Pusat Cluster ke-1 Iterasi – 19

Tabel 4. 15 Perhitungan iterasi cluster 1 iterasi ke 19

Derajat Keanggotaan Pada Cluster Ke-1	Data yang diCluster			$(\mu_{i1})^2$	$(\mu_{i1})^2 * X_{i1}$	...	$(\mu_{i1})^2 * X_{i5}$
	X_{i1}	…	X_{i5}				
0.000	1	…	32.00	0.000	0.000	…	0.000
0.924	3	…	48.05	0.853	2.560	…	41.004
0.970	1	…	46.60	0.942	0.942	…	43.884
0.911	0	…	43.30	0.830	0.000	…	35.927

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

0.921	6	…	43.75	0.848	5.089	…	37.110
			∑	3.473	8.591		157.926
∑ [(μ_{i1}^2)*X_{ij}] / ∑ (μ_{i1}^2)					2.474		45.472

Keterangan:

X_{i1} – X_{i5} = Data variabel X pada cluster ke 1 iterasi 19

$(\mu_{i1})^2$ = Hasil Nilai derajat keanggotaan dipangkat dua

$(\mu_{i1})^2$*X_{i1} - $(\mu_{i1})^2$*X_{i5} = Hasil Nilai derajat keanggotaan dipangkat dua data ke 1 sampai ke 5.

4.2.2.11 Perhitungan Pusat Cluster ke-2 Iterasi – 19

Tabel 4. 16 Perhitungan iterasi cluster 2 iterasi ke 19

Derajat Keanggotaan Pada Cluster Ke-2	Data yang diCluster			$(\mu_{i1})^2$	$(\mu_{i1})^2$*X_{i1}	...	$(\mu_{i1})^2$*X_{i5}
	X_{i1}	…	X_{i5}				
1.000	1	…	32.00	0.999	0.999	…	31.973
0.076	3	…	48.05	0.006	0.017	…	0.279
0.030	1	…	46.60	0.001	0.001	…	0.041
0.089	0	…	43.30	0.008	0.000	…	0.344
0.079	6	…	43.75	0.006	0.037	…	0.273
			∑	1.020	1.055		32.909
∑ [(μ_{i1}^2)*X_{ij}] / ∑ (μ_{i1}^2)					1.034		32.264

Keterangan:

X_{i1} – $X_{i\,5}$ = Data variabel X pada cluster ke 2 iterasi 19

$(\mu_{i1})^2$ = Hasil Nilai derajat keanggotaan dipangkat dua

$(\mu_{i1})^2$*X_{i1} - $(\mu_{i1})^2$*X_{i5} = Hasil Nilai derajat keanggotaan dipangkat dua data ke 1 sampai ke 5.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4.2.2.12 Menghitung Pusat Cluster Iterasi ke- 19

Tabel 4. 17 Hasil pusat cluster iterasi 19

		HASIL PUSAT CLUSTER (V)				
V=	C_1	2.474	8.924	0.239	3.283	45.472
	C_2	1.034	0.201	0.008	2.997	32.264

Keterangan:

V = Hasil pusat cluster

C_1 = Hasil pusat cluster ke 1 iterasi 19 dari jumlah data $(\mu_{i1})^2 * X_{i1}$ sampai dengan $(\mu_{i1})^2 * X_{i5}$

C_2 = Hasil pusat cluster ke 2 iterasi 19 dari jumlah data $(\mu_{i1})^2 * X_{i1}$ sampai dengan $(\mu_{i1})^2 * X_{i5}$

4.2.2.13 Penghitungan Fungsi Objektif Iterasi ke- 19

Tabel 4. 18 Perhitungan fungsi objektif iterasi 1

PENGHITUNGAN FUNGSI OBJEKTIF				
Derajat Keanggotaan data ke-i		$\left[\sum_{j=1}^{m}(X_{kj}-V_{1j})^2\right](\mu_{i1})^2$	$\left[\sum_{j=1}^{m}(X_{kj}-V_{2j})^2\right](\mu_{i2})^2$	L_1+L_2
$\mu_{i1}{}^2$	$\mu_{i2}{}^2$	L_1	L_2	
0.000	0.999	0.000	0.111	0.111
0.853	0.006	19.537	1.610	21.148
0.942	0.001	7.268	0.221	7.489
0.830	0.008	21.646	2.120	23.767
0.848	0.006	21.571	1.852	23.423
		Fungsi Objektif		75.939

Keterangan:

L_1 = Hasil pengurangan antara data variabel X (X_{i1} – X_{i5}) dan hasil pusat cluster (C_1) dan kemudian dikali dengan derajat keanggotaan di pangkat 2.

L_2 = Hasil pengurangan antara data variabel X (X_{i1} – X_{i5}) dan hasil pusat cluster (C_2) dan kemudian dikali dengan derajat keanggotaan di pangkat 2.

4.2.2.14 Menghitung Perubahan Matriks Partisi Iterasi ke- 19

Tabel 4. 19 Perhitungan perubahan matriks U iterasi 19 ke iterasi selanjutnya

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Matriks Partisi U				
$\sum_{j=1}^{m}(X_{kj}-V_{1j})^2$	$\sum_{j=1}^{m}(X_{kj}-V_{2j})^2$	$\sum_{i=1}^{2}\sum_{j=1}^{m}(X_{kj}-V_{2j})^2$	μ_{i1}	μ_{i2}
L_1	L_2	$L_T = L_1+L_2+L_3$	L_1/L_T	L_2/L_T
0.004	8.980	8.984	0.000	1.000
0.044	0.004	0.047	0.924	0.076
0.130	0.004	0.134	0.970	0.030
0.038	0.004	0.042	0.911	0.089
0.039	0.003	0.043	0.921	0.079

Keterangan:

L_1 = Hasil perubahan matriks U baru dari pengurangan antara data variabel X (X_{i1} – X_{i5}) dan hasil pusat cluster (C_1) di pangkat 2.

L_2 = Hasil perubahan matriks U baru dari pengurangan antara data variabel X (X_{i1} – X_{i5}) dan hasil pusat cluster (C_2) di pangkat 2.

4.2.2.15 Cek Kondisi Berhenti pada Iterasi ke- 19

Pada iterasi ke- 19 hasil telah ditemukan, karena selisih fungsi objektif sudah lebih kecil dari epsilon (error terkecil yang diharapkan).

Tabel 4. 20 Cek kondisi berhenti iterasi 19

P_{19}	75.939
P_{18}	75.9388
$\lvert P_{19} - P_{18}\rvert$	0.000282

Keterangan :

P_{18} = Fungsi objektif ke 18

P_{19} = Fungsi objektif ke 19

Hasil dari perhitungan FCM menurut derajat keanggotaan terakhirnya, yakni iterasi ke- 19 menghasilkan sebagai berikut

Tabel 4. 21 Hasil derajat keanggotaan pada iterasi ke 19

Data Ke	Derajat keanggotaan (m) data pada Cluster ke-		Data Cenderung Masuk ke Cluster dengan Derajat keangotaan
	1	2	

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

1	0.000423	0.999577	2
2	0.923683	0.076317	1
3	0.970369	0.029631	1
4	0.911003	0.088997	1
5	0.921082	0.078918	1

Maka bisa diambil kesimpulan pada data ke 1 akan masuk ke cluster 2, data ke 2 akan masuk ke cluster 1, data ke 3 masuk ke cluster 1, data ke 4 masuk ke cluster 1, data ke 5 masuk ke cluster 1. Untuk lebih jelasnya akan dipresentasikan pada Tabel 4.22.

4.2.2.16 Pemilihan Centroid

Tabel 4. 22 Pemilihan data masuk ke dalam cluster 1 atau 2

No.	Nama Siswa	Cluster
1	KHOIRUL MULYANTO	2
2	NAFI`A WILDA ZARKASI	1
3	AGUSTIN WIDYASTUTI RHODIAH	1
4	SITI UMROTUS SA'DIYAH	1
5	AISATUL CHOLIFAH	1

Hasil pada Tabel 4.22 menyatakan bahwa data lebih cenderung ke anggota cluster 1 maka, jarak centroid terdekat akan di optimalkan pada cluster ke 1. Tahap selanjutnya yakni masuk ke perhitungan KNN. Berikut merupakan pseudocode algoritma KNN.

Import library :

```
from sklearn.neighbors import KNeighborsClassifier
from scipy.spatial import distance
```

Ambil indeks data berdasarkan cluster terdekat:

```
index_xclust = np.where(fcm_labels == min_index)

    list_index_xclust = index_xclust[0].tolist()
```

mencari tetangga terdekat dari x train dan y train yang sudah diseleksi berdasarkan cluster

```
knn = KNeighborsClassifier(n_neighbors=n_n)

    knn.fit(x_train_new, y_train_new.values.ravel())
```

Output:

```
hasil = knn.predict([baris_yang_diperiksa])[0]

predicted = hasil
```


digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Gambar 4. 4 Pseudocode algoritma KNN

4.2.3 Perhitungan KNN

Pada tahapan ini, anggota cluster yang sudah diketahui dari tahap FCM akan diberi label data yang ditentukan oleh nilai k (tetangga) terdekat pada kelompok data yang serupa. Tahap awal, tetapkan nilai k. Pada perhitungan manual kita tetapkan K (tetangga) =3. Kemudian hitung pusat cluster dari pemberhentian iterasi terakhir pada Tabel 4.17 dan data test pada Tabel 4.24.

Tabel 4. 23 Data test

No.		Data tes					
	Nama	X1	X2	X3	X4	X5	Z
x	BAHARUDIN YUSUF	1	32	1	2	31.60	

Setelah mengambil nilai pusat cluster iterasi terakhir, maka tahap selanjutnya yaitu menghitung nilai distance antara pusat cluster dengan data test digunakan rumus persamaan 2.13. Berikut merupakan perhitungan jarak data menggunakan cluster ke 1.

$$C1_x = \sqrt{(2.474-1)^2 + (8.924-32)^2 + (0.239-1)^2 + (3.283-2)^2 + (45.472-31.60)^2} = 27$$

Berikut merupakan perhitungan jarak data menggunakan cluster ke 2

$$C2_x = \sqrt{(1.034-1^2 + (0.201-32)^2 + (0.008-1)^2 + (2.997-2)^2 + (32.264-\ 31.60)^2} = 32$$

Tabel 4. 24 Hasil distance pemilihan centroid terdekat

Centroid	Jarak data x
C_1	27
C_2	32

Pada tabel diatas, dapat disimpulkan bahwa, data baru cenderung masuk pada centroid ke 1. Kemudian data train akan dihitung berdasarkan anggota centroidnya saja. Disini digunakan data train ke 2 sampai 5 karena termasuk ke centroid ke 1. Berikut adalah perhitungan jaraknya.

$$C1_1 = \sqrt{(2.474-3)^2 + (8.924-5)^2 + (0.239-0)^2 + (3.283-4)^2 + (45.472-48.05)^2} = 5$$

$$C1_2 = \sqrt{(2.474-1)^2 + (8.924-7)^2 + (0.239-0)^2 + (3.283-4)^2 + (45.472-46.60)^2} = 3$$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$$C1_3 = \sqrt{(2.474-0)^2 + (8.924-12)^2 + (0.239-1)^2 + (3.283-1)^2 + (45.472-43.30)^2} = 5$$

$$C1_4 = \sqrt{(2.474-6)^2 + (8.924-12)^2 + (0.239-0)^2 + (3.283-4)^2 + (45.472-43.75)^2} = 5$$

Jika nilai K (tetangga terdekat)=3 maka nilai jarak data dapat diambil dari data nilai terkecil, yakni 3,5,5. Untuk jarak data ke 1,3 dan 4 memiliki hasil yang sama. maka diambil data pertama sebagai jarak terdekatnya. Untuk lebih jelasnya akan dipaparkan pada Tabel 4.25 dan Tabel 4.26.

Tabel 4. 25 Hasil jarak data cluster 1

	jarak data1	jarak data2	jarak data3	jarak data4
C_1	5	3	5	5

Tabel 4. 26 Hasil data terkecil dan data sebenarnya

Jarak data ke- n	Distance	Label sebenarnya
Jarak data ke 2	3	Lulus
Jarak data ke 1	5	Lulus
Jarak data ke 3	5	Lulus

Karena ketiga data berlabel sebenarnya “Lulus”, Maka dari tabel diatas dapat diambil kesimpulan yakni data baru akan ber label lulus.

4.3 Pelatihan Model

Berdasarkan Gambar 2.1 mengenai skenario pelatihan model. Dilakukan variasi model pelatihan yang bertujuan untuk membuat model dan skenario terbaik. Sehingga pada tahap pelatihan selanjutnya, cukup dilakukan dengan satu kali pelatihan model saja yang mana menggunakan skenario terbaik yang telah dirancang sebelumnya.

Pada pelatihan model FCM-KNN ini menggunakan bahasa python dengan bantuan library *scikit learn* yaitu library berbasis python untuk membangun model pembelajaran mesin yang menyediakan banyak algoritma didalamnya untuk klasifikasi, regresi maupun pengelompokan. Kemudian memakai library *fcmeans* untuk mengimport modul algoritma FCM, library *scipy* digunakan untuk menangani matriks dan operasi matematika yang lebih kompleks, library *numpy* yang merupakan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

library gratis untuk menyediakan fungsi matematika untuk mengelola array multi dimensi dan library *pandas* yang digunakan untuk membaca file dalam format .csv, .txt dan lainnya untuk memuat tabel virtual seperti *spreadsheet*.

Pada penelitian ini menggunakan penggabungan algoritma Fuzzy C-Means dan K-Nearest Neighbors. Untuk setiap model yang sudah dibuat maka akan dihitung menggunakan skor *confusion matrix* yang sudah dijelaskan pada rumus persamaan 2.26, 2.27, 2.28 dan pengujian *k-fold cross validation* guna mengetahui model yang terbaik. Untuk lebih jelasnya akan di ilustrasikan seperti Tabel 4.27.

Tabel 4. 27 Skenario pembagian 2500 data.

	2500 data									
	Percobaan A (250)	Percobaan B (250)	Percobaan C (250)	Percobaan D (250)	Percobaan E (250)	Percobaan F (250)	Percobaan G (250)	Percobaan H (250)	Percobaan I (250)	Percobaan J (250)
Percobaan 1	A	B	C	D	E	F	G	H	I	J
Percobaan 2	A	B	C	D	E	F	G	H	I	J
Percobaan 3	A	B	C	D	E	F	G	H	I	J
Percobaan 4	A	B	C	D	E	F	G	H	I	J
Percobaan 5	A	B	C	D	E	F	G	H	I	J
Percobaan 6	A	B	C	D	E	F	G	H	I	J
Percobaan 7	A	B	C	D	E	F	G	H	I	J
Percobaan 8	A	B	C	D	E	F	G	H	I	J
Percobaan 9	A	B	C	D	E	F	G	H	I	J
Percobaan 10	A	B	C	D	E	F	G	H	I	J
Keterangan :										
	Data Training									
	Data Testing									

Langkah pertama yakni menentukan jumlah cluster yang akan dibentuk dan tetapkan nilai K (tetangga terdekat). Maka pada Tabel 4.28 akan dijelaskan skenario perbandingan akurasi pada jumlah cluster dan tetangga terbaiknya tanpa menggunakan pengujian *cross validaton.*

Tabel 4. 28 Hasil pengujian akurasi nilai cluster dan K tanpa *cross validation* (train 60% dan test 40%)

	Skenario k	Accuracy
Cluster 2	k1	0.7709
	k3	0.7669
	k5	0.7909
	k7	0.7862
	Skenario k	Accuracy
	k1	0.7769

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Cluster 3	k3	0.7809
	k5	0.8049
	k7	0.8070
Cluster 4	Skenario k	Accuracy
	k1	0.7769
	k3	0.7809
	k5	0.8049
	k7	0.8061

Tabel 4. 29 Hasil pengujian akurasi nilai cluster dan K menggunakan *cross validation*.

Cluster 2	Skenario k	Accuracy
	k1	0.6907
	k3	0.6967
	k5	0.7007
	k7	0.7011
Cluster 3	Skenario k	Accuracy
	k1	0.7440
	k3	0.7011
	k5	0.7051
	k7	0.7007
Cluster 4	Skenario k	Accuracy
	k1	0.7435
	k3	0.7131
	k5	0.7107
	k7	0.7119

Dari penjelasan Tabel 4.28 dan Tabel 4.29, cluster 3 menghasilkan akurasi yang cukup baik dibandingkan cluster lainnya. Pada skenario pengujian tanpa k-*fold*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

didapatkan akurasi tertinggi cluster 3 pada k=7. Sedangkan skenario pengujian menggunakan 10-*fold cross validation* didapatkan akurasi tertinggi cluster 3 pada k=1. Kemudian dari hasil diatas akan dihitung menggunakan skor *confusion matrix.* Untuk lebih jelasnya akan dipaparkan pada Tabel 4.29.

Tabel 4. 30 Confusion matrix model FCM-KNN dengan skenario k=7

Confusion Matrix		Kelas hasil prediksi	
		Positif	Negatif
Kelas sebenarnya	Positif	72	31
	Negatif	260	1139

Kesimpulan yang didapatkan pada Tabel 4.29 adalah didapatkan data lulus 72, data lulus yang teridentifikasi tidak lulus 31, data tidak lulus yang teridentifikasi tidak lulus 1139, dan tidak lulus yang teridentifikasi lulus 260.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB V
PENUTUP

5.1 Kesimpulan

Dari proses penelitian yang telah dilakukan sudah dijelaskan tahapan pada masing-masing sub bab. Maka, didapatkan kesimpulan serta dapat menjawab dari rumusan masalah pada bab 1.2 sebagai berikut:

1. Hasil dari membangun model FCM-KNN menggunakan 5 variabel (jalur penerimaan, jenis program studi, asal smu, pernah ponpes dan nem SMA) didapatkan dengan pengujian *10-fold cross validation* adalah k=1 dengan menggunakan 3 cluster sedangkan jika menggunakan pengujian training 60% dan testing 30% didapatkan hasil 3 cluster dengan k=7.
2. Performa metode FCM-KNN dengan pengujian *10-fold cross validation* diperoleh tingkat rata rata akurasinya adalah 71%.

5.2 Saran

Dalam serangkaian penelitian Prediksi Kelulusan Tepat Waktu Berdasarkan Data Registrasi Mahasiswa Dengan Algoritma *Fuzzy C-Means* dan *K-Nearest Neighbors* yang telah dilakukan tentunya masih banyak kekurangan. Oleh itu, penulis memberikan beberapa saran serta rekomendasi untuk tahap penelitian yang akan datang. Yaitu:

1. Model prediksi FCM-KNN dioptimasi lagi dengan melihat varian jumlah data agar tidak terjadi *class* yang tidak seimbang (*imbalance*). Salah satunya yakni menggunakan algoritma *Adaboost.*
2. Mengembangkan penelitian yang serupa dengan menambahkan beberapa *feature* serta dapat menggunakan metode pengelompokan lainnya yakni seperti K-Means, K-Modes, Bisecting K-Means

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR PUSTAKA

Afrianto, R. B., & Kurniawati, L. Y. (2013). Kategorisasi Dokumen Teks Secara Multi Label Menggunakan Fuzzy C-Means Dan K-Nearest Neighbors Pada Artikel Berbahasa Indonesia. *JUTI: Jurnal Ilmiah Teknologi Informasi*, *11*(1), 23. https://doi.org/10.12962/j24068535.v11i1.a17

Agustini, F. (2017). *Implementasi Algoritma Fuzzy C-Means Studi Kasus Penjualan Di Sushigroove Restaurant*. *3*(1), 6.

Ahmadi, A., & Hartini, S. (2013). *Penerapan Fuzzy C-Means dalam Sistem Pendukung Keputusan untuk Penentuan Penerima Bantuan Langsung Masyarakat (BLM).pdf*.

Aradea, Satriyo, A., Ariyan, Z., & Yuliana, A. (2011). *Penerapan Decision Tree Untuk Penentuan Pola Data Penerimaan Mahasiswa Baru*. 7.

Banjarsari, M. A., Budiman, H. I., & Farmadi, A. (2015). *Penerapan K-Optimal Pada Algoritma Knn untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan IP Sampai Dengan Semester 4*. *02*, 15.

BANPT. (2009). https://belmawa.ristekdikti.go.id/wp-content/uploads/aturanbelmawa/2009/permen_73_2009.pdf

Darujati, C., & Gumelar, A. B. (2012). *Pemanfaatan Teknik Supervised Untuk Klasifikasi Teks Bahasa Indonesia*. 9.

Falahi, F. M. (2019). *Penerapan Metode Clustering Untuk Pengelompokan Mahasiswa Potensial Drop Out Menggunakan Algoritma K-Means ++*.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

http://digilib.uinsby.ac.id/32990/2/Fakhri%20Mohammad%20Falahi_H02215004.pdf

Fansuri, F. (2012). *Jurnal Komputer dan Informatika (KOMPUTA)*. *1*, 7.

García, S., Luengo, J., & Herrera, F. (2015). *Data Preprocessing in Data Mining* (Vol. 72).

General Python FAQ. (2019). https://docs.python.org/3/faq/general.html#what-is-python

Han, J. (2011). *Data Mining: Concepts and Techniques*. 128.

Krisandi, N., Prihandono, B., & Helmi. (2013). *Algoritma K-Nearest Neighbor Dalam Klasifikasi Data Hasil Produksi Kelapa Sawit Pada Pt. Minamas Kecamatan Parindu*.

Mariana, N., Redjeki, R. S., & Razaq, J. A. (2015). *Penerapan Algoritma K-NN (Nearest Neighbor) Untuk Deteksi Penyakit (Kanker Serviks)*. *Vol. 7 No.1*.

Maulana, A., & Fajrin, A. A. (2018). Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Transaksi Penjualan Spare Part Motor. *KLIK - KUMPULAN JURNAL ILMU KOMPUTER*, *5*(1), 27–36. https://doi.org/10.20527/klik.v5i1.100

Natalius, S. (2011). *Metoda Naïve Bayes Classifier dan Penggunaannya pada Klasifikasi Dokumen*. 5.

Nugraheny, D. (2017). Metode Nilai Jarak Guna Kesamaan Atau Kemiripan Ciri Suatu Citra (Kasus Deteksi Awan Cumulonimbus Menggunakan Principal Component Analysis). *Angkasa: Jurnal Ilmiah Bidang Teknologi*, *7*(2), 21. https://doi.org/10.28989/angkasa.v7i2.145

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Nugroho, Y. S. (2015). *Klasifikasi Dan Klastering Mahasiswa Informatika Universitas Muhammadiyah Surakarta*. 10.

Pitria, P. (2014). *Analisis Sentimen Pengguna Twitter Pada Akun Resmi Samsung Indonesia Dengan Menggunakan Naïve Bayes*. 8.

Priandini, N., Zaman, B., & Purwanti, E. (2017). Categorizing Document by Fuzzy C-Means and K-Nearest Neighbors Approach. *AIP Conference Proceedings*. https://doi.org/10.1063/1.4994415

Priyatman, H., Sajid, F., & Haldivany, D. (2019). *Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa*. *5*.

Putra, M. I. (2019). *Sistem Rekomendasi Kelayakan Kredit Menggunakan Metode Random Forest Pada BRI Kantor Cabang Pelaihari*.

Ren, Q., Li, M., & Han, S. (2019). Tectonic discrimination of olivine in basalt using data mining techniques based on major elements: A comparative study from multiple perspectives. *Big Earth Data*, *3*(1), 8–25. https://doi.org/10.1080/20964471.2019.1572452

Satriya, R. H. D., & Santoso, E. (2018). *Implementasi Metode Ensemble K-Nearest Neighbor untuk Prediksi Nilai Tukar Rupiah Terhadap Dollar Amerika*. 8.

Turban, E., Aronson, J. E., & Liang, T.-P. (2005). *Decision Support Systems and Intelligent* (7 th Ed.).

Wijaya, A. P., & Santoso, H. A. (2016). *Naive Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government*. *1*, 8.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Yusuf, A., Ginardi, H., & Arieshanti, I. (2012). *Pengembangan Perangkat Lunak Prediktor Nilai Mahasiswa Menggunakan Metode Spectral Clustering dan Bagging Regresi Linier*. *1*, 5.

Zehansyah, M. B., Sari, Y. A., & Nugroho, S. A. (2019). Rekomendasi Multilabel Otomatis Pada Artikel Dengan Algoritme Fuzzy CMeans dan K Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *3*. http://j-ptiik.ub.ac.id

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id