

## **Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Local Mean Based K-Nearest Neighbor**

Diajukan Sebagai Syarat Untuk Menyelesaikan  
Pendidikan Program Starata-1 Pada  
Jurusan Teknik Informatika



Oleh:

Muhamad Reza Kurniawan  
NIM: 09021281823051

**Jurusan Teknik Informatika  
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA  
2022**

## LEMBAR PENGESAHAN SKRIPSI

Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode  
Local Mean Based K-Nearest Neighbor

Oleh:

Muhamad Reza Kurniawan  
NIM: 090212818123051

Indralaya, 24 Juni 2022

Pembimbing I,

Dian Palupi Rini, S.Si., M.Kom  
NIP. 198410012009121005

Pembimbing II,

Anggina Primanita, S.Kom.,M.IT  
NIP. 198908062015042002

Mengetahui,  
Ketua Jurusan



Alvi Syahrin Utami, M.Kom.  
NIP. 19781222200642003

## TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari Jumat tanggal 24 Juni 2022 telah dilaksanakan ujian sidang ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Muhamad Reza Kurniawan

NIM : 09021281823051

Judul : Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Local Mean Based K-Nearest Neighbor

dan dinyatakan **LULUS**

1. Ketua Pengaji

Dr. Abdiansah, S.Kom., M.CS.

NIP. 198410012009121005

2. Pembimbing I

Dian Palupi Rini, S.Si., M.Kom

NIP. 198410012009121005

3. Pembimbing II

Anggina Primanita, S.Kom.,M.IT

NIP. 198908062015042002

4. Pengaji I

Mastura Diana Marieska, M.T.

NIP. 198603212018032001

5. Pengaji II

Kanda Januar Miraswan, S.Kom., M.I.

NIP. 1671080901900006



## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Muhamad Reza Kurniawan

NIM : 09021281823051

Jurusan : Teknik Informatika

Judul Skripsi : Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Local Mean Based K-Nearest Neighbor

Hasil pengecekan *Software iThenticate/Turitin* : 8 %

Menyatakan bahwa Laporan Proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapa pun.



Indralaya, 24 Juni 2022



Muhamad Reza Kurniawan  
NIM. 09021281823051

## **Motto dan Persembahan**

*“When you decide to go to the sea, it was your own decision. Whatever happen to you on the sea, it depends on what you’ve done”*

– Roronoa Zoro

Kupersembahkan karya tulis ini kepada:

- Allah Subhanahu Wa Ta ’ala dan Rasulullah SAW
- Ibu dan Bapak
- Adikku Arman
- Dosen pembimbing Akademik dan Skripsi Saya
- Universitas Sriwijaya
- Sahabat dan teman-teman seperjuangan

## **ABSTRACT**

Student graduation is an indicator or a measurement of a university's success in conducting an education and learning process. Student Graduation is one way to determine university accreditation. Technology that can predict student graduation time is urgently needed to provide recommendations to universities or faculties so that they can take actions to reducing late-graduation students. Special software has been developed to analyze whether students have completed the course on time. The developed software can predict student graduation based on K-Nearest Neighbor (KNN) and Local Mean K-Nearest Neighbor (LMKNN). Then compare the results of the two methods to get the best one. The data used in this study was data from 127 students of the Informatics Engineering Department of Sriwijaya University. It consisted of 61 students who finished on time and 66 students who finished late. This research used the Cross-Validation method with 10 folds to find the performance value of the LMKNN method. The results of this research indicate that the accuracy of the LMKNN method reaches the highest accuracy value in the  $k = 9$  configuration with 0.737. The LMKNN algorithm can generate better accuracy values than the KNN algorithm with an increase of 2% to 5%.

*Key Word: Local Mean Based K-Nearest Neighbor, Data Prediction, Data Mining Student Graduation*

## **ABSTRAK**

Tingkat kelulusan mahasiswa adalah indikator atau tolak ukur kesuksesan universitas dalam pelaksanaan proses pendidikan dan pembelajaran. Kelulusan Mahasiswa adalah salah satu perangkat untuk menentukan akreditasi universitas. Teknologi yang dapat memprediksi waktu kelulusan mahasiswa sangat dibutuhkan untuk memberikan rekomendasi kepada Universitas atau Fakultas agar dapat melakukan tindakan-tindakan untuk mengurangi mahasiswa yang lulus terlambat. Untuk menganalisis kelulusan siswa tepat waktu, perangkat lunak khusus dikembangkan. Perangkat lunak yang dikembangkan mampu memprediksi kelulusan mahasiswa menggunakan *K-Nearest Neighbor* (KNN) dan *Local Mean K-Nearest Neighbor* (LMKNN). Hasil dari kedua metode tersebut kemudian dibandingkan untuk mendapatkan metode yang terbaik. Data yang digunakan dalam penelitian ini adalah 127 data mahasiswa yang diambil dari jurusan Teknik Informatika Universitas Sriwijaya, terdiri dari 61 data mahasiswa lulus tepat waktu dan 66 data mahasiswa tidak lulus tepat waktu. Pada penelitian ini menggunakan metode *Cross-Validation* dengan 10 *fold* untuk mencari nilai performa dari metode LMKNN. Hasil penelitian ini menunjukkan bahwa akurasi metode LMKNN mencapai nilai akurasi tertinggi pada konfigurasi  $k=9$  sebesar 0,737. Algoritma LMKNN dapat menghasilkan nilai akurasi yang lebih baik dari pada algoritma KNN dengan peningkatan sebesar 2% - 5 %.

Kata Kunci: *Local Mean Based K-Nearest Neighbor*, Klasifikasi Data, Penggalian data, Kelulusan Mahasiswa

## **KATA PENGANTAR**

Puji syukur penulis panjatkan kehadirat Allah SWT, Tuhan Semesta Alam atas berkat, rahmat, rahim dan karunia-Nya yang telah diberikan kepada penulis sehingga Tugas Akhir berjudul “Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Local Mean Based K-Nearest Neighbor” dapat disusun dengan baik sebagai syarat dalam menyelesaikan studi Strata-1 program studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Sriwijaya. Pada kesempatan ini penulis ingin mengucapkan ucapan terima kasih yang luar biasa besar kepada semua pihak yang memberikan dukungan, motivasi dan bimbingan selama penyusunan Tugas Akhir dan penelitian. Secara khusus ucapan terima kasih ini ditujukan kepada:

1. Allah Subhanallahu Wa Ta’ala atas segala berkah, rahmat, rahim dan karunia-Nya.
2. Keluarga tercinta, Ibu dan Bapak yang selalu memberikan doa serta semangat dan dukungan postive.
3. Adikku Arman yang selalu menemani dan membantu dalam banyak hal.
4. Bapak Jaidan Jauhari, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Ibu Alvi Syahrini Utami, M.Kom. selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Ibu Dian Palupi Rini, S.Si., M.Kom. selaku Dosen Pembimbing I yang senantiasa membagikan ilmu, membimbing, memberi arahan dan memotivasi saya untuk menyelesaikan Tugas Akhir ini dengan baik.

7. Ibu Anggina Primanita, S.Kom.,M.IT. yang dalam hal ini selaku Dosen Pembimbing II yang senantiasa membagikan ilmu, membimbing, memberi arahan dan memotivasi saya untuk menyelesaikan Tugas Akhir ini dengan baik.
  8. Dr. Abdiansah, S.Kom., M.Cs selaku ketua penguji Tugas Akhir yang telah memberikan nasihat dan saran yang membangun
  9. Mastura Diana Marieska, M.T. selaku penguji I Tugas Akhir yang telah memberikan nasihat dan saran yang membangun.
  10. Kanda Januar Miraswan, S.Kom., M.T. selaku penguji II Tugas Akhir yang telah memberikan nasihat dan saran yang membangun.
  11. Ibu Novi Yusliani S.Kom., M.T. selaku Dosen Pembimbing Akademik yang senantiasa memberikan saran, arahan dan bimbingan kepada saya selama masa perkuliahan.
  12. Seluruh Dosen Jurusan Teknik Informatika, juga Dosen Fakultas Ilmu Komputer yang telah banyak memberikan bekal, arahan, dan saran selama masa perkuliahan.
  13. Seluruh Staf Administrasi dan Pegawai yang telah membantu dalam urusan administrasi.
  14. Sahabat – sahabatku yang selalu sedia menjadi rekan diskusi baik dalam penggerjaan Tugas Akhir maupun selama masa perkuliahan.
- Penulis secara penuh menyadari akan kekurangan dalam penyusunan Tugas Akhir ini. Kekurangan ini semata – mata adalah karena keterbatasan pengetahuan juga pengalaman penulis. Oleh karena itu, kritik dan saran yang membangun sangat

diharapkan untuk menyempurnakan Tugas akhir ini sehingga dapat membawa manfaat lebih banyak dan luas. Akhir kata, semoga Tugas Akhir ini bermanfaat bagi kita semua.

Indralaya, 24 Juni 2022

Muhamad Reza Kurniawan

## DAFTAR ISI

**Halaman**

LEMBAR PENGESAHAN SKRIPSI .....	i
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI.....	ii
HALAMAN PERNYATAAN .....	iii
Motto dan Persembahan.....	iv
ABSTRACT.....	v
ABSTRAK .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR .....	xviii
DAFTAR LAMPIRAN .....	xxi
BAB I PENDAHULUAN .....	I-1
1.1. Pendahuluan .....	I-1
1.2. Latar Belakang .....	I-1
1.3. Rumusan Masalah .....	I-3
1.4. Tujuan Penelitian.....	I-4
1.5. Manfaat Penelitian.....	I-4
1.6. Batasan Masalah.....	I-4
1.7. Sistematika Penulisan.....	I-5
1.8. Kesimpulan.....	I-6
BAB II KAJIAN LITERATUR .....	II-1
2.1. Pendahuluan .....	II-1
2.2. Landasan Teori.....	II-1
2.2.1. Kelulusan Mahasiswa.....	II-1
2.2.2. <i>Data Mining</i> .....	II-2
2.2.3. <i>Machine Learning</i> .....	II-3
2.2.4. K-Nearest Neighbor .....	II-4

2.2.5.	Local Mean Based K-Nearest Neighbor .....	II-7
2.2.6.	Pengukuran Hasil Klasifikasi .....	II-8
2.2.7.	Rational Unified Process .....	II-10
2.3.	Penelitian Lain yang Relevan.....	II-12
2.4.	Kesimpulan.....	II-15
 BAB III METODE PENELITIAN.....		III-1
3.1.	Pendahuluan .....	III-1
3.2.	Pengumpulan Data .....	III-1
3.2.1.	Jenis dan Sumber Data .....	III-1
3.2.2.	Metode Pengumpulan Data .....	III-1
3.3.	Tahapan Penelitian .....	III-2
3.3.1.	Menentukan Kerangka Kerja Penelitian.....	III-3
3.3.2.	Menentukan Kriteria Pengujian.....	III-5
3.3.3.	Menentukan Format Data Pengujian .....	III-5
3.3.4.	Alat yang Digunakan dalam Pelaksanaan Penelitian .....	III-6
3.3.5.	Pengujian Penelitian .....	III-6
3.3.6.	Analisis Hasil Pengujian dan Membuat Kesimpulan .....	III-7
3.4.	Metode Pengembangan Perangkat Lunak .....	III-7
3.5.	Kesimpulan.....	III-8
 BAB IV PENGEMBANGAN PERANGKAT LUNAK .....		IV-1
4.1.	Pendahuluan .....	IV-1
4.2.	Fase Insepsi .....	IV-1
4.2.1.	Pemodelan Bisnis .....	IV-1
4.2.2.	Kebutuhan Sistem.....	IV-2
4.2.3.	Analisis dan Perancangan.....	IV-4
4.2.3.1.	Analisis Kebutuhan Perangkat Lunak .....	IV-4
4.2.3.2.	Analisis Data .....	IV-4
4.2.3.3.	Analisis Pre-Processing.....	IV-5
4.2.3.4.	Analisis Proses Klasifikasi .....	IV-9
4.2.3.5.	Analisis Hasil Klasifikasi .....	IV-19

4.2.4.	Implementasi .....	IV-19
4.2.4.1.	<i>Diagram Use Case</i> .....	IV-20
4.2.4.2.	Skenario Use Case.....	IV-21
4.3.	Fase Elaborasi .....	IV-25
4.3.1.	Pemodelan Bisnis .....	IV-25
4.3.1.1.	Perancangan Data.....	IV-25
4.3.1.2.	Perancangan Antarmuka Pengguna.....	IV-25
4.3.2.	Kebutuhan .....	IV-26
4.3.3.	Analisis dan Perancangan.....	IV-27
4.3.3.1.	Diagram Aktivitas .....	IV-27
4.3.3.2.	Sequence diagram .....	IV-31
4.4.	Fase Konstruksi .....	IV-34
4.4.1.	Kebutuhan .....	IV-34
4.4.2.	Implementasi .....	IV-35
4.4.2.1.	Implementasi Kelas .....	IV-36
4.4.2.2.	Implementasi Antarmuka Pengguna.....	IV-37
4.5.	Fase Transisi.....	IV-38
4.5.1.	Pemodelan Bisnis .....	IV-38
4.5.2.	Kebutuhan .....	IV-38
4.5.3.	Analisis dan Perancangan.....	IV-38
4.5.3.1.	Rencana Pengujian .....	IV-39
4.5.4.	Implementasi .....	IV-40
4.6.	Kesimpulan.....	IV-43
 BAB V HASIL DAN ANALISIS PENEITIAN .....		V-1
5.1.	Pendahuluan .....	V-1
5.2.	Hasil Perangkat Lunak .....	V-1
5.3.	Data Hasil Penelitian.....	V-3
5.3.1.	Konfigurasi Percobaan .....	V-3
5.3.1.1.	Data Hasil Konfigurasi I .....	V-4
5.3.1.2.	Data Hasil Konfigurasi II .....	V-22
5.3.1.3.	Data Hasil Analisis Klasifikasi.....	V-39

5.4. Analisis Hasil Penelitian .....	V-40
5.5. Kesimpulan.....	V-49
<b>BAB VI HASIL DAN ANALISIS PENEITIAN.....</b>	<b>VI-1</b>
6.1. Pendahuluan .....	VI-1
6.2. Kesimpulan.....	VI-1
6.3. Saran.....	VI- 2
<b>DAFTAR PUSTAKA .....</b>	<b>xxii</b>
<b>LAMPIRAN .....</b>	<b>xxviii</b>

## DAFTAR TABEL

Tabel II- 1. <i>Confusion Matrix</i> .....	II-9
Table III- 1. Rancangan Tabel <i>Confusion Matrix</i> .....	III-5
Table III- 2. Rancangan Tabel Hasil Pengujian.....	III-6
Table III- 3. Tabel Hasil Analisis Klasifikasi .....	III-7
Tabel IV- 1. Kebutuhan Fungsional Perangkat Lunak .....	IV-3
Tabel IV- 2. Kebutuhan Non Fungsional Perangkat Lunak .....	IV-3
Tabel IV- 3. Data Indeks Prestasi Semester Mahasiswa Teknik Informatika Reguler .....	IV-5
Tabel IV- 4. Hasil <i>Cleaning</i> .....	IV-7
Tabel IV- 5. Hasil <i>Case Folding</i> .....	IV-8
Tabel IV- 6. Hasil <i>Labeling</i> .....	IV-9
Tabel IV- 7. Data Hasil Pengurutan Jarak Metode KNN .....	IV-13
Tabel IV- 8. K-Data Pertama Metode KNN .....	IV-14
Tabel IV- 9. Hasil Klasifikasi Metode KNN .....	IV-14
Tabel IV- 10. Data Hasil Pengurutan Jarak Metode KNN .....	IV-18
Tabel IV- 11. K-Data Pertama Setiap Kelas Metode LMKNN .....	IV-18
Tabel IV- 12. Hasil Klasifikasi Metode KNN .....	IV-19
Tabel IV- 13. Definisi Aktor.....	IV-20
Tabel IV- 14. Definisi <i>Use Case</i> .....	IV-21
Tabel IV- 15. <i>Use Case</i> Klasifikasi Menggunakan KNN.....	IV-21
Tabel IV- 16. <i>Use Case</i> Klasifikasi Menggunakan LMKNN.....	IV-23
Tabel IV- 17. <i>Use Case</i> Klasifikasi dari Input Pengguna.....	IV-24
Tabel IV- 18 Implementasi Kelas.....	IV-36

Tabel IV- 19. Rencana Pengujian Use Case Proses Klasifikasi KNN .....	IV-39
Tabel IV- 20. Rencana Pengujian Use Case Proses Klasifikasi LMKNN .....	IV-39
Tabel IV- 21. Rencana Pengujian Use Case Klasifikasi Data Masukan Pengguna .....	IV-39
Tabel IV- 22. Pengujian Use Case Proses Klasifikasi KNN .....	IV-40
Tabel IV- 23. Pengujian Use Case Proses Klasifikasi LMKNN .....	IV-41
Tabel IV- 24. Pengujian Use Case Klasifikasi Data Masukan Pengguna .....	IV-42
Tabel V- 1. Hasil <i>Confusion Matrix</i> Klasifikasi KNN k = 1 .....	V-4
Tabel V- 2. Hasil Performa Klasifikasi KNN K = 1 .....	V-5
Tabel V- 3. Hasil <i>Confusion Matrix</i> Klasifikasi KNN k = 2 .....	V-6
Tabel V- 4. Hasil Performa Klasifikasi KNN K = 2 .....	V-7
Tabel V- 5. Hasil <i>Confusion Matrix</i> Klasifikasi KNN k = 3 .....	V-8
Tabel V- 6. Hasil Performa Klasifikasi KNN K = 3 .....	V-8
Tabel V- 7. Hasil <i>Confusion Matrix</i> Klasifikasi KNN k = 4 .....	V-9
Tabel V- 8. Hasil Performa Klasifikasi KNN K = 4 .....	V-10
Tabel V- 9. Hasil <i>Confusion Matrix</i> Klasifikasi KNN k = 5 .....	V-11
Tabel V- 10. Hasil Performa Klasifikasi KNN K = 5 .....	V-12
Tabel V- 11. Hasil <i>Confusion Matrix</i> Klasifikasi KNN k = 6 .....	V-13
Tabel V- 12. Hasil Performa Klasifikasi KNN K = 6 .....	V-13
Tabel V- 13. Hasil <i>Confusion Matrix</i> Klasifikasi KNN k = 7 .....	V-14
Tabel V- 14. Hasil Performa Klasifikasi KNN K = 7 .....	V-15
Tabel V- 15. Hasil <i>Confusion Matrix</i> Klasifikasi KNN k = 8 .....	V-16
Tabel V- 16. Hasil Performa Klasifikasi KNN K = 9 .....	V-17

Tabel V- 17. Hasil <i>Confusion Matrix</i> Klasifikasi KNN k = 9 .....	V-18
Tabel V- 18. Hasil Performa Klasifikasi KNN K = 9 .....	V-19
Tabel V- 19. Hasil <i>Confusion Matrix</i> Klasifikasi KNN k = 10 .....	V-20
Tabel V- 20. Hasil Performa Klasifikasi KNN K = 10.....	V-20
Tabel V- 21. Hasil <i>Confusion Matrix</i> Klasifikasi LMKNN k = 1 .....	V-22
Tabel V- 22. Hasil Performa Klasifikasi LMKNN K = 1 .....	V-23
Tabel V- 23. Hasil <i>Confusion Matrix</i> Klasifikasi LMKNN k = 2 .....	V-24
Tabel V- 24. Hasil Performa Klasifikasi LMKNN K = 2.....	V-25
Tabel V- 25. Hasil <i>Confusion Matrix</i> Klasifikasi LMKNN k = 3 .....	V-26
Tabel V- 26. Hasil Performa Klasifikasi LMKNN K = 3 .....	V-26
Tabel V- 27. Hasil <i>Confusion Matrix</i> Klasifikasi LMKNN k = 4 .....	V-27
Tabel V- 28. Hasil Performa Klasifikasi LMKNN K = 4.....	V-28
Tabel V- 29. Hasil <i>Confusion Matrix</i> Klasifikasi LMKNN k = 5 .....	V-29
Tabel V- 30. Hasil Performa Klasifikasi LMKNN K = 5.....	V-29
Tabel V- 31. Hasil <i>Confusion Matrix</i> Klasifikasi LMKNN k = 6 .....	V-30
Tabel V- 32. Hasil Performa Klasifikasi KNN K = 6.....	V-31
Tabel V- 33. Hasil <i>Confusion Matrix</i> Klasifikasi LMKNN k = 7 .....	V-32
Tabel V- 34. Hasil Performa Klasifikasi LMKNN K = 7.....	V-33
Tabel V- 35. Hasil <i>Confusion Matrix</i> Klasifikasi LMKNN k = 8 .....	V-34
Tabel V- 36. Hasil Performa Klasifikasi LMKNN K = 9.....	V-34
Tabel V- 37. Hasil <i>Confusion Matrix</i> Klasifikasi LMKNN k = 9 .....	V-35
Tabel V- 38. Hasil Performa Klasifikasi LMKNN K = 9.....	V-36
Tabel V- 39. Hasil <i>Confusion Matrix</i> Klasifikasi LMKNN k = 10 .....	V-37

Tabel V- 40. Hasil Performa Klasifikasi LMKNN K = 10.....V-38

Tabel V- 41. Nilai Rata-Rata Peforma Klasifikasi Menggunakan KNN dan LMKNN

.....V-39

## DAFTAR GAMBAR

Gambar II- 1. Ilustrasi klasifikasi <i>k-nearest neighbor</i> . Sumber : Liu (2011).....	II-5
Gambar II- 2. Arsitektur <i>Rational Unified Process</i> . Sumber : Anwar (2014)....	II-11
Gambar III- 1. Diagram Tahapan Penilitian.....	III-2
Gambar III- 2. Diagram Alir Perangkat Lunak .....	III-3
Gambar IV- 1. Use Case .....	IV-20
Gambar IV- 2. Rancangan Antar Muka Perangkat Lunak .....	IV-26
Gambar IV- 3. Diagram Aktifitas Klasifikasi Menggunakan KNN .....	IV-28
Gambar IV- 4. Diagram Aktifitas Klasifikasi Menggunakan LMKNN .....	IV-29
Gambar IV- 5. Diagram Aktifitas Klasifikasi dari Input Pengguna .....	IV-30
Gambar IV- 6. <i>Sequence Diagram</i> Klasifikasi Menggunakan KNN .....	IV-31
Gambar IV- 7. <i>Sequence Diagram</i> Klasifikasi Menggunakan LMKNN .....	IV-32
Gambar IV- 8. <i>Sequence Diagram</i> Klasifikasi Data dari Input Pengguna .....	IV-33
Gambar IV- 9. Diagram Kelas Perangkat Lunak .....	IV-35
Gambar IV- 10 Implementasi Antarmuka Pengguna .....	IV-37
Gambar V- 1. Hasil Tampilan Awal Perangkat Lunak.....	V-2
Gambar V- 2. Tampilan Fitur Klasifikasi Perangkat Lunak.....	V-2
Gambar V- 3. Tampilan Fitur Klasifikasi Data Masukkan Oleh Pengguna .....	V-3
Gambar V- 4. Grafik Hasil Performa Klasifikasi KNN $k = 1$ .....	V-5
Gambar V- 5. Grafik Hasil Performa Klasifikasi KNN $k = 2$ .....	V-7
Gambar V- 6. Grafik Hasil Performa Klasifikasi KNN $k = 3$ .....	V-9
Gambar V- 7. Grafik Hasil Performa Klasifikasi KNN $k = 4$ .....	V-10
Gambar V- 8. Grafik Hasil Performa Klasifikasi KNN $k = 5$ .....	V-12
Gambar V- 9. Grafik Hasil Performa Klasifikasi KNN $k = 6$ .....	V-14

Gambar V- 10. Grafik Hasil Performa Klasifikasi KNN k = 7 .....	V-15
Gambar V- 11. Grafik Hasil Performa Klasifikasi KNN k = 8 .....	V-17
Gambar V- 12. Grafik Hasil Performa Klasifikasi KNN k = 9 .....	V-19
Gambar V- 13. Grafik Hasil Performa Klasifikasi KNN k = 10 .....	V-21
Gambar V- 14. Grafik Hasil Performa Klasifikasi LMKNN k = 1 .....	V-23
Gambar V- 15. Grafik Hasil Performa Klasifikasi LMKNN k = 2 .....	V-25
Gambar V- 16. Grafik Hasil Performa Klasifikasi LMKNN k = 3 .....	V-27
Gambar V- 17. Grafik Hasil Performa Klasifikasi LMKNN k = 4 .....	V-28
Gambar V- 18. Grafik Hasil Performa Klasifikasi LMKNN k = 5 .....	V-30
Gambar V- 19. Grafik Hasil Performa Klasifikasi LMKNN k = 6 .....	V-31
Gambar V- 20. Grafik Hasil Performa Klasifikasi LMKNN k = 7 .....	V-33
Gambar V- 21. Grafik Hasil Performa Klasifikasi LMKNN k = 8 .....	V-35
Gambar V- 22. Grafik Hasil Performa Klasifikasi LMKNN k = 9 .....	V-36
Gambar V- 23. Grafik Hasil Performa Klasifikasi LMKNN k = 10 .....	V-38
Gambar V- 24. Grafik Perbandingan Performa Hasil Klasifikasi k : 1 .....	V-41
Gambar V- 25. Grafik Perbandingan Performa Hasil Klasifikasi k : 2 .....	V-42
Gambar V- 26. Grafik Perbandingan Performa Hasil Klasifikasi k : 3 .....	V-43
Gambar V- 27. Grafik Perbandingan Performa Hasil Klasifikasi k : 4 .....	V-44
Gambar V- 28. Grafik Perbandingan Performa Hasil Klasifikasi k : 5 .....	V-44
Gambar V- 29. Grafik Perbandingan Performa Hasil Klasifikasi k : 6 .....	V-45
Gambar V- 30. Grafik Perbandingan Performa Hasil Klasifikasi k : 7 .....	V-46
Gambar V- 31. Grafik Perbandingan Performa Hasil Klasifikasi k : 8 .....	V-46
Gambar V- 32. Grafik Perbandingan Performa Hasil Klasifikasi k : 9 .....	V-47

Gambar V- 33. Grafik Perbandingan Performa Hasil Klasifikasi k : 10.....V-48

## **DAFTAR LAMPIRAN**

Lampiran 1 .....	xxviii
Lampiran 2 .....	xxxv

## **BAB I**

### **PENDAHULUAN**

#### **1.1. Pendahuluan**

Bab ini memberikan gambaran tentang gagasan utama penelitian ini. Pada bab pendahuluan akan membahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

#### **1.2. Latar Belakang**

Angka kelulusan tepat waktu mahasiswa sangat penting untuk suatu perguruan tinggi. Menurut Novianti & Prasetyo (2017) Tingkat kelulusan merupakan indikator atau ukuran tingkat keberhasilan suatu perguruan tinggi dalam melaksanakan proses belajar mengajarnya. Akreditasi universitas meningkat seiring dengan jumlah lulusan. Menurut Badan Akreditasi Nasional, kelulusan mahasiswa merupakan salah satu instrumen untuk menentukan akreditasi suatu universitas. Terdapat dua poin profil lulusan menurut Badan Akreditasi Nasional Perguruan Tinggi, yaitu persentase kelulusan tepat waktu dan persentase mahasiswa yang *Drop Out* (DO) atau berhenti. Namun menurut Saputra (2021), tidak semua mahasiswa dapat menyelesaikan studinya tepat waktu. Oleh sebab itu, diperlukan suatu teknik untuk memprediksi waktu kelulusan mahasiswa untuk memberikan rekomendasi kepada Universitas atau Fakultas demi mengurangi mahasiswa yang lulus tidak tepat waktu.

Algoritma pembelajaran mesin dapat digunakan untuk memperkirakan apakah siswa akan lulus tepat waktu. Dalam berbagai penelitian sebelumnya, pendekatan *K-Nearest Neighbor* (K-NN), *Support Vector Machine* (SVM), dan

*Naive Bayes* sering digunakan untuk meramalkan waktu kelulusan mahasiswa.

Dalam penelitian yang menggunakan KNN untuk meramalkan waktu kelulusan mahasiswa, Rohman (2015) melakukan penelitian tersebut. Pendekatan KNN dalam penelitian ini menghasilkan nilai akurasi terbesar yaitu 85,15%.

Nilai akurasi pendekatan KNN masih kurang baik menurut berbagai penelitian, terutama jika dibandingkan dengan metode klasifikasi lainnya (Syaliman, 2018). Studi Widaningsih (2019) membandingkan algoritma C4.5, Naive Bayes, KNN, dan SVM untuk memperkirakan kapan mahasiswa akan lulus. Akurasi teknik KNN dalam penelitian ini hanya 68,5 persen, sedangkan akurasi ketiga metode lainnya, Naive Bayes, SVM, dan C4.5, masing-masing adalah 76,79 persen, 74,04 persen, dan 75,96 persen.

Studi lain oleh Danades, *et al.* (2016), di mana pada penelitian tersebut mereka membandingkan akurasi metode KNN dan metode SVM dalam penelitian ini. Hasil dari penelitian ini adalah nilai rata-rata akurasi metode KNN hanya sebesar 71,28% sedangkan nilai rata-rata akurasi metode SVM sebesar 92,40%.

Rendahnya nilai akurasi dari metode KNN disebabkan oleh beberapa hal. Menurut Lidya, *et al* (2015) penentuan kelas data baru pada metode KNN didasarkan pada sistem *vote majority*, di mana yang bergantung pada jumlah K yang digunakan sehingga dapat mengabaikan kedekatan antar data. Kemungkinan munculnya kelas mayoritas ganda adalah masalah lain menurut Ertugul dan Tagluk (2017) dari metode KNN, karena sistem penentuan kelas untuk data baru berdasarkan *vote majority*, maka kelas mayoritas yang ganda dapat mempengaruhi hasil dari penentuan keputusan.

Berdasarkan permasalahan sebelumnya metode *Local Mean Based K-Nearest Neighbor* (LMKNN) diusulkan Mitani & Hamamoto (2006) untuk meningkatkan hasil klasifikasi dari metode KNN. Pemilihan kelas data baru didasarkan pada perhitungan jarak terdekat dengan *local mean* masing-masing kelas data, penentuan kelas untuk data baru tidak lagi berdasarkan sistem *vote majority* seperti metode KNN. Selain itu, LMKNN dapat mengatasi efek negatif dari outlier dalam data pelatihan. LMKNN juga dapat mengatasi terabaikannya kedekatan antar data yang dialami oleh KNN.

Berdasarkan uraian dan referensi dari penelitian sebelumnya, algoritma *Local Mean Based K-Nearest Neighbor* (LMKNN) digunakan untuk memprediksi kelulusan mahasiswa tepat waktu. Hasil penelitian ini diharapkan dapat lebih akurat mengidentifikasi mahasiswa yang lulus terlambat.

### 1.3. Rumusan Masalah

Berdasarkan masalah yang telah diuraikan sebelumnya, maka rumusan masalah penelitian ini adalah:

1. Bagaimana membangun sistem prediksi kelulusan mahasiswa tepat waktu menggunakan metode *Local Mean Based K-Nearest Neighbor*?
2. Bagaimana hasil akurasi dari sistem prediksi kelulusan mahasiswa tepat waktu menggunakan metode *Local Mean Based K-Nearest Neighbor*?
3. Bagaimana hasil perbandingan akurasi Metode *K-Nearest Neighbor* dan *Local Mean Based K-Nearest Neighbor*?

#### **1.4. Tujuan Penelitian**

Tujuan dari penelitian ini adalah sebagai berikut:

1. Membangun sistem prediksi kelulusan mahasiswa tepat waktu menggunakan metode *Local Mean Based K-Nearest Neighbor*.
2. Mengetahui tingkat akurasi dari sistem prediksi kelulusan mahasiswa tepat waktu menggunakan metode *Local Mean Based K-Nearest Neighbor*.
3. Mengetahui hasil perbandingan Metode *K-Nearest Neighbor* dan *Local Mean Based K-Nearest Neighbor*.

#### **1.5. Manfaat Penelitian**

Manfaat dari penelitian ini adalah sebagai berikut:

1. Memberikan informasi akurasi dan hasil prediksi kelulusan mahasiswa tepat waktu menggunakan *Local Mean Based K-Nearest Neighbor* sebagai sebagai rekomendasi untuk mengurangi mahasiswa yang lulus terlambat.
2. Dapat mendeteksi mahasiswa yang terlambat lulus dengan akurasi yang lebih baik dibanding dengan metode *K-Nearest Neighbor* konvensional.
3. Hasil penelitian dapat menjadi rujukan untuk penelitian terkait dimasa mendatang.

#### **1.6. Batasan Masalah**

Batasan masalah dalam penelitian ini sebagai berikut:

1. Data yang digunakan merupakan data mahasiswa jurusan Teknik Informatika Reguler Universitas Sriwijaya angkatan 2015 dan 2016, yang berjumlah 127.
2. Data yang digunakan adalah Indeks Prestasi Semester (IPS) 1 sampai dengan Semester 6.
3. Klasifikasi terdiri dari 2 kelas, yaitu lulus tepat waktu (1) dan lulus tidak tepat waktu (0).
4. Nilai k tetangga terdekat adalah 1 (satu) sampai 10 (sepuluh).

### **1.7. Sistematika Penulisan**

Sistematika penulisan dalam penelitian ini adalah sebagai berikut:

#### **BAB I. Pendahuluan**

Bab ini menjelaskan tentang latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian dan batasan masalah. Dasar-dasar penelitian dibahas dalam bab ini, akan menjadi dasar dan acuan pengembangan penelitian pada bab selanjutnya.

#### **BAB II. Kajian Literatur**

Bab ini dibahas landasan teori yang digunakan di dalam penelitian, termasuk di dalamnya pemrosesan pembelajaran mesin, algoritma *Local Mean Based K-Nearest Neighbor*, dan penelitian terkait.

#### **BAB III. Metodologi Penelitian**

Bab ini dijelaskan proses pengumpulan data dan perancangan pembangunan perangkat lunak. Tahapan penelitian dibahas lebih rinci berdasarkan kerangka kerja yang dibuat.

## **BAB IV. Pengembangan Perangkat Lunak**

Bab ini berkaitan dengan analisis dan rancangan pengembangan perangkat lunak. Dimulai dengan analisis kebutuhan, desain dan konstruksi perangkat lunak, dan diakhiri dengan pengujian untuk memastikan bahwa pengembangan sistem yang memenuhi persyaratan telah dibangun.

## **BAB V. Hasil dan Analisa Penelitian**

Bab ini menjelaskan hasil pengujian. Berisi tabel dengan hasil evaluasi dan analisis percobaan, serta grafik, akan berfungsi sebagai referensi untuk menarik kesimpulan dari studi.

## **BAB VI. Kesimpulan**

Pada Bab ini membahas mengenai kesimpulan dari semua uraian pada bab sebelumnya serta saran yang diambil dari hasil penelitian.

### **1.8. Kesimpulan**

Berdasarkan uraian di atas, penelitian ini akan mengembangkan perangkat lunak untuk memprediksi kelulusan tepat waktu dengan metode *Local Mean Based K-Nearest Neighbor*. Metode ini diharapkan dapat diimplementasikan dengan baik dan menghasilkan akurasi yang lebih baik.

## DAFTAR PUSTAKA

- Ahmad, A. (2017). Mengenal artificial intelligence, machine learning, neural network, dan deep learning. *J. Teknol. Indones.*, no. October, 3.
- Akbulut, Y., Sengur, A., Guo, Y., & Smarandache, F. (2017). Ns-k-nn: Neutrosophic set-based k-nearest neighbors classifier. *Symmetry*, 9(9), 179.
- Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.
- Anwar, A. (2014). A review of rup (rational unified process). *International Journal of Software Engineering (IJSE)*, 5(2), 12-19.
- Cholissodin, I., Soebroto, A. A., Hasanah, U., & Febiola, Y. I. (2020). AI, MACHINE LEARNING & DEEP LEARNING.
- Danades, A., Pratama, D., Anggraini, D., & Anggriani, D. (2016, October). Comparison of accuracy level K-nearest neighbor algorithm and support vector machine algorithm in classification water quality status. *2016 6th International Conference on System Engineering and Technology (ICSET)* (pp. 137-141). IEEE.
- Devasia, T., Vinushree, T. P., & Hegde, V. (2016, March). Prediction of students performance using Educational Data Mining. *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)* (pp. 91-95). IEEE.
- El Naqa, I., & Murphy, M. J. (2015). What is machine learning?. In *machine learning in radiation oncology* (pp. 3-11). Springer, Cham.
- Ertuğrul, Ö. F., & Tağluk, M. E. (2017). A novel version of k nearest neighbor: Dependent nearest neighbor. *Applied Soft Computing*, 55, 480-490.

- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Himawan, D. (2014). Aplikasi Data Mining Menggunakan Algoritma ID3 Untuk Mengklasifikasi Kelulusan Mahasiswa Pada Universitas Dian Nuswantoro Semarang. *Semarang: Universitas Dian Nuswantoro*
- Hulukati, W., & Djibrin, M. R. (2018). Analisis tugas perkembangan mahasiswa fakultas ilmu pendidikan universitas negeri gorontalo. *Jurnal Bikotetik (Bimbingan Dan Konseling: Teori Dan Praktik)*, 2(1), 73-80.
- Irawan, Y. (2019). Penerapan Data Mining Untuk Evaluasi Data Penjualan Menggunakan Metode Clustering Dan Algoritma Hirarki Divisive Di Perusahaan Media World Pekanbaru. *Jurnal Teknologi Informasi Universitas Lambung Mangkurat (JTIULM)*, 4(1), 13-20.
- Jawthari, M., & Stoffová, V. (2021). Predicting students' academic performance using a modified kNN algorithm. *Pollack Periodica*, 16(3), 20-26.
- Karamouzis, S. T., & Vrettos, A. (2009, March). Sensitivity analysis of neural network parameters for identifying the factors for college student success. *2009 WRI World Congress on Computer Science and Information Engineering* (Vol. 5, pp. 671-675). IEEE.
- Kurnawan, I., Marisa, F., & Purnomo, P. (2018). Implementasi Data Mining dengan Algoritma Apriori untuk Memprediksi Tingkat Kelulusan Mahasiswa. *Jurnal Teknologi dan Manajemen Informatika*, 4(1).
- Kramer, O. (2013). K-nearest neighbors. In *Dimensionality reduction with unsupervised nearest neighbors* (pp. 13-23). Springer, Berlin, Heidelberg.

- Kruchten, P. (2004). *The rational unified process: an introduction*. Addison-Wesley Professional.
- Lidya, S. K., Sitompul, O. S., & Efendi, S. (2015, March). Sentiment Analysis Pada Teks Bahasa Indonesia Menggunakan Support Vector Machine (SVM) Dan K-Nearest Neighbor (K-NN). *Seminar Nasional Teknologi Informasi dan Komunikasi* (pp. 1-8).
- Liu, B. (2011). *Web data mining: exploring hyperlinks, contents, and usage data* (Vol. 1). Berlin: springer.
- Mitani, Y., & Hamamoto, Y. (2006). A local mean-based nonparametric classifier. *Pattern Recognition Letters*, 27(10), 1151-1159.
- Nababan, A. A., & Sitompul, O. S. (2018, April). Attribute weighting based K-nearest neighbor using gain ratio. *Journal of Physics: Conference Series* (Vol. 1007, No. 1, p. 012007). IOP Publishing.
- Novianti, A. G., & Prasetyo, D. (2017, November). Penerapan Algoritma K-Nearest Neighbor (K-NN) untuk Prediksi Waktu Kelulusan Mahasiswa. In *Seminar Nasional APTIKOM (SEMNASTIKOM)* (pp. 108-113).
- Oyelade, O. J., Oladipupo, O. O., & Obagbuwa, I. C. (2010). Application of k Means Clustering algorithm for prediction of Students Academic Performance. *arXiv preprint arXiv:1002.2425*.
- Pan, Z., Wang, Y., & Ku, W. (2017). A new k-harmonic nearest neighbor classifier based on the multi-local means. *Expert Systems with Applications*, 67, 115-125.

- Papilaya, J. O., & Huliselan, N. (2016). Identifikasi gaya belajar mahasiswa. *Jurnal Psikologi*, 15(1), 56-63.
- Qian, Y., Zhou, W., Yan, J., Li, W., & Han, L. (2015). Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. *Remote Sensing*, 7(1), 153-168.
- Rohman, A. (2015). Model Algoritma K-Nearest Neighbor (K-Nn) Untuk Prediksi Kelulusan Mahasiswa. *Neo Teknika*, 1(1).
- Rohman, A., & Rochcham, M. (2019). Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Mahasiswa. *Neo Teknika*, 5(1).
- Saputra, A. I. (2021). *Penerapan Algoritma Modified K-Nearest Neighbor pada Klasifikasi Masa Studi Mahasiswa* (Doctoral dissertation, UNIVERSITAS MUHAMMADIYAH JEMBER).
- Santosa, B. (2007). Data mining teknik pemanfaatan data untuk keperluan bisnis. *Yogyakarta: Graha Ilmu*, 978(979), 756.
- Saputra, A. I., Oktavianto, H., & Al Faruq, H. A. (2021). Penerapan Algoritma Modified K-Nearest Neighbor (MKNN) Pada Klasifikasi Masa Studi Mahasiswa Teknik Informatika.
- Setiowati, Y., & Helen, A. (2018). Klasifikasi Analisis Sentimen Mengenai Hotel di Yogyakarta. *SCAN-Jurnal Teknologi Informasi dan Komunikasi*, 13(1), 1-10.
- Suhada, K., Elanda, A., & Aziz, A. (2021). Klasifikasi Predikat Tingkat Kelulusan Mahasiswa Program Studi Teknik Informatika dengan Menggunakan

- Algoritma C4. 5 (Studi Kasus: STMIK Rosma Karawang). *Dirgamaya: Jurnal Manajemen dan Sistem Informasi*, 1(2), 14-27.
- Syaliman, K. U. (2018). Peningkatan Akurasi pada Metode Klasifikasi K-Nearest Neighbor Menggunakan Local Mean Based dan Distance Weight K-Nearest Neighbor.
- Tu, L., Wei, H., & Ai, L. (2015, May). Galaxy and quasar classification based on local mean-based k-nearest neighbor method. In *2015 IEEE 5th International Conference on Electronics Information and Emergency Communication* (pp. 285-288). IEEE.
- Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437-444.
- Wibowo, I. C., Fauzan, A. C., Yustiana, M. D. P., & Qhabib, F. A. (2019). Komparasi Algoritma Naive Bayes dan Decision Tree Untuk Memprediksi Lama Studi Mahasiswa. *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, 1(2), 65-74.
- Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4, 5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, 13(1), 16-25.
- Xu, J., Zhang, Y., & Miao, D. (2020). Three-way confusion matrix for classification: A measure driven view. *Information sciences*, 507, 772-794.

