# PENERAPAN *SUPERVISED MACHINE LEARNING ALGORITHM* UNTUK PREDIKSI PERFORMA MAHASISWA

**PROPOSAL TESIS**



**Oleh :**

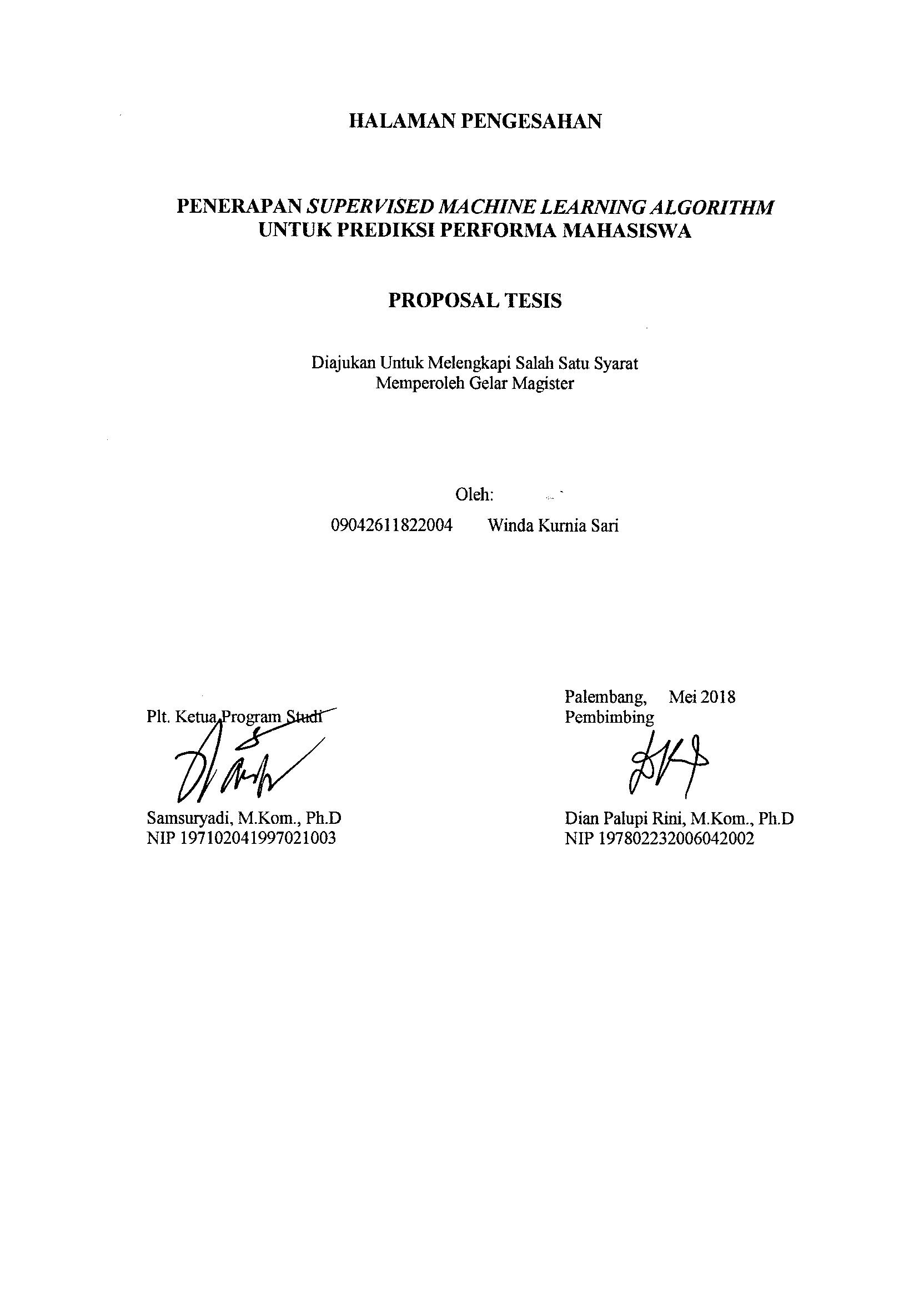
09042611822004 Winda Kurnia Sari

# MAGISTER TEKNIK INFORMATIKA

# FAKULTAS ILMU KOMPUTER

# UNIVERSITAS SRIWIJAYA

# 2018



ii

DAFTAR ISI

[**HALAMAN JUDUL** i](#_Toc507588184)

[**LEMBAR PENGESAHAN** ii](#_Toc507588185)

[**DAFTAR ISI** iii](#_Toc507588187)

[**DAFTAR GAMBAR** v](#_Toc507588188)

[**DAFTAR TABEL** v](#_Toc507588189)i

[**BAB I PENDAHULUAN**](#_Toc507588190)

[1.1 Latar Belakang 1](#_Toc507588191)

[1.2 Rumusan Masalah 2](#_Toc507588192)

[1.3 Batasan Masalah 2](#_Toc507588193)

[1.4 Tujuan Penelitian 3](#_Toc507588194)

1.5 Manfaat Penelitian 3

[1.6 Sistematika Penulisan 3](#_Toc507588196)

[**BAB II TINJAUAN PUSTAKA**](#_Toc507588197)

[2.1 Penelitian Terkait 4](#_Toc507588198)

[**BAB III METODE PENELITIAN**](#_Toc507588201)

[3.1 Metode Penelitian](#_Toc507588202) 6

[3.2 Metode Pengumpulan Data 6](#_Toc507588203)

[3.3 Metode Pengembangan Perangkat Lunak](#_Toc507588204) 7

[3.4 Algoritma Machine Learning 11](#_Toc507588205)

[3.4.1 Random Forest 11](#_Toc507588203)

[3.4.2 K-Nearest Neighbor 11](#_Toc507588203)

[3.4.3 Performa Akademik 11](#_Toc507588203)

[3.5 Performa Akademik 11](#_Toc507588202)

[3.6 Jadwal Penelitian 12](#_Toc507588204)

# [DAFTAR PUSTAKA 13](#_Toc507588201)

iii

**BAB I**

**PENDAHULUAN**

**1.1 Latar Belakang Masalah**

Peramalan kinerja mahasiswa sangat penting bagi pendidik untuk mendapatkan *feedback* awal dan mengambil tindakan segera atau tindakan pencegahan dini jika diperlukan untuk meningkatkan kinerja mahasiswa (Al-Shehri *et al*, 2017).

Teknik yang digunakan untuk mengelompokkan performa akademik mahasiswa menggunakan algoritma statistik dan agar mendapatkan hasil yang baik, ada beberapa tehnik yang dapat meningkatkan akurasi hasil tes pada metode data proses. Tehnik *Information Gain* diuji coba pada Decision Tree, Random Forest, JST, SVM dan Naïve Bayes untuk memprediksi performa akademik mahasiswa yang dapat meningkatkan performa algoritma (Rahman et al, 2017). Penelitian dengan objek performa akademik dilakukan dengan membuat monitoring performa akademik mahasiswa dengan indikator IPK menggunakan perbandingan metode Naïve Bayes dan Decision Tree (C. E. López G et al, 2014).

Penelitian prediksi performa akademik mahasiswa pernah dilakukan sebelumnya menggunakan beberapa metode klasifikasi, yaitu prediksi performa mahasiswa menggunakan SVM dan K-Nearest Neighbor dengan pemecahan masalah metode regresi (Al-Shehri *et al*, 2017), menghitung akurasi menjadi tiga kelompok Low Performa, Medium Performa dan High Performa (C. J. V. A-Arnedo et al, 2016). Penelitian berikutnya, menerapkan Algoritma *Machine Learning*  untuk memprediksi memprediksi IP kumulatif akhir dari Mata kuliah inti pada bidang tertentu setiap term (Jie Xu *et al*, 2017), melihat motivasi akademik dengan dibuat dua analisi yaitu Pre-Inovasi dan Pos-Inovasi dengan Algoritma *Machine Learning* (A. Rahal et al, 2016).

. Berdasarkan uraian diatas bahwa Algoritma *Machine Learning* digunakan dalam proses klasifikasi data di berbagai bidang. Oleh karena itu, akan diteliti pengelompokkan performa akademik mahasiswa dengan menggunakan metode klasifikasi Algoritma *Machine Learning* yaitu Random forest dan K-Nearest Neighbor dengan menghitung akurasi masing-masing.**1.2 Rumusan Masalah**

Untuk mengelompokkan performa akademik mahasiswa dibutuhkan algoritma yang sederhana, mudah diimplementasikan dan proses yang cepat. Algoritma yang digunakan dalam proses pengelompokkan performa akademik mahasiswa adalah algoritma statistik, dengan metode klasifikasi algoritma *Machine Learning* Random Forest dan K-Nearest Neighbor untuk menghitung akurasi yang dibagi menjadi tiga pengelompokkan performa mahasiswa yaitu *Low* Performa (Bakal D.O), *Medium* Performa (Semester 8 dan keatas) dan *High* Performa (6 – 7 semester). Maka dari itu perumusan masalah dalam penelitian ini adalah apakah metode klasifikasi algoritma *Machine Learning* Random Forest dan K-Nearest Neighbor dapat digunakan untuk melakukan pengelompokkan performa akademik mahasiswa dan bagaimana akurasi Random Forest dan K-Nearest Neighbor dalam mengelompokkan performa akademik mahasiswa.

**1.3 Batasan Masalah**

Adapun batasan masalah penelitian ini:

1. Melakukan pengelompokkan berdasarkan Mata kuliah wajib, pekerjaan orang tua, IP, IP Kumulatif dan SKS berjalan mahasiswa;
2. Performa akademik mahasiswa dikelompokkan dalam 3 kelompok yaitu *Low* Performa (Bakal D.O), *Medium* Performa (Semester 8 dan keatas) dan *High* Performa (6 – 7 semester);
3. Data Mata kuliah wajib, pekerjaan orang tua, IP, IP Kumulatif dan SKS berjalan yang diambil adalah data semester 1 sampai 4 mahasiswa jurusan Teknik Informatika, Sistem Informasi dan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.

**1.4 Tujuan Penelitian**

Adapun tujuan penelitian ini:

1. Mengelompokkan performa akademik mahasiswa menjadi tiga kelompok yaitu *Low* Performa (Bakal D.O), *Medium* Performa (Semester 8 dan keatas) dan *High* Performa (6 – 7 semester).
2. Membuat fitur dan label pada dataset dengan penerapan metode klasifikasi algoritma *Machine Learning*.
3. Mengetahui akurasi metode klasifikasi algoritma *Machine Learning* Random Forest dan K-Nearest Neighbor dengan melihat Mata kuliah wajib, pekerjaan orang tua, IP, IP Kumulatif dan SKS berjalan mahasiswa 2 tahun pertama.

**1.5 Manfaat Penelitian**

Manfaat penelitian ini adalah untuk membantu program studi dalam memonitor performa tinggi akademik mahasiswa untuk menghasilkan lulusan yang tercepat dan terbaik.

**1.6 Sistematika Penulisan**

Untuk mempermudah dalam penyusunan tesis dan memuat uraian secara garis besar isi dari setiap bab, maka dibuatlah sistematika penulisan:

**1. BAB I - PENDAHULUAN**

Bab ini berisi tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

**2. BAB II - TINJAUAN PUSTAKA**

Bab ini berisi tentang seluruh penjelasan mengenai tinjauan pustaka yag berhubungan dengan permsalahan yang dibahas pada penulisan tesis ini.

**3. BAB III – METODOLOGI PENELITIAN**

Bab ini berisi penjelasan secara bertahap dan terperinci tentang langkah-langkah (metodologi) yang digunakan untuk membuat kerangka berfikir dan kerangka kerja dalam meyelesaikan tesis.

**4. BAB IV – ANALISA DAN PEMBAHASAN**

Bab ini berisi tentang analisa dan pembahasan dari tiap-tiap blok diagram perencanaan rangkaian data-data hasil pengukuran.

**5. BAB V – KESIMPULAN (SEMENTARA)**

Bab ini berisi kesimpulan tentang hasil yang telah diperoleh serta merupakan jawaban dari tujuan yang ingin dicapai pada BAB I (Pendahuluan), akan tetapi masih bersifat sementara.

**BAB II**

**TINJAUAN PUSTAKA**

**2.1 Penelitian Terkait**

Terdapat beberapa penelitian mengenai metode klasifikasi Random Forest dan K-Nearest Neighbor *machine learning* yang telah dilakukan sebelumnya:

1. Penerapan *machine learning* untuk memprediksi performa mahasiswa di Perguruan Tinggi, Penelitian ini juga membangun empat basis prediksi untuk setiap kuarter implementasi dengan empat klasik Machine Learning Algoritma: Linear regression, Logistic regression, Random forest, dan K-Nearest Neighbour dengan menggunakan dataset matakuliah inti saja. Hasil penelitian ini, diantara basis prediksi yang diimplementasikan Random forest adalah yang terbaik dan KNN adalah yang terburuk (Jie Xu *et al*, 2017).
2. Algoritma *Machine Learning* juga diterapkan pada kasus peningkatan performa mahasiswa dalam matakuliah kuantitatif untuk menganalisis motivasi akademik. *Research question* pada studi ini untuk menentukan apakah inovasi yang baru dilaksanakan berdampak pada kinerja mahasiswa dengan dibuat dua analisi yaitu Pre-Inovasi dan Post Inovasi. Hasil menunjukkan kelompok post-inovasi mengungguli kelompok pra-inovasi dengan lebih banyak As (+43%), Bs (+35%) dengan Cs yang lebih sedikit (-20%) mendukung hipotesis bahwa inovasi yang disarankan meningkatkan performa mahasiswa secara signifikan (A. Rahal *et al*, 2016).
3. Prediksi performa mahasiswa dengan menggunakan *Support Vector Machine* dan *K-Nearest Neighbor* dengan metode regresi. Pada penelitian ini menerapkan keduanya SVM dan KNN pada dataset untuk memprediksi nilai mahasiswa dan selanjutnya dibandingkan akurasinya. Hasilnya, SVM mendapatkan hasil yang cukup baik dengan korelasi koefisien dengan 0.96, sementara KNN mendapatkan korelasi koefisien dengan 0.95 (Al-Shehri *et al*, 2017).
4. *Random forest classifier* untuk menilai dan memprediksi belajar mahasiswa (SETAP). Penelitian ini menggunakan Random Forest metode Machine Learning untuk memprediksi efektifitas kerja tim RPL berdasarkan data yang dikumpulkan selama pengembangan proyek tim mahasiswa. Penelitian ini juga menunjukkan hasil dari analisi Random Forest untuk full dataset. Hasilnya SETAP dapat mengenali tim mahasiswa yang pasti gagal atau membutuhkan pengawasan pada awal kelas dengan akurasi yang baik yaitu sekitar 70% (Petkovic *et al*, 2016).
5. *Robust Random Forest-based Tri-Training Algorithm* pada prediksi awal mahasiswa dalam masalah. Tugas prediksi pada penelitian ini diperiksa dalam konteks yang lebih praktis di mana kekurangan data dipertimbangkan. Oleh karena itu, algoritma yang diusulkan telah memasukkan metode imputasi data yang tidak lengkap berdasarkan *k* tetangga terdekat dalam proses pembelajaran semi-supervised yang iteratif. Diukur akurasi dengan persentase berbeda dari beberapa metode dari tidak berlabel data dalam dataset 2 dan 3 tahun. Sebagai hasilnya, algoritma *Robust Tri-training (Random Forest)* mengungguli yang sudah ada dengan pendekatan preprocessing tradisional untuk tugas (V.T.N Chau et al, 2017).
6. Teknik evaluasi *Machine Learning* dalam tugas prediksi reprovasi siswa SMA Teknik. Dalam penelitian ini, studi kasus dan hasil yang diperoleh menunjukkan bahwa penggunaan *Machine Learning* lebih spesifik dengan pembelajaran *supervised* klasifikasi, sangat berguna untuk pembuatan alat bantu yang dapat memantau dan memprediksi kinerja siswa. Terdapat empat metode perbandingan yaitu Naïve Bayes, SMO, SimpleCart, OneR. Ini juga menunjukkan bahwa alat yang dikembangkan menggunakan algoritma yang didasarkan pada metode *tree-based* dapat membuat prediksi dari 89% hingga 94% akurasi (G.D.M Junior et al, 2017).
7. Deteksi awal mahasiswa *at-risk* menggunakan *Machine Learning* dalam LMS Log Data. Penelitian ini memeriksa teknik *machine learning* khas untuk mahasiswa berdasarkan data log aktual LMS dan menyelidiki kinerja mereka. Random Forest menunjukkan perilaku yang paling stabil dan keseimbangan presisi dan recall yang baik. Model ini dapat mendeteksi sekitar 40% siswa berisiko pada akhir minggu ketiga semester pertama dengan hanya data log LMS (N. Kondo et al, 2017).

BAB III

METODOLOGI PENELITIAN

**3.1 Metode Penelitian**

Langkah-langkah dalam penelitian ini, adalah sebagai berikut:

1. Studi literature mengenai metode klasifikasi Algoritma Machine Learning Random Forest dan K-Nearest Neighbor;
2. Menganalisis penelitian yang telah ada sebelumnya;
3. Mengumpulkan data untuk penelitian dengan cara mengumpulkan Data Matakuliah wajib, pekerjaan orang tua, IP, IP Kumulatif dan SKS berjalan mahasiswa yang didapat dari bagian Administrasi Akademik Fasilkom;
4. Mengubah data yang telah dikumpulkan sehingga dapat memenuhi kebutuhan penelitian;
5. Memberi fitur dan label data;
6. Melakukan perhitungan akurasi metode Random Forest dan K-Nearest Neighbor;
7. Melakukan analisis dan pembahasan hasil penelitian;
8. Membuat kesimpulan terhadap hasil penelitian.

**3.2 Metode Pengumpulan Data**

3.2.1 Sumber Data

Penelitian ini menggunakan data berupa Matakuliah wajib, pekerjaan orang tua, IP, IP Kumulatif dan SKS berjalan yang didapatkan dari bagian administrasi akademik Fakultas Ilmu Komputer Universitas Sriwijaya.

* + 1. Jenis Data

Data yang digunakan pada penelitian ini adalah data primer yang berupa Matakuliah wajib, pekerjaan orang tua, IP, IP Kumulatif dan SKS berjalan mahasiswa. Data tersebut akan diberi fitur dan label numerik dan dalam format numpy array.

* + 1. Teknik Pengumpulan Data

Data penelitian ini dikumpulkan dengan cara mengajukan permohonan permintaan data Matakuliah wajib, pekerjaan orang tua, IP, IP Kumulatif dan SKS berjalan mahasiswa ke bagian Akademik Fakultas Ilmu Komputer Universitas Sriwiya.

**3.3 Metode Pengembangan Perangkat Lunak**

Metode yang diterapkan dalam pengembangan perangkat lunak ini berorientasi objek dengan menggunakan *Rational Unified Process* (RUP). Tabel I-1 mendeskripsikan hubungan alur kerja dan fase RUP yang dilakukan pada penelitian ini.

Tabel 1. Hubungan alur kerja dan fase RUP

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Alur Kerja | Fase | | | |
| Insepsi | Elaborasi | Konstruksi | Transisi |
| 1 | Pemodelan Bisnis | * Menganalisis deskripsi umum sistem; * Menentukan aktor yang terlibat yaitu user; * Membuat daftar use case; * Membuat skenario dari masing-masing use case tersebut; * Fitur utama sistem adalah pengujian akurasi terhadap metode RF dan KNN. | * Mengembangkan atau menambah use case jika ada; * Memperbaiki skenario use case sesuai fitur yang tersedia. | Membuat use case menggunakan Astah. | Membuat dokumentasi use case beserta skenarionya. |
| 2 | Kebutuhan | * Merumuskan kebutuhan perangkat lunak; * Mengumpulkan data matakuliah, pekerjaan orang tua, IP, IP Kumulatif dan sks berjalan mahasiswa Fakultas Ilmu Komputer yang berasal dari bagian adminitrasi akademik Fakultas Ilmu Komputer, Universitas Sriwijaya; * Melakukan persiapan hardware Laptop yang digunakan * Melakukan pencarian jurnal-jurnal yang berhubungan dengan metode klasifikasi algoritma machine learning |  | Melakukan proses prapengolahan yaitu seleksi data mahasiswa nonaktif, membagi fitur dan melabeli data. 2. Data yang telah di Fitur dan di Label-i harus numerik; 3. Fitur dan Label harus format numpy array; 4. Fitur dan label harus dalam bentuk matriks yang spesifik. | Membuat dokumentasi analisis kebutuhan perangkat lunak. |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 3 | Analisis dan Desain | * Melakukan studi analisis mengenai metode klasifikasi Random Forest dan K-Nearest Neighbor Maching Learning; * Melakukan pembuatan arsitektur sistem yaitu use case, skenario use case, kelas analisis, kelas diagram, dan diagram sekuens berdasarkan fitur yang telah dirancang. | * Melakukan perbaikan dan pengecekan kembali kesesuaian use case, skenario use case, kelas analisis, kelas diagram, dan diagram sekuens yang telah dibuat; * Selanjutnya menentukan keterhubungan antar kelas pada kelas diagram. | Membuat diagram UML yaitu kelas analisis, kelas diagram, dan diagram sekuens dengan menggunakan Astah. | Membuat dokumentasi diagram UML yaitu kelas analisis, kelas diagram, dan diagram sekuens. |
| 4 | Implementasi | Melakukan pemodelan untuk implementasi metode klasifikasi Random Forest dan K-Nearest Neighbor Maching Learning | Studi perencanaan integrasi metode metode klasifikasi Random Forest dan K-Nearest Neighbor Maching Learning untuk pengelompokkan performa akademik mahasiswa. | Mengimplementasikan hasil analisis dan desain ke dalam bahasa pemrograman (coding) menggunakan Python | Melakukan penyempurnaan pemrograman sampai pada tahap akhir dan pengintegrasian masing-masing subsistem yang telah di-coding. |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 5 | Pengujian | Melakukan perencanaan pengujian sistem pengelompokan performa akademik mahasiswa menggunakan data primer dari yang berasal dari bagian adminitrasi akademik Fakultas Ilmu Komputer, Universitas Sriwijaya; | Melakukan pengujian sistem pengelompokan performa akademik mahasiswa menggunakan fitur dan label dataset yang sudah diolah | Melakukan pengujian tingkat lanjut dengan memperhitungkan tingkat akurasi sistem pengelompokan performa akademik mahasiswa. | * Membuat dokumentasi terhadap evaluasi pengujian terkait kekurangan sistem pengelompokkan performa akademik mahasiswa. * Membuat kesimpulan. |

**3.4 Algoritma Machine Learning**

**3.4.1 Random Forest**

Leo Brejman yang memperkenalkan ide algoritma Random Forest. Algoritma Random Forest dapat diterapkan dalam Regresi dan klasifikasi. Dasar pembuatan blok Random Forest adalah Decision Tree. Pada Random Forest, setiap decision tree dilatih menggunakan subset data *tranining* dan persamaan yang dipilih secara random , pemilihan subset fitur secara random digunakan untuk latihan setiap decision tree. Selama proses tes, setiap decision tree memproses vektor input fitur dan membuat keputusannya sendiri. Pada akhirnya skema voting mayoritas digunakan untuk menggabungkan semua hasil dari decision tree untuk menghasilkan keputusan akhir (Manzoor et al, 2018). Klasifikasi Random Forest berkinerja cepat bahkan dengan dataset yang besar karena pemilihan acak dari subset data training dan dapat mengatasi kehilangan data dalam dataset dan menghasilkan tingkat pengenalan yang lebih tinggi (V. F. Rodriguez-Galiano et al, 2013).

**3.4.2 K-Nearest Neighbor**

Metode K-Nearest Neighbor adalah sebuah contoh *classifier* yang membandingkan contoh-contoh baru dengan contoh-contoh yang terlabel pada set latihan. Sebagai algoritma pembelajaran yang malas, alih-alih membuat model klasifikasi eksplicit, KNN menolak semua perhitungan sampai fase klasifikasi. Aturan pengklasifikasiannya berdasarkan pada pengambilan *k*  yang paling mirip contoh pada set latihan untuk contoh yang tidak diketahui. Walaupun demikian, algoritma KNN memberikan kepentingan yang sama untuk semua tetangga, berasumsi bahwa batas-batas antara kelas-kelas didefinisikan dengan sempurna, yang bearti tidak semua benar (Maillo et al, 2017).

**3.5 Performa Akademik**

Performa akademik merupakan peran penting dalam menghasilkan lulusan berkualitas terbaik yang akan menjadi pemimpin besar dan tenaga kerja untuk negara sehingga bertanggung jawab untuk pembangunan ekonomi dan social negara.

Pengukuran performa akademik mahasiswa telah menerima banyak perhatian dalam penelitian sebelumnya, itu adalah aspek menantang literature akademik, dan kinerja mahasiswa ilmu dipengaruhi karena faktor sosial, psikologis, ekonomi, lingkungan dan pribadi (W. W. F. Lau et al, 2017). Faktor-faktor diatas dapat mempengaruhi performa mahasiswa, akan tetapi tiap orang di negara berbeda memiliki faktor yang bervariasi. Performa akademik akan dikelompokkan menjadi 3 kelompok klasifikasi.

Tabel 2. Tabel Klasifikasi Performa Akademik

|  |  |
| --- | --- |
| **Masa Studi** | **Klasifikasi** |
| 6-7 Semester | High Performa |
| 8-12 Semester | Medium Performa |
| Bakal *Drop Out* | Low Performa |

**3.6 Jadwal Penelitian**

Berikut adalah jadwal penelitian yang akan dilakukan pada penelitian ini:

Tabel 3. Jadwal penelitian

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No | Kegiatan | Jan 2019 | | | Feb 2019 | | | | Mar 2019 | | | | Apr 2019 | | | |
| 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 1 | Pemodelan Bisnis |  | | |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | Kebutuhan |  |  |  | | | | |  |  |  |  |  |  |  |  |
| 3 | Analisis dan Desain |  |  |  |  |  | | | | | | |  |  |  |  |
| 4 | Implementasi |  |  |  |  |  |  |  | | | | | |  |  |  |
| 5 | Pengujian |  |  |  |  |  |  |  |  |  |  |  |  | | | |

**DAFTAR PUSTAKA**

A. Rahal, M. Zainuba, Improving students' performance in quantitative courses: The case of academic motivation and predictive analytics, The International Journal of Management Education, pp. 1-10, 2016 Elsevier.

C. E. López G., E. L Guzmán, F. A. González, A Model to Predict Low Academic Performance at a Specific Enrollment Using Data Mining, LACCEI special IEEE-RITA Publication, pp. 1-8, 2015.

C. J. V. A-Arnedo, F. J. Gallego-Dur an, F. O. Llorens-Largo, P. Compan-Rosique, R. Satorre-Cuerda, R. Molina-Carmona, Improving the expressiveness of black-box models for predicting student performance, Computers in Human Behavior, pp. 1-11, 2016 Elsevier.

D. Petkovic, M. Sosnick-Perez, K. Okada, R. Todtenhoefer, S. Huang, Using the Random Forest Classifier to Assess and Predict Student Learning of Software Engineering Teamwork, IEEE, pp 1-7, 2016.

G. D. M. Junior, S. M. Oliveira, C. C. Ferreira, Evaluation techniques of Machine Learning in task of reprovation prediction of Technical High School students, IEEE, pp. 1-7, 2017.

H. Al-Shehri, A. Al-Qarni, L. Al-Saati, A. Batoaq, H. Badukhen, S.Alrashed, J. Alhiyafi, S. O. Olatunji, Student Performance Prediction Using Support Vector Machine and K-Nearest Neighbor, Canadian Conference on Electrical and Computer Engineering, pp. 1-4, 2017.

Jie Xu, K.H Moon, M. van der Schaar, A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs, Journal of Selected Topics in Signal Processing, pp. 1-12, 2016.

J. Maillo, J. Luengo, S. Garcıa, F. Herrera, I. Triguero, Exact Fuzzy k-Nearest Neighbor Classification for Big Datasets, IEEE, pp. 1-7, 2017.

L. Rahman, N. A. Setiawan, A. E. Permanasari, Feature Selection Methods in Improving Accuracy of Classifying Students’ Academic Performance, International Conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), pp. 1-5, 2017.

N. Kondo, M. Okubo, T. Hatanaka, Early Detection of At-Risk Students Using Machine Learning Based on LMS Log Data, IIAI International Congress on Advanced Applied Informatics, pp. 1-4, 2017

V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, J. P. Rigol-Sanchez, An assessment of the effectiveness of a random forest classifier for land cover classification, ISPRS Journal of Photogrammetry and Remote Sensing, 67, pp.93-104, 2013.

V. T. N. Chau, N. H. Phung, A Robust Random Forest-based Tri-Training Algorithm for Early In-trouble Student Prediction, NAFOSTED Conference on Information and Computer Science, pp. 1-6, 2017.

W. W. F. Lau, Effects of social media usage and social media multitasking on the academic performance of university students, Computers in Human Behavior, pp. 1-6, 2017 Elsevier.