**MODEL ALGORITMA K-NEAREST NEIGHBOR (K-NN) DAN NAÏVE BAYES UNTUK PREDIKSI KELULUSAN MAHASIWA.**

Ryan Situmorang, Woro Isti Rahayu, Rd Nuraini Siti Fatonah

Universitas Logistik Bisnis Internasional; Jl. Sariasih No.54 Bandung, 4015. Telp.(022) 2009562, 2009570, Fax.(022) 2009568, 2011089 Program Studi D4 Teknik Informatika

[ryanface290300@gmail.com](mailto:ryanface290300@gmail.com) , [woroisti@ulbi.ac.id](mailto:woroisti@ulbi.ac.id), [nurainisf@ulbi.ac.id](mailto:nurainisf@ulbi.ac.id)

***Abstract***

In the education system students are an important asset for an educational institution and for that it is necessary to pay attention to the graduation rate of students on time. The percentage increase in students' ability to complete their studies on time is one element of the university's accreditation assessment. For this reason, it is necessary to monitor and evaluate student graduation using data mining classification. By processing student data, students will get important things for higher education purposes, especially for student graduation. In this study, students graduated using the K-Nearest Neighbor data mining classification algorithm by clustering the data k=1, k=2, k=3, k=4, and k=5. The results obtained with data cluster k=5 accuracy is 85.15% and the AUC value is 0.888 which is the highest accuracy.

***Keywords: Student, K-Nearest Neighbor, Naïve Bayes***

**Abstrak**

Dalam sistem pendidikan mahasiswa adalah aset penting bagi sebuah institusi pendidikan dan untuk itu perlu diperhatikan tingkat kelulusan mahasiswa tepat pada waktunya. Presentase naik turunnya kemampuan mahasiswa untuk menyelesaikan studi tepat waktu merupakan salah satu elemen penilaian akreditasi universitas. Untuk itu perlu adalah pemantauan dan evaluasi terhadap kelulusan mahasiswa dengan menggunakan klasifikasi data mining. Dengan mengolah data mahasiswa akan memperoleh hal penting untuk keperluan perguruaan tinggi terutama untuk prediksi kelulusan mahasiswa. Dalam penelitian ini dalam memprediksi kelulusan mahasiswa dengan menggunakan algoritma klasifikasi data mining K-Nearest Neighbor dengan mengklaster data k=1, k=2, k=3, k=4, dan k=5. Hasil yang diperoleh dengan cluster data k=5 accuracy adalah 85,15% dan nilai AUC adalah 0.888 adalah akurasi paling tinggi.

**Kata Kunci: Mahasiswa, *K-Nearest Neighbor, Naïve Bayes***

1. **Pendahuluan**

Lembaga pendidikan tertinggi adalah perguruan tinggi yang merupakan penyelenggara pendidikan akademik bagi mahasiswa. Mahasiswa sering disebut kelompok masyarakat yang memiliki ciri intelektualitas yang lebih luas dibandingkan dengan kelompok seusia mereka yang bukan mahasiswa ataupun kelompok usia lain yang dibawah mereka. Dengan intelektualitasnya mahasiswa akan mampu menghadapi dan mencari permasalahan secara sistematis yang nantinya diterapkan dalam kehidupan sehari-hari agar bisa bersaing dalam dunia kerja (Azwar, 2004).

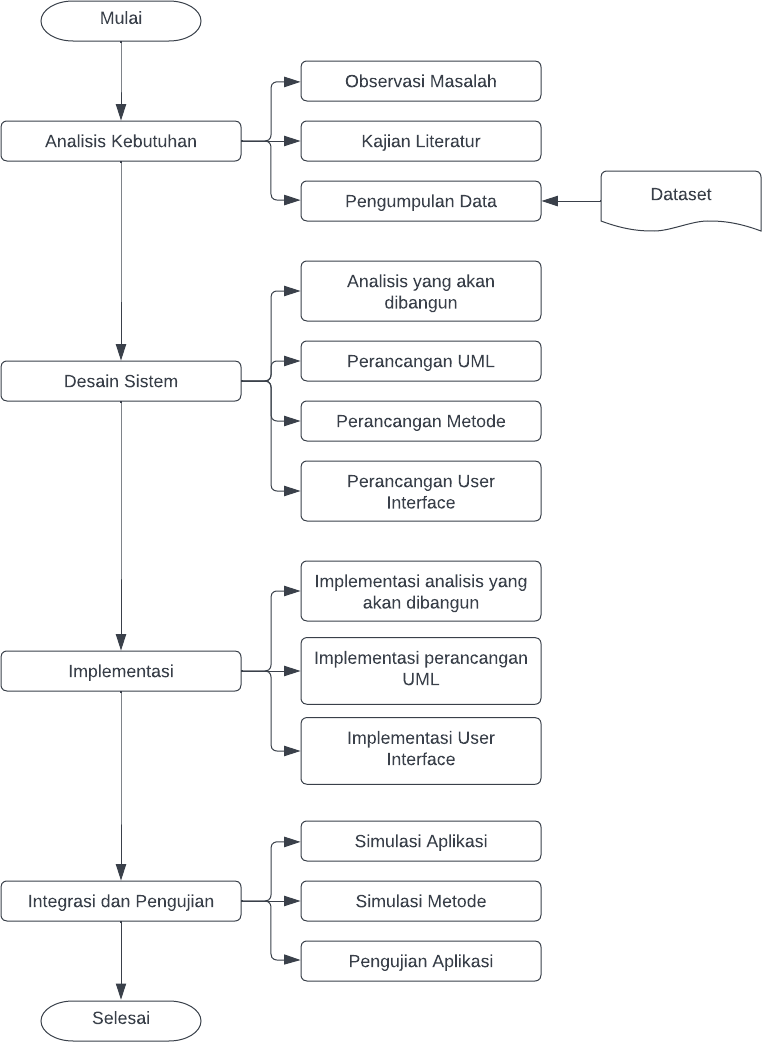
Perguruan tinggi dituntut untuk menyelenggarakan pendidikan yang berkualitas bagi mahasiswa sehingga menghasilkan sumber daya manusia yang berilmu, cakap, kreatif dan bersaing. Dalam sistem pendidikan mahasiswa adalah aset penting bagi sebuah institusi pendidikan dan untuk itu perlu diperhatikan tingkat kelulusan mahasiswa tepat pada waktunya. Prosentase naik turunnya kemampuan mahasiswa untuk menyelesaikan studi tepat waktu merupakan salah satu elemen penilaian akreditasi Prodi D-4 Tekini Informatika. Untuk itu perlu adanya pemantauan maupun evaluasi terhadapat kecenderungan mahasiswa lulus tepat waktu atau tidak. Pemantauan atau evaluasi terhadap kinerja akan menghasilkan informasi yang bermanfaat untuk membantu mahasiswa, dosen, administrator, dan pembuat kebijakan (Ogor, 2007).

Dengan demikian jelas bahwa memprediksi kelulusan mahasiswa adalah hal yang penting bagi penyelenggara pendidikan untuk menentukan strategi bagi institusinya. Suatu perguruan tinggi menyimpan data akademik, administrasi, biodata mahasiswa dan lain-lain. Data tersebut akan sangat bermakna jika digali dengan tepat sehingga dapat diketahui pola atau pengetahuan untuk mengambil keputusan (Ogor, 2007).

Dengan menerapkan data mining dapat memecahkan masalah dengan menganalisis data yang dimiliki perguruan tinggi. Dalam mengolah data mahasiswa untuk prediksi telah diselesaikan telah oleh Karamouiz dan Vrettos dengan menggunakan metode neural network (Vrettos, 2009), Qudri dan Kalyankar dengan metode decision tree (Kalyankar, 2010), Suhartini dan Ernastuti dengan metode C4.5 dan naïve bayes, (Ernastuti, 2010)

Hastuti dengan komparasi metode Logistic Regression, Decision Tree, Naïve Bayes, Neural Network (Hastuti, 2012) dan Tahyudin, Utami dan Amborowati dengan mengkomparasi algoritma decision tree, naïve bayes, ANN, Support Vector Machine (SVM) dan Logistic Regression (LR) (Tahyudin, 2013).

Tetapi belum ada yang menggunakan metode klasifikasi data mining lainnya seperti K-Nearest Neighbour, ID3, CHAID dan Learning Discriminant Analysis (LDA). Untuk itu dalam penelitian ini akan dilakukan prediksi kelulusan mahasiswa dengan metode K-Nearest Neighbour.

1. **Metode Penelitian**

**Gambar 1 Alur Metodologi Penelitian**

**2.1 Analisis kebutuhan yang digunakan**

Dalam tahap ini, dilakukan analisis kebutuhan yang diperlukan seperti mencari data. Penelitian ini mengumpulkan data dengan cara sebagai berikut :

1. Observasi

Dalam tahap ini, penulis perlu melakukan analisis masalah yang ada dalam penelitian seperti melakukan *research* terhadap identifikasi masalah, tujuan penelitian, serta manfaat penelitian. Dalam tahap identifikasi masalah dilakukan dengan melihat kondisi dan situasi saat ini yang menjadi topik masalah dalam penelitian ini. Dalam tahap tujuan penelitian dan manfaat penelitian dilakukan dengan mencari tahu mengenai hal-hal yang ingin dicapai dalam penelitian ini dan manfaat yang dapat dirasakan oleh peneliti maupun orang lain.

1. Kajian literatur

Dalam tahap ini akan melakukan kajian pustaka yang berkaitan dengan topik.

1. Pengumpulan data

Dalam tahap ini akan dilakukan pengambilan data Mahasiswa yang langsung dari Prodi D-4 Teknik Informatika.

* 1. **Desain System**

Desain sistem dalam penelitian ini, tidak hanya tampilan saja tetapi yang dimaksud desain dalam metode ini adalah desain sistem yang meliputi:

1. Analisis proses yang akan dibangun

Analisis yang akan dibangun merupakan aplikasi gambaran alur yang nantinya akan ada pada aplikasi prediksi kelulusan Mahasiswa. Dalam Aplikasi ini juga dapat melihat perbandingan nilai akurasi dari metode *K-Nearest Neighbor* dan *Naive Bayes.*

1. Perancangan *Unified Modelling Language* (UML)

Umumnya dibuat pada saat akan membangun aplikasi. Dalam penelitian ini, perancangan UML dibutuhkan untuk mengetahui ruang lingkup yang akan dibangun dan juga proses bisnis secara keseluruhan. Perancangan *Unified Modelling Languange* yang akan dibuat ini menggunakan *object oriented Unified Modelling Language* (UML) yang terdiri dari *use case diagram* dan *activity diagram*.

1. Perancangan Algoritma K-NN dan *Naive Bayes*

Dalam aplikasi yang akan dibangun terdapat metode,K-NN dan *Naive Bayes* untuk menyelesaikan permasalahan yang telah dianalisis pada sistem yang diusulkan.

**2.3 Proses Implementasi**

Proses implementasi ini berisi simulasi metode K-NN dan *Naive Bayes* untuk mengetahui apakah metode yang digunakan dapat diimplementasikan atau tidak dalam aplikasi yang akan dibangun. Imeplementasi metode ini akan diterapkan dengan menambahkan algoritma metode yang digunakan dalam aplikasi yang akan dibuat.

**2.4 Integrasi dan Pengujian**

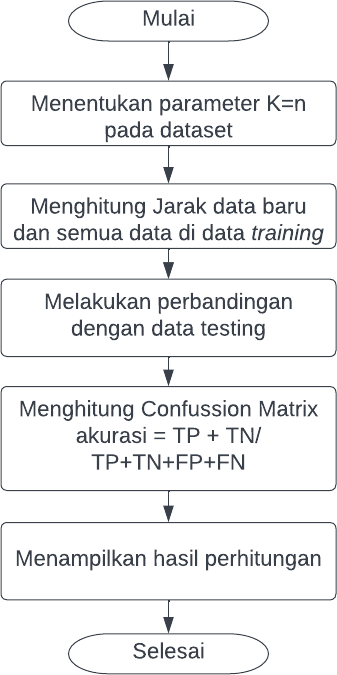
Tahap ini mengintegrasi sistem secara keseluruhan setelah seluruh untuk atau modul yang dikembangkan dan diuji ditahap implementasi. Apabila proses integrasi selesai, selanjutnya akan dilakukan pemeriksaan dan pengujian sistem secara keseluruhan untuk mengidentifikasi kemungkinan adanya kegagalan dan kesalahan sistem. Dimana pengujian berdasarkan spesifikasi, kebenaran perangkat lunak yang diuji hanya dilihat berdasarkan keluaran yang dihasilkan dari data atau kondisi masukan yang diberikan untuk fungsi yang ada tanpa melihat bagaimana proses untuk mendapatkan keluaran tersebut.

Dalam tahap ini, penulis melakukan testing pada aplikasi yang telah dibuat untuk menguji apakah sistem telah berjalan sesuai dengan yang diinginkan.

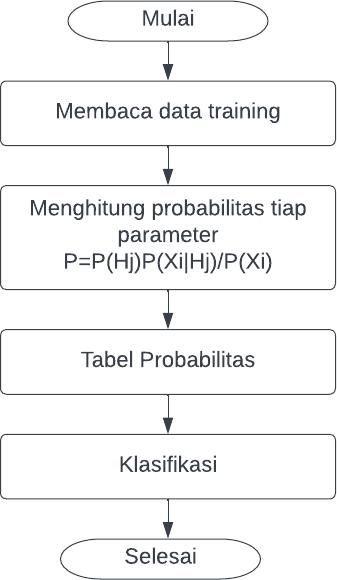
**2.5 Penerapan Metode Yang Digunakan**

Dalam tahap ini berisi simulasi algoritma yang digunakan dalam penelitian untuk mengetahui apakah algoritma yang digunakan dapat diimplementasikan dengan baik atau tidak. Algoritma yang digunakan adalah algoritma, K-NN dan *Naive Bayes*, lalu dilanjutkan klasifikasi data dan perbandingan nilai akurasi menggunakan algoritma K-NN dan algoritma *Naive Bayes.*

1. Algoritma *K-Nearest Neighbor*.



Dalam penelitian ini, penerapan metode K-NN dilakukan dengan menentukan nilai parameter K. Langkah selanjutnya yaitu menghitung jarak data baru dan semua data yang ada dalam data *training*. Setelah itu dilakuan perbandingan dengan data *testing* dan ditentukan *confussion matrix*. Setelah itu aplikasi akan menampilkan data yang telah di prediksi.

1. Algoritma Naïve Bayes

Dalam proses penerapan metode *Naïve Bayes* dengan membaca *training*  terlebih dahulu. Setelah itu dilakukan perhitungan probabilitas tiap parameter, lalu dihitung juga perhitungan. Hasilnya akan dibandingkan dengan metode KNN