

1. Pendahuluan

Latar Belakang

Universitas umumnya dituntut untuk menghasilkan lulusan yang baik dan memiliki daya saing di dunia kerja. Namun pada kenyataannya, terdapat banyak mahasiswa gagal memenuhi *learning outcome* dari mata kuliah yang diambil dan mengakibatkan mahasiswa tersebut harus mengulang mata kuliah. Adapun umumnya mahasiswa yang mengulang berasal dari jurusan teknik [1].

Terdapat berbagai faktor yang mengakibatkan mahasiswa mengulang mata kuliah, seperti faktor motivasi untuk lulus, keluarga, kesulitan di mata kuliah yang berhubungan matematika, dan sebagainya [2,3]. Mengulang mata kuliah umumnya dapat memberikan dampak pada akademik mahasiswa, seperti tidak lulus tepat waktu, atau bahkan mahasiswa tersebut harus keluar dari universitas tempat mahasiswa belajar [4,5]. Mahasiswa yang memiliki probabilitas tinggi mengalami kegagalan akademik ini disebut sebagai *at-risk student* [6].

Penelitian-penelitian untuk memprediksi *at-risk student* telah dilakukan beberapa kali sebelumnya dengan metode-metode machine learning, parameter, dan tujuan yang berbeda-beda. N. Güner, et.al. [1], misalnya, mengimplementasikan *machine learning* untuk memprediksi mahasiswa yang akademiknya akan bermasalah di fakultas teknik Universitas Pamukkale menggunakan metode *Support Vector Classification (SVC)*, *Least-Square Support Vector Classification (LSSVC)*, dan *Radial Basis Function Neural Networks (RBFNN)*. Dari pengujian yang dilakukan, didapatkan hasil RBFNN memiliki akurasi tertinggi dengan persentase 77.17% di *train*, 78% di *validation*, dan 77.60% di *test*.

Riset lainnya [4] menggunakan metode *Random Forest*, *AdaBoost*, *Logistic Regression*, *Support Vector Machine (SVM)*, dan *Decision Trees* untuk mengidentifikasi mahasiswa yang tidak akan lulus tepat waktu dari sekitar 200.000 siswa sekolah di Amerika Serikat. Hasil uji coba menunjukkan metode *Random Forest* memberikan hasil terbaik dengan nilai dari *identification before off-track* = 0.8, yang berarti jika ada 100 siswa yang gagal lulus tepat waktu, dengan metode *random forest* dapat dihasilkan prediksi bahwa 80 orang yang akan gagal lulus tepat waktu.

Machine learning juga digunakan untuk mendeteksi *at-risk student* pada mahasiswa tahun pertama di jurusan teknik elektro di Eindhoven University of Technology [7]. Penelitian tersebut menggunakan metode *Decision Trees*, *CART (SimpleCart)* dan *C4.5 (J48)*, *Bayesian classifier (BayesNet)*, *Logistic Model (SimpleLogistic)*, *Rule-Based Learner (JRip)* dan *the Random Forest*. Hasil uji coba menunjukkan metode *Decision Trees* memberikan hasil yang paling baik dengan akurasi 75% dan 80%.

Riset lainnya [8] memprediksi kegagalan pada pembelajaran jarak jauh dengan menerapkan teknik pembelajaran aktif. Penelitian ini dilakukan dengan menggunakan metode *Naïve Bayes*, *Bayes Net*, *J48*, *Logistic Regression*, *Multilayer Perceptrons (MLP)*, *Random Forest* dan *Sequential Minimal Optimization (SVM's representative)*. Hasil pengujian yang dilakukan menunjukkan *Naïve Bayes* memperoleh akurasi tertinggi dengan jarak antar 66.29% hingga 84.56%.

Prediksi kegagalan pada pembelajaran jarak jauh juga diteliti oleh S. B. Kotsiantis, et al [9]. Penelitian dilakukan di Hellenic Open University dengan menggunakan metode *C45*, *Back Propagation (Artificial Neural Network (ANN)'s representative)*, *3-NN*, *Naive Bayes*, *Sequential Minimal Optimization*, *Maximum Likelihood Estimation*. Dari pengujian yang dilakukan, metode *Naïve Bayes* dianggap metode paling sesuai dengan akurasi 63% di awal dan melebihi 83% sebelum pertengahan periode akademik. Selain akurasi yang lebih baik dibanding metode-metode lainnya, waktu training lebih singkat, sehingga *Naïve Bayes* menjadi metode paling baik untuk digunakan dalam prediksi kegagalan pada pembelajaran jarak jauh siswa.

Topik dan Batasannya

Dari penelitian-penelitian sebelumnya didapatkan metode-metode yang memberikan hasil terbaik, yaitu RBFNN [4], *Decision Trees* [7], dan *Naïve Bayes* [8,9]. Penelitian-penelitian tersebut sangat bergantung pada prestasi dan kemampuan akademik, yaitu nilai matematika, nilai bahasa Inggris, nilai ilmu pengetahuan sosial, nilai ilmu pengetahuan alam, literasi komputer, GPA tahun pertama kuliah, GPA kumulatif, jumlah jawaban benar di ujian matematika-2 UEE, nilai ujian lisan, dan ranking kelulusan sekolah.

Di sisi lain, berdasarkan riset [1,2,5], faktor-faktor yang berpengaruh pada kesuksesan akademik, selain kemampuan akademik adalah motivasi belajar, kesesuaian minat dengan jurusan kuliah yang diambil, sosial, psikologi, ekonomi, dan keluarga. Faktor-faktor selain kemampuan akademik ini disebut faktor non-akademik.

Pada tugas akhir ini telah dilakukan studi mengenai prediksi *at-risk student* dengan metode *Naïve Bayes* untuk mengetahui pengaruh faktor non-akademik terhadap hasil prediksi. Adapun *at-risk student* yang dimaksud pada tugas akhir ini adalah mahasiswa yang memiliki kemungkinan tinggi akan mengalami kegagalan pada suatu mata kuliah yang menyebabkan mahasiswa harus mengulang mata kuliah tersebut.

Pertanyaan yang dijawab pada tugas akhir ini adalah:

1. Bagaimana pengaruh faktor non-akademik pada hasil klasifikasi prediksi *at-risk student*?
2. Faktor atau parameter non-akademik apa saja yang memberikan hasil prediksi *at-risk student* terbaik ketika digabungkan dengan parameter akademik?

3. Berapakah akurasi tertinggi yang bisa didapat dengan menggunakan metode *Naïve Bayes* pada klasifikasi dengan menggunakan kombinasi parameter non-akademik dan seluruh parameter akademik?

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. *At-risk student* pada tugas akhir ini adalah mahasiswa yang berkemungkinan tinggi mengalami kegagalan pada mata kuliah Dasar Algoritma dan Pemrograman (DAP) dengan indeks nilai E atau nilai akhir kurang dari sama dengan 40 pada mata kuliah tersebut, sehingga mahasiswa tersebut harus mengambil kembali mata kuliah tersebut.
2. *Data set* merupakan data yang diambil dari 110 mahasiswa yang sedang mengambil mata kuliah DAP di kelas IF-42-01, IF-42-04, dan IF-42-11 pada jurusan S1 Informatika Universitas Telkom.

Tujuan

Tujuan yang ingin dicapai dari tugas akhir ini adalah menganalisa pengaruh faktor non-akademik terhadap hasil prediksi *at-risk student*. Selain itu, tujuan lainnya adalah mencari faktor atau parameter non-akademik yang memberikan akurasi prediksi *at-risk student* terbaik ketika dikombinasikan dengan parameter akademik.

Organisasi Tulisan

Penulisan tugas akhir ini tersusun dalam beberapa bagian, antara lain sebagai berikut:

1. Pendahuluan

Menjelaskan mengenai latar belakang, rumusan masalah dan tujuan dari topik yang diambil.

2. Studi Terkait

Berisi tinjauan pustaka dan penelitian yang pernah dilakukan sebelumnya terkait dengan topik yang diambil sebagai literatur dan acuan dalam membuat tugas akhir.

3. Sistem yang Dibangun

Menjelaskan dan menggambarkan alur kerja sistem, dimulai dari penginputan data mahasiswa DAP hingga didapat akurasi ketika menggunakan kombinasi parameter terbaik.

4. Evaluasi

Menjelaskan mengenai pengujian yang dilakukan, melakukan perbandingan akurasi prediksi *at-risk student* ketika menggunakan kombinasi parameter terbaik, seluruh parameter, dan parameter akademik, analisis terhadap hasil yang dicapai dan menjelaskan skenario yang dibangun untuk melakukan pengujian.

5. Kesimpulan dan Saran

Memberikan kesimpulan mengenai penelitian yang telah dilakukan beserta saran untuk penelitian lebih lanjut.