

**PENERAPAN DATA MINING UNTUK MEMPREDIKSI  
KETERLAMBATAN PENYELESAIAN STUDI MAHASISWA S1  
UNIVERSITAS PENDIDIKAN GANESHA**

**Oleh**

**Ni Made Yunitasari Maharani Pricilla Murti, NIM 2113101018**

**Program Studi S1 Matematika**

**Jurusan Matematika**

**ABSTRAK**

Penelitian ini bertujuan untuk memprediksi keterlambatan penyelesaian studi mahasiswa S1 di Universitas Pendidikan Ganesha. Keterlambatan penyelesaian studi merupakan masalah yang umum terjadi di perguruan tinggi dan dapat berdampak negatif pada reputasi institusi serta kualitas pendidikan. Metode yang digunakan dalam penelitian ini adalah regresi linier untuk memprediksi Indeks Prestasi Kumulatif (IPK) semester 8 dan regresi logistik untuk mengidentifikasi status keterlambatan mahasiswa. Data yang digunakan berasal dari UPA TIK Universitas Pendidikan Ganesha, mencakup informasi akademik mahasiswa dari tahun 2018 hingga 2021. Hasil penelitian menunjukkan bahwa model regresi linier memiliki akurasi yang sangat baik dengan nilai MAPE sebesar 1,5% dan koefisien determinasi ( $R^2$ ) sebesar 0,883, yang menunjukkan bahwa 88,3% variasi nilai IPK semester 8 dapat dijelaskan oleh variabel input yang digunakan. Sementara itu, model regresi logistik menunjukkan nilai AUC sebesar 0,984 dan akurasi klasifikasi sebesar 94,7%, menandakan kemampuan yang sangat baik dalam membedakan antara mahasiswa yang terlambat dan tidak terlambat. Atribut yang berpengaruh dalam memprediksi keterlambatan penyelesaian studi adalah IPK semester 8 dan jumlah SKS. Penelitian ini diharapkan dapat memberikan wawasan bagi pihak universitas dalam mengambil langkah-langkah preventif untuk mengurangi keterlambatan penyelesaian studi mahasiswa.

**Kata kunci :** Data Mining, Keterlambatan Penyelesaian Studi, Regresi Linier, Regresi Logistik, Universitas Pendidikan Ganesha.

***APPLICATION OF DATA MINING TO PREDICT THE DELAY IN  
COMPLETION OF STUDY OF UNDERGRADUATE STUDENTS OF  
UNIVERSITAS PENDIDIKAN GANESHA***

**By**

**Ni Made Yunitasari Maharani Pricilla Murti, NIM 2113101018**

**Study Program S1 Mathematics**

**Department of Mathematics**

***ABSTRACT***

*This research aims to predict the delay in completion of undergraduate student studies at Universitas Pendidikan Ganesha. Delay in study completion is a common problem in higher education and can have a negative impact on the reputation of the institution and the quality of education. The method used in this study is linear regression to predict the Grade Point Average (GPA) of the 8th semester and logistic regression to identify the delay status of students. The data used comes from UPA TIK Universitas Pendidikan Ganesha, covering student academic information from 2018 to 2021. The results show that the linear regression model has excellent accuracy with a MAPE value of 1.5% and a coefficient of determination ( $R^2$ ) of 0.883, which indicates that 88.3% of the variation in semester 8 GPA values can be explained by the input variables used. Meanwhile, the logistic regression model shows an AUC value of 0.984 and a classification accuracy of 94.7%, indicating an excellent ability to distinguish between students who are late and not late. The influential attributes in predicting late completion of studies are GPA in semester 8 and the number of credits. This research is expected to provide insight for the university in taking preventive measures to reduce the delay in completion of student studies.*

**Keywords:** *Data Mining, Study Completion Delay, Linear Regression, Logistic Regression, Universitas Pendidikan Ganesha.*