ABSTRACT

Abiella Andrea N.A.P. Panggabean (01679210009)

PREDICTING KINDERGARTEN, PRIMARY, AND SECONDARY LEVEL STUDENT RETENTION AT A PRIVATE SCHOOL IN INDONESIA USING MACHINE LEARNING

(xi + 45 pages; 16 figures; 14 tables; 5 appendixes)

Customer retention is important for most companies from various industries, including the field of education. Student retention prediction is necessary for schools because it can identify students who are most likely to retain in the next academic year and help school management to evaluate and improve its service and strategy. This research raises a case study of a private school in Jakarta, Indonesia, that has a target of student retention.

A tool called Orange3 was used to execute this experiment. The data was collected from a private school. After carrying out the data pre-processing stage, the models were trained using Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), and Neural Network (NN). This experiment also searched for the best predictor using Gini impurity.

Although generally all machine learning models in this research showed good results, the analysis results reveal that Random Forest was the best machine learning model to predict student retention. Random Forest was able to achieve an accuracy score of 0.891 in the training phase and 0.917 in the testing phase. For the best predictor, the results showed that the most important feature and the best predictor for student retention prediction is the customer service team's observation on parents' daily behavior.

Keywords: Student Retention, Educational Data Mining, Random Forest, Logistic Regression, Support Vector Machine, Neural Network

References: 61 (1999 – 2023)

ABSTRAK

Abiella Andrea N.A.P. Panggabean (01679210009)

PREDICTING KINDERGARTEN, PRIMARY, AND SECONDARY LEVEL STUDENT RETENTION AT A PRIVATE SCHOOL IN INDONESIA USING MACHINE LEARNING

(xi + 45 halaman; 16 gambar; 14 tabel; 5 lampiran)

Retensi pelanggan adalah hal yang penting bagi banyak perusahaan dari berbagai industri, termasuk di bidang pendidikan. Prediksi retensi murid dibutuhkan oleh sekolah karena hal tersebut dapat mengidentifikasi murid yang kemungkinan besar akan mendaftar kembali di tahun ajaran selanjutnya dan membantu manajemen sekolah untuk mengevaluasi dan meningkatan layanan dan strategi perusahaan. Penelitian ini mengangkat studi kasus tentang sebuah sekolah swasta di Jakarta, Indonesia, yang memilik target retensi murid.

Data diperoleh dari sebuah sekolah swasta. Proses pengolahan data dilakukan menggunakan alat bernama Orange3. Model *machine learning* yang dipakai dalam penelitian ini adalah *Random Forest* (RF), *Logistic Regression* (LR), *Support Vector Machine* (SVM), and *Neural Network* (NN). Penelitian ini juga berusaha untuk menemukan variabel prediktor terbaik menggunakan *Gini impurity*.

Meskipun secara umum semua model *machine learning* dalam penelitian ini menunjukkan hasil yang baik, analisa dari hasil tiap model menunjukkan bahwa *Random Forest* adalah model terbaik untuk memprediksi retensi murid. *Random Forest* mampu mencapai skor akurasi sebesar 0,891 pada fase pelatihan (*training*) dan 0,917 pada fase pengujian (*testing*). Untuk prediktor terbaik, hasil menujukkan bahwa variable terpenting dan predictor terbaik untuk prediksi retensi murid adalah pengamatan tim layanan pelanggan terhadap perilaku sehari-hari orang tua murid.

Kata kunci: Student Retention, Educational Data Mining, Random Forest, Logistic Regression, Support Vector Machine, Neural Network