Can machine learning catch a falling student?

A data-driven approach towards school online learning decision-support

PROBLEM

Facilitating personalised support for individual students is challenging in online learning, as teachers are not able to observe them as they would in traditional classrooms.

REQUIREMENT

Curro Holdings Ltd. required a decision support system, with predictive capabilities, to assist teachers to implement strategic interventions and deliver timely feedback.

PROJECT APPROACH

In order to predict which students are at risk of academic failure, learners' online behaviour had to be analysed and the results summarised on an interactive dashboard.

OUTCOMES ACHIEVED

71% Improvement in average intervention time.

Learner's summary statistics

Academic risk categorisation

SYSTEM COMPONENTS

To facilitate the analysis of data using the decision-support system, 3-components were developed.

- Data cleanup algorithm
- Predictive analytical model
- Graphical user interface

SYSTEM FLOW

The data inputs that were transformed into output insights on the user interface are:

Activity logs → Clickstream data → Term marks → **Decision Support System** → Risk level → Mark comparison → Performance indicators

1 DATA CLEANUP ALGORITHM

- Preprocess the data collected on the Learner Management System
- Remove duplicate data and transform raw data into the right formats
- Remove all data irrelevant to the outcomes of the project

2 FEATURE EXTRACTION

Features were extracted from the cleaned data, based on three criteria. Features identified were:
- Relevant to academic performance
- Reflect adaptable student behaviour
- Was populated across the dataset

3 FEATURE IMPORTANCE

The selected features' importance was evaluated to analyse their contribution in predicting a learner's final results.

4 MACHINE LEARNING ALGORITHMS

Machine learning was then used to predict a learner's term marks. The Scikit library in Python was used as the modeling software. Three algorithms' predictive capabilities were tested on the data set.

5 TRANSLATION INTO INSIGHTS

The learners' behaviour was categorised based on the predicted term mark from the machine learning algorithm into a potential academic risk category (low, medium or high).

6 USER INTERFACE

The insights, academic risk category and the learner's behavioural summary statistics were integrated into an interactive dashboard.

PREDICTIVE ALGORITHM COMPARISON

Alternative algorithms were applied and the prediction capabilities were compared using the Mean Squared Error (MSE). The Random Forest Regressor outperformed the alternatives.
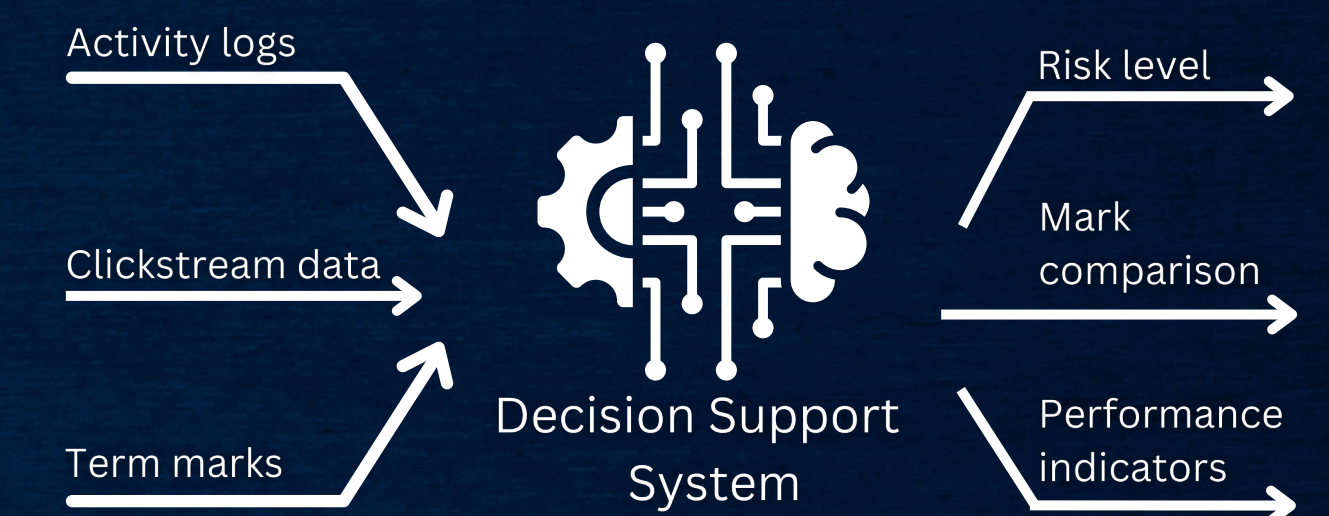
$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$y_i = i^{th}$ actual value

$\hat{y}_i = i^{th}$ predicted value

	Mean Squared Error	Mean Absolute Error
Linear regressor	320	14%
LASSO algorithm	238	12%
Random Forest Regressor	208	11%

SIGNIFICANT INDICATORS OF ACADEMIC RISK

The important factors in predicting a learner's success, as identified by the Random Forest Regressor algorithm, were translated to the interactive dashboard on PowerBI.

Legend: Grade 10 2022, Grade 11 2022

Categories: Activity Marks, Number of Learning Activities, Number of Supporting Activities, Number of Clicks, Total Zero-Score Activities, Exam Attendance

A PROJECT BY
TANÉ E. COETZER
u20420082