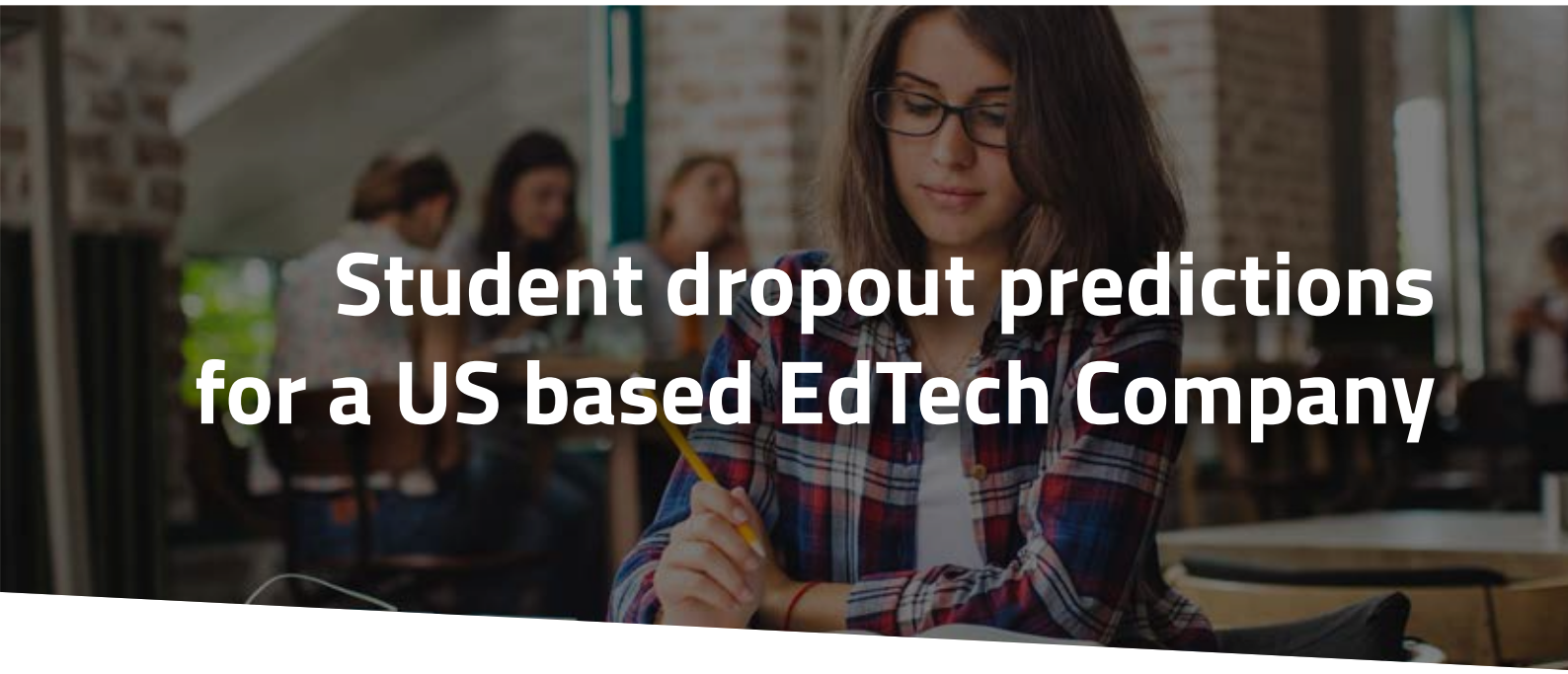
INDIUM
Digital | QA | Gaming

Success Story



Student dropout predictions for a US based EdTech Company

Business:
Data Analytics & Data Visualization

Domain:
EdTech

Tools:
PostgreSQL, MongoDB, R, R Shiny, etc.

Key Highlights

Key Success:
» Dropout rates significantly decreased by 12%.
» The model achieved an accuracy of 85% at predicting drop out cases.

Algorithms (implemented in R)
» Logistic Regression
» Random Forests
» XGBoost

Data Visualization
» R Shiny
» ggplot2
» Highcharts
» DYgraph libraries

Client
Our client provides an intelligent Learning Management System (LMS) platform for Universities in the US to enable collaborative learning.

Overview
One factor in university rankings and funding qualifications is the student dropout rate- the percentage of students who do not complete college requirements and obtain their final degrees. The key proposition from the client was a platform that empowers professors/ instructors to assess, track and predict student's performance in real time. Indium's engagement was to develop a feature, along with the related models and analytics algorithims, which could predict which students were likely to drop out before, in some cases, even the student was aware.

1 Status Quo

Our client provides an intelligent Learning Management System (LMS) platform for Universities in the US to enable collaborative learning. The ecosystem is a marketplace for content/ course transactions, tutoring services and reporting on learning metrics.

The key problem that the platform addresses is to empower the teachers/ instructors using the platform to assess, track and predict performance of learners in real time. While the instructors try different ways to engage students in their learning processes, there is always an obscured risk of student dropouts and withdrawals.

The client wanted to enable the platform to forestall the risks associated with the following:

» Failure to complete courses
» Students withdrawal from a course
» Student dropouts

2 Business Requirements

The platform required actionable intelligent insights to detect dropout risks early and mitigate them.

» Analyse the data set of student demographics and academic history to predict the probability of drop outs over a term/ period.
» Instructors need data driven approach to a proactive guidance/ mentoring.
» Make predictions of dropout intentions and performance facts at student level.
» Visualize the interpretation of results.

3 Solution

Indium Software conducted a due diligence of the current platform features and the problems to be addressed. This was identified as a typical classification problem. Our approach to the problem spanned Data Handling, Solution implementation/ Machine Learning Model and Data Visualization.

Data Handling

Data from Learning Management Systems was obtained in large volumes. MongoDB was used to handle the load.

» Student demographic data: socio-economic, education, academic achievements, previous grades, high school performance etc. of each student.
» Performance: monthly test scores, assignment scores, project scores, class participation scores etc.

Data Preparation

The number of data variables available were not all logical to feed to the machine learning model. Indium Software conducted exploratory data analysis to rationalize the variables for the defined use cases.

» Used Chi-squared test and Correlation techniques to select and also eliminate redundant variables.
» The student dropout results with the data set revealed a 10% dropout rate. This data was inadequate and to feed the attrition model. To overcome the biased data distribution, Indium Software implemented SMOTE (Synthetic Minority Oversampling Technique) to ingest synthetic data to tune the model to an efficient one.

Drop out Probability Prediction Modelling

Indium Software set out to build an intelligent system using robust classification algorithms to identify risks of Dropouts, Failures and Withdrawals (DFW). The primary goal to model the drop out probability was to calculate two kind of risks which yielded two predictive models.

» Classification of inherent risks – Day 0 Module: Modelling the data on Day-0 of the student picking up a course. This is modeled using the academic history data and available demographic data. Both the data sets were combined to form derived variables such as – total courses a student dropped in the past, total courses a student completed successfully in the past, total credits attempted and earned by the student till date, total credits a student has enrolled for in that term.

» Performance Risks- student performance metrics which are subject to daily changes for DFW. Choice of quality data: Student's performance risk predicting model uses daily scores of assignments/quiz which was dispensable to train the drop out model. Indium Software adopted a different modelling strategy by using the cumulative sum data that will sustain the 'recency' factor to the data.

We calculated the probability of student's DFW on a daily basis and it revealed if he is going to dropout by the increase in the probability value. We zeroed on to XGboost algorithm which gave a better recall and sensitivity, after trying out the models in Logistic regression, Random forest.

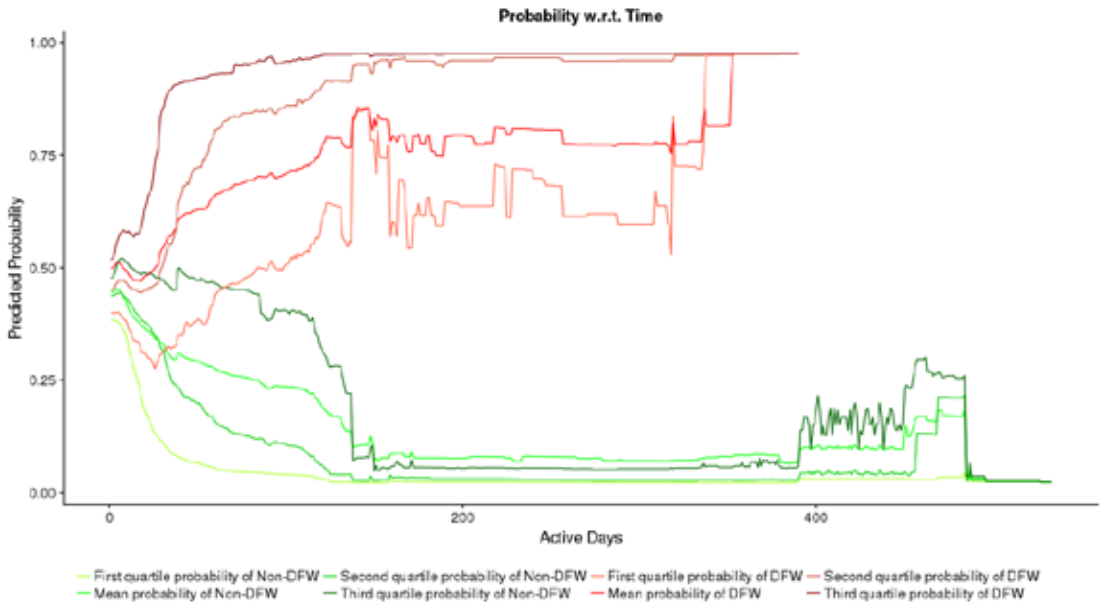
4 Business Impact

» The predictive model fared well with the recall and sensitivity of training data to achieve a prediction of 85% of the drop out cases.

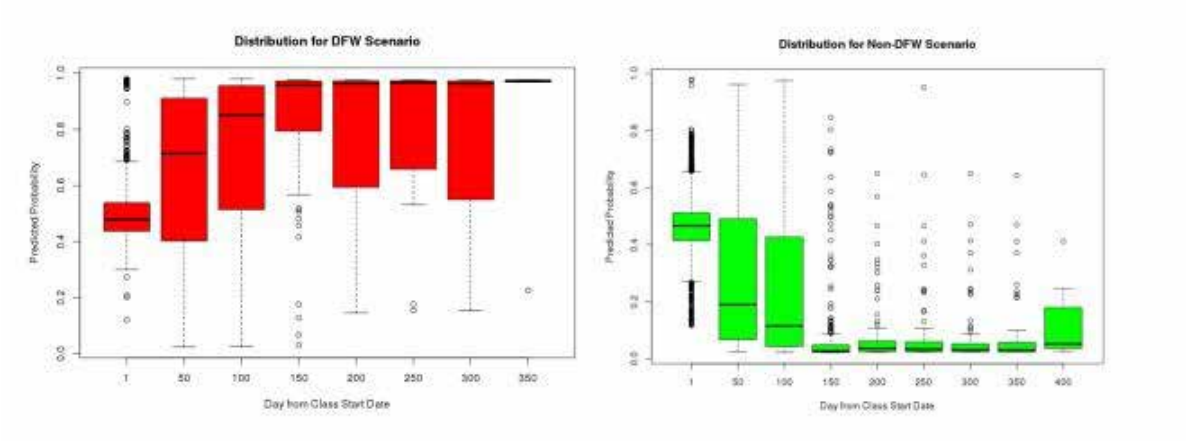
» Instructors were able to track risks on a daily basis and better manage at-risk students.

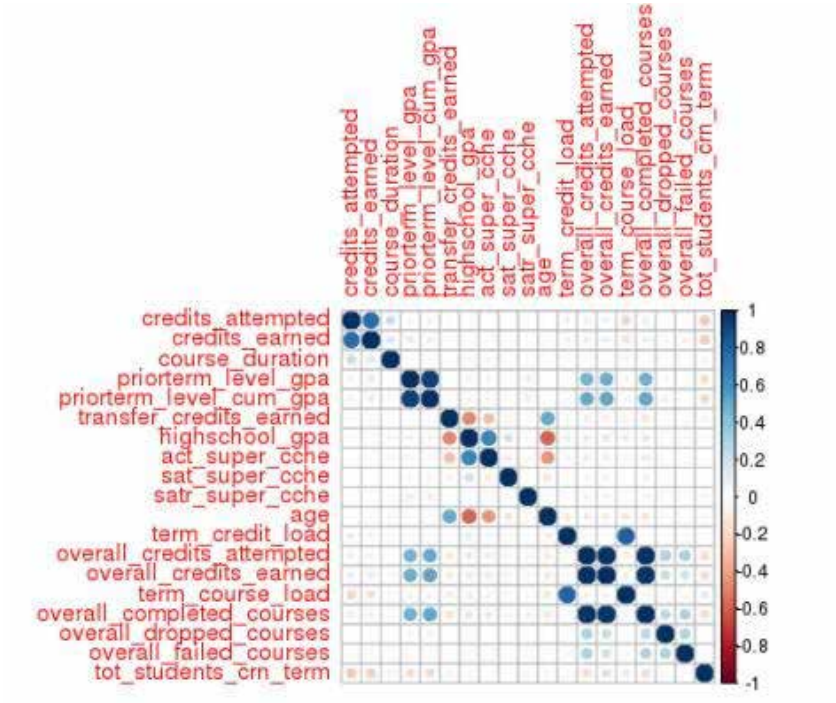
» Dropout rates significantly decreased by 12%.

5 Sample Visuals



The green lines indicate the probability for students who will not dropout and the red line indicates the probability of students who will dropout, plotted against the number of days





Correlation matrix of all the variables used which helped in reducing the variables

4 Additional Insights Provided

» Students who are going to dropout characterized by high probability, took more time to complete the assignments. This indicator helped in estimating the dropouts at the beginning.

» The number of Day0 courses was directly proportional to the dropout probability. This showed an easy method to monitor suspicious students.

» Students who are going to dropout also missed the due dates for the submission. This helped the business to keep a tracker for the due dates and initiate a monitoring activity

INDIUM
Digital | QA | Gaming

INDIA
Chennai
+91 44 6606 9100
Bengaluru
+91 80 4645 7777
Mumbai
+91 022 6215 4028

USA
Cupertino | Princeton
Toll-free: 1 888 207 5969

SINGAPORE
+65 9630 7959

UK
London
+44 773 653 9098

General Inquiries : info@indiumsoftware.com
Sales Inquiries : sales@indiumsoftware.com