



## Student dropout predictions for an US based EdTech Company

**Application:** Intelligent Learning Management System (LMS)

**Services Offered:** Data Visualization

**Solution:** Classification Algorithms, R Shiny Visualizations

### Key Highlights

**Domain:**

Education

**Duration:**

3 months

**Team:**

1 Data Scientist

1 Data Engineer

1 Program Manager

**Technology:**

Data stored in MongoDB & PostgreSQL

R on R server

### Client

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

## Business Requirements

- ✓ The platform required actionable intelligent insights to early detect these risks and mitigate them
- ✓ 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 the results

## Our Approach

Indium 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 is illustrated from Data Handling perspective, Solution implementation and Data Visualization outlook.

Data from Learning Management Systems is in large volumes

- *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.

High volume data was stored in MongoDB.

The primary goal was to calculate two kind of risks:

- 1) *Classification of inherent risks* – Day 0 Module and
- 2) *Performance Risks* (student performance metrics which are subject to daily changes) for
  - Dropouts (D)
  - Fails (F)
  - Withdrawals (W)

Of students from a class

Indium set out to build an intelligent system using robust classification algorithms to identify risks of DFW.

## Algorithms

- ✓ Logistic Regression
- ✓ Random Forests
- ✓ Xgboost

All implemented in R (R server)

## Data Visualization

- ✓ We used R Shiny / ggplot2 / highcharts / DYgraph libraries on R Server to visualize the result interpretations of student behavior trends in dropouts and withdrawals

## Value Delivered

- ✓ 85% accurate models integrated in LMS product. Helps instructor better **manage at-risk students**.
- ✓ Instructors were able to track risks on a daily basis
- ✓ Dropout rates significantly decreased by 12%

# Sample Visuals



