Microsoft Partner Microsoft Microsoft Gold Cloud Platform Gold Data Analytics Gold Data Platform Gold DevOps Gold Datacenter





### **King's College London and Adatis explore how Data Science can help to predict Student Withdrawals**

An innovative pilot project confirms the transformational impact of an Azure-based data lake and machine learning on predicting potential student course withdrawals.

#### At a glance.

- Ability to highlight students with a high probability of withdrawing to allow targeted support
- Proof of concept built by Adatis with close collaboration from King's College London Analytics Team
- Model trained with data feeds from two source systems, VLE and Student Records
- Supports predictive analytics with machine learning
- Utilising pan-university data lake, data warehouse and analytics platform
- Models trained and evaluated using Azure Databricks, scikit-learn library and Imblearn library
- Model baseline created with Auto-ML from Azure Machine Learning Services
- Full experiment tracking implemented with MLFLow
- Databricks Model registry used to version each trained model
- Microsoft Azure, Databricks and Power BI
- The King's / Adatis project team now exploring how to implement and manage production solution

#### The challenge.

#### Student withdrawal is a big challenge for the higher education sector.

Today, on average 7 out of every 100 undergraduate students in the UK drop out of their higher education course before their second year. As more students embrace remote learning, it is becoming increasingly difficult for higher education institutions to identify students who are struggling. It is crucial to be able to identify these students and to make timely interventions to improve student retention, performance, and wellbeing.





Establishing a reliable solution to this problem that doesn't infringe upon the privacy of individuals is challenging, as is predicting the complexities of when and why students will decide to leave their courses. The decision is often influenced by several contributing factors, including academic (progress, feedback), social (peer group, engagement) and external (family, health, financial).

For King's College London there are significant student, reputational and commercial benefits in being able to establish an early warning system that would enable staff to identify students who require additional support. King's and Adatis have a successful and long-standing relationship developing Data Platforms. Based on their partnership, they agreed to work together to see if training a machine learning model using data within the platform could provide an accurate and effective predictive solution.

King's Director of Analytics, Richard Salter, understands that early identification is the key to success.

'King's wants to support every student to reach their potential and achieve their ambitions. Where students are disengaging or experiencing difficulties, speed is of the essence. If we can identify the issue early and ideally predict it before it happens, then we are much better able to support the student. If we are slow to identify and respond, then the chances of redressing the situation are drastically reduced.'

#### An effective proof of concept (POC) exercise is a necessary first step to prove or disprove whether a machine learning solution is viable.

It is advantageous to prove (or disprove) whether required predictions can be accurately generated from the available data quickly, and with a minimum of cost, before committing to a full investment in a machine learning solution. It can often turn out that data isn't accessible, there isn't enough of it or that even if the data does support the required predictions, that this can't be generated in actionable timescales.

The project team had to be focussed, efficient in their use of time and effective in delivery of reliable results. Adatis experience and proven machine learning development approach offered the highest likelihood of success.

## What ethical considerations need to be made before embarking on such a project?

Often where a model's focus relates to individuals, questions of privacy and consent must be considered. Do subjects consent to the use of their data for the intended purpose?

# 🖪 adatis

Also, do any of the key features of the model involve sensitive data about individuals? Student activity and demographics provide key data points to a predictive model of this type. Achieving the fine balance to utilise these in an ethical way that complies with privacy guidelines like GDPR, whilst also meeting with the consent of subjects, would be a critical success factor.

#### The solution.

#### **Project Approach**

Adatis applied their proven AI POC approach to collect, analyse, and prepare data and then to train and evaluate candidate models quickly and rigorously.

The required data was identified, sampled, cleaned, and analysed to identify any anomalies or gaps. Once the data was understood, work began to identify and trial candidate features which might help to accurately predict a withdrawal outcome. Different combinations of feature sets and algorithms were explored, analysed, and evaluated.

The Adatis team worked very closely with their King's counterparts to ensure that ethical considerations were sufficiently considered during the scoping and planning of the work, and measures were agreed and implemented to ensure the privacy and security of data processing throughout.

'The approach was about building on the work already completed around data and analytics, harnessing the expertise that Adatis possessed to enable us to make rapid progress in developing a solution that could be upscaled and operationalised for the whole institution.'

**Richard Salter, Director of Analytics** 

#### **Student Withdrawal Project Approach**

🔲 adatis

This is our standard approach for a data science project that covers all they key components to ensure a successful implementation.



The Adatis AI team conducted two, iterative model build phases. The first proved that machine learning could accurately predict student withdrawal, however the predictions it delivered weren't timely. Meaning they couldn't highlight the risk of a student withdrawing with sufficient lead time to allow King's to make a meaningful intervention.



Figure 1 - Delivery process of Student Withdrawal phase 1 focussing on Student data

Adatis and King's reviewed the situation and quickly agreed the scope of a second, short POC phase. For the second iteration, an additional data source was made available that offered insight into a student's engagement with King's online learning system. Data points from this system included number of logins, interactions with forums and groups and assessment submissions. These additional activity-related data points provided far greater insight into student engagement throughout the academic year. The model that was trained subsequently delivered both accurate and more timely withdrawal indicators.



Figure 2 - Delivery process of Student Withdrawal phase 2 focussing on online engagement data

# 🖪 adatis

#### The results

Adatis proved that probability indicators of student withdrawal could be delivered in a timely manner to key King's staff members with a degree of accuracy that provides the confidence for them to act upon the insights. The most recently trained classification model reached an overall accuracy of 92 percent. Rarely though is Accuracy alone an indicator of success and so Adatis optimised the models for Precision, which aims to reduce the number of false positives. This is so that the King's team could begin to focus on a condensed cohort of students with the final model producing a Precision score of 98 percent.



The solution incorporated a trained model as well as the data pipelines to enable predictions to be inferred on an ongoing basis. Furthermore, a way of working on future machine learning problems was proven. This was all achieved in a few weeks for minimal investment.

'The results of the different models from the proof of concept were well beyond our expectations. They unambiguously affirmed the potential of this solution and very quickly we moved into thinking about how we could bring it into full production and leverage the value from the insights the model provided.' **Richard Salter, Director of Analytics - King's College London** 

#### Where are we now?



## adatis



The resulting benefits case was sufficiently compelling for King's to request further support from Adatis to implement the model on a production footing.

### What are the benefits?

## 

King's have the insight to identify specific individuals who may be at risk of withdrawing with time to intervene.

### STUDENT

King's can inform students of their probability to succeed based on the behaviours they exhibit.

## 

Courses with high completion probability can be used as models for courses with lower completion probabilities.

### CAPABILITY

King's have tools and skills to develop future Data Science solutions and a working example to emulate.

### E PLANNING

With a better understanding of student withdrawal likelihood, planning can be done more accurately.

Adatis offer professional services specialising in data analytics, from data management strategy and consultancy through to world class delivery and managed services.

With offices in London, Farnham and Sofia they have been delivering innovative and successful solutions for clients since 2006.



We love helping people unlock the power and value of their data. If you are interested in finding out more please contact us on **+44 (0)1252 267 777** or via email at **enquiries@adatis.co.uk** 

You can also check out our website at **www.adatis.co.uk** or come and see us at numerous global conferences and speaking events.

# 🖪 adatis