

## InsPIRE Insight

# From Research to Practice: Translating a Dissertation-based solution into an Enterprise-level System

By

- Wei Qiu (Centre for the Applications of Teaching and Learning Analytics for Students (ATLAS), InsPIRE)
- Andy W. H. Khong (School of Electrical and Electronic Engineering; Lee Kong Chian School of Medicine)
- Fun Siong Lim (Centre for the Applications of Teaching and Learning Analytics for Students (ATLAS), InsPIRE)

Nanyang Technological University  
Singapore

## KEY IMPLICATIONS

1. Translating academic research into sustainable enterprise systems requires a robust, automated data pipeline to ensure scalability and data security.
2. Automated grade prediction models can successfully identify at-risk students early in the semester, enabling timely human intervention.
3. Collaboration between researchers, IT services, and student care units is essential to move learning analytics from theoretical models to practical application.

## BACKGROUND

Reducing student dropout rates and enhancing academic success are critical challenges in higher education. While predictive machine learning models developed in research (such as PhD dissertations) show promise in identifying at-risk students, these models often remain theoretical or pilot-based. They rarely transition into deployed, enterprise-level systems that benefit the wider student body due to challenges in scalability, integration, and data governance.

## FOCUS OF INITIATIVE

This initiative aimed to operationalize a suite of grade-prediction machine learning models, originally developed during a PhD program, into a scalable, university-wide early alert system. The goal was to create a fully automated pipeline that could process student data and deliver actionable insights to student care managers to support struggling students.

## KEY OUTCOMES

- **Automated Data Pipeline:** The team established a robust infrastructure integrating

Denodo (data virtualization), Snowflake (data warehousing), Dataiku (machine learning and version control), and QlikSense (visualization) to automate the flow from raw data to actionable dashboards.

- **High Accuracy:** In deployed trials, the system achieved a true-positive rate of over 70% in identifying students at risk of academic challenges.
- **Effective Intervention:** Student care managers reported that the system allowed them to identify struggling students earlier than before, providing time to help them "turn things around".
- **Student Impact:** Students identified by the system reported increased confidence and motivation after receiving support, moving from failing grades to passing averages.

## SIGNIFICANCE OF OUTCOMES

- **Implications for practice:** This initiative demonstrates that "at-risk" prediction models can be effectively integrated into daily university operations. It highlights that the value of learning analytics lies not just in the algorithm, but in the *actionability* of the data provided to support staff. The automated pipeline ensures that insights are delivered securely and reliably without constant manual oversight.
- **Implications for policy and future initiatives:** Universities should consider investing in enterprise infrastructure that supports the translation of internal research into institutional tools. This case study suggests that data governance policies and student care procedures must evolve alongside technological implementation to ensure interventions are ethical and effective. It serves as a model for preventing valuable dissertation research from being "shelved".

## PARTICIPANTS/SCOPE

The system has been implemented across five schools within the university. Since January 2021, six trials have been conducted involving 13,944 students.

## METHODOLOGY/APPROACH

The project utilized a "human-in-the-loop" transition from research to enterprise. The technical approach involved a four-stage pipeline:

1. **Extraction:** Raw data is integrated via Denodo from sources like Oracle Cloud.
2. **Processing:** Data is transformed and processed in Snowflake for scale.
3. **Modeling:** Dataiku hosts machine learning models that retrain on historical data and update automatically every semester.
4. **Visualization:** Results are pushed to Qlik Sense dashboards for student care managers to review.

## References

- Qiu, W. (2023). Academic achievement inspired machine learning methods for student grade prediction and at-risk detection. Singapore: Nanyang Technological University.
- Qiu, W., Khong, A., Supraja, S., & Tang, W. (2023). A Dual-Mode Grade Prediction Architecture for Identifying At-Risk Students. *IEEE Transactions on Learning Technologies*, 17, 803-814.

Link to full article at

<https://www.solaresearch.org/wp-content/uploads/2025/02/LAK25-CompanionProceedings.pdf>

Contact Wei Qiu at [qiuwei@ntu.edu.sg](mailto:qiuwei@ntu.edu.sg) for more information about the initiative.