



# Data Science II - MBAN 6120

## Group Project Report

**Instructor:**

Delina Ivanova  
delina@schulich.yorku.ca

**Report by:**

Anmol Mittal - 221059688  
Divyansh Gupta - 221297213  
Joe  
Priya Chaudhuri - 221060736  
Sanya Kapoor -

December 4th, 2024

## **Contents:**

1. Executive Summary
2. Opportunity
3. Data Source
4. Model Selection
5. Business Case
6. Conclusion

# **Course Recommendation System**

**Business Report**

# 1. Executive Summary

The Course Recommendation System is a groundbreaking solution designed to simplify elective course selection for postgraduate students by providing curated recommendations based on their preferences. Addressing a widespread challenge in higher education, the system uses machine learning, specifically **content-based filtering** powered by **TF-IDF vectorization** and **Cosine Similarity**, to analyze course descriptions and attributes like program eligibility, focus area, and credit hours. By reducing the time and effort required to navigate extensive catalogs, this tool enhances decision-making and significantly improves the student experience. Its scalable architecture also ensures seamless integration with institutional systems and flexibility for future enhancements like collaborative filtering and user feedback.

The financial viability of this solution is compelling. With an initial investment of \$50,000, covering model development, data pipeline setup, and infrastructure, and annual operational costs of \$15,000, the system can be deployed as a **Software-as-a-Service (SaaS)** platform. By targeting 50 institutions in the first year at a subscription fee of \$10,000 per institution, it is expected to generate \$500,000 in revenue, achieving profitability within the first year. Scaling to 200 institutions by Year 3 can drive annual revenues of \$2 million, with high margins due to minimal operational costs.

This solution has immense potential to scale globally, addressing the needs of over 20,000 universities worldwide offering graduate programs. Beyond immediate benefits to students, it positions institutions as leaders in leveraging technology to enhance academic outcomes. The Course Recommendation System is not just a tool—it is an opportunity to revolutionize the education sector, making it more student-centric, efficient, and future-ready.

## Athena Course Recommendation System

Find similar courses from over 300 courses from Athena.

Type or Select a Course:

BUSINESS ANALYTICS

Get Recommendations

### Recommended Courses:

User Experience Design

Innovation Methods

Problem-Solving Strategies

### User Experience Design

**Description:** Crises in organizations are inevitable, and whether an organization survives them is due in large part to the actions of its managers and leaders. This course provides students with an overview of key factors in organizational crisis management, including (1) Assessing Crisis Vulnerabilities; (2) Planning for Crises; and (3) Managing during Crises. Prerequisites: All 5000-series Required Foundations of Management Core Courses.

**Professor:** Dr. Rachel Gray

**Credits:** 3

**Rating:** 4.8

## 2. Opportunity

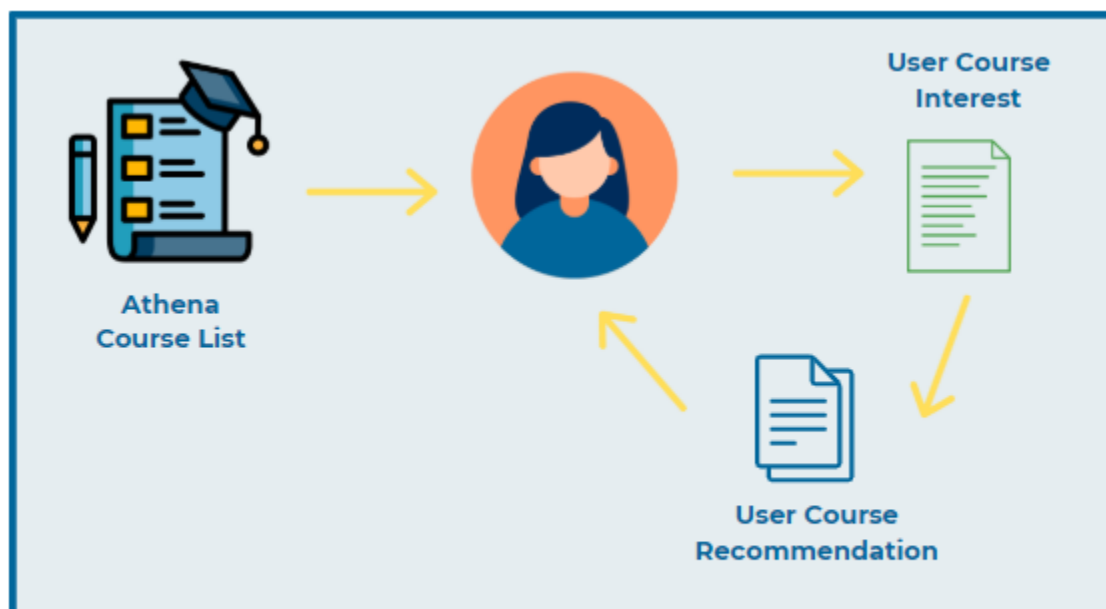
### The Problem

Navigating through elective course options is a universal pain point for students across higher education programs in Schulich. Current systems, often clunky and outdated, require students to sift through extensive catalogs, leading to suboptimal decisions and frustration. This problem is particularly pronounced in postgraduate programs where course alignment with career aspirations is critical.

### Proposed Solution

Our Course Recommendation System streamlines this process by:

- Allowing students to input a course of interest.
- Analyzing course descriptions and attributes (e.g., program, area, credits).
- Recommending top-N similar courses in real-time.



*Content Based Filtering*

### Market Potential

The system addresses a common challenge across all postgraduate institutions globally, making it a highly scalable solution. With an estimated 20,000+ universities worldwide offering graduate programs, the adoption potential is immense. By implementing a SaaS platform, universities can

integrate this system with their existing course portals, offering tailored solutions to their student populations.

### 3. Data Source

#### Data Provenance

The data powering this prototype is sourced from the Schulich School of Business course catalog. It includes:

- Textual Data: Course descriptions.
- Categorical Data: Program, area, and term.
- Numerical Data: Credit hours.

#### Data Sanctity

The data's credibility stems from its official source, ensuring accuracy and reliability. However, for scalability:

- The data pipeline should be automated to ingest updates directly from institutional systems.
- Quality checks must be implemented to handle inconsistencies across institutions.

#### Ethical Considerations

The model operates on anonymized course data, ensuring compliance with data privacy standards. Any future integration with user interaction data will require explicit consent.

### 4. Technical Approach

The recommendation engine employs:

1. **TF-IDF Vectorization:** Captures semantic similarity in course descriptions.
2. **Cosine Similarity:** Ranks courses based on their textual and categorical closeness. It further calculates pairwise cosine similarity between courses based on their combined feature vectors. And, produces a similarity matrix where higher values indicate more similar courses.
3. Recommendations are generated dynamically and are curated to closely match the chosen course in terms of course content, subject area, and credits.

## Future Enhancements

To improve accuracy and scalability:

- **Hybrid Filtering:** Combine content-based and collaborative filtering.
- **Advanced NLP Models:** Incorporate transformers like BERT for deeper semantic understanding.
- **Explainable AI:** Ensure recommendations are transparent and interpretable.
- **Feedback System:** The feedback system will track user interactions (likes/dislikes) with recommended courses. This data will adjust the similarity matrix by boosting the scores for liked courses and reducing scores for disliked ones. Over time, the system will refine recommendations to better align with user preferences, creating a personalized experience similar to Netflix's approach.

## Validation and Evaluation

The model's performance is evaluated using metrics like precision and recall, ensuring the relevance and accuracy of recommendations.

## 5. Business Case

### Cost Analysis

Developing a scalable SaaS platform involves:

- Initial Development Costs:
  - Model training and optimization: \$15,000
  - Data pipeline setup: \$10,000
  - Backend infrastructure: \$20,000
  - Misc: \$5000
- Operational Costs:
  - Cloud hosting (AWS/GCP): \$5,000/year
  - Maintenance and updates: \$10,000/year

Total Initial Investment: \$50,000

Annual Operational Costs: \$15,000

### Revenue Potential

The SaaS model can charge institutions:

- Annual Subscription Fee: \$10,000 per institution
- Target Clients (Year 1): 50 institutions  
Revenue (Year 1): \$500,000

By Year 3, scaling to 200 clients can generate **\$2M annually**, making this venture highly profitable with minimal ongoing costs.

## **Profitability**

With a break-even point achievable within the first year, the model ensures strong margins. Expanding features (e.g., user reviews and interactive dashboards) can unlock additional revenue streams.

## **6. Conclusion**

The Course Recommendation System addresses a critical gap in postgraduate education by enabling informed decision-making and enhancing user experience. With a robust technical foundation and significant market potential, this solution offers societal impact and business profitability.

We can serve thousands of institutions globally by transitioning this prototype into a SaaS platform. The scalability and a compelling ROI position this project as a transformative innovation in education technology.