Project Proposal Form MCST1043
Sem: 2 Session: 2024/25

SECTION A: Project Information.

Program Name:	**Masters of Science (Data Science)**
Subject Name:	**Project 1 (MCST1043)**
Student Name:	Li Xinya
Metric Number:	**MCS241029**
Student Email & Phone:	lixinya@graduate.utm.my & 0175703279
Project Title:	Intelligent Prediction of University Course Satisfaction Using Text Mining and Machine Learning
Supervisor 1:	
Supervisor 2 / Industry Advisor(if any):	

SECTION B: Project Proposal

Introduction:

In the current quality assurance system of higher education institutions, student evaluations of teaching have emerged as a crucial means for gauging course quality, optimizing instructional design, and assessing teachers' performance. Traditional evaluation approaches typically rely on structured rating forms, which facilitate statistical processing but struggle to comprehensively mirror students' genuine experiences and subjective emotions. By contrast, open-ended text comments within the evaluation system, as unstructured data, can offer deeper insights into students' individualized feedback regarding course content, teaching methods, and classroom ambiance, and they encompass a substantial amount of valuable information.

However, the analysis of these text data poses challenges due to their complex content and diverse emotional expressions. Traditional manual analysis methods are ill-suited to achieving efficient and systematic information extraction.(Wang, Y., Liu, X., Zhang, H., Wang, T., & Xu, J. 2019) To leverage such data more effectively, it is imperative to draw on techniques such as natural language processing (NLP) and information extraction and emotion recognition from student comments, thereby elevating the machine learning to conduct automated information extraction and emotion recognition from student comments, thereby elevating the intelligence level of the teaching feedback system.

In recent years, sentiment analysis technology, as a significant research direction within NLP, has been

extensively utilized in domains such as e-commerce reviews, public opinion monitoring, and the financial market, demonstrating considerable prowess in identifying the emotional tendencies of text and quantifying user attitudes. (Elnagar, A., Al-Debsi, R., & Einea, O. 2020) Its introduction into the teaching scenarios of higher education institutions not only aids in extracting emotional features from student comments but also enables integration with structured information like ratings and course difficulty to construct more explanatory and predictive satisfaction models, furnishing data support and decision-making grounds for teaching management in higher education.

Problem Background:

With the rapid advancement of higher education, the content of courses, teaching methodologies, and the modes of interaction between teachers and students have been increasingly diversified. How to assess teaching quality scientifically and impartially has emerged as a core topic in the teaching management and educational quality guarantee of higher education institutions. Within the quality assessment system of higher education, student evaluation of teaching constitutes a significant basis for gauging teaching quality, optimizing course design, and enhancing teachers' instructional approaches. Currently, the majority of higher education institutions employ structured questionnaires (such as Likert scales) to collect students' evaluation data of courses. Although quantitative indicators facilitate statistical analysis, they often fail to comprehensively reflect students' authentic learning experiences. Nevertheless, this structured scoring approach frequently suffers from issues such as limited expression dimensions and the absence of subjective experiences, making it challenging to fully represent students' genuine learning perceptions and course participation experiences.

In recent years, a growing number of higher education institutions have incorporated open-ended text evaluations into their teaching evaluation systems. Compared to rating data, these unstructured text analysis proves to be rather challenging. Concurrently, the rapid evolution of sentiment analysis and machine learning technologies has furnished new technical support for the intelligent mining of teaching evaluation data in higher education institutions. Sentiment analysis, as one of the core directions in the domain of natural language processing (NLP), has been extensively applied in scenarios such as public opinion analysis, e-commerce reviews, and financial sentiment prediction. However, its application in the field of educational evaluation remains in the exploratory stage. How to leverage these technologies to automatically extract sentiment tendencies, key themes, and satisfaction-influencing factors from the vast volume of student evaluation texts and construct high-precision prediction models has become a crucial issue in elevating the intelligent level of teaching evaluations in higher education institutions.

Problem Statement:

Traditional teaching evaluations in higher education institutions primarily rely on structured scoring forms, conducting quantitative assessments from dimensions such as teaching attitude, course arrangement, and learning difficulty. Though this modality is conducive to processing and statistics, it fails to comprehensively reflect students' genuine sentiments and individualized requirements, disregarding "soft feedback" like emotions.

In contrast, the open-ended text comments composed by students in teaching evaluations, as unstructured data, frequently encompass more specific and emotional feedback. They can mirror the evaluations of course design, teaching methods, and classroom experiences from multiple perspectives, and disclose the emotional fluctuations and satisfaction levels during their learning processes.

With the advancements of natural language processing (NLP) and machine learning (ML) techniques, the automatic processing of unstructured text has become feasible. Sentiment analysis is capable of identifying the emotional tendencies and intensities within the text, while machine learning can integrate multiple features to construct satisfaction prediction models. However, a systematic study on effectively integrating text emotional features with structured scores and applying them to the modeling of teaching satisfaction in higher education institutions is still lacking.

Aim of the Project:
This project is intended to select the student course evaluation data offered by the open education platform RateMyProfessor as the research object and construct a satisfaction prediction model integrating text mining and machine learning. It will carry out emotion recognition and feature extraction on student text comments and conduct modeling analysis by combining multi-dimensional information such as course grades and teaching difficulty. Through this research, it is anticipated to enrich the means of teaching evaluation in higher education institutions and explore a scalable and extensible intelligent teaching feedback mechanism, providing theoretical references and practical approaches for the scientific and individualized development of educational evaluation.

Objectives of the Project:
1. Utilizing natural language processing techniques, the texts of students' course evaluations undergo preprocessing, word segmentation, word vector representation, and feature extraction. An emotion analysis model is established to automatically identify and classify the emotional tendencies in students' evaluations.

2. In combination with the dataset, machine learning algorithms (such as Random Forest) or deep learning techniques (such as LSTM) are employed to construct prediction models, which can automatically predict the satisfaction level or continuous score of the course.

3. Explanation tools like SHAP and LIME are utilized to determine the key features influencing the

prediction results, assisting managers in understanding the impact of teaching factors on satisfaction.

Scopes of the Project:

1. Data Collection: Select the student teaching evaluation data from public educational evaluation platforms (such as RateMyProfessor), encompassing structured data (quantitative indicators like course ratings, teacher ratings, and course difficulty) and unstructured data (students' text comments on courses and teachers).

2. Data Preprocessing: Handle and eliminate missing values and outliers in the collected initial data to ensure the quality and reliability of the data.

3. Feature Engineering: Text Feature Extraction: Key word frequency information and sentiment features.

4. Model Construction: Construct prediction models via machine learning algorithms (such as random forest, deep learning, etc.).

5. Model Evaluation and Optimization: Assess the performance of the model using metrics such as Accuracy and F1-score.

Expected Contribution of the Project:

1. An intelligent teaching evaluation analysis system based on deep learning will be developed to automatically handle students' text evaluations and scoring data, construct a multimodal fusion model, enhance the efficiency and intelligence level of teaching evaluation analysis, and offer data support for teaching optimization.

2. Explainable AI techniques such as SHAP and LIME will be incorporated to increase model transparency, identify the crucial factors influencing student satisfaction, assist teachers in precisely improving the teaching process, and augment the interpretability and application value of the model.

3. A modeling framework integrating structured and unstructured data will be established to provide a novel approach for modeling university course satisfaction and promote the research and development in the field of educational data mining in aspects of sentiment computing and multi-source data fusion.

Project Requirements:

Software:	Python, R
Hardware:	Personal computer with sufficient processing power and storage capacity
Technology/Technique/ Methodology/Algorithm:	Machine Learning Deep Learning Statistical Analysis Natural Language Processing (NLP) Accuracy; F1-score

Type of Project (Focusing on Data Science):

[√]　Data Preparation and Modeling

[√]　Data Analysis and Visualization

[　]　Business Intelligence and Analytics

[√]　Machine Learning and Prediction

[　]　Data Science Application in Business Domain

Status of Project:

        [√]   New

        [   ]   Continued

If continued, what is
the previous title?

SECTION C: Declaration

I declare that this project is proposed by:

        [√]   Myself

        [   ]   Supervisor/Industry Advisor (   )

Student Name:

Li Xinya	15/4/2025
Signature	**Date**

SECTION D: Supervisor Acknowledgement

The Supervisor(s) shall complete this section.

I/We agree to become the supervisor(s) for this student under aforesaid proposed title.

Name of Supervisor 1:

Signature	**Date**

Name of Supervisor 2 (if any):

Signature	**Date**

SECTION E: Evaluation Panel Approval

The Evaluator(s) shall complete this section.

Result:

[   ] FULL APPROVAL   [   ] CONDITIONAL APPROVAL (Major)*

[   ] CONDITIONAL APPROVAL (Minor)   [   ] FAIL*

* Student has to submit new proposal form considering the evaluators' comments.

Comments:

Name of Evaluator 1:

Project1 Proposal Form MSc (Data Science)

Signature   **Date**

Name of Evaluator 2:

Signature   **Date**