SCHOOL OF COMPUTING
Faculty of Engineering

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:	Gao Jingkai
Metric Number:	MCS241032
Student Email & Phone:	gaojingkai@graduate.utm.my & +60167101780
Project Title:	Knowing our choices: unveiling true voting patterns through machine learning (ML) and natural language processing (NLP) in European Parliament
Supervisor 1:	
Supervisor 2 / Industry Advisor(if any):	

SECTION B: Project Proposal

Introduction:

Recently, there has been a marked increase in the use of computing technology to solve problems related to political elections and legislative votes. This tendency is also evident in the exploration of the decisions of the European Parliament (EP). However, most of the research conducted on the European Parliament tends to be interpretative and descriptive, rather than focusing on forecasting. Therefore, this study aims to study political data by using Machine Learning (ML) technology and Natural Language Processing (NLP) method. Enhance our understanding of government systems by assessing the predictive power of EP legislative texts.

Problem Background:

Anticipating the diverse voting behavior of European Parliament Members (MEPs) is complex due to their diverse backgrounds, affiliations, countries, and ideological stances. These factors result in varied preferences, posing a substantial analytical challenge in understanding and identifying significant patterns.

Problem Statement:

Most current methods of studying voting in the European Parliament are limited to observing the results and then describing them briefly. This kind of analysis has a single perspective and does not utilize advanced data science techniques. Therefore, in general, this makes it difficult for policy analysts and political researchers to analyze real political issues.

Aim of the Project:

The project aims to forecast MEPs' voting behavior using semantic descriptors of legislative proposals through Machine Learning (ML), contributing insights to both Machine Learning and political science domains.

Objectives of the Project:

1. To build a voting prediction model based on machine learning algorithms.
2. To extract key textual features from legislative documents using natural language processing techniques, enhancing the predictive capability of the model.
3. To evaluate and compare the performance of various machine learning models, and analyze the importance of textual features influencing voting predictions.

Scopes of the Project:

1. Only real European Parliament voting data and corresponding bill text data were used from 2004 to 2019.
2. NLP analysis focused on the text extraction and application of semantic features.
3. Research scope explicitly exclude the parliament or other organizations, other datasets, and the relative contents of political decisions.

Expected Contribution of the Project:

1. A model based on machine learning and natural language processing (NLP) is developed to predict the voting behavior of members of Parliament effectively.
2. Hope this model can effectively reveal the law text semantic structure and the dynamic relationship between voting decisions.
3. Hope this model can help policy researchers more accurate understanding of European political council decisions.

Project Requirements:

Software: Python programming language, Jupyter notebook, Anaconda

Hardware: CPU: at least an Intel i5 (or equivalent)

RAM: At least 16 GB (32 GB recommended for large datasets)

Storage: Minimum of 200GB of free space

GPU: Use it for training complex ML models

Technology/Technique/ Methodology/Algorithm: Machine Learning Algorithms: Random Forest, SVM, Logistic Regression.

NLP methods: semantic feature extraction, text preprocessing, word embedding, and feature importance analysis (SHAP).

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: Gao Jingkai

Signature **Date** April 7, 2025

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: .....

Signature **Date**

Name of Evaluator 2: .....

Signature **Date**