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:	An Interpretable Hybrid BERT-ML Framework for Twitter Sentiment Classification
Supervisor 1:	
Supervisor 2 / Industry Advisor(if any):	

SECTION B: Project Proposal

Introduction:

Social media (particularly Twitter) contains a vast amount of emotionally charged text, and sentiment analysis of this content has significant value in areas such as business, public opinion monitoring, and social governance (Rodríguez-Ibáñez et al., 2023). However, the subjectivity and non-normative nature (slang, sarcasm) of tweets limit the semantic capture capabilities of traditional NLP techniques (Sadia & Basak, 2021). In recent years, research has shown that Transformer-based BERT models can more effectively capture contextual semantics and significantly outperform traditional methods in short text sentiment classification (Devlin et al., 2019; Rogers et al al., 2020). However, existing research indicates that BERT belongs to a type of model that is structurally complex and lacks interpretability, which limits its application in high-risk scenarios such as financial risk control (Rogers et al., 2020; Gurrapu et al., 2023). Model explanation techniques such as SHAP have been proposed to improve the understandability of its decision-making process (Lundberg & Lee, 2017). Therefore, this paper proposes a Twitter sentiment prediction framework that uses SVM as the main classifier, combines BERT semantic embeddings, and introduces the SHAP method to enhance model interpretability. To further verify the model's effectiveness, the study will also use RF and LR as comparative evaluation baselines.

Problem Background:

Social media (especially Twitter) is an important data source for sentiment analysis, reflecting individual emotions and collective attitudes (Rodríguez-Ibáñez et al., 2023). However, the unstructured nature, semantic ambiguity, and linguistic diversity (slang, abbreviations, etc.) of its text limit the effectiveness of traditional NLP methods (Mutanov et al., 2021; Sadia & Basak, 2021). Although Transformer models such as BERT have improved semantic understanding capabilities (Devlin et al., 2019; Rogers et al., 2020), they still have limitations in handling informal expressions (such as sarcasm) in tweets (Sadia & Basak, 2021). Furthermore, BERT lacks interpretability, which limits its application in critical scenarios (Lundberg & Lee, 2017). Therefore, the key challenge at present is to build a tweet sentiment recognition framework that possesses both strong semantic understanding capabilities and provides decision interpretability.

Problem Statement:

Although Transformer models such as BERT have demonstrated excellent performance in sentiment classification (Devlin et al., 2019; Rogers et al., 2020), their direct application to Twitter faces two major challenges. First, the unstructured and informal language of tweets (slang, sarcasm, etc.) makes it difficult for models to accurately capture complex emotions (Sadia & Basak, 2021; Mutanov et al., 2021). Second, BERT's lack of interpretability limits its practicality in high-risk scenarios such as financial risk control (Lundberg & Lee, 2017). Current research focuses on performance improvement but neglects the need to integrate semantic adaptability and interpretability. Therefore, there is an urgent need for a sentiment prediction framework that can accurately process complex social media text and provide a transparent decision-making process to enhance the credibility of practical applications.

Aim of the Project:

This research aims to construct a Twitter sentiment analysis framework combining BERT semantic embeddings with an SVM classifier, introduce the SHAP method to enhance model interpretability, and evaluate the advantages of the proposed framework in terms of decision transparency and practical application reliability through performance comparison with traditional machine learning models (Random Forest and Logistic Regression).

Objectives of the Project:

1. To collect and pre-process Twitter data covering informal expressions, and to utilize BERT to extract contextual semantic information.
2. To construct an SVM sentiment classifier and conduct comparative analysis with RF and LR to evaluate the performance of each model in terms of accuracy, F1-score, and interpretability.
3. To apply SHAP for model explanation (local/global), and comprehensively verify the effectiveness and transparency of the framework in conjunction with performance metrics.

Scopes of the Project:

1. Focusing on English Twitter text containing informal expressions (slang, abbreviations, etc.) (excluding emoticons), and performing binary (positive/negative) sentiment analysis.
2. Using only BERT to extract semantic embeddings, with SVM as the primary classifier (comparing with RF, LR), and using only SHAP for model explanation.
3. Not involving other Transformer models (such as RoBERTa), deep classifiers (such as LSTM/CNN), or other explanation methods (such as LIME).
4. Evaluating based on existing publicly available English tweet datasets (such as Sentiment140), excluding real-time scraping or cross-lingual tasks.
5. Not covering multi-modal, multi-lingual sentiment analysis and deployment optimization.

Expected Contribution of the Project:

1. Methodological Innovation: Proposing a hybrid framework combining BERT embeddings with traditional classifiers (SVM/RF/LR), balancing semantic understanding and classification efficiency.
2. Enhanced Interpretability: Applying SHAP for local/global explanations of sentiment models, enhancing transparency and credibility.
3. Application Validation: Demonstrating the framework's adaptability to informal Twitter text containing slang, irony, etc., providing practical reference.
4. Comparative Cognition: Deepening the understanding of the performance of lightweight classifiers combined with semantic embeddings through comparison with RF/LR.
5. Practical Bridge: Providing an interpretable social sentiment analysis model paradigm, connecting academic research and engineering applications.

Project Requirements:

Software:	Python programming language, Google Colab
Hardware:	CPU: Minimum Intel i5 or AMD Ryzen 5
	RAM: ⩾ 16GB
	Storage: ⩾ 10GB available disk space
	GPU (Optional): CUDA-compatible GPU (e.g., NVIDIA GTX 1660 or better) recommended for BERT processing acceleration
Technology/Technique/ Methodology/Algorithm:	Natural Language Processing (NLP), Traditional Machine Learning Model Integration
	BERT: Used for extracting contextual semantic embeddings from tweets
	SVM (Primary Model): Main classifier for sentiment prediction

RF and LR: Baseline models for performance comparison

SHAP: Explanation technique for analyzing influential features behind sentiment classification

Evaluation Metrics: Accuracy, F1-score, SHAP-based interpretability visualizations

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

April 13, 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: .....

Signature **Date**

Name of Evaluator 2: .....

Signature **Date**