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:	Yang Mu
Metric Number:	MCS241045
Student Email & Phone:	yangmu@graduate.utm.my number:1162302346
Project Title:	Big data driven: Forecast of global real estate market ups and downs in some regions
Supervisor 1:	
Supervisor 2 / Industry Advisor(if any):	

SECTION B: Project Proposal

Introduction:

The global real estate market is a complex system that is affected by many factors, including economic growth, interest rate changes, demographic changes, and government policies. In recent years, with the rapid development of big data and artificial intelligence technologies, it has become possible to use big data analysis technology to predict the trend of the global real estate market, and it has gradually become a research hotspot. However, most existing studies focus on the prediction of a single region or a single indicator, making it difficult to gain a comprehensive insight into the overall trend and potential risks of the global real estate market.

Problem Background:

The global real estate market data is rich, but there are problems such as scattered in data sources, inconsistent formats, and uneven quality, which lead to many challenges predicting and analyzing the global real estate market. For example, the lack of unified predicting and analyzing the global real estate market. For example, the lack of unified standards to describe and manage real estate data makes it difficult to effectively integrate and analyze data from different sources. In addition, most existing

forecasting models are based on a single region or a single indicator, which makes it difficult to accurately reflect the overall trend of the global real estate market.

Problem Statement:

How to utilize big data technology, build a unified framework to describe and manage global real estate market data, and develop an effective forecasting model based on this framework, so as to more accurately predict the ups and downs of global real estate markets?

Aim of the Project:

This project aims to use big data analysis technology to build a global real estate market forecast model based on partial regional data, to improve the accuracy and reliability of forecasts, to provide investors and decision makers with a more effective reference basis, and to provide more in-depth insight into the development of the global real estate market.

Objectives of the Project:

Data collection and sorting: Collecting and sorting global real estate transaction data, including price, transaction volume, area, geographic location, house type, etc. information, and performing data cleaning and preprocessing, ensuring data integrity and consistency.

Data framework construction: Build a unified data framework, describe and manage collected real estate data, ensure data comparability and analysis, and provide a standardization basis for data integration and analysis.

Forecasting model development: Develop effective forecasting models, such as time sequence models, machine learning models, etc., to predict the ups and downs of the global real estate market in some areas, and to evaluate and verify the model, analyze the accuracy and reliability of the model.

Analysis and visualization of results: Visualize forecast results, analyze and interpret forecast results, put forward some recommendations, such as forecasting and risk analysis of future real estate market trends, as well as investment strategy recommendations, etc.

Scopes of the Project:

Data source: Mainly uses public global real estate transaction data, such as data from government agencies, real estate websites, research institutions, etc.

Prediction indicators: Mainly focus on the rise and fall trend of housing prices, and may include other indicators, such as transaction volume, rental level, etc.

Prediction method: Mainly uses time series analysis, machine learning models and other methods for prediction.

Result analysis: Compare and analyze the prediction results with the actual situation, explain the reasons behind the prediction results, and put forward some suggestions.

Expected Contribution of the Project:

Promote standardization of global real estate market data management: By establishing a unified metadata framework, this project will provide standardized specifications for global real estate market data management, promote the integration analysis of data from different sources, improve data quality and credibility, and provide the basis for more accurate market forecasting. Improve the accuracy and reliability of global real estate market forecasting: This project will develop a forecasting model based on big data analytics and combine multiple data sources and analysis methods to improve the accuracy and reliability of forecasting, provide investors and decision makers with more effective reference basis to help them better understand market trends and potential risks. Provide new perspectives and tools for business analysis: The research results of this project can provide new perspectives tools for business analysis. For example, enterprises can use the metadata framework and prediction model built by this project to analyze real estate market trends, formulate more effective investment strategies, optimize business processes, and improve profitability. Urban planning and development: It can provide data support for urban planning and development, help urban planners better understand real estate market trends, and formulate more reasonable urban planning and development strategies.

Project Requirements:

Software: Python, Pandas

Hardware: Computer, network, data

Data collection and preprocessing

Technology/Technique/ Methodology/Algorithm:

- Collect and organize metadata from COVID-19 scientific datasets.
- Construct a metadata framework for COVID-19 scientific datasets,encompassing external features, content features, and sharing features.
- Utilize Protégé software to build an ontology for scientific datasets, defining core concepts and attribute relationships.
- Store the constructed knowledge graph using the Neo4j graph database, enabling query retrieval and reasoning for entities and their relationships.
- Data analysis
- Model Evaluation

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: Yang Mu

Yang Mu — **Signature**

07/04/2025 — **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**