

Abstract

In the current situation where everything is running at a fast pace, the real estate market is a complex system that is highly affected and closely related to people. Not only under the regulation of the economy, but also including the influence of policies, geographical factors, and even wars and natural disasters. (Mariia I. Ermilova, 2021) Despite the progress made in big data and artificial intelligence, the existing research mainly focuses on local predictions and lacks an overall framework for global market analysis. And mostly analyze the impact of a single condition on the real estate market. This study aims to address this gap by constructing a unified data framework through the utilization of big data analysis and developing a predictive model for predicting regional real estate market trends. This project emphasizes the application and prediction of business analysis, the preprocessing and integration of diverse data, and the application of machine learning models. The expected contributions include improving the accuracy of predictions, standardizing data management practices, and providing actionable insights for investors, policymakers, and urban planners. (Petros et al, 2024)

Chapter 1 Introduction

1.1 Introduction

The real estate market plays a very important role in economic stability and the development of people's livelihood. However, its volatility has brought significant challenges to almost all people, so accurate prediction tools are extremely necessary. However, the current research is still fragmented, with limited integration of cross-regional data and standardized methods under a single condition. However, with the development of society and technology, the progress of big data is also advancing rapidly. Therefore, accurately predicting the global real estate market is not out of reach. This study proposes a new approach to unifying global real estate data management and developing a novel comprehensive predictive model, with the aim of providing comprehensive insights into market trends and informing strategic decisions. (Ruoying Tan & Tze-Haw Chan, 2021)

1.2 Background of the Problem

The global real estate market has generated a large number of datasets, but their different conditions, scattered sources and inconsistent standards have hindered effective analysis. Existing models often rely on data from a single region or isolated indices, and are unable to capture interrelated global trends. For instance, China's building height limit regulations almost affect different Southeast Asian countries, and the high housing management fees in the United States do not exist in many countries either. Furthermore, the traditional single-condition prediction method is also difficult to accurately deal with and adapt to the

real estate market that is dynamically influenced by multiple conditions. These restrictions emphasize the need for standardized data infrastructure and advanced analytical models to address the complexity of global market forecasting.

1.3 Statement of the Problem

The core challenge lies in the lack of a unified framework to integrate and analyze heterogeneous global real estate data. The current papers and studies have inconsistent conditions and data, regional biases and insufficient prediction accuracy. (Bo et al, 2022) This study aims to answer: How to utilize big data technology to construct a standardized data framework, develop reliable prediction models, and predict fluctuations in the global real estate market.

1.4 Research Questions

How can the scattered and inconsistent global real estate data be standardized and integrated into a unified framework?

How to determine representative and effective market data for research under the vast amount of global data?

Which predictive models (such as time series and machine learning) are the most effective in predicting the trends of the real estate market in multiple regions?

How to unify the influencing factors and make comparisons among the data?

How can the accuracy and reliability of these models be verified based on real market data?

1.5 Objectives of the Research

Data collection and standardization: Aggregate and preprocess global real estate data (such as prices and transaction volumes) to ensure consistency and reliability.

Framework development: Design a standardized metadata framework to support the integration and analysis of diverse and multi-regional data.

Model construction: Develop and validate predictive models using time series analysis and machine learning algorithms.

Result interpretation: Visualize predictions, analyze model performance, and provide actionable suggestions for stakeholders.

1.6 Scope of the Study

Data source: Public datasets from government agencies, real estate platforms and research institutions.

Forecast indicators: Focus on the trend of housing prices, supplemented by transaction volume and rental levels.

Methods: Time series analysis (such as ARIMA) and machine learning techniques (such as LSTM, regression models).

Geographical focus: Select regions with different market dynamics to test the universality of the model.

1.7 Significance of the study

The contributions of this research to the academic and industrial fields are as follows:

Promote data standardization: Establish a unified framework to enhance data comparability and quality.

Enhance predictive capabilities: Provide scalable models for accurate global market forecasting and reduce investment risks.

Inform decision-making: Provide effective tools for residents, investors, policymakers and urban planners to enhance the quality of life and work.

Support interdisciplinary applications: Support business intelligence, risk analysis, and sustainable urban development through data-driven insights. (Mariia I. Ermilova,2021)