



黑龙江大学

HEILONGJIANG UNIVERSITY

ADD: Heilongjiang University, No.74, Xuefu Rd, Harbin, China, 150080

TEL: +86-451-86609033 / 86608417

## PhD Thesis Evaluation Report

**Name of Candidate:** Qi Dongfang

**Title of Thesis:** Research on Investment Attractiveness and Environmental Safety in China and Southeast Asia: Empirical Models and Data Analysis

**Supervisor:** Prof. Bure Vladimir Mansurovich, Saint Petersburg State University, D.Tex.

**Reviewer:** Prof. Tang Xiaomin, Heilongjiang University, D.Sc.

### Introduction

The research presented in this thesis addresses critical issues that are highly relevant in today's society, namely regional economic and environmental challenges. These two factors are pivotal in determining a region's development trajectory: economic strength dictates the level of growth, while environmental health ensures sustainability. Economic investment attractiveness, explored in the first two chapters, is crucial for fostering regional prosperity. By employing statistical methods such as multiple linear regression and cluster analysis, the study identifies key factors like per capita income and the cost of fixed assets, which play significant roles in different economic contexts. The subsequent chapters focus on air quality, a pressing environmental concern with direct implications for public health and well-being. The application of advanced machine learning techniques, including neural networks and ensemble models, showcases cutting-edge approaches to time series forecasting. The identification of significant factors like PM10 and CO through the use of SHAP values enhances the interpretability of these models, offering valuable insights into pollution dynamics. This research contributes meaningful solutions to ongoing economic and environmental debates, demonstrating how statistical and machine learning methods can be harnessed to address complex societal challenges effectively.

### Definition of the Research Problem, Scope, and Goals

The research problem addressed in this thesis revolves around the intertwined issues of regional economic development and environmental sustainability, as these challenges are crucial for determining both the growth potential and sustainable future of regions. This study primarily focuses on two key areas: economic analysis and environmental analysis.

The economic aspect examines investment attractiveness by applying statistical methods such as multiple linear regression and cluster analysis. This approach helps identify influential factors like per capita income and the cost of fixed assets across different economic contexts, providing insights for policy and strategic decisions. On the environmental front, the research delves into air quality issues using advanced machine learning models to forecast air quality and identify key pollutants affecting public health and environmental longevity. By employing neural networks and ensemble techniques, the study develops accurate predictive models that can aid in addressing pollution impacts. A significant goal of this research is to enhance the interpretability of machine learning models through SHAP values, which clarify the influence of critical factors like PM10 and CO on air quality predictions. By solving these interconnected problems, the research aims to offer methodologies and insights that effectively tackle regional economic and environmental challenges.

### **Review of Relevant Literature**

This thesis builds upon a foundation of existing research in the fields of economic development, environmental sustainability, and advanced analytical techniques. In examining regional economic issues, prior studies have extensively explored investment attractiveness, often highlighting factors such as per capita income and fixed asset costs as significant determinants. These elements are crucial for understanding how regions attract and sustain economic growth. In the realm of environmental studies, air quality has been a focal point due to its direct impact on public health and ecosystem stability. Previous research has identified pollutants like SO<sub>2</sub>, PM<sub>10</sub>, and CO as key contributors to air pollution, necessitating accurate modeling and prediction to mitigate their effects. Advancements in statistical methods, such as multiple linear regression and cluster analysis, have traditionally been employed to dissect these complex issues. However, recent literature emphasizes the growing importance of machine learning techniques for tackling multi-dimensional data problems. Neural networks, including ANN, RNN, LSTM, GRU, and their bidirectional variants, have shown promising results in time series forecasting, offering enhanced accuracy and efficiency. Moreover, ensemble models like LightGBM are gaining traction due to their superior predictive performance and speed, providing robust frameworks for environmental data analysis. The challenge of interpretability in machine learning models is well-documented, with SHAP values emerging as a valuable tool to elucidate the influence of individual predictors within these "black box" systems.

### **Presentation, Language, and Style**

The presentation of the thesis is well-executed, exhibiting minimal grammatical or spelling errors. The writing style maintains coherence and is suitably aligned with the standards of scientific research. Citations are accurately provided, and the references comprehensively address the current state of knowledge. The work is clearly articulated, effectively organized, and enriched with tables and diagrams to support the content.

### **Thesis Structure and Methods Used**

The thesis is structured into five chapters, employing a combination of statistical methods and machine learning techniques to explore regional economic and environmental issues. The first two chapters focus on economic topics, specifically investment attractiveness. Chapter one utilizes multiple linear regression to create annual models, followed by stepwise regression to identify key influencing factors, such as per capita income. Building on this, chapter two incorporates additional data using cluster analysis to divide regions into four groups. It then models each group separately, identifying significant factors like the cost of fixed assets for economically advanced areas and per capita income for less developed ones.

Chapters three through five concentrate on air quality studies. In chapter three, multiple linear regression is used to establish yearly models, with stepwise regression pinpointing SO<sub>2</sub> as a principal factor affecting air quality. Chapter four employs seven neural network models — ANN, RNN, LSTM, GRU, Bi-RNN, Bi-LSTM, and Bi-GRU — for time series forecasting of air quality. Among these, Bi-RNN demonstrates the highest accuracy and second-best timeliness, with all models showing excellent predictive performance aligning well with data trends. Chapter five uses an ensemble model approach with three different models to forecast air quality, highlighting LightGBM's superior accuracy and efficiency, making it the optimal model for this task.

### **Contribution to Knowledge and Practice**

This thesis makes significant contributions to both academic knowledge and practical applications in the fields of regional economics and environmental science. By integrating statistical methods and machine learning techniques, the research provides a comprehensive framework for understanding and addressing the dual challenges of economic development and environmental sustainability. In terms of economic analysis, the study enhances our understanding of investment attractiveness by identifying critical factors such as per capita income and the cost of fixed assets. This insight is invaluable for policymakers and stakeholders aiming to optimize investment strategies and boost regional economic growth. The innovative use of cluster analysis further refines these insights by tailoring strategies to different economic contexts. On the environmental front, the application of advanced neural network models offers robust tools for forecasting air quality. The identification of key pollutants like PM<sub>10</sub> and CO through interpretable machine learning models provides actionable data for environmental management and policy formulation. The introduction of SHAP values significantly improves model transparency, enabling better-informed decision-making. Overall, this research bridges theoretical advancements with practical implementations, offering methodologies that can be directly applied to enhance economic resilience and environmental health in various regions.

### **Recommendations**

1. Clarity and Structure: Ensure each chapter clearly outlines its objectives, methodologies, and findings to enhance readability and coherence.

2. Literature Integration: Expand the literature review to include more recent studies, providing a comprehensive context for your findings and demonstrating the research's relevance.

### **Overall Result**

The thesis is an original work that presents thorough research on regional economic and environmental issues through the application of statistical and machine learning methods. The author effectively analyzes how economic factors influence development height and environmental aspects impact sustainability. Demonstrating high competence in both areas, the candidate explores investment attractiveness and air quality with precision. The language and style are appropriate for scientific discourse, providing clarity and insight into complex data analysis and methodological approaches.

### **Questions to the defendant:**

1. How did author choose the specific economic factors for the regression models in chapters one and two?
2. In chapter 4, what methods did author employ to prevent overfitting in the neural network models?

### **Conclusion**

The dissertation titled "Research on Investment Attractiveness and Environmental Safety in China and Southeast Asia: Empirical Models and Data Analysis" meets the criteria for a comprehensive study in the fields of economic and environmental analysis. The candidate has effectively employed statistical and machine learning methods to explore the dual issues crucial to regional development: economic growth and environmental sustainability. This study significantly contributes to understanding the interplay between investment attractiveness and environmental safety, providing valuable insights with potential applications in policymaking and strategic planning for sustainable development in China and Southeast Asia. Based on these evaluations, I recommend the acceptance of this PhD thesis.

Prof. Dr. Tang Xiaomin  
Heilongjiang University



2024.10.10