Tomas Bata University in Zlín
Faculty of Applied Informatics

OPPONENT´S EVALUATION OF THE MASTER´S THESIS

Student: Pierre Pascal Bian Theke   **Opponent:** Assoc. Prof. Martin Kotyrba, Ph.D.

Study program: **Engineering Informatics**
Study course/Specialization: **Information Technologies**
Academic year: **2023/2024**

Master's Thesis topic: **Estimating Taxi Fares Using Machine Learning**

Evaluation of the thesis:

This master's thesis delves into the application of Artificial Intelligence (AI) algorithms for estimating taxi fares. By leveraging AI to estimate fares across various city zones, the synchronization of taxi fleets can be enhanced, thereby reducing waiting times for passengers. The thesis focused on the use of the Random Forest machine learning algorithm for this purpose. It includes a literature review on the evolution of AI in market analysis and price estimation, highlighting the transition to machine learning for comprehensive market surveys. The primary aim of this thesis is to develop a robust machine-learning model for accurately estimating taxi fares based on a variety of dynamic and static inputs.

The thesis is methodically structured into six core sections. It begins with an introductory chapter that sets the stage for the subsequent discussion. The second chapter elaborates on the fundamentals of machine learning. The third section offers an extensive literature review, analysing previous studies and models pertinent to the framework of this research. The fourth section details the techniques employed in fare estimation. The fifth section demonstrates the implementation of these theoretical constructs, while the final section presents empirical results, a thorough analysis, and the proposed taxi fare estimation algorithm. The final section also includes a critical summary of the findings, conclusions, implications for the study, and suggestions for future research. The work is 84 pages long, 64 sources of literature that are correctly cited. The work contains everything important that can be expected from a high-quality thesis. **I dare say that the work in this state definitely has a publication potential.**

The thesis results are very interesting. Random Forest Model displays a varied spread of predictions across actual fare amounts. It generally aligns with the actual fares but tends to underpredict higher fare amounts. The model shows superior performance metrics, including MSE, MAE, R-squared, MAPE, and overall accuracy. Neural Network Model exhibits tightly clustered predictions around zero, regardless of the actual fare amounts, indicating potential issues with underfitting or NN model parameters. This model does not perform as well as the Random Forest model.
The analysis highlights the strengths of the Random Forest model in handling large datasets and its ability to deal with non-linearity, which improves the predictability of taxi fares. However, it also notes areas for potential improvement, such as further tuning to reduce errors at higher fare ranges.

Questions for defense:
1. What were the main challenges encountered in adapting the Random Forest model for taxi fare estimation, and how were they addressed?

2. Is possible that external factors, such as weather conditions or special events, influence the accuracy of fare predictions, if yes, how?
3. What future improvements do you suggest for enhancing the model's accuracy and applicability?

In conclusion, this thesis of Pierre Pascal Bian Theke successfully demonstrates the potential of AI algorithms, particularly the Random Forest model, in accurately estimating taxi fares. The findings underscore the importance of robust machine learning models in optimizing pricing strategies and improving service quality in the taxi industry. Future research should focus on integrating additional data sources and refining model parameters to enhance predictive accuracy and operational efficiency. Thesis has highly level of processing and definitely I recommend this thesis for defense

Overall evaluation of the thesis:
The Opponent shall grant a mark according to the ECTS classification scale:
A – Excellent, B – Very Good, C – Good, D – Satisfactory, E – Sufficient, F – Insufficient
An "F" grade also means "I do not recommend the thesis for defence."

I recommend this thesis to be defended and suggest the following evaluation:

A - Excellent

In the case of an evaluation grade of "F – Insufficient", please supply the main shortages and reasons for this assessment.

Date: 24.5.2024   Thesis Opponent's Signature: