Jikhan Jeong

(Address) Hulbert Hall 313, Washington State University, Pullman, WA 99164-6210 **(E-mail)** jikhan.jeong@wsu.edu
(Homepage) https://sites.google.com/view/jikhan-jeong/main **(Github)** https://github.com/jikhans

1. A short description of the course of graduate study pursued, including major and minor fields of study, and the years of enrollment at relevant institutions.

I am a third year Econ Ph.D. student at Washington State University. My research fields are:

- ✓ Applied Microeconomics for Business and Energy Economics,
- ✓ Applied Microeconometrics for Discrete Choice and Deep Learning
- ✓ Natural Language Processing for Estimating Heterogeneous Consumer Preferences, Promotion Targeting, Financial Market Risk Prediction

2. A short description of general research interests

Innovation in end-to-end Artificial Intelligence (AI) will raise new questions about how to handle sequential choice problems in high dimensional feature space,[i] how to diminish the problem of market power,[ii] how to deal with bias in machine learning,[iii] how to design learning algorithms for smart device.[iv] In this regard, my research focuses are on how to apply Economics and Machine Learning (ML) to these research projects:

1. Estimating Individual Consumer Preferences on Smart Device Using Amazon Online Review Data
2. Applying ML to predict US Stock Market Risk Using New York Times News and Financial Indicator.
3. Using Game Theory to Determine How Big-data and AI Can Function as Entry Barriers.
4. Personalized Prediction for Coupon Redemption for Non-Habitual Purchase

3. Any research projects that are in process, and a dissertation title, short description,

- ✓ **"Estimating Individual Consumer Preferences on Smart Device Using User-Generated Content"**

Due to the increasing volume of consumer online review data, a firm may be able to identify individual-level heterogeneous consumer needs and preferences from user-generated online review data. Review data are usually unstructured data generated by experienced consumers to provide feedback about particular products. This paper will estimate individual consumer preferences regarding a programmable thermostat by using consumer review data found in the Amazon online shopping website. The purpose of this study is to build a predictive machine learning model that can accurately estimate an individual reviewer's overall evaluation of a purchased programmable thermostat. In contrast to the computer engineering approach for product recommendation, this study will focus on the relationship between user-generated content and consumer preference. This study will be useful for industry leaders and policy makers in the smart device market because user-generated data may reveal concerns and needs of experienced consumers. In particular, when a retail firm sells a new product such as a smart device that is different from the typical product, reviews by experienced consumers will provide useful product information for other consumers who are in the market for new smart devices. This review data is valuable because it contains feedback from actual user experiences rather than recommendations from focused user

groups and experts, and therefore reveals the actual preferences of experienced consumers of the product. Therefore, unexperienced and potential consumers can receive useful insights about a new product from the review data. In addition, the smart device firm can estimate individual consumer preferences, concerns, and needs from user-generated review data and use it in the process of new product design, product quality improvements, and consumer service upgrades. In this regard, this study analyzes unstructured user-generated content to draw useful insights for a smart device firm.

- ✓ **"Predicting Daily Stock Market Risk Using News "**

There has been a trend applying diverse machine learning methods to predict the stock market movement, in spite of high volatility and uncertainty of the stock market. In particular, many scholars have tried to implement end-to-end machine learning approach to predict the Dow Jones Industrial Average (DIJA) from Reddit WorldNews Headline Data by using a Natural Language Process (NLP). However, the data structure of this study is a time series data. In addition, previous studies have focused on predicting the stock price index using New Headline Data rather than using additional information from label data. Therefore, by using a time series analysis, this study predicts the daily movement of DIJA not only applying NLP using Reddit News Headline data but also using the lagged differenced opening stock price index as an independent variable based on time series analysis. By using Regularized Logistic Regression, Random Forest and Ada-boosting with Bayesian Hyperparameter Optimization, this study evaluates the effect of combining lagged difference opening stock price index with New Headline data on the predictive performance for the daily stock price movement compared with the predictive performance using only New Headline data. As a result, the predictive performance of each model using the lagged differenced opening stock price index is better than that of each model without using the lagged differenced opening stock price index. This fact indicates that by combing time series analysis with the machine learning approach, the predictive performance of machine learning for stock market movement can be improved.

- ✓ **“AI and Big Data in an Entry Game”**

Artificial intelligence (AI) can be used to differentiate products in the external business environment and to reduce costs in the internal business environment (with and without automation). AI in both the external and internal business environments is an increasing concave function of big data. In the entry game, the incumbent firm generates both shareable and un-shareable data in the external environment. However, the internal business environment data is not shared by either the incumbent or the entry firm because it is almost always confidential. Above all, this study assumes that the incumbent firm in the market has better AI algorithms than the entry firm. In addition, the entry firm must decide whether it will enter the market and compete with the incumbent firm on quantity. Using the above hypothetical setting, this study demonstrates

that when the entry firm penetrates the market, the amount of external business data shared by the incumbent with the entry firm affects the profitability of both firms. These results indicate how AI algorithms and big data can function as entry barriers and how sharing data from the external business environment can relax those barriers. Therefore, the results of this study may help industry leaders and policymakers understand how AI and big data can affect the market structure.

- ✓ **"Personalized Prediction for Coupon Redemption for Non-Habitual Purchase"**

I am working on data preprocessing with offline retail data.

4. Listing of committee members.

Advisor: Prof. Alan Love

5. ETC

- ✓ **(Homepage)** https://sites.google.com/view/jikhan-jeong/main

I wrote my current research projects and past research projects in my homepage.

- ✓ **(Github)** https://github.com/jikhans

Scratch version of deep learning and other discrete choice model code are in my Github, excluding the private version of the on-going research project code. I am using R for discrete choice and Python for Machine Learning, including Pytorch and Keras. For NLP, PyTorch is more suitable. For computing, I use Kamiak HPC rather than clouding services such as AWS. Kamiak HPC partitions contain GPU and can load PyTorch, Cuda, and Cudnn. One merit of HPC is that we can allocate enough memory for the certain sbatch submission, which often causes memory issues.

---

[i] Francisco J. R. Ruiz, Susan Athey, David M. Blei. SHOPPER: A Probabilistic Model of Consumer Choice with Substitutes and Complements. ArXiv 1711.03560. 2017.

[ii] Michael Jordan and Tom Mitchell. Machine learning: Trends, perspectives, and prospects. Science 17 Jul 2015: Vol. 349, Issue 6245, pp. 255-260. http://science.sciencemag.org/content/349/6245/255

[iii] The Trouble with Bias. Invited Talk at the NIPS Conference, 2017. Video:https://www.youtube.com/watch?v=fMym_BKWQzk

[iv] Urieli, D., & Stone, P. (2013, May). A learning agent for heat-pump thermostat control. In Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems (pp. 1093-1100). International Foundation for Autonomous Agents and Multiagent Systems.