Myeongsup Kim

+82 10-5597-6807 | msub0310@gmail.com

• Research Interests

Recommendation System

Natural Language Processing, Adversarial Training

Time Series Forecasting, Anomaly Detection

• Educations

Korea University Mar 2020 - Feb 2022

Master of Science in Industrial and Management Engineering
Data Science and Business Analytics Lab. [Hompage]
Research Field: Natural Language Processing, Anomaly Detection, Smart Manufacturing
Thesis: Text Embedding Augmentation based on Retraining with Pseudo-Labeled Adversarial Embedding
Advisor: Prof. Pilsung Kang

KonKuk University Mar 2015 - Feb 2020

Bachelor of Science in Industrial Engineering
Research Itern at Business Intelligence Lab. [Hompage]
Research Field: Text Analytics, Social Media and Patent Data Analytics
Advisor: Prof. Janghyeok Yoon

• Careers

Dable (Ynolja Y-Community) Mar 2022 - Present

Machine Learning Engineer

Dable is a company that provides personalized recommendations for content-type advertisement to the nation's largest media network, as well as provides a service that recommends news tailored to each user's interests so that they can stay for a long time.

I worked primarily as a machine learning engineer managing large-scale model training using over tens of millions data instances and machine learning researcher creating better performing models in low-latency environments.

• Publications

Myeongsup Kim, Pilsung Kang. (2022). "Text Embedding Augmentation Based on Retraining with Pseudo-Labeled Adversarial Embedding". *IEEE Access*. 9. 8363-8376. (SCIE) [Full Paper]

Myeongsup Kim, Jaeeun Choi, Huisik Min, Seounghyun Lee, Jongmin Lee, Janghyeok Yoon. (2020). "Identifying Determinants of Foreign Guests' Satisfaction and Management Strategies for Hotels in Seoul Using Online Data Mining". *The Academy of Customer Satisfaction Management*. 22(1). 1-24. (KCI) [Abstract] [Full Paper]

• Presentations

Myeongsup Kim, Pilsung Kang. (2021). "Text Embedding Augmentation based on Adversarial Training". *Spring Conference of Korean Institute of Industrial Engineers*. Jeju International Convention Centor, Jeju. 2nd June [Slide]

Myeongsup Kim, Youngjae, Park, Seongju Lee, Kwon-neung Lee, Jaeeun Choi. (2019). "Detecting Emerging Customer Needs Using Social Media Time Series Prediction: Emerging Keyword Detection Approach Based on Word Embedding, Network Analysis, and LSTM". *Student Project Competition of Korean Institute of Industrial Engineers Fall Conference*. Seoul National University, Seoul. 8th November [Slide]

Myeongsup Kim, Huisik Min, Jaeeun Choi, Jongmin Lee, Seoughyeon Lee. (2019). "Identifying Determinants of Foreign Guests' Satisfaction and Management Strategy for Hotels in Seoul Using Online Data Mining". *KMAC Management Innovation Research Paper and Case Study Competition*. Yeouido CCMM Building. Seoul. 2nd November [Slide]

Youngjae Park, **Myeongsup Kim**, Kisoo Um. (2019). "Opportunity Analysis for Smart Speaker Products Using Social Media Data Mining". *Society of Korea Industrial and Systems Engineering Project Competition*. Kongju National University. Cheonan, 3rd May [Slide]

• Projects

Developing DSP based on Time Series CTR Prediction May 2020 - Aug 2020

Affiliated by Dable | **Project Lead of 6 Members**

Predicted the probability that users click on an advertisement (click throught rate; CTR) using time series model.

Developed statistics-based models, DLinear, NLinear, PatchTST, Prophet, and etc.

Improved revenue per 1000 impressions (revenue per mille; RPM) through a combination of statistics-based models and ML/DL-based models.

Developed a demend-side platform (DSP) with RPM that exeeds the RPM of existing DSP by 7.73%.

Expected 2.78% increase to company-wise revenue through the project.

Tools: PyTorch, AWS Athena, AWS S3, Ray Tune, Redash

Models: Statistic-based model, DLinear, NLinear, PatchTST, Prophet

Developing Pytorch-based Deep Learning Model Training Pipeline Mar 2020 - May 2020

Affiliated by Dable | **Project Lead of 1 Member and Assisted by 1 Member**

Developed a workflow pipeline to train a pytorch-based deep learning model.

Developed click through rate predict deel learning models including MaskNet, AutoInt, DeepFM, DLRM, and etc.

Developed pipeline employing tens of millions of training instances using ray and athena.

Applied optimization so that the training process can be completed within 10 hours.

Improved offline gAUC performance by 0.05 point compared to previous model.

Tools: PyTorch, AWS Athena, AWS S3, Ray Data, Redash

Models: MaskNet, AutoInt, AutoInt+, DeepFM, DLRM, MLP, Transformer, FT-Transformer

Developing External SSP Jan 2020 - Apr 2020

Affiliated by Dable | **Project Member**

Developed a supply-side platform (SSP) that enables bidding on external media such as Kakao, Google, MSN and etc.

Developed LGBM, a small-sized machine learning model for low latency.

Developed automatic tuning process based on ray tune.

Developed machine learning-based calibration model to make the output of model reflect the actual confidence.

Achieved revenue of over X0,000,000₩ on Kakao media during April 2020.

Tools: LightGBM, Ray Tune, Scikit-learn, AWS Athena, AWS S3, Redash

Models: LightGBM, Isotonic Regression

Participation in Team Kaggle and DSP Development based on Kaggle Result Sep 2022 - Dec 2022

Affiliated by Dable | **Project Member**

Building a baseline model for Kaggle and established kaggle rules and construct dataset.

Submitted more than 10 models during the 1-month kaggle period, tied for 2nd place of 9 members. (0.09 Point Improvement in gAUC)

Developed ensemble of ML/DL, MaskNet, Focal Loss, and etc.

Developed deep learning model training pipeline using Tensorflow without Tensorflow Extended (TFX).

Tools: Tensorflow, TFRecord, Ray Tune, LightGBM, Redash

Models: DeepFM, MLP, MaskNet, LightGBM, Transformer

Improving Deep Learning based DSP May 2022 - Aug 2022

Affiliated by Dable | **Project Member**

Detecting appropriate offline evaluation metric for maximizing RPM (e.g. gAUC, rAUC, sAUC, and etc).

Logged training information to MLflow, versoned model, and managed training parameters.

Checking the degree of model performance degradation according to the training interval.

Created and managed dashboards to check model performance.

Logged SHAP values to check model status.

Tools: Tensorflow, Tensorflow Extended, TFRecord, MLflow, Redash, AWS EC2, AWS Route 53, AWS ECS, AWS ECR

Models: DeepFM, Focal Loss, gAUC Metric, SHAP

Detecting defective equipment in the semiconductor manufacturing process Jul 2021 - Nov 2021

Granted by SK Hynics | **Project Lead of 1 Member**

Developed a regression model to predict semiconductor yield.

Calculate the SHAP value of the equipment for each predicted value of the model.

Using the SHAP value, calculate the influence of each equipment and detect the abnormal equipment.

Golden Path Detection in Manufacturing Process Mar 2020 - May 2021

Granted by Samsung Electronics | **Project Member**

Developed sequential pattern mining based model to detect combinations of process equipment that negatively affected semiconductor yield.

Developed metrics for validating and evaluating model performance.

• Managements

Enterprise Level Presentation Dec 2022 & Feb 2020

Affiliated by Dable

Enterprise-level sharing of DSP development based on AI (Feb 2020).

Engineering organization level sharing of DSP development based on AI (Dec 2022).

Onboarding New Members Sep 2022 - Oct 2022 & Jul 2020

Affiliated by Dable

Renewed onboarding materials and documents

Managed the onboarding of two members and supported the onboarding of two members.

Advancing Recruitment Process and Participating in Technical Interviews Jun 2022 - Sep 2022

Affiliated by Dable

Designed recruitment question for machine learning.

Participated in 50+ recruitment document screenings.

Participated in 10+ recruitment technical interviews.

Questioned about details of career and machine learning fundabmentals.

• Teaching Experiences

SK Hynics Jul 2021 - Nov 2021

Project Assistant
Developed manufacturing equipment detection model that caused low wafter yield

LG Innotec Sep 2020 - Nov 2021

Project Assistant (Two Teams)
Developed manufacturing defect type prediction model
Developed manufacturing process yield prediction model

Hyundai Steel Jul 2021

Programming Practice Assistant
Assisted python programming practice course on anomaly detection

LG Innotek Jul 2021

Programming Practice Assistant
Assisted python programming practice course on machine learning calssifier

LG Energy Solution Jun 2021

Programming Practice Assistant
Assisted python programming practice course on anomaly detection

LG Chemical Jul 2021

Programming Practice Assistant
Assisted python programming practice course on ensemble, DNN, and CNN

LG Innotek Jul 2021

Programming Practice Assistant
Assisted python programming practice course on machine learning classifier

• Knowledge Sharing

Prompt-Based Learning | *Seminar* [Video] [Slide]

What Changes Can Large-scale Language Model Bring? Intensive Study on HyperCLOVA: Billions-scale Korean Generative Pretrained Transformers | *Group Study* [Video] [Slide]

Transformer: Complexity, Parameters, and Scaling | *Group Study* [Video] [Slide] [Github]

Transformer Survey: Architecture Level Variants | *Group Study* [Video] [Slide] [Github]

Learning to Perturb Word Embeddings for Out-of-distribution QA | *Group Study* [Slide]

Adversarial Examples Improve Image Recognition | *Seminar* [Video] [Slide]

SMART: Robust and Efficient Fine-Tuning for Pre-trained Natural Language Models through Principled Regularized Optimization | *Seminar* [Video] [Slide]

FreeLB: Enhanced Adversarial Training for Natural Language Understanding | *Seminar* [Video] [Slide]

Syntax and Semantics in Language Model Representation | *Seminar* [Video] [Slide]

Semantics Aware BERT for Language Understanding | *Seminar* [Video] [Slide]

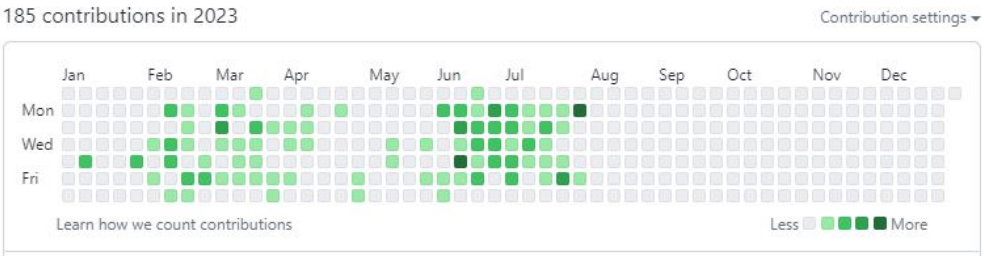
Vokenization: Improving Language Understanding with Contextualized, Visual-Grounded Supervision | *Group Study* [Slide]

• Skills

Programming	Python, SQL
Libraries	Pytorch, Tensorflow, Huggingface, Scikit-Learn, Numpy, Pandas, Matplotlib
Clouds	AWS, Athena, EC2, S3, ECS, ECR, Route 53
Others	Git, Docker, Ray, MLflow, Redash, Prefect, Airflow

• Source Code Contribution Footprints

185 contribution to Dable in 7 months of 2020
34% of contribution belongs to code review in this year
Commits made during this year are squashed



631 contribution to Dable in 10 months of 2022
3 month from march corresponds to the onboarding period