Myeongsup Kim

+82 10-5597-6777 | msub0310@gmail.com | LinkedIn

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
Researcher at 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 (Yanolja Y-Community) Mar 2022 - Present

Dable is a startup 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.

Part Lead Mar 2024 - Present

I worked as a part lead managing AI team of 7 members. My responsibilities included communicating with other teams, code implementation, code reviews, setting work direction, prioritizing tasks, and motivating team members.

Machine Learning Engineer Mar 2022 - Mar 2024

I worked primarily as a machine learning engineer managing large-scale model training pipeline 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) [Link]

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) [Link]

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

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

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

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

Projects

Advancing Mobile App DSP July 2024 - Present

Affiliated by Dable | **Project Lead of 5 Members**

This project was organized directly under CEO and was bi-weekly reported to C-level including CEO and CTO.
Increased advertiser return on ad spend (RoAS) by making use of user information in the mobile environment.
Discussed about the initial data management platform (DMP).
Developed ML/DL models and algorithm to aquire last clicks in the mobile advertising environment.
Performed A/B test on weekly basis to increase advertisement performance.

Tools: PyTorch, Redis, AWS Athena, AWS S3, FastAPI, Airflow
Models: DCN, ESMM, ESCM^2, LightGBM
Impact: Achieved gross profit over tens of millions KRW during September 2024.

Advancing Deep Learning CPA DSP May 2024 - Present

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

This project was organized directly under CEO and was bi-weekly reported to C-level including CEO and CTO.
Set target cost per action (CPA) for each advertisers and developed system to achieve CPA below target CPA.
Designed and operated ML/DL models that reflecting the importance of conversion.
Developed logic to set bid price considering the competitive environment of real-time bidding.
Developed system that bid considering the advertiser's target CPA.
Developed server to reduce system complexity and latency.
Established collaboration system between members in the business division and ML Engineers.

Tools: PyTorch, Redis, AWS Athena, AWS S3, FastAPI, Airflow
Models: DCN, ESMM, ESCM^2, LightGBM
Impact: Achieved 45% of advertisers meeting target CPA during October 2024.

Feature Engineering Taskforce Jan 2024 - Mar 2024

Affiliated by Dable | **Project Member**

This project was organized directly under CEO and was weekly reported to C-level including CEO and CTO.

Processed features to improve the click and conversion prediction performance of deep learning models.

Was responsible for feature processing for NLP domain , developed in-house fine-tuned language models for feature generation.

Refined unsupervised text representation and text category features.

Developed wrapping server to use the external LLM API for feature processing.

Tools: PyTorch, HuggingFace, Trainer, Peft, OpenAI, Cohere, AWS Athena, AWS S3, FastAPI

Models: BERT, RoBERTa, SimCSE, LoRA, Adversarial Training

Impact: Established of a company-wide collaboration system to discover features.

Developing DSP for Mobile Application Environment Sep 2023 - June 2024

Affiliated by Dable | **Project Lead of 2 Members**

Launched product named ***wheres.ai*** [link]

Developed ML/DL models to predict customer purchase for mobile app advertisement.

Developed LightGBM for rapid initial development, later developed deep learning models for higher performance.

Processed the features and datasets necessary for model development from the original logs.

Developed initial server required for serving ML/DL models.

Developed and operated an algorithm that efficiently captures clicks over time.

Tools: PyTorch, LightGBM, AWS Athena, AWS S3, Ray Data, Ray Tune, Redash, FastAPI

Models: LightGBM, DLRM, ESMM, ESCM^2

Impact: Achieved 8,000% RoAS for the Yanolja advertiser in May 2024.

Developing DSP based on Time Series CTR Forecasting May 2023 - Aug 2023

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

This project was organized directly under CEO and was weekly reported to C-level including CEO and CTO.

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

Developed statistics-based models, DLinear, NLinear, 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%.

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

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

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

Developing Pytorch-based Deep Learning Model Training Pipeline Mar 2023 - May 2023

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

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

Developed click through rate prediction deep 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.

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

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

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

Developing External SSP Jan 2023 - Apr 2023

Affiliated by Dable | **Project Member**

This project was organized directly under CEO and was weekly reported to C-level including CEO and CTO.

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.

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

Models: LightGBM, Isotonic Regression

Impact: Achieved revenue of over tens of millons KRW on Kakao media during April 2020.

Participating in Team Kaggle and Developing DSP 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.

Detecting Golden Path 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

Team Management Mar 2024 - Present

Affiliated by Dable
Responsible for communication with other teams and management within the company.
Adjusting work priorities and motivating team members.

Enterprise Level Presentation Dec 2022 & Feb 2023

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).

Advancing Recruitment Process and Participating in Technical Interviews Jun 2022 - Mar 2024

Affiliated by Dable
Designed recruitment question for machine learning.
Participated in 200+ recruitment document screenings.
Participated in 10+ recruitment online phone interview.
Participated in 30+ recruitment on-site 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

The Generation of Spark Vol.1 | *Blog Posting* [Blog]

Improving Fine-tuning Performance of Language Model | *Blog Posting* [Blog]

Personalized Recommendation System based on Machine Learning | *Blog Posting* [Blog]

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]

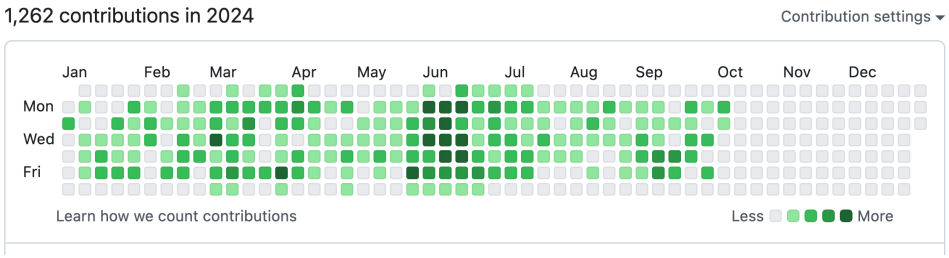
Vokenization: Improving Language Understanding with Contextualized, Visual-Grounded Supervision | *Group Study* [Slide]

Skills

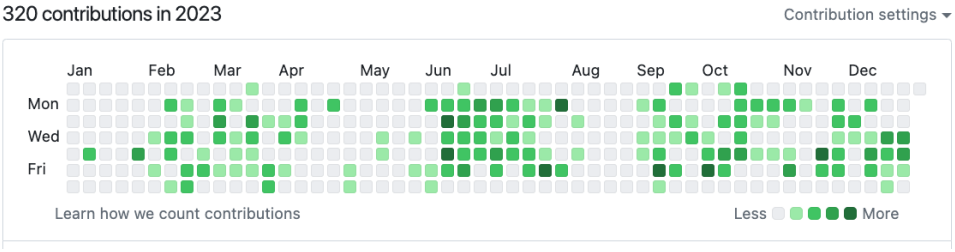
Programming	Python, SQL
Libraries	Pytorch, Tensorflow, Huggingface, Scikit-Learn, Numpy, Pandas, Matplotlib
Clouds	AWS, Athena, EC2, S3, ECS, ECR, Route 53
Others	Git, Docker, Ray, MLflow, Redash, Prefect, Airflow, FastAPI

Source Code Contribution Footprints

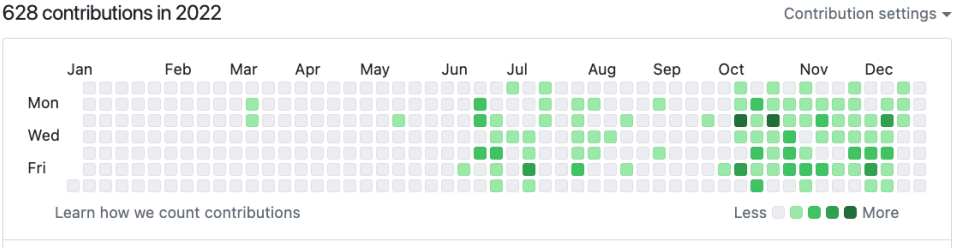
1,262 contribution to Dable in 10 months of 2024
18% of contribution belongs to code review in this year



185 contribution to Dable in 2023
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 months from march corresponds to the onboarding period



lastest update: 2024-10