DONGWON KIM

SOFTWARE ENGINEER

| dkim010@gmail.com | dkim010 | dongwon-kim-68235a35 | Google Scholar

Summary

- **Fields**: Distributed computing, data engineering, recommender system
- **Programming Languages**: Python, C, Bash, Scala, Java
- **Skills**: Aerospike, HBase, Spark, Impala, Hive, Presto, AWS, Docker

Work Experience

Senior Software Engineer   May 2022 - Present

TOSS OF VIVA REPUBLICA,   Seoul, Korea

Viva Republica operates financial super app Toss including payment services, financial services, and stock brokerage with over 15 million active users in Sep. 2023.

- **Machine Learning Platform**
  - Improving the development culture of my team

    Bringing together projects where individuals were working on different roadmaps, different code styles, and different patterns so that they were unified and on the same page.

  - A DAG factory framework to easily develop DAGs for big data processing without data engineering knowledge

    Many Airflow operators are supported like Redis operator with YAML-formatted configurations for low code. : Airflow, Impala, Spark, Kubernetes

  - An in-house feature store which utilizes a variety of storage to store and retrieve features for offline/online environments

    It is being used by ad services for real-time inference. : Spark, Kubernetes, HBase, Aerospike, Kotlin, Spring

  - A recommendation platform that allows service providers to define the scoring formula based on various types of items

    : Spark, Redis, Kotlin, Spring

  - Design/Developed Multi-armed Bandits platforms for recommendations of shopping.

    : Spark, Redis, HBase

Technical Leader   Mar. 2020 - May 2022

FINANCE SERVICE OF NAVER FINANCIAL CORP. IN NAVER CORP.,   Sungnam, Korea

NAVER FINANCIAL operates financial platform including online payments, loan services from Nov. 2019, and the number of monthly active users is over 13 million in Mar. 2021.

- **Data Pipeline / Analysis**
  - Building and sustaining a secure Hadoop environment

    The Hadoop has about two hundreds user : Hadoop, Kerberos, Ranger, Ansible, OpenWhisk

  - Automatically collecting the massive and various formatted data from diverse data sources

    Only configurations are required when ingesting new data without source codes : Hive, Compactor, OracleDB, MySQL, MongoDB, Hive, Kafka Connect, Airflow, Spark

  - For target marketing, identifying users on specific conditions such as vehicle owners, potential customers, and so on

    We profile users for preparing diverse finance services : LightGBM, Lift Score, Bayesian Optimization

  - Analyzing financial and commercial data for promotion activities of NAVER FINANCIAL services

    : Propensity Score Analysis, Casual Discovery

Technical Leader
FLO Music Service of Dreamus Company in SK Telecom,

Dec. 2018 - Mar. 2020

Seoul, Korea

SK Telecom is the largest mobile operator in Korea, and FLO music service has 1.3 million memberships.

- **Music Recommender System**
  - Migration to AWS from on-premise infrastructure

  We want to get scalable big data platform, under 100 engineers were using this infrastructure : EC2, VPC, S3, EMR, Aurora, Presto, Redash, CloudWatch
  - Design/development a data pipeline including log data and music metadata

  There are about 10 GB of daily log data and 20 GB of music metadata : S3, Glue, Hive, Spark, Presto, Sqoop, Docker
  - Personalized playlist generation with trained models

  27 percentages of users listen with these playlists in Aug. 2019, though under 5 percentages of users listened in Jan. 2019 : Fasttext, Faiss, Spark
  - Continuous extraction of mood features from songs with music information retrieval techniques

  Mood information has been used for the recommendation based on users' mood preference : Tensorflow, Lambda, ECS, SQS, DynamoDB, S3, Glue

Software Engineer
Music Streaming Service in NAVER Corp.,

Aug. 2015 - Nov. 2018

Sungnam, Korea

NAVER is Korea's top search portal, and NAVER music streaming service has a membership of over 0.8 million in 2018.

- **Music Recommender System**
  - Development of a pipeline to extract mood features from song

  That pipeline extract those features as soon as a song is released newly : Python, PyArrow, RabbitMQ, Nomad, Consul
  - Development of a real-time recommender system (ratio station) applying user's listening feedback

  About 20 percentage of users were listening with this continuous ratio station in Dec. 2018 : Python, Redis, Similarity Search, Matrix Factorization, Gensim, Fasttext, Annoy, Faiss
  - Recommendation songs not listened yet applying a mixture of collaborative/content-based filtering

  About 10 percentage of users were listening discovered songs in Dec. 2018 : Spark, Hive, Matrix Factorization
  - Design/development of A/B testing platform *: HBase, Spark, SpringBoot*

Education

M.Sc. in Computer Science
Yonsei University, *one of Korea's top-three universities, considered the most prestigious in the country.*

Mar. 2011 - Aug. 2015

Seoul, Korea

"Content-centric energy management for mobile devices" (Advisor: Prof. Hojung Cha)

B.S. in Computer Science
Yonsei University

Mar. 2007 - Feb. 2011

Seoul, Korea

Selected Publications

Deep Learning (3 papers)

- B. Jeon, A. Kim, C. Kim, **D. Kim**, J. Park, J. Ha, "Music Emotion Recognition via End-to-End Multimodal Neural Networks," *Proceeding of the 11th ACM Conference on Recommender Systems* (RecSys 2017), Aug. 2017

System Engineering (12 papers)

- **D. Kim**, N. Jung, Y. Chon, H. Cha, "Content-Centric Energy Management of Mobile Displays," *IEEE Transactions on Mobile Computing* (TMC), DOI: 10.1109/TMC.2015.2467393
- C. Yoon, **D. Kim**, W. Jung, C. Kang, H. Cha, "AppScope: Application Energy Metering Framework for Android Smartphone using Kernel Activity Monitoring," *Proceeding of 2012 USENIX Annual Technical Conference* (USENIX ATC), Jun., 2012.