Tae Hyon Lee

San Francisco, CA   (415) 832 - 9291
https://leeth7830.github.io/   leeth7830@gmail.com

Skills

Languages: Java, Python, SQL(Hive, Presto, PostgreSQL)
Tools: AWS (S3, EMR, DynamoDB, Athena, Lambda, Sagemaker), Docker, Oozie, Jupyter, Zeppelin
Others: Spark, Hadoop, Hive, Kafka, NoSQL, Machine learning, Tableau, RESTful API, Communication

Experience

Associate Data Engineer, Lending Club, San Francisco, CA   Jan 2020 - May 2020
- **(Data Pipeline)** Designed, implemented, and maintained efficient & reliable & self-serviceable data streaming and batch process pipelines to move and transform data using Java, Python, Kafka, and Spark and later injected into graphs and machine learning models.
- **(Self-service & Automation)** Developed internal automation applications and frameworks to unlock access to data and enable self-service for stakeholders across Marketing, Product, Data Scientist & Engineering teams.
- **(Design and Architecture)** Worked cross-functionally with business, data science, and engineering teams to define problem statements, collect data, and make recommendations.

Internships

Data Engineering Intern, Lending Club, San Francisco, CA   June 2019 - August 2019
- **(Data Pipeline Optimization)** Optimized and maintained existing pipelines using Spark, built data quality measures using dashboards and data validator, and set up pipeline integrity checks.
- **(Adoption and Migration)** Worked with different engineering teams to enable adoption and migration of our platforms and integration of our systems and platforms with other products, and manage upstream and downstream dependencies.

Data Science Intern, MonJa, Oakland, CA   July 2017 - December 2017
- **(ETL, Data Mining, Data Modeling)** Built and extended MonJa's data stores using Spark scripts and proposed new data structures for data storage and code that will lead to more effective and efficient use of market data by MonJa's product software and modeling toolkit.
- **(Statistical Predictive Modeling)** Prepared clean data and trained a machine learning model that identifies significant variables that impact risk and determines whether a borrower will go delinquent in the first six months.
- **(Report Generation, Analysis)** Prepared slice-and-dice Tableau reports of modeled data to support investors and senior research members' deep analysis of the lending marketplace.

Database Administrator Intern, Moneycorp, San Francisco, CA   May 2017 – June 2017
- **(Data Quality Assurance)** Wrote scripts that transform and filter raw data into a usable format.
- **(Optimization)** Restructured and optimized data storage to eliminate redundant data and achieve scalability and efficiency.

Education/Training

MS, Computer Science University of San Francisco   2018 - 2019
GPA: 3.7, Coursework: Big Data, Algorithms, Software Development
MBA, Information Systems San Francisco State University   2016 - 2017
Coursework: Behavior, Organization, and Management, Managerial Decision Making
Training: SuperDataScience - Machine Learning A-Z™: Hands-On Python & R In Data Science
BA, Economics (w/ specialization in programming) University of California, Los Angeles   2012 - 2015
Coursework: Intermediate Programming, Discrete Math, Algorithms Statistics for Economists