Haneul Kim

24, Imok-ro, Jangan-gu, Suwon-si, Gyeonggi-di, Republic of Korea
Email: Haneulkim214@gmail.com
Website: https://haneulkim214.github.io
LinkedIn: https://www.linkedin.com/in/haneul-kim-8b8720164/

Specialization

- Python (sklearn, nltk, pandas, numpy, plotly, pymysql, boto3 and more)
- SQL and NoSQL (MySQL, Oracle, MongoDB, DynamoDB)
- AWS (Elasticsearch, Kibana, and DynamoDB)
- Mathematics (Linear Algebra, Chaos Theory)
- Statistics (Probability, Bayes statistics)

Experience

Olulo (Micro-mobility service) **Sep.2019 - Current**
Data Engineer

- Implemented machine learning ensemble models to optimize operation costs to more than 20%.
- Data mining, analysis and visualization to create statistical experiment about whether giving out coupon at certain time increase re-use rate the most. Proved experiment to be significant using hypothesis testing and with the conclusion increased over 25% re-use rate.
- Established efficient data pipeline using AWS services for data analysis.
- Providing fully automated ELK service dashboard for daily report and error tracking.

Akashiro Japanese Restaurant **Aug.2016 - Jun.2017**
Customer Service

- Served around 150 customers daily, significantly improved communication skills and successfully increased rating by 50% from beginning.
- Organized daily sales report and calculated total cash/card amount in MS Excel.

Projects

Air Quality Visualization: Input a country, receive summarized information and map with air quality index.

- Automated data mining in Wikipedia and summarize using natural language technique term frequency-inverse document frequency.
- Each city in given country calls API and store result into MySQL database using sqlAlchemy.
- Created flask-scheduler for parallel programming that updates MySQL database in the background while maintaining successful running app.
- Developed using MySQL, Python (sqlAlchemy, flask, pandas, and nltk), html, CSS, JavaScript (d3, jQuery, and leaflet).

Keyword Finder: Project has been built out of own laziness of reading through multiple job descriptions and figuring out what skills are needed.

- Given job title and location program will visualize using charts, most frequent words that appeared in all the job descriptions.
- Built fully automated data mining, data transformation, and data visualization program.
- Administered natural language processing tools to clean and tokenize data.
- Machine learning algorithms such as count vectorizer to design bag of words and considered both unigram and bigram words for most frequent words' visualization.
- Link: https://github.com/HaneulKim214/Keyword-Finder

Amazon sentiment analysis

- Attained csv data, standardized, cleaned data.
- Machine learning algorithms with full understanding of its statistical model and accuracy models.
- Developed using Python (sklearn, pandas, and numpy) statistical models (Logistic regression and AUC_ROC curves)
- Link: https://haneulkim214.github.io/templates/amazon_review.html

Education

University of Toronto **Jan.2019 – July.2019**

Data Analytics Boot Camp

- Intensive 6-month boot camp focused on state-of-the-art data analytics and visualization tools.
- Attained knowledges in Hadoop, SQL, NoSQL, Cloud services, Tableau, and Power BI.
- Successfully built full-stack data visualization apps using Python, SQL, JavaScript, and Hadoop.

University of Toronto **Sep.2013 – May.2018**

Bachelor of Science: Mathematics, Psychology

- Relevant Courses: Linear Algebra I&II, Chaos theory, Differential equations, Statistics, Probability, Combinatorics, Differential equation, and Abstract Mathematics.

Chinese University of Hong Kong **Sep.2015 – May.2016**

Bachelor of Science: Mathematics, Psychology

- Full scholarship exchange student.