

Big Data Engineer

Sunghwan Ki (Danny)

Phone : 205-734-9654
E-Mail : kish1919@gmail.com
Tech_Blog : kish191919.github.io
Portfolios : gooddatadanny.com/
Linked-In : linkedin.com/in/danny-ki/
Git Hub : github.com/kish191919
Address : Birmingham, AL

Experience

- PNC Bank | Big Data Engineer** *Birmingham, AL* **Jul. 2019 – Present**
- Develop the Data Warehouse ETL process by using Python and Linux Shell
 - Work on analyzing Hadoop cluster and different big data analytic tools including Hue, Hive, and Impala
 - Implement and automate job work-flow scheduling and monitoring in test and production environment by CA7
 - Work on the production support for scheduled ETL daily and monthly jobs and investigate and resolve any process errors or code defects
 - Work closely with the data architecture team for the creation, maintenance, and validation of data dictionaries and schemas to improve the data quality
 - Responsible for data ingestion, manipulation, and transformation of raw and master files per business use case using several application tools such as Jira, Bitbucket, Jenkins, and IntelliJ
- Hyundai Powertech America | Production Control Specialist** *West Point, GA* **Jul. 2018 - Jul. 2019**
- Created a machine learning model with Python that optimizes the manufacturing process
 - Conducted a data regression analysis of the relationship between irregular usage of parts and production processing condition, achieving a 6% cost reduction than 2017
 - Analyzed machine recalls and attribute them to specific part failures and find methods to avoid such defects
- Donwon Autopart Technology Georgia LLC | Sales and Analytics Manager** *Hogansville, GA* **Jul. 2015 - Dec. 2017**
- Analyzed vast amounts of complex pricing-related data
 - Demonstrated efficiency of 2D barcode adoption in logistics management through data analysis
[Since setting up the 2D Barcode system, the incidence of mislabel and late delivery reduced by 94%]
- Kukdo Chemical Co.,Ltd | Purchasing Assistant Manager** *Seoul, South Korea* **Dec. 2013 - Jul. 2015**
- Analyzed price proposals, financial reports, market data, and other data in the chemical industry using Excel
 - Utilized web scraping techniques to extract and organize competitor and supplier data using Python
- Lotte Chemical Alabama Corp | Purchasing Specialist.** *Auburn, AL* **Oct. 2011 -Nov. 2013**
- Analyzed ICIS and Platts data to negotiate with suppliers to get a fair price and conditions
 - Analyzed cost of material to estimate alternatives that satisfied with a limited budget (\$800K)

Education

- University of Wisconsin-La Crosse | Master of Science - Data Science** **Jan. 2019 - Present**
- Cumulative GPA: 3.5 / 4.0
 - Relevant Coursework: Data Warehousing, Machine Learning, Data Analysis, Data Mining, Prescriptive Analytics
- Fast Campus | Data Science Intensive** **Jan. 2018 - May. 2018**
- Completed intensive, five-month full-time career accelerator course with 800+ hours of professional training and practicum experience covering the full range of data science practices and tools
 - Relevant Coursework: AL/Machine Learning, NLP, AWS, Hadoop, Hive, Statistics
- Udacity | Programming Nanodegree - Data Analyst** **May. 2017 - Aug. 2017**
- Project-based programming credential
 - Learned the practice of determining patterns from data to predict future outcomes and trends
- Myoungji University | Bachelor of Business Administration** **Mar. 2000 - Feb. 2009**
- GPA: 3.6 / 4.5

Key Skills

- Programming Language: Python, SQL, Linux
- Big Data Platform: AWS, Spark, Hadoop, Hue, Sqoop, Oozie
- Application: Jenkins, Bitbucket, Control-M
- Database: MySQL, Google Big Query
- Analysis : Tableau, Hive, Impala
- Toolkits: Git, IntelliJ, JIRA

Data Projects [Portfolio Website: <http://gooddatadanny.com/>]

Vehicle Value Forecast Web Services [<http://usedcar.ga/>]

- Built a web scraper to gather data and stored in AWS and XGBoost model to predict the vehicle value
- Tech: Data Scraping & Cleansing, Machine Learning, AWS, MySQL, Flask, Bootstrap, High Chart, Nginx

Predict House Prices | Kaggle Competition

- Found features that have the influence on house values and then predicted it by Ordinary Least Squares Model
- Tech: Data Cleansing & Visualization, Linear Regression, Statistics, Python (Numpy, Pandas, Statsmodel)

Spooky Author Identification | Kaggle Competition

- Text analysis competition. Identify an author from sentences by Natural Language Process and Naive Bayes Model
- Tech: NLTK (Natural Language Toolkit), Naive Bayes Classification, Python (Numpy, Pandas, Scikit-learn)

Digit Recognizer | Kaggle Competition

- Image analysis competition. Identify digits from a dataset of handwritten images by Neural Network using Keras
- Tech: Keras, Convolutional Neural Network, Python (Numpy, Pandas, Scikit-learn)

Certificates

- AWS Certified Developer – Associate (Issue date : Jul. 2021)
- AWS Certified Cloud Practitioner (Issue date : Apr. 2021)

Additional Training

- Udemy : Snowflake - The complete masterclass (Nikolai Schuler / Dec.2021)
- Udemy : Data Analyst using Sqoop, Hive and Impala (Mohamed F./ Nov.2021)
- Coursera : Data Structures (*Alexander S. Kulikov* / Jul. 2018)
- Coursera : Java Programming (*Susan H Rodgers* / Jul. 2018)
- Coursera : Probability and data (*Mine* / Aug. 2017)
- Coursera : Machine learning (Andrew Ng / Jul. 2017)
- Coursera : Machine learning foundation (*Carlos, Emily* / Jul. 2017)
- Coursera : Python programming (*Charles Severance* / Jul. 2017)