Benjamin DK Luong

https://benjamindkluong.github.io/
benjamindkluong@gmail.com | 913.244.9507

EDUCATION

UNIVERSITY OF MISSOURI, COLUMBIA
MS IN DATA SCIENCE – 3.9/4.0
May 2020 | Columbia, MO

UNIVERSITY OF MISSOURI, KANSAS CITY
BA IN FINANCE – 3.73/4.0
Graduated Dec 2015 | Kansas City, MO

SKILLS

DATA SCIENCE
TensorFlow • Keras • Scikit-Learn • Numpy • Pandas
•Experience in machine learning and deep learning techniques: regression and classification analyses, clustering, neural networks, etc.
• PCA, k-NN, Naive Bayes, SVM, Decision Forests, CNN, RNN

VISUALIZATION
Matplotlib • Seaborn • ggplot • Tableau

PROGRAMMING
Python • R • HTML • CSS • JavaScript • VBA

DATABASE
SQL • MongoDB • Postgres • SSIS

TECHNOLOGY
Hadoop • Docker • Git • Bitbucket

CHARACTERISTICS:
Problem solver • Predictive analytics
• Using machine learning, statistical modeling to solve business problem
• Strategic thinker, creative, self-motivated and entrepreneur mindset
• Researching and building machine learning application
• Providing data-driven recommendations and insights

LINKS

Github:// BenjaminDKLuong
LinkedIn:// benjamin-dk-luong
Site:// benjamindkluong.github.io

CERTIFICATES

Data Analyst Nanodegree
Machine Learning Nanodegree

WORK EXPERIENCE

DTE ENERGY | SENIOR DATA SCIENTIST
Sep 2022 – Present | Detroit, MI
- Working on various machine learning projects to prevent outages

CHARTER COMMUNICATIONS
DATA SCIENCE ANALYST | Billing Assurance Department
Aug 2021 – Sep 2022 | Maryland Heights, MO
- Worked on data validation and data migration from ICOMs (DB2), CSG (ORACLE), Teradata, and DataLake (SQL SERVER) to DART (hive/impala). This pipeline is prerequisite for biller accuracy project below
- Led biller accuracy project which contains +100 issued cases. He also created score cards and dashboards in Tableau
- Converted SQL to pySpark to run on Edge Node. This process keeps all tables up to date for Tableau dashboard
- Built open order error predictive model which can tell us if an open order is likely to have an error and what kind of error it is

CHARTER COMMUNICATIONS
DATA SCIENCE ANALYST | Fraud Department
Mar 2020 – Feb 2021 | Kansas City, MO
- Built Risk Recommendation Engine that detects frauds, compared performance between machine learning models (Logistic regression, XGBOOST)
- Created a web app to visualize all network linkages that help analysts detect frauds (NetworkX + Dash)
- Applied Decision Tree on order dataset to create a set of rules that are implemented in FraudNet system
- Performed A/B testing to find best attribute combinations for detecting new fraud trends
- Applied facial recognition to detect fraudsters who use different IDs

SPRINT | PROJECT SPECIALIST II | IT Compliance Department
Jan 2020 – Mar 2020 | Overland Park, KS
- Automated manual processes using VBA-Excel Marcos and Python
- Built a script that can get 7000+ IP Addresses from hostnames using NSLOOKUP
- Performed data integrity check on vulnerability scan results against Vulnerability Management Tool Database
- Developed a pipeline that combined, transformed multi-source data and outputted current state of all prepaid applications

HAVI | DATA SCIENTIST | Data Science Department
May 2019 – August 2019 | Downers Grove, IL
- Implemented deep learning techniques on multivariate time series transactional data
- Used LSTM, GRU models on Keras-Tensorflow backend to predict unit-sold at McDonalds stores

- Performed Market Basket Analysis to see relationships between items in transactional data, and quantitative research to understand store sales related to marketing campaigns
- Compared performance of GBRT, XGBOOST, and LIGHTGBM on data that contains time components
- Exposed to Azure Data Lake, PySpark, Hadoop

LE-REVE (T-NAILS) | BUSINESS INTELLIGENCE ANALYST | Part time
Jan 2015 – May 2019 | Overland Park, KS
- Monitored advertisements performance: impressions, clicks, amount spent, verified calls, and map actions
- Suggested customer reviews management and customer reward program to business's owner

EXTRA PROJECTS

CLOUD PROJECTS
- AWS project – Image resizing: Create a lambda function that listens to a S3 bucket. If users upload new photos, the lambda function will resize the photos smaller (using EC2 Instance) and upload them to another S3 bucket.
- Google Cloud Project – Reddit Sentiment Analysis: Scrape reddit newsfeeds, and store them in a Google Storage. Another function cleans up the raw data, and uses Google Natural Language API to get sentiment scores. The final data will be uploaded to a PostgreSQL database for end users analyzing.

ENGINE LIFE (LINK | CODE)
- Predict how long the engines will last based on measurements in time series

LOAN REPAYMENT (LINK | CODE1 | CODE2)
- Connect multiple tables into a big dataset to predict if a borrower is able to repay the loan

DOG BREED RECOGNITION | RECOGNITION SYSTEMS (CODE)
- Design and build a convolutional neural network to identify dog breeds from scratch.
- Implement the model along with face detection algorithm to build dog identification app

BASEBALL | TABLEAU VISUALIZATION (SHOWCASE | WRITEUP)
- Visualize the players population via height, weight, and homeruns.
- Create a quadrant that distinguishes good and bad players

WORLD CUP 2018 | ANALYSIS AND VISUALIZATION (CODE)
- Analyze and visualize the data that has 18,000 FiFa18 players using **Seaborn**, and **Matplotlib**.
- Predict the outcomes of semi-finals and final game of World Cup 2018 based on teams' forward, midfield, defend, goalkeeper, value and overall score.

MEDICAL COMPARISON | DATA TRANSFORMATION AND ANALYSIS (CODE)
- Combine and merge data files into dataframes, used **numpy** and **pandas** to transform from messy datasets to tidy datasets
- Create extra variables from existing one to find insights from the datasets, and visualized the insights with graphs.