MATTHEW LEE

DATA SCIENTIST

Contact

matt.eungoo.lee@gmail.com
2147557255
New York, NY
matthew-eungoo-lee/
silvernine209

Summary

Creative and persistent problem solver with professional engineering background. Enthusiastic about applying data science and machine learning techniques on real world problems.

Skills

PYTHON

Python
Scikit-Learn
Pandas
Numpy
TensorFlow
Keras
Selenium
Beautifulsoup
spaCy
Nltk
Gensim
Librosa

MACHINE LEARNING

Regression
Classification
Clustering
Natural Language Processing
Signal Processing
Neural Network (Convolutional & Recurrent)

DATA VISUALIZATION

Tableau
Matplotlib
Seaborn
Flask

PROJECT AND DATA MANAGEMENT

Github
MongoDB
PostgresSQL

TOOLS

AWS
SAP
Excel
Powerpoint
Fusion360

Experience

Metis — Sept. 2019 to Nov. 2019
Data Scientist Teaching Assistant — NYC

- Assisted Metis instructors in the classroom
- Conducted code reviews and supported data scientist students.

Metis — July 2019 to Sept. 2019
Data Scientist — New York, NY

Metis is a full-time 12 week ACCET-accredited immersive data science program which focuses on project design, data acquisition/cleaning/visualization, and communication using statistical modeling and machine learning. **Projects highlighting key skills are:**

EDM Generator Using LSTM [GitHub] [Presentation]

- Generated LSTM models that generate endless Classical music infused EDM songs.
- Designed a pipeline that would allow the model to blend any genre of music instead of just EDM and Classical music.
- Built a Flask app to utilize pre-trained model to generate and play EDM music easily [Demo].

Bitcoin Trader Using Sentiment from News Articles [GitHub] [Presentation]

- Web scraped news articles from Google search to generate sentiment features to be fed into neural network (LSTM) for price prediction.
- Used unsupervised learning(clustering) to discover types of topics and their trends around Bitcoin over time.

Voice Emotion Classifier [GitHub] [Presentation]

- Worked with audio signals to train gradient boosting machine(LightGBM) to perform multi-class classification on speaker's emotions (neutral, happy, sad, angry, fearful, disgust, and surprised).
- Deployed pre-trained model in Flask for MVP demonstration [Demo].

Chicago Daily Crime Count Prediction [GitHub] [Presentation]

- Web scraped daily weather data, unemployment rate, public transportation ridership, and Chicago's historical crime data to perform EDA and find correlations.
- Performed iterative modeling and feature engineering to build a linear regression model to predict daily crime count in Chicago.

Hanwha Azdel, Inc. — July 2015 to July 2019
Facilities Engineer & Process Engineer — Forest, VA

Facilities Engineer (Aug 2016 - Jul 2019)

- Designed and executed $1M project to establish new process for new product launch. Project involved system designing, budgeting, managing contractors and timeline, overseeing installations, and final commissioning.
- Completed 30+ end-to-end projects in varying sizes with total spending of $1.3M.
- Worked with cross functional teams to transform initiatives into executable projects.
- Expanded plant's capabilities (HVAC, safety, equipment upgrade, etc..) through executing capital projects.

Process Engineer (Jul 2015 - Aug 2016)

- Supported one of five production line that operated 24/7.
- Achieved monthly peak production efficiency (94.3% versus 4 year average of 89.7%).
- Resolved customer complaints for products from assigned production line through Six Sigma, corrective actions, automation, and Poka-yoke solutions.
- Collected and wrangled process data to discover optimal process settings to boost profit.

Projects

Stock Movement Predictor — Sept. 2018 to Oct. 2018

- Used market data from 2007 to present provided by Two Sigma, Kaggle competition host, to predict movement of various stocks.
- EDA was performed to select features and implemented neural network and XGBoost for modeling.
- Final rank of 49th/2927 (top 2%) in Kaggle competition : Two Sigma: Using News to Predict Stock Movements.

Object Detection in Images — July 2018 to Aug. 2018

- Utilized AWS, Darknet, and YoloV3 to perform image object detection task on 1.7 million images over 500 different classes.
- Transfer learning and ensemble of models were used for final submission.
- Final rank of 100th/454 (top 23%) in Kaggle competition : Google AI Open Images - Object Detection Track.

Power Line Fault Detection — Jan. 2019 to Mar. 2019

- Performed feature engineering and signal processing (800,000 time steps) on 3 phase power line signals to detect faults (anomaly).
- CNN and LSTM architectures were utilized to detect anomalies in power line signals.
- Final rank of 130th/1451 (top 9%) in Kaggle competition : VSB Power Line Fault Detection.

Education & Certifications

The University of Texas at Austin
BS Mechanical Engineering 2015

Stanford University on Coursera
Machine Learning

Udacity
Machine Learning Engineer Nanodegree

Activities

3D Modeling & Printing

Drew models of each major equipment in a produciton line using 3D CAD (Fusion360) based on 2D drawings. Then, 3D models were optimized for 3D printing in a 100:1 scale and 3D printed [Project Link].