

SKILLS

METHODS: Data Analysis and Visualization, Regression, Classification, Dimension Reduction, Natural Language Processing, Supervised/Unsupervised Learning, A/B Testing, Object Detection, Deep Learning

PROGRAMMING: SQL, Python

OTHER TOOLS: Amazon Web Services, Google Cloud Platform, Git, TensorFlow, Scikit-learn, Tableau, Agile, Cloud AutoML

EXPERIENCE

Semios Jan. 2018 to Current
Data Scientist Vancouver, BC

- Developed a data conditioning pipeline to flag and fix environmental data reporting over 1 million data points every 10 minutes.
- Maintained the codebase which performs scheduled hourly and daily cron jobs of the data conditioning, model calculation, and data storage (in Google BigQuery and Datastore) pipeline using AWS Elastic Beanstalk.
- Developed an object detection model using TensorFlow, GCP AI Platform and GCP Cloud AutoML capable of localizing and identifying six different insects from images obtained from traps.
- Utilized data collected from field observations to create models to predict crop phenology and pest pressure within an orchard to assist farmers in making key data-driven decisions to protect their crops and increase yield.
- Developed internal tools and dashboards to track KPIs using Tableau

Metis Sept. 2017 to Dec. 2017
Data Scientist San Francisco, CA

Developed five end-to-end data projects in a 12-week immersive data science training program focused on developing skills in Statistics, Machine Learning, Programming, Design, and Visualization. Using Python as a primary programming language, explored data sets and machine learning algorithms through course and project work, including:

- Classification of Wildfire Causes in the US - developed a classification model trained on 1.8 million cases of wildfires in the US over a 26 year period to aid authorities determine the cause of a wildfire based on its location and fire characteristics using scikit-learn, including data storage and extraction using SQL and data visualization in D3.js, and created a predictor application using Flask and the Google Maps API;
- Sentiment Analysis of YouTube Comments - aggregated and analyzed consumer sentiment on the features of the iPhone X based on YouTube comments of iPhone X videos, utilizing the YouTube API and MongoDB to extract and store text data, and a variety of natural language machine learning methods including text processing using NLTK, topic modelling with GENSIM (latent semantic indexing and latent Dirichlet allocation), clustering with scikit-learn, and sentiment analysis using TextBlob and VADER;
- Predictive Model for NBA Rookie Effectiveness - developed a regression model that can aid NBA General Managers predict the effectiveness of a player in their rookie season based on their college statistics, NBA team statistics, and physical characteristics using scikit-learn, web scraping with BeautifulSoup, and data visualization with Seaborn;
- Classification of Architectural Styles Using Neural Networks - applied transfer learning to train a convolutional neural network to classify a building's architectural style based on labelled Google images of buildings using Keras and Amazon Web Services EC2, trained a variational autoencoder and applied t-SNE for dimension reduction to visualize building similarities, and created a predictor/visualization application using Flask.

EDUCATION

University of British Columbia
Bachelor of Applied Science - Mechanical Engineering 2010