SAMY PALANIAPPAN, PHD

DATA SCIENTIST

drsamypal@gmail.com (408)561-2888 San Jose, CA. linkedin.com/in/samy-palaniappan-phd github.com/SamyPal

Skills

Software Tools	Python, Java, MATLAB, Maple, FORTRAN, PostgreSQL, MongoDB, PyMongo, Google Cloud, Amazon Web Services
Python Libraries	Scikit-learn, Pandas, Numpy, StatsModels, NLTK, Spacy, Scipy, LIME, Surprise
Data Viz	Tableau, Plotly, Bokeh, FlaskApp, Matplotlib, Seaborn
Machine Learning Models	Linear Regression, Logistic Regression, K-Nearest Neighbors, Naive-Bayes, Decision Trees, Random Forest, Neural Networks, Time-Series Forecasting, RNN, LSTM, LDA, NMF, PCA, SVD

Education

Ohio University
PhD Chemical Engineering 2013

Anna University
BS Chemical Engineering 2005

Stanford University
Certificatevia Coursera Machine Learning 2020

IBM Data Science
Certificate via Coursera Python for Data Science, AI, and Data Viz 2019

IBM Data Science
Certificate via Coursera SQL for Data Science 2020

Professional Affiliations

Society of Petroleum Engineers, AIChE

Sigma Xi, The Electrochemical Society

Publications and Presentations

Google Scholar:
https://scholar.google.com/citations?user=gepnCCEAAAAJ&hl=en

Experience

Metis — Chicago, Il
Data Scientist — 2020

Participated in a full-time, immersive data science bootcamp focused on machine learning, programming, statistical analysis, and data communication

Honed skills in data acquisition, database management, exploratory data analysis, supervised and unsupervised learning, data visualization, and natural language processing through five end-to-end machine learning projects

Katholieke Universiteit — Leuven, Belgium
Research Fellow — 2016 to 2019

Developed experiments and machine learning models in python to predict the performance of novel technologies for metal recovery from real geothermal conditions.

Leadership Activities: Mentored graduate students in developing research skills by discussing research methodology and project. Developed data management protocol for laboratory, managed equipment, involved in writing proposals.

Stantec Consulting / Kore Infrastructure LLC., — Los Angeles, CA
Lead Process Engineer — 2013 to 2016

Forecasted H2S content in Produced gas, conceptualized process, completed efforts for FEED, basic, and detailed engineering to build a sour gas treatment facility to handle 3.5 MMSCFD of produced gas with 25000 ppm of H2S. (Project cost: $15M).

Supervised and reviewed the detailed engineering for treatment plant to process 144 tons per day solid waste treatment. (Project cost: $39M).

Ohio University — Athens, Ohio
Graduate Research Fellow — 2006 to 2013

Investigated the electrodeposition of metallic and bimetallic catalysts and their electrocatalytic activity as cathodes for hydrogen production using ammonia and urea electrolysis.

Achievements: Published 3 articles in major peer-reviewed journals. Presented talks at several international conferences. Winner of several poster sessions.

Leadership Activities: Co-founder and treasurer for Ohio University ECS Student Chapter, 2011 – 2012. Mentor for OU Chem-E-Car team received 2010 National AIChE Inherent Safety in Design Award 2010.

Selected Projects

Cogs in Amazon E-Customer Relations (NLP + Recommender System)
Used Natural Language Processing topic modeling to develop a model to generate customer predilections and built a collaborative-filtering based recommendation engine for repeat customers and cosine similarity from a popular list for first-time customers. Used SQL and Python installed on Google cloud Virtual Machine and Dask to handle data.

Stock Superclusters (Unsupervised Learning + Web Scraping + Sentiment Analysis)
Developed model to combine news frequency to price fluctuation. Used clustering techniques to separate high-performing S&P500. News articles were scraped from major financial journals with Reddit, Twitter, and Newsio API. Used MongoDb for storing and handling data.

Heart Health Inspector (Classification + Flask WebApp +SQL)
Used CART models and ensemble models on patient physiological data to predict the risk of coronary heart disease. Developed a flask app to accept individual health data visualize the results of the test. Used AWS and SQL for handling data.

Forecasting Annual Auto-Sales - (Linear + Time Series Regression + Web Scraping)
Used linear and time-series modeling to predict total annual car sales along with the top-selling models based on macro-economic factors. Predicted MSRP based on car features respectively. Data was scraped from https://carsalesbase.com using Beautiful Soup and Selenium.