RAMASAMY PALANIAPPAN, PHD

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

Summary

● Data scientist with strong mathematics, statistics, and programming skills with a passion for deriving insights from data.
● I have a Ph.D. in engineering, 10 years of experience with data analytics, and Tableau Desktop Specialist, Lean Six-Sigma Green Belt certification.
● Rated by LinkedIn as top 5% for my skill in MS Excel and Machine Learning out of the 6.5 million and 108k test-takers, respectively.

Experience

Data Scientist

Metis · Chicago, Il   01/2020 to Current

Completed six end-to-end data science projects with machine learning (ML) to build pipeline frameworks to address key business questions in the auto, health, e-commerce, and finance sectors. Handled data mining, database management, exploratory data analysis, data visualization, handled big data with **pySpark** and Dask. All the projects used **GIT** for version control.

● <u>Stability Prediction for a Binary Compound of a Certain Composition</u> (Clustering + Dimensionality reduction (PCA, Autoencoders, Genetic Algorithm) + RandomForest): Given a pair of elements (A and B), I developed a multi-labeled classifier model to predict the stability of 11 of its compounds varying in the composition of B ranging from 10% to 90% with a recall of 97% and a precision weighted f1-score of 90%.

● **<u>Forecasting e-Scooter Daily Demand for NYC</u>** *(AWS + PySpark + Sequence Modeling + CNN)*: Preprocessed 80 million e-scooter rental data using **PySpark** on AWS using Facebook Prophet and WaveNet, with MAPE (mean absolute percentage error) of 35 and 44, respectively. Also predicted the likely destination for scooters, based on their rental time and location.

● **<u>Cogs in Amazon e-Customer Relations</u>** (*GCP + NLP - Topic Modeling + Recommendation System*) Developed a model pipeline for authors and sellers based on customer reviews for books. Associated the role of Vine-reviewer's high rating to the increased purchase despite poor customer ratings.

● **<u>Smart Ad Recommendation</u>** (*Web Scraping + GCP + Image Classification + Dimensionality Reduction*): Developed an image-based advertisement recommendation system with Convolutional Neural Networks for image classification with 90% accuracy, and PCA for dimensionality reduction by data manipulation techniques.

● **<u>Stock Superclusters</u>** (*Web Scraping + Mongo DB on AWS EC2+Clustering + Sentiment Analysis*): Separated highly performing stocks from poorly performing stocks that formed the SP500 index. NY Times, and Twitter data for the stocks were scoured using their **<u>REST API's</u>,** which was then used to deduce sentiment to predict day-to-day stock movement. **<u>REST API's</u>** from Quandl and Yahoo Finance was used to download data on the daily stock movement for S&P500 stock tickers. Text documents were stored in, and queried from a NOSQL MongoDB database.

● **<u>Heart Health Check App</u>** (*Classification + Flask WebApp +SQL*): Classified risk of contracting coronary heart disease in the next 10 years with 89% f1-score using a Random Forest algorithm. Deployed model using a live Flask-Web App via., Heroku.

Process Engineer/ Data Scientist

Beacon Greentech (Contract) · Chennai, India   07/2019 to 12/2019

● Used **SQL** to query data from existing solar power grids, residential, and commercial solar panel databases. Used data wrangling to clean and prepare data for modeling. Developed sequence models to forecast solar energy production.
● Developed predictive analytical model for customer churn from, customer support, billing, solar energy, and health of solar panels. Reduced customer churn by 25 % within 3 months, by moving focus areas, and scheduling automated maintenance of panels.

Postdoctoral Fellow (Data Science)

Katholieke Universiteit · Leuven, Belgium   12/2016 to 07/2019

● Developed a **probabilistic model pipeline** to simulate the integrated energy requirement for metal extraction from deep geothermal systems. **Monte Carlo model** was used for **bootstrapping a logistic regression model** to obtain a result with a **mean absolute error of 0.7 kWh/kg**. Project was done in **R**.
● Designed a lab-scale reactor prototype and experiments for the data collection on metal recovery from deep geothermal fluids. Designed and executed several *A/B testing* experiments for increased metal recovery.

Lead Chemical and Materials Process Engineer

Kore Infrastructure · Los Angeles, CA   02/2016 to 12/2016

● Supervised the detailed process modeling and process engineering for the construction of a treatment plant to process 144 tons per day solid waste. (Project cost: $39M).
● Lead research team, designed and supervised the execution of experiments, performed data analysis. Presented research to leadership. Interfaced with engineering & construction team.

Chemical Process Engineer II
Stantec Consulting · Bakersfield, CA   11/2013 to 02/2016
● Developed process models for the construction of a Flue Gas Desulfurization Plant for Freeport McMoRan O&G. Conceptualized process, completed efforts for FEED, basic, and detailed engineering to build a sour gas treatment facility. (Project cost: $15M).
● Forecasted, with time series modeling, H2S content in produced gas from oil wells.. Performed process modeling for industries ranging from upstream oil & gas to renewables fuels, and chemicals to food processing.

Research Fellow
Ohio University · Athens, OH   03/2006 to 11/2013
● Improved the efficiency by 15% for producing hydrogen by from ammonia electrolysis by optimizing parameters for hydrogen production on a range of metallic and bimetallic electrocatalysts using Design of Experiments (multifactor A/B testing)

Technical Skills

Software Tools: Python, R, Java, PostgreSQL, PyMongo, Google Cloud, AWS, MongoDB, Github, GitLab, Jupyter Notebook, VS Code
Python Libraries: Scikit-learn, Pandas, Numpy, Scipy, Keras, Tensorflow, PlaidML
Data Viz: Tableau, Seaborn, Matplotlib, FlaskApp
Machine Learning Models: Linear Regression, Logistic Regression, ARIMAX, Naive Bayes, SVM, Decision Trees, Random Forest, KNN, K-means, DBSCAN, HCA, SVD, PCA, NMF, LSA, LDA, RNN, LSTM, CNN, XGBoost, AdaBoost
Statistical Tools: Descriptive Statistics, Predictive Modeling, Bayesian Statistics, Probability Distributions, Dimensionality Reduction, Sampling Methods, Time Series Modeling, A/B Testing

Education

Stanford University & DeepLearning.ai
Certificate Machine Learning, Deep Learning, and TensorFlow 2020

IBM
Certificate Data Science Specialization 2020

KU Leuven
Postdoctoral Fellowship Materials Engineering 2019

Ohio University
PhD Chemical Engineering 2014

Awards

• Postdoctoral fellowship (EU funded), KU Leuven. 2016 to 2019.
• Graduate student fellowship (NSF funded), OU. 2006 to 2013.
• Board Member and Treasurer for Electrochemical Society Student Chapter at OU. 2011 to 2013.
• Recipient of Best Poster Awards at Research and Creative Activity Expo, OU. 2010, 2011.

Supporting Work

Freelance Data Scientist @ Upwork and CrowdDoing   05/2020 to Current
● Contributed to _7 machine learning projects_ ranging from _classifying bank loan approvals,_ detecting _credit card fraud detection_ to _online sales prediction_, and _product price estimation_.
● I organized the data science team and activities into talent verticals and subprojects horizontals with _Agile framework_, which facilitated easy collaboration among new and experienced volunteers.

Affiliation
● Sigma Xi, Scientific Research Society.
● Applied Computing Machinery - SF Bay Area Chapter