

# 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 programming skills and a passion for deriving insights from data. I have a Ph.D. in engineering, a Lean Six-Sigma Green Belt certification, and 10 years of experience with data analytics (statistical, mathematical, and process modeling)

## 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 Spark and Dask.

- **Forecasting e-Scooter Daily Demand for NYC (GCP + Dask + Additive Regression Model + CNN)**: Preprocessed 80 million e-scooter rental data using Dask on GCP 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.
- **Cogs in Amazon e-Customer Relations (GCP + NLP - Topic Modeling + Recommendation System)**: Developed a feedback 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.
- **Smart Ad Recommendation (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.
- **Network Attack Detection (Clustering + Dimensionality reduction + Autoencoders + Classification)**: Used CICIDS2017 dataset to identify 14 different types of network anomalies with an MCC score of 95%.
- **Stock Superclusters (Web Scraping + Mongo DB on AWS EC2 + Clustering + Sentiment Analysis)**: Separated highly performing stocks from poorly performing stocks that formed the SP500 index. Leveraged daily news sentiment analysis to predict day-to-day stock movement.
- **Heart Health Check App (Classification + Flask WebApp + SQL)**: Classified people's 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.

### Postdoctoral Fellow

**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**.
- 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

02/2013 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 to handle 3.5 MMSCFD of produced gas with 25000 ppm of H<sub>2</sub>S. (Project cost: \$15M).
- Forecasted, with time series modeling, H<sub>2</sub>S content, conceptualized process, completed efforts for FEED, basic, and detailed engineering to build a facility to treat 3.5 MMSCFD of sour produced-gas.
- Performed process modeling for industries ranging from upstream oil & gas to renewables fuels, and chemicals to food processing.

# RAMASAMY PALANIAPPAN, PHD

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

## Research Fellow

**Ohio University** · Athens, OH

03/2006 to 11/2013

- Developed a mechanistic model to evaluate electrochemical kinetic rate constants for hydrogen production from experimental data and Levenberg-Marquardt algorithm to solve the non-linear system of equations.
- Optimized parameters for hydrogen production on a range of metallic and bimetallic electrocatalysts using Design of Experiments (multifactor 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

## Technical Skills

**Software Tools:** Python, R, Java, PostgreSQL, PyMongo, MATLAB, Google Cloud, AWS, Maple, FORTRAN, MongoDB, C++, Github, GitLab

**Python Libraries:** Scikit-learn, Pandas, Numpy, StatsModels, NLTK, Spacy, Scipy, Keras, Tensorflow, PlaidML, Gensim

**Data Viz:** Tableau, Seaborn, Bokeh, Plotly, 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

**Chemical Modeling and Simulation:** Mathematical Modeling, Molecular Modeling, CASTEP, VASP, Gaussian, Process Modeling, Aspen, PRO/II, PipePhase, DYNsim, Finite Element Modeling, Materials Modeling, COMSOL, HSC, FactSage

### Communication Skills:

Excellent written and verbal communication skills: 15 publications at peer-reviewed journals with 203 total citations and 8 presentations at international conferences

## Volunteering

### Data Scientist @ CrowdDoing

05/2020 to Current

I organized data science team and activities. I also currently lead the Metaanalysis team (Team of 6 data scientists) - to extract information from journal articles using BioNLP and bioPandas for topic modeling.

### Data Scientist @ KU Leuven

12/2016 to 06/2019

- **Consulted on 7 machine learning projects** ranging from classifying bank loan approvals, detecting credit card fraud detection to online sales prediction, and product price estimation.
- **Developed data collection and management protocol** for the laboratory. Mentored graduate students with project direction.

### Activities @ Ohio University

Spring/2006 to Winter/2010

- **Teaching Assistant** for ChE 1800 – **Approaches to Engineering Problem Solving**, and ChE 3800 – **Numerical Modeling and Analysis**. Coordinated with the faculty to teach, organize laboratory work, graded assignments, and supervised exams.
- **Founding member** and Treasurer for ECS Ohio University Chapter. Organized invited lectures and seminars for the department.

## 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.