# Sean Conroy: Data Science

address: Dallas, TX 75228 phone: 412-855-7246 email: seantconroy@gmail.com
Github Profile StackOverflow Profile Kaggle Profile

*overview*

Analytics professional with strong data science, engineering & manufacturing background and 12 years of experience writing code in production and research environments. Known for capacity to quickly gain mastery of a subject and build analytical tools to support top decision makers.

*key skills: machine learning, data engineering, BI template development, translating engineering & scientific algorithms into R / Python workflows, designing and building analytics tools*

coding: R, Python, VBA, SQL, HTML, JavaScript, Matlab
platforms: RStudio, Spotfire, Anaconda, Data Robot, GitHub, Spyder, Atom, Access, Excel
Python libraries: scikit-learn, TensorFlow / Keras, pandas, numpy
R libraries: data.table, tidyverse, doparallel
machine learning / data science:

- Coursera:
  - Machine Learning with Andrew Ng, Stanford University
  - Data Science Specialization with R, Johns Hopkins
  - How to Win a Data Science Competition: Learn from Top Kagglers
  - Bayesian Statistics: From Concept to Data Analysis
  - Bayesian Methods for Machine Learning
- Deep Learning Bootcamp (week-long intensive on neural networks using TensorFlow / Keras)
- Experience with Neural Networks, XGBoost, SVM, and other scikit-learn algorithms
- GitHub: Owner/Maintainer of current organization's Corporate GitHub account
  - ~300 contributions to current org's repositories in the last year
- Statistical Process Control (week-long training on SPC by UT professor)

*experience*

**Covey Park Energy (Oil & Gas E&P)** January 2018 – current

*Data Analyst, Business Development Team*

Principal developer for a team that created analytical tools for the Business Development, Operations and Finance departments using R, Python, SQL and Spotfire:

- Built machine learning models in Python to predict Haynesville gas well performance
  - Included over 100 features, including geology, well design and completion parameters
  - Automated ETA, Hyperparameter Tuning and
  - Deployed models for Clustering, Classification & Regression using XGBoost, DBSCAN & ANN
- Automated Decline Curve Analysis:
  - **Deliverable**: Spotfire template for automated forecasting of oil & gas wells
  - Custom algorithms for noise removal, well segmentation, best forecast detection
  - Non-linear regression, parallel processing, ~ 5000 lines of code, 7 libraries built
- Pressure Normalized Rate (PNR) Forecasting with Aries Comparison:
  - **Deliverable**: Spotfire template that is now widely used by organization's engineers and analysts
  - Iterative calculation of Bottom Hole Pressure using complex petroleum engineering models
  - Functionality to load, translate, plot, and adjust "Aries"-style forecasts

- (Pason Drilling Analytics) Automated Analysis of 10-second Drilling Data:
  - Multi-step workflow for cleaning data, automated joins, detecting rig-states, and analyzing data for automated KPI analysis

**Diab Americas (Aerospace Foam Manufacturing & Analytics)** **June 2016 – January 2018**

*Process Engineer*

- Built (for fun) entirely 4 new functional databases using (MS Access / SQL / VBA) for
  - Production Management Database (what was produced, when, how, where is it now)
  - Laboratory Database (automated data loading from multiple data sources)
  - Quality Database (is it good to use for this or that customer)
  - Inventory Database (where is it now, how did it get here, where should it go)
- Built numerous analytical tools w/ "dashboards" for automated data processing / report generation
- Process Engineer for 4 product lines (provided in-depth engineering analysis of extrusion data)

**Poly-America (Blown Film Manufacturing Analytics)** **February 2013 – May 2016**

*Assistant Laboratory Manager*

- Regularly perform extensive statistical analysis for internal corporate sales & marketing clients
- Develop numerous Visual Basic automation tools / "dashboards" / lab software
- Completed Projects:
  - (R) Process Variation analysis tool: used to analyze years of 24/7 production data
  - (R) Production/QC/Lab Data System Aggregation: product-based datasets for analysis
  - (VBA) Materials Traceability Solution: provided complete traceability for two mfg. locations consuming over 1 billion lbs. of polyethylene per year
  - (VBA): Automated Film Statistics Toolbox; ANOVA-style analysis for arbitrary datasets
  - (VBA): Lab Information Management System; enable sample tracing and data analysis for all testing data throughout the laboratory

**Natural Composites (Material Science Startup)** **August 2011 – February 2013**

*Research and Development Engineer*

- Among many other hats, oversaw quality control process for all raw materials:
  - (MATLAB): Automated Fiber Length Measurement: length distribution by image analysis
  - (MATLAB): Quality Control Image Analysis: solution for evaluating raw material quality

**Antioch Ministries International, Kurdistan, Iraq** **September 2007 – August 2008**

*Missionary / English Professor, Koya University*

**Sandia National Laboratories, Livermore, CA** **May 2005-August 2007**

*Research Intern, DOE Q Clearance*

- Modeling, simulation, and algorithm development
  - (MATLAB): Mid-Fidelity Physics Simulation of Radioactive Isotopes at a Border Crossing
  - (MATLAB) Automated Isotope Classification Package using Wavelets

*(Education)*

- **Masters of Engineering** — **Baylor University**
  - GPA: 3.66 — Graduated 2011
- **Bachelor of Science, Physics** — **Carnegie Mellon University**
  - GPA 3.22 — Graduated with honors, 2007