

# Mark Pedigo, PhD

Data Scientist

[mark.pedigo@gmail.com](mailto:mark.pedigo@gmail.com) • 314.775.9142  
[LinkedIn](#) • Hopkins, MN, US

Accomplished PhD data scientist with extensive problem-solving, modeling, and mentorship experience. Excellent understanding of predictive modeling, statistical analysis, dashboard design and development, data visualizations, database queries, coding, and code-testing techniques. Strong interpersonal, verbal, and written communication skills have facilitated engagement with management, clients, and co-workers to help foster a productive team environment.

## Skills Overview

Business	<i>Communication.</i> Presentation of complex ideas to non-technical audiences, stakeholder engagement. <i>Leadership.</i> Team management, mentorship.
Teaching/Mentoring	<i>Teaching.</i> Data science/programming/math courses at Washington University and DataCamp. <i>Mentoring.</i> Mentoring junior developers and students on Python, data, machine learning. <i>Writing.</i> Articles, team presentations on ML and Python.
Technical	<i>Machine Learning.</i> Predictive models for procurement approval timelines, infant mortality. <i>Operations Research.</i> Clinic location optimization and rationalization. <i>Statistics.</i> A/B testing, regression analysis, Monte Carlo simulations. <i>Visualization.</i> Dashboards and financial planning tools. <i>LLMs.</i> Tools for intent classification and performance evaluation using AWS, OpenAI API.
Programming/Tools	<i>Python + libraries:</i> data manipulation (NumPy, pandas), ML (scikit-learn, XGBoost), visualization (Seaborn, Plotly), UI (Streamlit); <i>Optimization</i> (Gurobi); <i>Database queries</i> (SQL); <i>Cloud</i> (AWS, Azure)

## Career Experience

### Freelance

May 2024 – present

- **Technical Author.** Create and publish articles on programming and mathematics.
- **Course Creator, DataCamp.** Collaborated with content managers and developers to define audience needs, create course specifications, and design detailed outlines. Developed coding exercises, wrote and recorded video scripts, and incorporated peer and student feedback to refine course content.

### UnitedHealth Group, Hybrid (Minnetonka, MN)

May 2024 – Jan 2025

#### Principal Data Scientist

- **Clinic Optimization and Rationalization.** Spearheaded the evaluation and selection of optimization solutions for clinical optimization. Organized software demonstrations with optimization software and real estate firms, ultimately partnering with IBM to develop a proof-of-concept (POC). Collaborated on requirements gathering and facilitated weekly sprint meetings.
- **Procurement ML models.** Supervised a direct report with the design and implementation of ML models to predict turn-around time and approval authorities for procurement requests.
- **A/B Testing.** Designed an A/B test to test the sentinel effect of travel procedures.
- **Team Development and Education.** Delivered presentations to the corporate services team regarding high-level machine learning approaches and their potential applications in procurement and real estate.

### 1904 Labs, Remote (St. Louis, MO)

Jul 2022 – Feb 2024

#### Senior Data Scientist

- **Healthcare Fraud Detection** (Kingdom Capital / PercayAI)  
Devised methods to detect potential FWA from the data, e.g., identify hospital stays and cross-check inpatient/outpatient claims for geographical or temporal inconsistencies. Implemented several of these methods using SQL, Python, and Spark. Found aberrations in the data indicating potential prescription fraud that became part of a report presented to the Missouri AG.
- **IVA Dispatcher Bot Performance** (Ring/Amazon)  
Developed code to garner intent predictions and report classification performance metrics. Created validation sets by manually expanding from an intent set generated by GPT. Employed AWS S3, Lex, Python + pandas + Seaborn.
- **LLM Intent Classification Evaluator**

Created a tool to measure the accuracy of LLM classifications. Given a prompt, input data, and a selection of LLM model, the tool generated a report of accuracy and classification errors. The tool was used to help select an LLM for a sales demo and used Python + StreamLit, OpenAI API, AWS Bedrock API.

- **Mentoring**

Mentored an individual for a summer internship and a senior capstone project, a key component of his successful admissions package to Purdue.

**Adjunct Professor, Washington University, University College**, Saint Louis, MO

Jan 2012 – May 2022

- Taught data science, programming, and mathematics courses to professionals, non-traditional, and undergraduate students of varying technical backgrounds.
- Designed and created “Intro to Data Science” from the ground up; created course materials for all courses.

**MedeAnalytics**, Remote (Richardson, TX)  
Senior Data Scientist

Mar 2021 – Jul 2022

- **Infant Mortality Prediction Model** (Mississippi Department of Medicaid)  
Created SQL code to feature engineer fields based on healthcare claims data (ICD-10, DRG, CPT/HCPCS, revenue codes, patient and provider geospatial data) to be consumed downstream by a predictive model.
- **Mentoring**  
Mentored a Junior Data Scientist to elevate his Python programming skills.

**Centene**, Hybrid (St. Louis, MO)  
Data Visualization Developer

Jul 2019 – Mar 2021

- **Special Investigative Unit Provider Pre-Pay Interface**  
Developed visualizations for a front-end interface for a project targeting healthcare fraud detection. Managed deployments into a Kubernetes hosted production microservices environment via a CI/CD pipeline.
- **Visualizations: Business Planning Application, Agile Charts**  
Developed visualizations for Agile charts (burn down, lag charts) and for financial variance calculations based on actual v predicted costs, budget, and fixed costs from rosters. Employed R/Shiny for R Studio Connect and Python Plotly Dash for containerized Docker/Linux environments.

**EPsi**, Remote (San Ramon, CA) / Hybrid (Chesterfield, MO)  
Data Scientist

Sep 2016 – Jun 2019

- Implemented financial forecasts for hospital administrators using statistical models (ARIMA, SARIMA, Holt Winters) and software tools (e.g., Python, statsmodels, Prophet, RNN on AWS SageMaker).
- Crafted visualizations in Tableau and SiSense regarding hospital operations data and clinical outcomes.
- Presented results to executive leadership for strategic review.

**Booz Allen Hamilton, Strategic Innovation Group**, Remote (Washington DC) / O’Fallon, IL  
Data Analyst / Developer

Mar 2014 – Sep 2016

- **Cost Estimation** (NASA)  
Co-developer of NASA’s NextGen Cost Estimating software, the Project Cost Estimating Capability (PCEC).

## Education

Doctor of Philosophy in Mathematics, Saint Louis University, Saint Louis, MO

Bachelor of Computer Science, University of Missouri-Rolla, Rolla, MO

Master of Natural Science, Emphasis in Mathematics, Southeast Missouri State University, Cape Girardeau, MO

Professional Certificate in AI, Stanford University

## Awards

- NASA (2015): Software of the Year Award Honorable Mention, Agency Honor Award, Marshall Space Flight Center Tech Transfer Team Award, Marshall Space Flight Center Software of the Year Award, Cost Estimating Team Award

## Published Works

### Articles

- Pedigo, Mark. "Python String Interpolation: A Beginner's Guide." DataCamp, Forthcoming.
- Pedigo, Mark. "Python Lambda Functions: A Beginner's Guide." DataCamp, [January 31, 2025](#).
- Pedigo, Mark. "ChatGPT's Influence on Industries." 1904 Labs Blog, [June 15, 2023](#).
- Pedigo, Mark. "1904 Labs Supports Personal Growth." 1904 Labs Blog, [October 6, 2023](#).
- Pedigo, Mark. "How to Import Data with the pandas read\_csv Command." Towards Data Science, Medium, [June 21, 2020](#).
- Pedigo, Mark. "Is God a Mathematician? – Book Review." Cantor's Paradise. Medium., [January 3, 2020](#).
- Pedigo, Mark. "The Unreasonable Effectiveness of Mathematics in the Natural Sciences – A Summary." Cantor's Paradise. Medium, [September 29, 2019](#).

### Online Course

- Pedigo, Mark. "Case Study: Building Software in Python" [MOOC]. DataCamp, [December 2024](#).

### Research

- "On the lower central series of the free nilpotent groups of finite rank." Pedigo, Mark, R. Blyth, Communications in Algebra, [May 29, 2019](#).
- Huang, Sidi, M. Pedigo, C Shaw, K. Odom, Data-Driven Guidelines for Correlation of Cost and Schedule Growth. Conference presentation to the International Cost Estimating and Analysis Association (ICEAA Conference), March 30, 2015.
- Ma, J., Pedigo, M., Blackwell, L., Hackworth, C., Holcomb, K., & Gildea, K. (2009). 20 Years of the Line Operations Safety Audit (LOSA) Program: From Flight Operations to Maintenance and Ramp Operations. Oklahoma City. OK: FAA Civil Aerospace Medical Institute.
- Ma, J., Pedigo, M., Gildea, K., & Holcomb, K., (2009) Proactive Safety Management: Maintenance and Ramp Line Operations Safety Audit (LOSA) Year 2. Oklahoma City, OK: FAA Civil Aerospace Medical Institute.
- Goldwasser, Michael, and Pedigo, Mark. (2008). Online, Non-preemptive Scheduling of Equal-Length Jobs on Two Identical Machines. ACM Transactions on Algorithms (TALG) archive, 5(1). A preliminary version appeared in Proceedings of the Tenth Scandinavian Workshop on Algorithm Theory (SWAT), 113-123.