

Jonathan Poeder

2347 W. Charleston St., Apt 2F, Chicago, IL, 60647 • 5133720261 • jonathan.poeder@gmail.com • jonathanpoeder.com

Certified data scientist with extensive experience developing actionable solutions for complex AI/ML projects and challenging coding problems in technology, advertising, banking, retail lending, and marketing. Passionate about generative AI, large language models, reinforcement learning, and what the transformer architecture has made possible.

HIGHLIGHTED ACHIEVEMENTS

- Successfully delivered models for ~\$3M in PoCs to domestic and international PwC clients in energy, food service, agriculture, and tech.
- Built time-series and customer level predictive stress testing models for Fifth Third's \$14B mortgage portfolio, supporting capital planning for the bank (CCAR / DFAST), and successfully delivered to the Federal Reserve.
- Developed and implemented predictive origination and fraud models that drove over \$20MM of incremental value in loss reduction. Models supported the underwriting of ~\$1 billion in consumer loans.
- Created a time-series auto-ML engine that generated \$1M+ in revenue by winning P&G's global forecasting business for The Nielsen Company. Platform deployed to Nielsen hubs globally, also winning non-P&G business in Western Europe.
- Developed ARIMA time series forecasts used in proposal to secure Nielsen's role as Walmart's primary insights partner.
- Developed methodology and architecture for an automated business alerting platform identified as one of Nielsen's top ten initiatives that year, on which they applied for 4 US patents.

EDUCATION

JOHNS HOPKINS UNIVERSITY, Whiting School of Engineering <i>Master of Science in Artificial Intelligence</i> **Program structured for engineering professionals.	Baltimore, MD 2028
UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE, Belk College of Business <i>Master of Science in Economics</i>	Charlotte, NC 2007
UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE, Belk College of Business <i>Bachelor of Science in Economics</i>	Charlotte, NC 2004

CERTIFICATIONS

Highlighted certifications and certificates:

Artificial Intelligence & Machine Learning, UT Austin
Data Scientist, datacamp

AI Fundamentals, datacamp
NLP Specialization, coursera

WORK EXPERIENCE

Protiviti AIML Architect	Chicago, IL 01/2026 - Present
◦ Architected a multi-tenant Azure Container Apps platform for an AI-driven financial close and business process assessment accelerator, with a management control-plane that auto-provisions a fully isolated per-client stack (Container App, Azure Files share, managed identity, RBAC, Front Door route) on demand via the Azure ARM SDK. Async idempotent provisioning, a PENDING-to-READY state machine, and janitor-based teardown reconciliation guarantee zero cross-tenant data collision. ◦ Engineered a 3,000+ line Azure OpenAI tool-calling agent exposing ~30 specialized functions that map raw client ERP and general ledger extracts to a canonical schema, infer chart-of-accounts tiers, generate KPI and variance SQL, and author financial-close narratives. KPI formula-locking and a self-review pass prevent generated-SQL drift, feeding a three-tier	

DuckDB layer and Power BI dashboards.

- Built a multimodal RAG accelerator that auto-drafts regulator-grade SR 11-7 / OCC 2011-12 model validation reports, fusing tri stream FAISS retrieval (narrative text, structured table-RAG, and GPT vision chart analysis via Azure Document Intelligence) into a single cited evidence layer, governed by a two-tier table-veracity validator and mandatory human-in-the-loop sign-off.
- Developed a full-stack demand forecasting application (FastAPI and React 19 / TypeScript) deployed to Azure Container Apps and Static Web Apps, pairing a parallelized model bank of SARIMAX, Holt-Winters, and ARIMA variants with a FRED macroeconomic leading-indicator overlay (relaxed-LASSO with BIC selection, VIF-guarded OLS contribution decomposition, cross-correlation lead-lag detection), plus an on-demand Anthropic Claude layer via Azure AI Foundry for natural-language-to SQL, diagnostic explanation, and executive narration.
- Designed and shipped a Model Context Protocol (MCP) server exposing the firm's SharePoint project archive as an agent queryable knowledge base for tools like Cursor and Claude Code, combining dense Azure OpenAI text-embedding-3-large vector search with keyword and metadata filtering over ChromaDB (HNSW cosine), fed by an incremental, self-healing, manifest-driven Playwright and Entra ID ingestion pipeline.
- Built an LLM-driven spreadsheet-to-warehouse accelerator for a global private equity client that auto-maps arbitrary investment cash-flow workbooks onto a star schema and traces each summary metric back through its source Excel formulas, generating a layered dbt (DuckDB) pipeline that reconciles fund-level Gross MOIC and IRR exactly to the legacy reporting.
- Served as technical lead across multiple AI Studio engagements, guiding junior team members' development work, and running internal-client discovery and SME requirements sessions to scope accelerators for internal stakeholders and external buyers.
- Selected as a judge for the firm-wide Protiviti Python Hackathon, evaluating competing submissions by rebuilding and running each end-to-end and scoring on reproduced runtime telemetry, surfacing blocking defects ahead of executive finalist selection.

PricewaterhouseCoopers

Chicago, IL

AIML Architect

05/2024 - 01/2026

- CNN classifier models for project margins. Focus on optimizing model performance via experimentation with custom and off-the-shelf loss functions like Cross-Entropy Loss and Mean Squared Error. Conducted a comprehensive exploration of optimizers including Adam, RMSprop, and SGD. Hyperparameter tuning for learning rate and momentum.
- Developed Gramian Angular Difference Field (GADF) techniques for feature extraction and prediction of energy consumption. Prophet, PyTorch, and TPOT, utilizing GPU acceleration for efficient model training.
- dbt pipeline development for Fortune 100 company data, forecasting a variety of targets from energy consumption and mineral demand to financial metrics like COGS and margin dollars. Created optimized SQL models, yaml configurations, ETL to support business intelligence, etc. Jinja to ensure scalable and efficient workflows.
- Engineered auto-ML pipelines including diagnostic tests, model specification, and time series forecasting. Implemented advanced diagnostics and clustering techniques to enhance model accuracy and performance, leveraging statsmodels, statsforecast, scipy, numpy, etc.
- Engineered a comprehensive suite of statistical tests and data transformations for our auto-ML architecture, including stationarity, normality, and autocorrelation tests, as well as PCA, UMAP, KMeans clustering, etc. for dimensionality reduction and pattern recognition.

Innov8iveLabs

Chicago, IL

Data Science Consultant, Owner

11/2023 – 05/2024

- Scoring models for collectability of delinquent accounts and detecting friendly fraud with data from Kount (Equifax), Emailage (LexisNexis), and Equifax Recovery Reports. Text pre-processing and cosine similarity for cross-data source PII validation
- Engineered robust WSGI APIs with detailed exception handling and seamless integration with AWS SageMaker for real time data processing and inference. Custom API key authentication to secure endpoints via AWS Secrets Manager as well.
- Streamlined deployment via shell scripting, Docker images, and AWS Elastic Container Registry (ECR).

Cross Screen Media

Alexandria, VA

Data Scientist Lead

07/2021 - 01/2024

- Used pandas, numpy, scipy, boto3, plotly express, and snowflake connector Python libraries to develop a Slack based cost alerting system for AWS and Snowflake services that helped identify anomalous spikes in compute to prevent over \$30k in compute costs from accruing in one month. AWS Cost Explorer API, Snowflake ACCOUNT_USAGE schema, Python, SQL, EC2 (Linux), S3, cron, git, Slack API, JSON objects, and plotly.

- Used pandas, numpy, scipy, sklearn, seaborn, matplotlib, plotly express and others to develop classification models that could identify whether an individual was a cable, broadcast, or non-audience member to extrapolate from our TiVo sample to the broader targeted segment for our measurement methodology.
- Developed classification model with similar Python libraries to reduce sample dimensionality for a survey of desktop, mobile, and CTV media consumption.
- Used json, pandas, and Google API Python modules for oauth and Sheets requests to automate Apache Airflow job file migration
- Updated legacy cable and broadcast consumption propensity models to comport with industry research at national and designated marketing area (DMA) level. ARIMA modeling, dbt (data build tool), Jinja, SQL CTEs & window functions.
- Led effort to identify and integrate AI/ML driven cost optimization with Keebo.ai to further reduce Snowflake compute costs 20-25%.
- Drove cross region Snowflake migration and replication effort. Detailed analysis of data assets for migration and S3 Glacier archival, creating and loading database / schema objects, cost monitoring and reporting for replication via Snowsight.

Clark Schaefer Hackett

Cincinnati, OH

Data Analytics Lead

03/2019 - 06/2021

- Drove successful outcomes for clients on projects related to financial analysis, data quality, data governance, and business intelligence for clients in healthcare, manufacturing, and retail in engagements worth over \$1.25M.
- Built and led data analytics service line responsible for incorporating analytics into CSH compliance and consulting lines of business as well as external data analytics consulting practice.
- Worked with Go-To-Market (GTM) services to create product and client development assets (proofs of concept, sell sheets, pitch decks, etc.) targeting small business, middle market, and large publicly traded companies.
- Created and supported data governance policy for CSH internally.
- Built an integrated development data warehouse and Linux environment for prototyping and Power BI dashboards.
- Established processes for recruitment, hiring, and mentorship of data science professionals.

Fifth Third Bank

Cincinnati, OH

Principal Data Scientist

10/2015 - 03/2019

- Developed an automated validation platform for DRM PD and LGD models in Fifth Thirds commercial portfolio. Scoring for over 30K commercial entities validated with multiple implementation issues identified across almost all portfolio segments.
- Optimized ETL data pipeline, developed time-series regression models, and reported on non-performing asset (NPA) performance for the 2016 Comprehensive Capital Analysis & Review (CCAR) submission to the Federal Reserve.
- Developed the loss given default (LGD) fractional logit model for Fifth Third's \$14B mortgage portfolio in their 2017 CCAR stress test submission. Full development cycle including ETL and preprocessing, feature engineering, modeling, back testing, and technical documentation. Additional models evaluated included nonlinear least squares regression.
- Successfully developed the probability of default (PD) model for Fifth Third's \$14B mortgage portfolio included in the 2017 CCAR submission. Full development cycle including ETL and preprocessing, feature engineering, modeling, back testing, and technical documentation. Models capture economic phenomena including deflation and negative interest rates.
- Worked with model risk management (MRM) to ensure all the requisite model validation materials were provided. Model successfully passed validation, and no MRAs from Federal Reserve.

Axcess Financial

Cincinnati, OH

Data Science Manager

03/2014 - 10/2015

- Developed and implemented ML classification models including random forest, decision tree, and logistic regression to predict loan applicant risk. Credit scoring supported underwriting Axcess' US retail product portfolio, the company's largest.
- Developed predictive classification models for credit scoring to support underwriting and mitigate risk in UK operations.
- Managed vendors and data acquisition for retro-appends supporting fraud and credit scoring initiatives.
- Created custom fraud models and evaluated 3rd party solutions to address retail portfolio's fraud detection requirements.
- Supported FP&A liquidity preservation and 2015 provision forecasting initiatives.
- Performed implementation testing and created model monitoring solutions for numerous retail portfolio product releases.
- Performed EDA on fraud activity for existing customers related to starter check usage.

- Recruited and led a team of 2 data scientists. Mentored on modeling, programming, and career development.

Senior Data Scientist

04/2013 - 03/2014

- Created and deployed an ETL platform in the SAS environment to streamline modeling and scorecard development.
- Led predictive modeling and analytics projects from inception to post-implementation monitoring for several loan products, incorporating segmentation, cohort analysis, and other quantitative analyses to drive decisions on the company's product portfolio. Specifically, the segmentation analysis drove decisions on the model architecture for the internal credit scoring solution behind their installment loan product.
- Worked with senior leadership and SAS support to assess server limitations and recommend hardware to meet the needs of a rapidly growing department. Performance testing and technical expertise provided to ensure successful deployment of new server environment.
- Provided leadership in the development of a lifetime value modeling framework that is being used to support risk-based pricing and campaign analysis for multiple marketing vehicles.

In4mation Insights

Needham, MA

Director, Marketing Science

07/2012 - 02/2013

- Developed algorithm in the SAS environment to estimate brand switching behavior.
- Trained analysts on and championed use of Mercurial distributed version control system.
- Utilized custom Bayesian modeling engine in UNIX environment.
- Data processing with MySQL and SAS, modeling with the MIXED procedure, and process automation with the macro facility for various price optimization studies.

The Nielsen Company

Stamford, CT

Senior Data Scientist

06/2009 - 07/2012

- Created a probability driven alerting engine using IML and the macro facility that analyzes the relationship between business targets and ARIMA based forecasts.
- Coded custom OS detection and ETL adjustment so modeling platform could operate in Windows and UNIX environment.
- Developed a dynamic internal forecast commoditization system in the SAS environment with FORECAST, ESM, and ARIMA for estimation, the macro facility for automation, and SQL for data processing.
- Leveraged modeling platform to generate ARIMA forecasts for Walmart RFP process, which Nielsen won to become their primary insight partner.
- Devised a hybrid neural network and time-series solution using MODEL and ARIMA to forecast sales for a new-to-market Dolce & Gabbana fine fragrance.
- Adaptation and development of European key business driver modeling framework to the SAS environment for Nielsen's Latin America clients.

Data Scientist

04/2008 - 05/2009

- Developed a scalable VARMAX (time-series) forecasting system with dynamic model optimization for brand and category level sales. Augmented Dickey-Fuller testing for stationarity, SQL for data processing, macro facility for automation.
- Developed generalized linear models (GLM) for analysis of covariance and MIXED models for the incremental volumetric effects of 'special pack' product releases. SQL for data management and processing, and various macro functionalities were incorporated.
- Created CPG price smoothing algorithm for data preprocessing and model prep.
- As a data scientist on a global team of R&D statisticians, performed ad hoc analyses using SQL and marketing mix modeling as well as multiplicative competitive interaction modeling techniques. Various linear and non-linear modeling.

PROJECTS

Full Stack Amazon & eBay Data Mining Web Application

Stood up a locally hosted LAMP (Linux, Apache, MySQL, PHP / Python) server to serve a web-based application for mining and analyzing ecommerce data. Driven for the most part by Django, Ajax, and Python it automatically collected product data from the Amazon Merchant Services and eBay seller APIs, performed ETL through a series of data pipelines, stored cleaned data in the data architecture (MySQL) I designed and built, and applied various proprietary algorithms for pricing and sales pattern recognition. For the front end, I created a form-based UI with JavaScript, Bootstrap, CSS, and HTML that allowed me to perform

ad hoc analyses to identify and track products, take notes, and store related manufacturer information. *The platform was used to identify products for my LLC that would eventually be sourced from China and Pakistan and sold through those channels, with over \$100k in revenue generated from 2 products identified.*

Instacart Basket Analysis

This analysis is focused on understanding customer purchasing patterns using a dataset containing 3 million grocery orders from over 200,000 Instacart users. Full cycle from extensive EDA, feature engineering, and data preparation through robust segmentation and basket analyses. Segmentation begins with crosstabulation, then principal components (PCA) for dimensionality reduction across products, and finally k-means for clustering customers. Concludes with a basket analysis to identify lift estimates from product associations. mlxtend for basket analysis, scikit-learn for PCA and KMeans, numpy for calcs and custom memory optimization, pandas, etc.

Custom Generative AI Tuning & Development

Created and tuned a custom generative AI large language model (LLM) w/ OpenAI for resume building. GPT tuned on academic transcripts, consulting projects, academic projects, and career related data. pydub and ffprobe libraries for audio processing, OpenAI API and OpenAI's Whisper for robust and accurate speech-to-text conversion.

Analytic Consulting - Apparel Manufacturer

Built models for sales analyses, customer segmentation, and price optimization. Performed data forensics for improved data quality, process automation for efficiency, and developed business intelligence dashboards for enhanced real-time insights. Projected savings of \$4M from eliminated cost inefficiencies and realized revenue growth of over 75% over a 2 year period.

Twitter Airline Sentiment Analysis

Identify the sentiment from tweets to understand an airlines' customer satisfaction. Vectorization (count and td-idf vectorizer), parameter tuning, confusion matrix model eval.

Image Classification w/ CNNs

Using computer vision tools, identify the plant seedlings species from 12 different species using a convolutional neural network. Keras, CNN, Jupyter notebooks.

Bank Customer Churn Prediction

Help an operations team identify customers that are more likely to churn by building an artificial Neural Network from scratch. Data mining, TensorFlow, Keras, artificial neural nets (ANN), Google Colab.

Credit Card Churn Prediction

Predictive modeling of customer attrition for credit card services and respective reasoning. Statistical modeling, cross validation, regularization, pipelines and hyperparameter tuning, up / down sampling with SMOTE.

Travel Package Purchase Protection

"Visit With Us" travel company dataset is used to analyze the customers' information and build a predictive model for customer acquisition on a new product. Customer profiling, bagging and random forest bagging classifiers, adaboost and gradient boosting classifiers, XGBoost, stacking classifier, hyperparameter tuning, GridSearchCV, RandomizedSearchCV.

Homelab

Created and manage infrastructure for an Ubuntu 22.04 LTS server edition rackmount. Use cases include local prototyping, model development, and experimentation. Runs CUDA, cuDNN, PyTorch, TensorFlow, etc. Some hardware details include 128GB of RAM, 8-core AMD Ryzen Threadripper 1900, NVIDIA GeForce RTX 3060.

CONTINUING EDUCATION

Post Graduate Program in AI/ML w/ UT Austin

Comprehensive 6-month curriculum with cutting-edge technologies that include deep learning, computer vision, NLP, and TensorFlow. Program includes 24 hands-on projects & 42 case studies.

Deep Learning for Coders with Fastai and PyTorch

Neural network fundamentals, computer vision, natural language processing (NLP), handling structured data, and ethical AI.

Natural Language Processing in Python w/ datacamp

Tokenization and regular expression to parse text, handling non-English text, topic identification, term frequencies, named-entity recognition (NER), and custom 'fake news' classifier. nltk, regex, matplotlib, bag-of-words, gensim, corpus creation, tf-idf, SpaCy, polyglot, count / tfidf vectorizers, scikit-learn.

Machine Learning with scikit-learn w/ datacamp

Using python for supervised learning to build predictive models, hyperparameter tuning, and evaluating holdout performance on real world data. Numerical / visual EDA, kNN fit & predict, overfitting, underfitting, sample splitting, lasso and ridge (regularized), ROC / AUC, data imputation, one hot encoding (OHE).

Data Science Toolbox w/ datacamp

Series of lessons that cover foundational python tools for data science practitioners. Nested functions, variable length arguments (positional & keyword), lambda functions with map / filter / reduce, error handling, enumeration, list comprehension, generators, streaming data.