

SUDHIR POL

📍 Seattle, Washington 📞 +1 (930) 904-4785 ✉️ sudhirpol522@gmail.com 🔗 [linkedin.com/in/sudhir-pol-a6b544179](https://www.linkedin.com/in/sudhir-pol-a6b544179) 🌐 [sudhirpol522.github.io](https://github.com/sudhirpol522)

Summary

Data Scientist with 3 years of industry experience at American Express, S&P Global, and Adobe, building predictive and language models that answer business questions and shipping them to production on AWS. Strength in probability calibration, evaluation design, and turning model output into decisions measured in analyst hours saved and risk correctly routed.

Education

Indiana University

Master of Science in Data Science

Aug. 2024 - May 2026

Bloomington, Indiana

Rajarambapu Institute of Technology

Bachelor of Technology in Electronics and Tele-Communications

Aug. 2017 - May 2021

Sangli, India

Experience

Adobe

May 2025 - Aug. 2025

Machine Learning Engineer Intern

San Jose, California

- Designed the scoring rubric and LLM as Judge evaluation harness that benchmarked Gemini 2.5 Pro, GPT 4.1, and Claude 4.5 on multimodal chart quality, raising offline F1 from 79 percent to 92 percent by calibrating judge scores against a human labeled reference set and tracking every run in Weights and Biases
- Converted a 48 hour manual review cycle into a 30 minute automated pipeline with a LangGraph agent over Model Context Protocol, and added a human approval gate so reviewers accepted or rejected each generated plan before it ran

S&P Global

Nov. 2022 - Aug. 2024

Data Scientist II

Hyderabad, India

- Fine tuned LayoutLMv3 for token classification over labeled spans in financial filings and served it on AWS EC2, versioning datasets and model artifacts with DVC and MLflow and releasing through a CI/CD pipeline so every result was reproducible and auditable
- Deployed Llama 2 on AWS SageMaker and Lambda with chain of thought prompt templates for information extraction across long documents, removing close to 20 minutes of manual data collection per analyst task
- Built a question generation and answering API on AWS EC2 around a T5 transformer, and engineered a Java parsing framework with named capture groups that replaced a paid third party vendor with an in house solution, saving roughly 10 minutes per analyst task

American Express

Dec. 2021 - Nov. 2022

Data Scientist I

Gurgaon, India

- Built a complaint risk classification model that sorted customer cases into 3 severity tiers using naive Bayes over TFIDF features, applying isotonic regression so predicted probabilities were calibrated and the routing threshold could be set on real risk rather than a raw score, deployed on AWS EC2
- Developed VIBE, a call quality scoring model combining XGBoost with BERT embeddings, containerized with Docker and served to the business as the measurement the annual representative bonus pool was decided on
- Modeled customer intent from call transcripts with an LSTM that mapped conversations to demand types, mining the transcript corpus with PySpark for feature generation and serving the model through Flask on AWS EC2 to prioritize incoming calls

Projects

Probabilistic Yard Gain Forecasting | *Python, Scikit-learn, XGBoost, Pandas, NumPy*

Nov. 2020

- Predicted the full cumulative distribution of yards gained per play rather than a point estimate, benchmarking logistic regression, SVM, random forest, gradient boosted trees, and XGBoost on held out data and applying isotonic regression so the forecast probabilities were reliable enough to act on

MultiDocChat: Document Question Answering System | *FastAPI, LangChain, FAISS, Kubernetes*

May 2025

- Built a retrieval augmented question answering service over multi document collections with FAISS vector indexing and MMR retrieval, scoring answer relevance with Gemini 2.5 Pro and tracing every query end to end in LangSmith to find where retrieval failed

Writing

- [Speculative Decoding: Make LLM Inference Faster](#) *Oct. 2025*
- [Demystifying GPTQ: From Lagrange Multipliers to Vectorized PyTorch](#) *Jun. 2026*
- [The Math of AWQ: Protecting Salient Channels from the Inside Out](#) *Jun. 2026*

Skills

Languages: Python, SQL, PySpark, Java, C++

Machine Learning: Statistical Modeling, Classification, Regression, XGBoost, Random Forest, SVM, Naive Bayes, Probability Calibration, Feature Engineering, Evaluation and Metric Design, Time Series

Deep Learning and NLP: PyTorch, TensorFlow, Scikit-learn, HuggingFace Transformers, BERT, T5, LayoutLMv3, LSTM, Large Language Models, RAG, Prompt Engineering

Data and Cloud: Pandas, NumPy, Apache Spark, AWS (SageMaker, EC2, Lambda, S3), Docker, Kubernetes, MLflow, DVC, Weights and Biases, CI/CD, Git, FastAPI, Flask