

Ritin Wadekar

8317202165 | ritinwadekar6@gmail.com | [linkedin.com/in/ritinwadekar](https://www.linkedin.com/in/ritinwadekar) | ritinwadekar.com | Pune, Maharashtra

PROFESSIONAL EXPERIENCE

Onpoint Insights

Jan 2026 - Jun 2026

AI Engineer

San Francisco, CA

- Served as the sole AI Engineer owning end to end R&D on a medical coding pipeline for a healthcare client, delivering a production ready, government compliant product within 2 months targeting a market projected to reach \$8.4B by 2033
- Conducted research on automated ICD10 prediction approaches, avoiding a flawed vanilla LLM baseline (less than 70% coding accuracy) and establishing a validated retrieval based architecture as the team's technical foundation
- Engineered a retrieval system over three authoritative government (CMS) coding sources to ground LLM predictions in clinically accepted standards, reducing error rate to 7.66% versus a 10-20% manual coding industry baseline
- Designed a two stage retrieve then rank AI pipeline (candidate generation to LLM classification) for medical coding (ICD10/CPT coding), achieving over 3x prediction accuracy versus single pass LLM baselines
- Incorporated Medicaid and Medicare compliance rules into the automated coding pipeline to handle government payer requirements, targeting a coding error rate that costs the healthcare industry over \$1B annually

Onpoint Insights

Jan 2025 - Dec 2025

Data Scientist

Andover, MA

- Delivered a multi agent LLM pipeline on Azure Databricks for natural language to SQL generation executed over 20K+ product records, containerizing with Docker for reproducible runs across dev/staging, returning insights and cutting analyst effort by 40%
- Designed a RAG based querying AI assistant over 20+ documents using transformer based embeddings, FAISS vector indexing, and LLM driven intent routing (GPT4 models), improving answer accuracy by 17% over an embedding only semantic search baseline
- Executed A/B experiments, performed hypothesis testing, and built causal regression models to quantify incremental campaign lift for a top-3 pharmaceutical client, enabling spend optimization that improved ROI by 4.2%
- Deployed an Optical Character Recognition (OCR) workflow in Power Automate to transform unstructured PDF invoice documents into structured Excel datasets for financial analysis, achieving 92% extraction accuracy

Onpoint Insights

May 2024 - Aug 2024

Data Science Intern

Andover, MA

- Extracted 10M+ transaction records from Snowflake via SQL to engineer a hybrid recommendation system combining Market Basket Analysis with Word2Vec embeddings, served through vector similarity search, driving a 6.7% lift in Average Order Size
- Engineered an entity resolution system using Named Entity Recognition and embedding driven similarity matching (cosine similarity) to reconcile partial company names from Azure SQL Database into canonical entities, reducing reporting inconsistencies
- Built sales forecasting models using additive and multiplicative time series methods via Snowpark in Snowflake, executing in warehouse hyperparameter tuning and rolling backtests to achieve 12% MAPE and supporting inventory and production planning

Creative Galileo

Apr 2023 - Jul 2023

Data Science Intern

Pune, India

- Leveraged user behavioral data to build churn prediction models, performing model evaluation in Python (pandas, scikit-learn) and AWS SageMaker to generate risk scores for retention initiatives at a Series A Ed-Tech platform with 10M+ app downloads
- Diagnosed user telemetry data stored in AWS S3 to identify latency-related funnel drop-offs and rank performance bottlenecks, enabling targeted system optimizations that reduced application load times by 10% and payment-page churn by 12%

TECHNICAL SKILLS

Programming Languages: Python (NumPy, Pandas, Matplotlib, Scikit-learn, PyTorch, TensorFlow, Keras, PySpark, Plotly), R, SAS

Tools: Tableau, PowerBI, Alteryx, Databricks, Mixpanel, Microsoft Excel, Power Automate, Git, Github Actions (CI/CD)

Database Technologies: MySQL, PostgreSQL, MS SQL Server, MongoDB (NoSQL), Oracle Database, Apache Hadoop, Snowflake

Cloud Platforms: AWS (S3, Redshift, QuickSight, Athena, Glue), Azure (Kubernetes Service, Blob Storage)

EDUCATION

The University of Texas at Dallas

May 2025

Master of Science, Business Analytics and Artificial Intelligence

GPA: 3.5

Savitribai Phule Pune University

May 2023

Bachelor of Engineering in Electronics and Telecommunication

Honors in Artificial Intelligence and Machine Learning

ACADEMIC PROJECTS

10-K Document Analysis with Retrieval-Augmented Generation | NLP, Vector Embeddings, Semantic Search, Langchain

- Engineered a RAG system over 10-K filings by implementing query decomposition and embedding based semantic search with FAISS vector indexing, improving retrieval precision and reducing hallucinated responses by 18% compared to an embedding only baseline

Multi-Label Classification Using Transformer based Models | Predictive Modeling, BERT, QLoRA, LoRA

- Fine tuned Transformer based BERT models for multi label classification on high dimensional social media text data, applying LoRA/QLoRA for parameter efficient adaptation that reduced GPU memory usage by 40% while sustaining an F1 score of 0.86