

PARTH SALKE

Philadelphia, 19104, USA | +1 (267)-787-9995 | psalke@seas.upenn.edu | [LinkedIn](#) | [Website](#)

EDUCATION

UNIVERSITY OF PENNSYLVANIA, Philadelphia, PA

May 2026

Master of Science in Data Science

GPA: 4.00/4.00

Coursework: Big Data Analytics, Statistics, Machine Learning, Forecasting Methods (Time Series), Advanced Deep Learning, NLP

MAHARASHTRA INSTITUTE OF TECHNOLOGY- WPU, Pune, India

May 2024

Bachelor of Technology in Electronics and Communication Engineering

GPA: 9.77/10.00

TECHNICAL SKILLS

Programming Languages: Python, R Programming, SQL, MongoDB, Java

Frameworks: Apache Spark, PyTorch, TensorFlow, Scikit-learn, Hadoop, Airflow, Tableau, Power BI, Git

Tools: Databricks, Snowflake, Azure (Data Factory, Synapse Analytics, DevOps), AWS (S3, Redshift, Glue, Sagemaker, Bedrock), GCP (BigQuery, Vertex), Statistics, Docker, MS Excel, Langchain

WORK EXPERIENCE

The Wharton School, Wharton AI and Analytics (UPenn)

August 2024 – Present

Research Assistant (Data Scientist)

- Currently working with **Humana** on optimizing ETL pipelines and forecast fraud and abuse detection.
- Worked with **IKEA**- Optimized (D-S CTR) for IKEA's lower funnel using Google Analytics data, unlocking **\$30M per 1%** improvement. Conducted data organization, cleaning, and statistical modelling using PySpark and Snowflake, leveraging Decision Trees and clustering to identify key conversion drivers.
- Utilized BigQuery for processing, aiming to increase customer conversion rate from 1.4% to **2.6%**.

Extentia Information Technology, Pune, India

July 2023 – January 2024

Data Science Intern

- Facilitated project by implementing **ETL** pipelines using Azure Data Factory, Databricks, Synapse Analytics and Power BI, thus reducing data processing time by **14%** and leveraged Python, SQL and R scripts for data transformation.
- Automated data migration from **Oracle** with Pyspark & reduced data validation effort by **60%** and managed Data Lake.
- Finetuned Falcon and GPT20B LLM, reducing SDLC time for chatbot creation using OpenAI and Cognitive Search.
- Analysed text and generated vector embeddings for LLM training, leading to **10%** increased accuracy in validation set.

The Sparks Foundation, Pune, India

April 2022 – June 2022

Data Analyst Intern

- Conducted EDA on diverse datasets, enhancing overall data quality by rectifying **25%** of data anomalies.
- Refined and deployed advanced regression and clustering ML algorithms to identify patterns & trends within datasets, A/B testing to evaluate data cleaning methods and scaled this solution to handle over **100K** data points per minute.

PROJECTS

Building Resilient Portfolio: Equity Portfolio Optimization and Dynamic Risk Management

[\[Link\]](#)

- Led an AWS project using Sagemaker, S3, and Lambda to analyse financial stock portfolio risk and returns through web scraping and data wrangling.
- Applied K-means clustering for optimal asset selection and generated risk-return graphs.
- Used ARIMA model for stock return forecasting and designed constrained portfolio optimization methods.

FinSight RAG - Financial RAG with Mistral-7B and Langchain

[\[Link\]](#)

- Developed platform for traders by fine-tuning the Mistral-7B model on financial data using RAG, including a Q&A Chatbot for real-time earnings call queries with context precision more than **90%**, quantised model to **8 bits**.
- Built a custom in-memory vector store with Sentence Transformers and leveraged Pinecone for a comprehensive vector knowledge graph. Fine-tuned for semantic search with a Streamlit app for interaction.

Loan Eligibility Prediction Using Supervised, Unsupervised and Reinforcement Learning

- Utilised unsupervised learning for clustering and anomaly detection during preprocessing. Developed multiple Supervised ML models; XGBoost achieved the highest accuracy of **83%** on validation data.
- Employed reinforcement learning to optimize the decision-making process and served using a Flask server.

RESEARCH & PUBLICATIONS

- Graduate Teaching Assistant- TA for CIS 5210 (Artificial Intelligence) and ESE 5420 (Statistics for Data Science)
- Enhancing Driving Safety by Real-time Posture Detection & Analysis Using Computer Vision, IEEE NMITCON, 2023 [\[DOI\]](#)
- Adaptive Cooling System for CVT of ATVs with Wireless Data Acquisition System, IEEE ICONAT, 2023 [\[DOI\]](#)
- Certifications: Microsoft AZ-900, AI-900, AWS ML and Cloud Practitioner, Mastering Tableau, Big Data & Apache Essentials