OJAS SHARMA

New Jersey ojassharma16@gmail.com 9296847733 Website LinkedIn Github

EDUCATION

Rutgers University – New Brunswick, NJ, USA – | **MS – Statistics-Data Science |** **Expected May 2025**
Relevant Coursework: Probability and Stat Inference, Regression and Time Series, Data Structure and Algorithm, Database Management System, Stat Model and Computing, Deep Learning, Financial Time Series, Algorithmic Trading & Portfolio Management

Manipal University Jaipur – Jaipur, INDIA – | ***B. Tech Computer* Engineering |** **May 2023**
Relevant Coursework: Engineering Mathematics, Data Structures, Data Science with Python, AI/ML, Deep Learning

WORK EXPERIENCE

Data Scientist | Center Of Gambling Studies, Rutgers University | New Brunswick, NJ **Apr 2024 – Present**

- Currently conducting extensive data analysis on a **4.5 TB** dataset encompassing **13.6 billion records** seven gambling operators to **identify high-risk gambling behaviors** and develop predictive models for early intervention.
- Built and deployed data-driven risk assessment models like **K-Means & GMM clustering models** to classify **1M+ gamblers** based on behavioral indicators, achieving a **Silhouette Score of 0.79** and provide insights for regulatory policy recommendations.
- Applied **A/B testing** with multi-class models (SVM, LightGBM, CatBoost) to improve risk classification of gamblers, boosting intervention accuracy by **33%.**
- Developed an **outlier detection pipeline** using **Z-Score, IQR filtering, and PCA (99% variance retained)** to detect irregular deposit patterns.
- Results drove a **30% increase** in early detection of risky financial behavior, informing state-level policy and modeling financial risk indicators across multi-billion-dollar transaction datasets.

Data Analyst | Omalco Extrusion | New Delhi, India **Aug 2022 – July 2023**

- Automated **financial reporting and market analysis** with **Power BI and D3.js**, improving business team efficiency by **80%** and reducing **ETL job runtime by 50%** through optimized reporting table logic.
- Integrated real-time data pipelines for dynamic market analysis and implemented time-series forecasting algorithms like **ARIMA and Prophet models**, driving actionable insights to support data-driven strategies across departments.
- Implemented **hybrid data architectures and real-time dashboards** using PostgreSQL, Couchbase and D3.js.
- Conducted statistical analysis using **linear regression** and **hypothesis testing** to uncover key trends like price elasticity trends, seasonal demand fluctuations, & profitability shifts across customer segments, directly informing pricing & inventory strategies.

TECHNICAL PROJECTS

Full List

Advanced RAG Pipeline for Question *Github*

LLaMA 3, LangChain, Hugging Face Transformers, FAISS, PyTorch, FastAPI

- Developed a high-performance **Retrieval-Augmented Generation (RAG) system** integrating **LLaMA 3, LangChain, and FAISS**, improving response accuracy by **70%** while optimizing retrieval efficiency.
- Engineered a low-latency document retrieval pipeline, leveraging **vector search with FAISS and optimized nearest neighbor lookup**, reducing query time from **90 ms to 30 ms (~80% speedup)**.
- Enhanced system reliability through **GitHub Actions, CI/CD pipelines**, enabling **automated deployment and model performance tracking** across multiple iterations.
- Built and deployed an **API-based interface** using **FastAPI**, enabling efficient querying and interaction with the RAG system.

LLM-Powered Product Recommendations *Github*

PyTorch, HuggingFace, RAG, Quantized Low-Rank Adaptation, NLP, Prompt Engineering

- **Fine-tuned LLMs (Mistral-7B, TinyLlama)** for product recommendation using Amazon Review Dataset.
- Achieved **98%+ accuracy** in next-purchase prediction through novel prompting and efficient fine-tuning (LoRA, QLoRA).

Comparative Analysis of Retrieval-Augmented Generation (RAG) Techniques *Github*

Comparative study on different RAG architectures

- Conducted end-to-end evaluation across latency, coherence, and accuracy to determine optimal RAG structure.
- Evaluated **multiple RAG** approaches across accuracy, latency, and coherence determining **Hybrid RAG** as the most effective.
- Identified key trade-offs in retrieval methods, optimizing contextual relevance while reducing hallucinations in generative outputs.

Campus Watch Crime Analytics *Github*

- Developed a real-time interactive dashboard to **analyze campus crime patterns** using daily data scraped from SpotCrime via **Selenium-based automated pipeline**.
- **Implemented Plotly library** to craft interactive charts and **geospatial maps**, significantly enhancing the user experience with rich data visualizations.
- Demonstrated ability to blend **data engineering, visualization, and automation** to solve public safety analytics problems mirroring customer-facing use cases in **banking dashboards and fraud detection systems**.

SKILLS

- **Languages/Framework:** Python, PySpark, SQL, R, C, C++, MATLAB, Selenium, Flask, SreamLit
- **Data Visualisation :** Microsoft Power BI, Tableau, D3.js, Seaborn, Matplotlib, GGPLOT
- **Database/Cloud:** Hive, Snowflake, Google Cloud (BigQuery,Pub/Sub) , AWS (SageMaker, Bedrock, Redshift), Microsoft Azure.
- **NLP:** RAG, Text Preprocessing & Modeling, Named Entity Recognition, Generative AI, Hugging Face Transformers, LangChain
- **Machine Learning:** KMeans, Boosting Algos, PyTorch, TensorFlow, Bagging Models, Keras, LLMs Prompting/Finetuning
- **Tools/Frameworks:** Apache Spark, SSIS, Jupyter Notebooks, Git, Docker, Kubernetes