Ojas Sharma

+1 (929) 684-7733 ojassharma16@gmail.com LinkedIn Github Website

Education

Rutgers University - New Brunswick, NJ | M.S. - Statistics Data Science Sep. 2023 - May 2025
Relevant Coursework: Probability & Stat Inference, Regression & Time Series, Data Structure & Algorithm, Database Management System, Stat Model & Computing, Deep Learning, Algorithmic Trading

Manipal Univeristy, India | B.Tech Computer Engineering Aug. 2019 - May 2023

Skills

Languages:	Python, PySpark, SQL, R, C++, C, Java, Flask, MATLAB
ML/DL:	Statistical & Regression Analysis, Clustering Algorithms, Boosting/Bagging Algorithms, Keras, PyTorch
Databases/Cloud:	Big Data Analytics, MySQL, PostgreSQL, Hadoop, Snowflake, AWS, Vector DBs, Pinceone, LanceDB
NLP:	LLMs, RAG, Text Preprocessing, Named Entity Recognition, Hugging Face, LangChain
Data Visualisation :	Power BI, Tableau, GGPLOT2, Plotly, Seaborn

Experience

Full list

Data Scientist | Center of Gambling Studies | New Brunwsick, NJ May 2024 - Present

- Engineered an in-depth analysis on a **4.5 TB dataset with 13.6 billion records** from seven gambling operators, identifying high-risk gambler profiles to enable timely early interventions.
- Built and deployed data-driven risk assessment models like **K-Means and GMM clustering models** to classify over **1 million** gamblers based on behavioral indicators, achieving a **Silhouette Score of 0.79**, improving precision of regulatory policy insights.
- Enhanced effectiveness of early gambling interventions by **33%** through improved multi-class risk classification **models (SVM, LightGBM)**, validated via **A/B testing** on intervention outcomes.
- Monitored an outlier detection pipeline using **Z-Score, IQR filtering, and PCA (95% variance retained)** to detect irregular deposit patterns, enhancing the accuracy of fraud detection systems.
- Accomplished high-risk gambler detection **by 6%** and boosted early-risk identification **by 30%, influencing multi-billion-dollar policy decisions** and state-level regulations.

Data Analyst | Omalco Extrusion | New Delhi, India Aug. 2022 - July 2023

- Automated financial reporting and market analysis using **Power BI and D3.js**, improving business team efficiency **by 80%** and reducing ETL job runtime **by 50%** through optimized reporting table logic.
- Developed real-time data pipelines and applied time-series forecasting **(ARIMA, Prophet)** to predict sales and demand, driving strategic planning and inventory management.
- Architected and deployed **hybrid data storage solutions** combining PostgreSQL and Couchbase, powering **real-time dashboards with D3.js** for up-to-the-minute market visualization and faster decision-making.
- Leveraged **linear regression and hypothesis testing** to pinpoint seasonal demand fluctuations, directly informing pricing and inventory strategies, **impacting 100+** products from the company's offerings.

Projects

Full list

Restaurant QA System | *LLaMA 3, LangChain, Hugging Face Transformers, RAG, Mistral, PyTorch*

- Established and implemented a hybrid NLP pipeline combining **Abstractive QA, Extractive QA, and Closed-domain QA** to answer nuanced, restaurant-specific queries from unstructured data (reviews, menus, ratings).
- Designed baseline using BiLSTM with Attention and advanced to **T5, LLaMA3, Mistral, and RAG** models.
- Achieved top performance with Mistral **(F1 0.65, BERT 0.99, SAS 0.94)**, surpassing baselines and validating RAG's semantic strength **(BERT 0.97)** & **LLaMA3's SAS 0.93.**
- Leveraged advanced evaluation metrics(Exact Match, F1, BERT Score, Bi-Encoder, SAS) to quantify, improvements and guide model selection.

Twitter Search Application | *PostgreSQL, Couchbase, Elasticsearch, Joblib, LRU Cache, ACID, NoSQL*

- Accelerated tweet processing by **60% using joblib**-based parallelization for large-scale data extraction.
- Built custom LRU cache with TTL eviction, slashing query latency **from 5s to 0.4s** and minimizing DB load.
- Redesigned a hybrid PostgreSQL–Couchbase architecture with **Elasticsearch** for fast, engagement-sorted tweet search.

Campus Watch Crime Analytics |*Selenium,Plotly, Python,Geospatial Mapping,Web Scraping,Live Dashboards*

- Innovated 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.