ANKIT AGRAWAL

United States | +1 (765) 543 1519 | agrawala.1697@gmail.com | linkedin.com/in/ankit-agrawal1697 | github.com/ankit1697

PROFESSIONAL SUMMARY

Data Scientist with 5+ years of experience in predictive modeling across Retail, E-commerce and FinTech industries, specialising in Geospatial Data Science. Proficient in AI/ML techniques, Python and Statistics with an ability to develop data-driven solutions and translate data into actionable business insights to drive strategic decision-making.

EDUCATION

University of Chicago — **Chicago, IL**
Master's of Science - Applied Data Science | GPA: 4.0 — **September 2024 - December 2025**
Coursework - Big Data and Cloud Computing, Statistical Modeling, Generative AI

BMS Institute of Technology and Management — **Bangalore, India**
Bachelor's of Engineering - Computer Science and Engineering — **August 2015 - June 2019**

WORK EXPERIENCE

GeoIQ — **Bangalore, India**
Lead Data Scientist — **March 2024 - August 2024**

- Enhanced credit risk detection by building an address validation tool using **NER and elastic search** to evaluate address quality for loan approvals and collections, achieving **~92% accuracy** across city tiers.
- Streamlined retail site selection by developing a chatbot using **RAG-based LLM, LangChain and OpenCV** that analyses property survey videos, and extracts relevant information, enabling faster property filtering for business development teams.
- Expanded the company clientele by acquiring **8 new logos** through leading data science proof of concepts across multiple industries, contributing to **~$60k** in monthly recurring revenue.

Data Scientist — **September 2022 – February 2024**

- Ideated and implemented a TAM (Total Addressable Market) prediction framework using **Lasso Regression** to identify potential high revenue locations, **adopted by over 30** leading Indian retail brands.
- Optimized the **product assortment** of a leading Indian eyewear brand by developing a custom scoring model using **XGBoost**, resulting in **85%** of underperforming stores meeting their annual revenue targets.
- Enabled market expansion for brands by identifying **~21,000 markets**, including high streets, malls, and local markets, using **DBSCAN** to create clusters of Google's points of interests.

Mu Sigma — **Bangalore, India**
Data Scientist — **June 2019 - September 2022**

- Facilitated **personalized marketing** for a sports retailer by creating customer personas using **PCA and K-Means**, clustering the users into **7 segments** based on their digital and purchase behavior.
- Led Data Science initiatives for a Big Four US Bank, driving the strategy of an **MLOps platform** that reduced the turnaround time for model development and deployment by **~90%**.
- Analyzed credit card transactions data to build a **fraud detection model** using **Isolation Forest, DBSCAN and Autoencoders**, resulting in **~4% decrease** in the annual losses from credit frauds.
- Streamlined the over and under-provisioning of HDFS storage by developing a **forecasting model using FBProphet** to predict utilization of multiple teams for 30, 60 and 90 days, improving cost savings by **~12%**.

ACADEMIC PROJECTS

Application Tracker

- Designed and engineered an end-to-end job application management system using LangChain, enabling users to efficiently track and manage their job applications.
- Integrated a Named Entity Recognition (NER) model to match the required job skills with resumes, providing actionable insights to help users optimize their profiles and improve the chances of being shortlisted.

Will Turing Bots Replace human software developers

- Leveraged Big Data Analytics using PySpark, and SparkML to analyze 1.36 TiB data from GitHub, consisting of repository data including metadata like commits, file contents, languages, and licenses.
- Evaluated key metrics from the data to gauge the potential of emerging Generative AI platforms in replacing human software developers through comprehensive analysis and insights gained from GitHub data.

SKILLS

- **Programming:** Python (Pandas, NumPy, Scikit-learn, NLTK), PySpark, R, SQL, Django
- **Tools & Technologies:** Machine Learning Algorithms (Regression, Classification, Clustering, Time Series Forecasting), Statistics, Natural Language Processing (NLP), Deep Learning (Neural Networks, Tensorflow, PyTorch), Computer Vision, Generative AI, Large Language Models (LLM), LangChain, Geospatial Analytics, Exploratory Data Analysis, Model Explainability, Jupyter, Tableau, GCP, AWS, H3, REST APIs, Git, MS Office, Docker