Anurag Hambir

(812) 606-3617 | anuraghambir@gmail.com | linkedin.com/in/ah10 | anuraghambir.github.io

SUMMARY

- Results-driven Data Science and Analytics professional with 3 years of experience in recommendation systems, natural language processing, deep learning, predictive modeling, big data analytics, time-series forecasting, and cloud platforms
- Worked on projects in Media, Banking, Insurance, Financial Services, Social Science, and Information Technology domains
- Strong analytical and problem-solving abilities combined with effective cross-functional collaboration and team leadership skills

EDUCATION

Indiana University Bloomington, Master of Science in Data Science; GPA: 3.9/4.0 — Bloomington, IN
Courses: Machine Learning, Artificial Intelligence, Graph Analytics, Cloud Computing — *May 2023*

Savitribai Phule Pune University, Bachelor of Engineering in Computer Engineering; GPA: 3.7/4.0 — Pune, India
Courses: Data Structures, Big Data Analytics, Data Management, Software Engineering, Business Intelligence — *June 2018*

SKILLS

Languages: Python, Java, R, C, C++
Databases: MySQL, PostgreSQL, MongoDB, Hive, Elasticsearch, BigQuery
Machine Learning: Scikit-learn, SpaCy, NLTK, CoreNLP, XGBoost, Pandas, Numpy, TensorFlow, PyTorch, Keras
Tools/Frameworks: AWS, GCP, Snowflake, Hadoop, Spark, Git, REST API, Kibana, JIRA, Docker, Tableau, Flask, Microsoft Office Tools
Certifications: Data Analysis using PySpark

WORK EXPERIENCE

Data Scientist, Credit One Bank (Contract), Las Vegas, NV — *November 2023 – Present*

- Translated Python code to PySpark for customer segmentation, optimizing campaign offer targeting using KMeans clustering on a **6.5M** dataset and achieved a remarkable processing time reduction from **4** hours to an impressive **3** minutes
- Leveraged SQL to extract **200+** features for deriving **150** KPIs, including percentage increase in users and revenue
- Engineered directed acyclic graphs (DAGs) using Apache Airflow and PySpark to seamlessly orchestrate workflows for daily KPI generation. Effectively stored results in tables and presented insightful Tableau Reports to the CEO

Data Scientist, ProMazo Inc., Bloomington, IN — *June 2022 – December 2022*

- Developed a customer retention model using Python for a Fortune 500 banking and insurance client, boosting the F-1 score by **70%** through the XGBoost model and genetic algorithm, resulting in projected savings of **$269,000** from retained customers
- Utilized SQL and Snowflake to add **50** new features to the Attrition model, resulting in a **5%** accuracy improvement
- Optimized LSTM hyperparameters via grid search and cross-validation, boosting churn rate prediction precision by **15%**

Data Scientist, HT Media Ltd, Pune, MH, India — *July 2020 – June 2021*

- Engineered a scalable recommendation system with Python, leveraging profile-based user segmentation and Elasticsearch, and orchestrated deployment using Docker on AWS, contributing to a **1M+** increase in OttPlay application downloads
- Implemented comprehensive data-driven and A/B testing strategies to evaluate the performance of the recommendation engine, resulting in a **15%** increase in click-through rates
- Collaborated with cross-functional teams, employing statistical modeling and data analysis, to extract insights from user preferences and improve content recommendations, resulting in a **20%** increase in average time spent on the platform
- Developed Tableau visualizations for competitor ad data analysis, generating **10%** more leads for new client acquisition
- Implemented software development best practices for a Python ETL pipeline, extracting data from **100+** OTT websites. Managed data storage and retrieval with MongoDB, reducing collection costs by over **40%**
- Migrated the data pipeline to AWS Cloud infrastructure, reducing processing costs by **20%** by eliminating the usage of on-premises servers
- Managed various big data projects involving large unstructured data sets, including prototyping different data cleaning, and processing workflows, while offering guidance and mentorship to a team of **four** engineers

Data Engineer, Persistent Systems, Pune, MH, India — *August 2018 – November 2019*

- Implemented a machine learning pipeline using Python for sentiment analysis on clients' email data, employing Stanford's CoreNLP library and SpaCy, and delivered weekly client satisfaction Tableau reports to the CEO
- Processed and analyzed email data stored in Hadoop, through PySpark and stored the results in Hive tables
- Optimized the performance of the data pipeline by integrating Kafka, reducing the total processing time by **50%**