Anurag Hambir

(812)606-3617 ahambir@iu.edu linkedin.com/in/ah10 anuraghambir.github.io

Education

Indiana University, Bloomington **August 2021 – May 2023**
Master of Science in Data Science; GPA : 3.9/4.0 *Bloomington, Indiana*

Savitribai Phule Pune University **August 2014 – June 2018**
Bachelor of Engineering in Computer Engineering; GPA : 3.7/4.0 *Pune, India*

Technical Skills

Languages: Python (Pandas, Numpy, Scikit-learn), Java, R, Shell Scripting
Databases: MySQL, PostgreSQL, MongoDB, Hive, Elasticsearch
Machine Learning: Recommendation Engine, Classification, Regression, Forecasting (LSTM, ARIMA), NLP (LLM), CV
Tools/Frameworks: Amazon Web Services (AWS), Google Cloud Platform (GCP), Databricks, Hadoop, GIT, REST API's, Kibana, JIRA, Docker, Tableau, Selenium
Certifications: Data Analysis using PySpark

Professional Experience

Data Scientist Intern, Social Science Consulting LLC **August 2023 – Present**
Technologies: Python, Pandas, NumPy, PyTorch, NLP, GNN *Bloomington, IN, USA*

- Utilizing advanced large language models such as BERT and RoBERTa on a large textual data set to generate contextual word embeddings to be used as input features for generating personalized recommendations.
- Leading Graph Neural Network-based recommendation engine development (GraphSAGE, GATConv), enabling personalized foundation suggestions for grant recipients and promoting reciprocal recommendations.

Data Scientist Intern, ProMazo **June 2022 – December 2022**
Technologies: Python, Pandas, NumPy, XGBoost, TensorFlow, SQL, Snowflake, Scikit-learn *South Bend, IN, USA*

- Developed a customer retention model for a Fortune 500 insurance client using XGBoost model and an evolutionary algorithm that increased the F-1 score by **70%**, with projected savings of **$269,000** per retained customer.
- Utilized SQL and Snowflake to add **50** new features to the Attrition model, resulting in a **5%** accuracy improvement.
- Maintained and functionalized an LSTM forecasting model using Keras to predict subscriber churn rates over a 3-month period.

Data Scientist, HT Media Ltd **July 2020 – June 2021**
Technologies: Python, Pandas, ElasticSearch, Docker, AWS (S3, Lambda, EC2, DynamoDB), Tableau *Pune, MH, India*

- Managed a variety of data projects involving large unstructured data sets, while offering guidance and mentorship to a team of **four** junior data scientists.
- Designed and implemented an ETL pipeline using AWS Lambda and S3 for crawling movies and shows from OTT websites reducing data collection costs by over **40%**.
- Designed and developed Tableau visualizations that displayed competitor ad data on HT Media websites, providing valuable market insights to the sales team and contributing to new client acquisition.
- Built a recommendation engine using user profile based segmentation and Elasticsearch database and deployed using Docker on AWS that contributed to the overall increase in the downloads of the **OttPlay application** by over **1 million**.

Software Engineer, Persistent Systems **August 2018 – November 2019**
Technologies: Python, Java, R, PySpark, NLP, Apache Hadoop, Hive, Kafka, Pandas, NumPy *Pune, MH, India*

- Implemented a data pipeline using PySpark to extract and analyze sentiments and entities from clients' email data, employing Stanford's CoreNLP library and SpaCy. Shared weekly reports on client satisfaction with the CEO.
- Processed and analyzed email data stored in Hadoop, through PySpark and stored the results in Hive tables.
- Improved the performance of the data pipeline by integrating Kafka, reducing the total processing time by **50%.**

Projects

Vehicle Detection and Counting in Images (Guided by Prof. David Crandall)

- Transfer learning of YOLOv7, RetinaNet and Faster R-CNN models using TensorFlow, PyTorch on a custom dataset.
- Achieved Mean Average Precision score of **0.76** with YOLOv7, outperforming all the other models.

Map Reduce operations on Google Cloud Platform (GCP) (Guided by Prof. Prateek Sharma)

- Built Map Reduce architecture for word count and inverted index problems using google cloud functions.
- Created an API to invoke the operations and built a web UI to showcase the output.