Shivam Vats

720-761-0049 | shivam.vats@colorado.edu | GitHub | LinkedIn | Website | Medium

EDUCATION

University of Colorado Boulder, **Masters in Data Science**   Aug 2022 – May 2024
Coursework: Statistics, Database Systems, Machine Learning, Big Data, Deep Learning   GPA: 4.0

GGSIPU, **Bachelor of Technology in Computer Science**   Aug 2016 – May 2020

SKILLS

- **Programming Frameworks:** Python, R, MySQL, Postgres, Mongo, Flask, Docker, Keras, Tensorflow, Pytorch
- **Data Processing:** PySpark, Dataproc, Athena, Redshift, S3, EC2, Lambda, Dbt, Tableau, Excel, Influx, Grafana, Plotly
- **Certifications:** Deeplearning.AI, AWS Solutions Architect Associate, Data Science and Machine learning in Python

PROFESSIONAL EXPERIENCE

Dish Networks, Colorado, United States   May 2023 – Aug 2023
Data Science Intern – Product Research and Development [link]

- Architected data pipeline to stream video playback key performance indicators (KPIs). Transmitted data from android application to an InfluxDB datalake. Queried data inside a Grafana dashboard for real time monitoring and analytics.
- Utilized tableau dashboards to communicate the summary statistics of KPIs to the networking teams reducing the response time to potential outages by 23%.
- Integrated datalake with AWS Data Catalog for improved organization and governance. Filed proof-of-concept for optimized transmission of data from the android devices to datalake leveraging Kafka for real-time streaming.

Leeds School of Business, University of Colorado Boulder   Jan 2022 – Present
 Research Assistant [link]

- Automated data collection workflow using Bureau of Labor Statistics (BLS) API, shrinking down the data ingestion time.
- Built econometrics dashboard for Colorado Restaurant Association via Tableau to communicate employment and wage statistics.
- Designed data pipeline to process AUBER (Association of University Business and Economic Research) survey results in R, delivering summary statistics to the stakeholders.

Amazon Web Services, Hyderabad, India   Feb 2020 – Aug 2022
Software Development Engineer

- Designed and developed React-Redux application for customer support engineers reducing average time taken to debug an AWS Organizations ticket by 18%.
- Created an NLP regression model to score customer's sentiment on a scale from 0 to 100 on data collected from support tickets. Reduced mean squared prediction error by 75% utilizing TF-IDF provided by NLTK.

PROJECT AND RESEARCH EXPERIENCES

Song Recommender System [link]   Aug 2022 – May 2023

- Engineered ETL pipeline with Spotify Web API to gather songs data. Modelled it in Dbt and conducted EDA in Tableau.
- Utilized Dataproc to perform distributed computing via PySpark to generate relevant song recommendations.
- Achieved scalability and reliability by CICD on a flask application via Docker.

CU Course Similarity [link]   Jan 2023 – May 2023

- Discovered similar courses offered at University of Colorado Boulder using semantic analysis of course descriptions, helping administration uncover connections across departments.
- Adapted BERT to compute sentence embeddings and stored it in Mongo NoSQL database. Visualized similarity relationships with NetworkX library.

Furnace Maintenance Prediction [link]   Jan 2022 – May 2023

- Programmed time series classification model to predict when an aluminium furnace requires maintenance with 93% accuracy using supervised learning with time-based cross-validation via logistic regression, decision trees, random-forest and xg-boost.
- Checked data stationarity and engineered time-lag features by determining the period of input data signals with Plotly. Devised time-based clustering for feature reduction of quantitative features and chi-squared test for qualitative features.

ACHIEVEMENTS

Winner – Smart India Hackathon   Mar 2019
Led a team of six persons to curate an application reducing exploitation faced by farmers due to middlemen. Developed mobile and web application with ReactJs and NodeJs to sell produce online, book crop storage facilities, and logistics services.