

Shouvik Sengupta

☎ +1 (720)-761-7639 | ✉ shouvik.sengupta@colorado.edu | 🌐 shouvik17 | 📱 Shouvik-7

SUMMARY

Data Scientist with expertise in advanced analytics, machine learning, and software engineering. Skilled at transforming raw data into actionable insights and communicating technical concepts to diverse audiences. Seeking a challenging role to apply advanced data science techniques, programming skills and problem-solving abilities to build data driven business solutions.

EDUCATION

University of Colorado, Boulder, USA

Aug 2022 - May 2024

Master of Science in Data Science | GPA: 3.93/4.0

Relevant Courses: Machine Learning, Statistical Methods, Data Mining, Natural Language Processing, Neural Networks

Vellore Institute of Technology, Vellore, India

Jul 2017 - Jun 2021

Bachelor of Technology in Electrical and Electronics Engineering | GPA: 3.53/4.0

TECHNICAL SKILLS AND CERTIFICATIONS

- **Programming Languages:** Python, R, SQL, Java, HTML, CSS, Javascript, MATLAB
- **Libraries/Frameworks:** Tensorflow, Scikit-learn, Keras, Pandas, Matplotlib, Streamlit, NLTK, plotly, PyTorch, pyspark, LangChain
- **Other Tools:** Databricks, MongoDB, Azure, Tableau, PowerBI, Excel, Docker, Amazon Web Services, Airflow, postgres, Git
- **Skills:** Machine Learning, Data Visualization, Hypothesis testing, Natural Language Processing, ETL pipeline, Agile, Gen AI
- **ML:** SVM, Neural Networks, KNN, K-Means Clustering, DBSCAN, Regression, Decision Trees, Naïve Bayes, XGBoost
- **Certifications:** [DeepLearning.AI TensorFlow Developer](#), [Machine Learning](#), [Data Analysis with Tidymverse](#)

PROFESSIONAL EXPERIENCE

Data Scientist | Virufy

Sep 2024 - present

- Developed and deployed AI models on **AWS** for COVID Detection from coughs for early diagnosis using audio files
- Researched and adapted technologies to build and scale seamless cost-effective **ETL** pipelines reducing processing time by 5%

Data Science Course Assistant | University of Colorado Boulder, Boulder, Colorado

Sep 2023 - May 2024

- Mentored, graded and developed tests for Machine Learning course, fostering supportive learning experiences for 60 students

Software Developer | Tata Consultancy Services, Bangalore, India

Jun 2021 - Aug 2022

- Orchestrated the migration of Restful microservices from **Python** to Java **Springboot**, optimizing the performance by 20%
- Optimized performance by 10% by implementing **Kafka** based messaging and **MongoDB** for data storage and backup
- Utilized **Mockito** for unit testing, and deployed the project seamlessly using **docker** containers over Amazon web services
- Automated build, testing and deployment to AWS using Jenkins, facilitating CI/CD workflows for robustness and scalability

Intern | Schneider Electric Infrastructure LTD, Vadodara, India

May 2019 - Jul 2019

- Built Arduino-based rack-in rack-out counter for switchgear, reducing cost by 50% compared to its mechanical counterpart
- Participated and delivered presentations, and automated updating of weekly FMEA **excel** and **powerpoint** using **Python**

PROJECTS

Driver Profiling (Honda Research Institute)

Jan 2024 - May 2024

- Collaborated in cross functional teams to develop a driving profiling model on **telematics** data with **CNN-LSTM** and **XGBoost**
- Implemented **isolation forest** to analyze driving patterns and anomalies to score drivers thus promoting safer driving practices
- Generated interactive **Tableau Dashboards** offering to communicate driver data insights to non-technical stakeholders

Conversation emotion-cause pair extraction

Aug 2023 - Dec 2023

- Led a team to model emotion-cause pair extraction, enhancing the team's productivity through delegation & clear goal setting
- Achieved **61%** accuracy by Finetuning **RoBERTa** model for emotion classification with **pytorch** strengthening my NLP knowledge
- Applied transfer learning on **BERT SQuAD** Hugging face model, achieving **62.2%** proportional F1 score on causal span detection

English Premier League Predictor

Aug 2023 - Dec 2023

- Built an ETL pipeline using **Airflow** and **AWS**, storing EPL data in **Amazon Redshift** for efficient **SQL** based extraction and analysis
- Trained machine learning models to predict match, and created **PowerBI** dashboards, enhancing data driven decision-making

Statistical Analysis of Greenhouse Gases

Jan 2023 - May 2023

- Performed **Regression Modelling** to forecast CO and NOx levels, drawing meaningful insights for emission management
- Utilized **Exploratory Data Analysis(EDA)**, **Data Cleaning**, **Feature Selection** & **diagnosis methods** to improve adjusted R² by 6%

Chest X-ray Pneumonia Detection

Jan 2023 - May 2023

- Deployed a **AWS EC2** web app using **Deep learning** to detect pneumonia from chest x-rays with **87%** accuracy helping diagnosis
- Integrated **MLflow**, **DVC**, **github actions** and **modular coding** for efficient experiment tracking and version control & CI/CD

EXTRACURRICULAR ACTIVITIES

Team Lead, Tenant Name Detector - 2nd Place, 2023 Data Science Hackathon DaSSA

Team Lead, Species Kingdom Classification - 3rd Place, 2023 Data Science Hackathon DaSSA