Shivansh Gupta

Graduate
IIT (ISM) Dhanbad
Homepage: gshivansh2011.github.io

+91-7725975890
gshivansh2001@gmail.com
linkedin.com/in/shivanshgupta2001
github.com/gshivansh2001

Education

Qualification	Institute	CGPA/Percentage	Year
B.Tech., Petroleum Engineering	Indian Institute of Technology (ISM), Dhanbad	7.83/10	2023
Class XII - Science , CBSE	Nalanda Academy, Kota	87.6%	2018
Class X, CBSE	St. Mary's Sen. Sec. School, Banda	10/10	2016

Experience

- **Indian School of Business,** **Pre - Doctoral Research Associate** *Dec'24 - Ongoing*
  Department: *Information Systems,* ***Advisor:*** *Prof. Vasundhara Sharma* Hyderabad, IN
  - **Research Focus:** Passionate about applying machine learning (ML) and data science to macroeconomic measures and public policy, aiming to evaluate and enhance their impact on society.
  - Conducting a research project on the impact of social movements in the digital world and another project simulating a virtual environment to curb the spread of misinformation in social networks, while supporting professors in academic research and data-driven analyses.
  - Currently working with web scraping tools like Selenium, applying natural language processing (NLP) techniques, and leveraging large language models (LLMs) for text data analysis and generating actionable insights.
  - **Methodologies:** Econometric Analysis, Structural Modeling & Survery
  - **Courses Taken:** Advanced Statistics, Econometrics - I
- **HCLTech,** **Data Scientist** *Jul'23 - Nov'24*
  Clients: *Chevron, ExxonMobil* Bangalore, IN
  - Working with the AIOps team to develop a supervised machine learning model which analyzes Azure Virtual Desktop log data to predict issues to automate and streamline L1 support.
  - Developed a Generative AI application for analyzing O&G daily drilling reports (DDRs) to convert them into OSDU-compliant JSON data for direct ingestion in the OSDU environment. Used Azure ML Studio with OpenAI GPT-3.5 Turbo API and LangChain in Python for developing the application and Gradio for developing the interface.
  - Created and managed data pipelines and dashboards for asset management across 800 O&G wells using Azure Databricks, Power BI and internal asset management tools.
  - Developed an application for registering new and reimaged remote desktop data, creating an SQL database for asset management with Azure Function and Power Apps in Python.
  - Contributed to OSDU production and logging practices with DataVedik.
- **Directorate General of Mines Safety (DGMS),** **Project Intern** *Aug'22 – May'23*
  Advisor: *Prof. Siddharth Aggarwal* Dhanbad, IN
  - Developed data scraping model using `re` & NLTK libraries for text mining & extracting numerical data from yearly mine fatality PDF reports.
  - Created a comprehensive MySQL database to store mine fatality records, enabling effective data processing with pandas.
  - Facilitated a comprehensive analysis of root causes of fatalities through data visualization using Plotly and Matplotlib in Python.
- **HCLTech,** **ML Intern** *May'22 – Aug'22*
  Client: *ExxonMobil* Noida (Remote)
  - Developed adaptive ML-based alarm system using Random Forest in Python for early detection of kicks and blowouts in O&G wells, using Equinor's Volve dataset.
  - Addressed imbalance in the dataset with SMOTE. Trained supervised ML models using Sklearn & TensorFlow for optimal prediction results.
  - Conducted extensive data preprocessing & EDA using seaborn & pandas. Implemented a method to label kicks in the dataset using DDRs & lithological parameters.
  - A patent was filed (in progress) for the data labelling mechanism used in the model. Received pre-placement offer upon completion.
- **IIM Udaipur,** **Research Intern** *May'21 – Jul'21*
  Department: *Marketing,* ***Advisor:*** *Prof. Ashish Galande* Udaipur (Remote)
  - Created a structured dataset of Fortune 500 companies' tweets through web scraping using Twint and Scrapy in Python to analyze CSR activities and study principles of conscious capitalism.
  - Utilized text analysis and NLP techniques to compute lexical diversity and sentiment scores using NLTK and VADER libraries.
  - Preprocessed and vectorized text, and leveraged sentiment scores to develop a sentiment analysis model using RNN in TensorFlow.

Projects

- **Face Mask Detection Using CNN | Dataset: Masked Images using Bing Search API** *May 2021 – Jul 2021*
  Tools: Python, CNN, VGG-19, ResNet50, TensorFlow, Keras Project Repository | Project Demo
  - Developed the model using CNN, applying image data preprocessing and augmentation techniques to enhance robustness.
  - Trained and fine-tuned VGG-19 and ResNet50 pre-trained models, achieving a high training accuracy of 0.994.
  - Demonstrated effectiveness in real-time mask detection scenarios by testing on diverse image sets.
  - Utilized Bing Search API to curate a dataset of masked and unmasked faces for supervised training.
- **Gender & Age Prediction Using Voice | Dataset: Common Voice Dataset** *May 2021 – Jul 2021*
  Tools: Python, TensorFlow, Librosa, Pandas, Seaborn Project Repository | Project Demo
  - Converted audio files and extracted features using MFCC with the Librosa library for audio preprocessing and exploratory data analysis.
  - Trained multiple ML algorithms to build two models for gender and age detection, achieving accuracies of 0.93 and 0.79 respectively.
  - Implemented data visualization and pattern analysis using Pandas and Seaborn to support model interpretation.

Skills

- **Programming Languages**: C, C++, Python, SQL, R
- **Framework**: Keras, TensorFlow, Flask, Lang Chain, Gradio, PyTorch, Selenium
- **Tools**: MS Excel, PowerBI, R Studio, STATA, Azure Cloud (OpenAI Studio, ML Studio, Databricks, Function and Power Apps)
- **Libraries**: matplotlib, seaborn, pandas, NumPy, stats models, NLTK, OpenCV, Plotly, spaCy, Scikit - learn

Certifications

- Azure Fundamentals (AZ 900), Microsoft *2023*
- Azure AI Fundamentals (AI 900), Microsoft *2024*
- Azure Power BI Data Analyst (PL 300), Microsoft *2024*
- Deep Learning Specialization, Deeplearning.ai *2021*

Achievements

- Achieved top 10%ile, Univ.AI Data Science hackathon, displaying analytical prowess with an AIR of 96.
- Headed the SPE IIT (ISM) Student Chapter Editorial, securing the 2022 Student Chapter Excellence Award, awarded to only the top 20% of chapters globally
- Earned a Bronze medal in the Otto Group Product Classification Kaggle Competition, showcasing expertise in ML and data classification.
- Secured 2nd position for writing and directing mono-act & 5th position in stage play as director at InterIIT Cultural Meet '22, outperforming 23 IITs

Positions of Responsibility

- Coordinator, aBhAy Dramatics Club *Apr'21 - Apr'22*
- Head Chapter Editor, SPE IIT (ISM) Student Chapter *Feb'21 - Feb'22*
- Volunteer, Fast Forward India *2019 - 2023*