Shreyas Aswar

317-464-7756| aswarshreyas@gmail.com| shreyasaswar.me| github.com/shreyasaswar |linkedin.com/in/shreyasaswar

Education

Master of Science in Applied Data Science – Indiana University, Indianapolis (GPA: 3.97/4.00) May 2024
Relevant Coursework: Data Analytics, Deep Learning , Data Visualization, Statistics, Informatics, Cloud Computing.

Bachelor of Engineering in Computer Engineering – Savitribai Phule University, India Jul 2018
Relevant Coursework: Data Structures and Algorithms, Mathematics, Database Management Systems.

Experience

Lead Research Assistant, DATA Lab at Indiana University Jun 2023 – May 2024
- Enhanced data consistency and quality by cleaning, aggregating, standardizing qualitative text data of 40,000 sentences.
- Developed classification models (Logistic Regression, LSTM, CNN) to classify English dialects for content localization, achieving 90% accuracy and 30% faster data processing, effectively reducing project timelines by 20%.
- Built and optimized neural network models for English dialect translation using PyTorch, benchmarking 54 BLEU score.
- Leveraged A/B testing frameworks to iterate and refine model parameters for improved predictive performance.
- Saved $30,000 in manual translation costs by using Generative models, delivering results 30 times faster at 0.2% of cost.

Graduate Research Assistant, PLHI Lab at Indiana University Aug 2022 – May 2024
- Implemented an optimized SQL data pipeline to extract and load patient data from MIMIC IV database into an in-memory database, reducing data processing time by 50% for real-time predictive modeling.
- Built real-time training and predictive modeling pipeline for forecasting diagnosis using scikit-learn and TensorFlow.
- Orchestrated the collection, preprocessing, and transformation of over 1 million time series data points from multiple sources on Parkinson's activities using Python, Pandas, NumPy, and Excel.
- Conducted exploratory data analysis, visualized key insights, and developed predictive models (Random Forest, Decision Trees), identifying tremor-reducing activities for Parkinson's patients with 88% accuracy.

Graduate Research Assistant, Indiana University Aug 2022 – May 2023
- Designed and implemented a MySQL relational database with data modeling for research grant proposal management.
- Streamlined workflows by automating data entry and enabling easier retrieval, reducing manual effort by 70%.
- Developed a comprehensive Tableau dashboard that tracked and visualized 10+ key performance indicators, leading to a improved decision-making across departments.

Data Analyst, Gut Lernen Technocraft Dec 2018 – Nov 2021
- Extracted and transformed unstructured data from ERP systems and APIs into structured format using Python and SQL.
- Examined behavior patterns with statistical analysis, resulting informed decision-making & 20% increase in sales.
- Conducted thorough data validation checks, resulting in 28% reduction in data errors and boosting data reliability.
- Collaborated in developing 10+ Power BI dashboards to visualize key findings and perform KPI analysis, driving cost savings of 1.5 lakh INR.
- Streamlined task management processes by implementing JIRA workflows, achieving a 30% faster task completion rate and driving a 15% improvement in team productivity.

Software Developer Intern, Mass Technologies Dec 2016 – May 2017
- Developed RESTful APIs for efficient data handling and integration across systems, leading to streamlined workflows.
- Implemented web scraping for data collection and supporting automated data analysis tasks.

Projects

PurnaData
- Developed a cloud-based ERP solution, enhancing enterprise data management across various sectors through integrated reporting tools for real-time insights and decision support.
- Employed AWS technologies including S3, Glue, Redshift to ensure scalable, secure, and efficient management of data.

Spotify Songs Popularity Analysis
- Analyzed the Spotify data to identify trends, patterns, and potential quality issues, using statistical analysis methods.
- Leveraged R to build linear & decision tree ML models for predicting song popularity, achieving a 64% accuracy rate.
- Conducted hypothesis testing, feature importance analysis, and significance tests to identify critical variables.

NLP - Machine Translation
- Led development of Bert infused Transformer based Machine Translation model for sequence-to-sequence translation, surpassing benchmarks with a 40 BLEU score.
- Optimized model efficiency by implementing 8-bit quantization, significantly reducing computational complexity.

Skills

Programming Languages:	Python, R, SQL, SAS, C++, JavaScript, Bash
Data Analysis:	Pandas, Numpy, Excel, Hypothesis Testing, A/B Testing, Time Series Analysis
AI/ML:	Scikit-learn, TensorFlow, PyTorch, Keras, Predictive modeling, Regression, Clustering
Data Management & Cloud:	AWS, Azure, MySQL, PostgreSQL, MongoDB, Airflow, Snowflake
Tools & Visualization:	Git, Docker, Jenkins, MLflow, Agile, Scrum, Tableau, PowerBI, Matplotlib
Certifications:	AWS Certified Cloud Practitioner, Machine Learning

Publications

- AAVE Corpus Generation and Low-Resource Dialect Machine Translation - **ACM'24**
- Generalizability of Human Activity Recognition ML Models from non-Parkinson's to Parkinson's Disease Patients - **IEEE'23**
- Event Information Extraction From Email And Updating Event In Calendar - **IJARIIE'18**

Professional References

- Sunandan Chakraborty - Program Director Data Science, Indiana University - sunchak@iu.edu
- Suvarna Kadam - VP (AI), KECILIN - suvarna.kadam@gmail.com
- Mathew Palakal - Former Dean, Luddy School, Indiana University - mpalakal@iu.edu
- Eric Graves - Project Lead, NSWC - etgraves@iu.edu
- Ming Jiang - Assistant Professor, Indiana University - mj200@iu.edu