

AMAN SRIVASTAV

Roll No : 2023/DSC/15

Email-id : amansrivastav7112@gmail.com

Mobile No.: 6200733384

[LinkedIn Profile](#)

[Github Profile](#)

EDUCATION

Year	Degree	Institute	CGPA/%
2023-2025	M.Tech(Data Science)	Delhi Technological University(DTU), New Delhi	7.4 CGPA
2019-2023	B.Tech (ECE)	Bharath Institutes of Higher Education and Research	7.6 CGPA
2019	12 th CBSE	Jeewan Public School, Motihari	57.4%
2017	10 th CBSE	Jeewan Public School, Motihari	8.6%

EXPERIENCE

Data Analyst Intern – RealPage *September 2021 – November 2021*

- **Project:** Revenue Analysis Conducted data analysis (**Univariate, Bi-Variate, Correlation**) on attributes of a predictive model and applied **feature transformation** and **imputation techniques** to improve data quality. Utilized **Apache Spark, Python (Pandas, NumPy), Power BI** for visualization, and **Git/GitHub** for version control.

PROJECTS

- **PDF Query System using LangChain, OpenAI LLM, and FAISS**
 - Built a **PDF querying system** using **LangChain** and **OpenAI's Large Language Model (LLM)** for efficient information retrieval.
 - Developed a pipeline to extract, process, and embed text from PDFs, enabling **contextual and accurate querying**.
 - **Tools & Technologies:** Python, LangChain, FAISS, OpenAI API, Pandas, NumPy.
- **Quora Question Pairing System Using NLP**
 - Built an NLP-based system to identify and pair duplicate questions from Quora.
 - Preprocessed data using **tokenization, stemming, and stop-word removal** for feature extraction.
 - Engineered features with **TF-IDF, word embeddings (Word2Vec, GloVe)**, and **cosine similarity** for similarity analysis.
 - **Tools & Technologies:** Python, Scikit-learn, NLTK, Spacy, TensorFlow.
- **Multiple Disease Prediction System – Web App**
 - Developed a **web application** using **Streamlit** to predict multiple diseases, including kidney, heart, and liver diseases.
 - Trained models using **Convolutional Neural Networks (CNN), Random Forest**, and datasets from **Kaggle**.
 - Achieved **96% accuracy** for kidney disease, **85%** for heart disease, and **88%** for liver disease predictions.
 - **Tools & Technologies:** Machine Learning, CNN, Random Forest, Streamlit, Kaggle, Python.

TECHNICAL SKILLS and TOOLS

Programming and Tools: Python, SQL, Power BI, Git, GitHub, FastAPI.

Data Science and Machine Learning: Probability and Statistics, Linear Algebra, Machine Learning, Deep Learning, NLP, GENAI, Hugging Face Transformers.

Cloud Computing and Model Deployment: Apache Spark, AWS (S3, EC2), CI/CD Pipelines.

COURSES AND CERTIFICATIONS

- Complete Generative AI Course With Langchain and Huggingface
- Core Java Programming from Scratch to Advance