Email: sonikumrai54@gmail.com
LinkedIn : www.linkedin.com/in/soni-72780a2a6
Github : https://github.com/s5021

SONI

EDUCATION

Sharda University, Gautam Buddha Nagar, India *August 2023 - May 2025*
Master of Science - Data Science and Analytics — CGPA: 8.978

MGKVP University Varanasi, India *July 2019 - August 2022*
Bachelor of Science - Mathematics — Percentage - 73

EXPERIENCE

Machine Learning Intern
Team Of Keys, Noida, Gautam Buddha Nagar, India, On-site *April 2025 - Present*
Data Science Intern
LearnNex, Tripura (West), India, Remote *Sep 2024 - Nov 2024*
Data Science Intern
Digisamaksh, Greater Noida, India, On-site *May 2024 - July 2024*
Data Science Intern
Oasis Infobyte, New Delhi, India, Remote *April 2024 - May 2024*
Mathematics Tutor
Qalp Educare, Varanasi, India, On-site *October 2022 - August 2023*

PROJECTS

- **Multi-Agent Chatbot Web App** — Developed a multi-agent chatbot using JavaScript, HTML/CSS, Tesseract.js, and Groq's LLaMA API with Agentic (document-based multilingual QA) and Global Chat modes.
- **Football Player Tracking** — YOLOv5/YOLOv8 + DIS Optical Flow with Roboflow; visualizations using Python, OpenCV, Matplotlib.
- **Smart Vision Face Recognition** — Live face detection using OpenCV, Haar Cascades; Tkinter GUI for real-time identification.
- **Community Connect (SDG-8)** — Conducted village survey, analyzed 100+ responses with SPSS and hypothesis testing.
- **Unemployment Analysis (Python)** — EDA on 500K+ entries using Pandas, NumPy, Seaborn; created visual reports and insights.
- **WhatsApp Chat Analyzer** — Python, NLP, Matplotlib; chat stats, sentiment, active times, and keyword analysis from chat files.
- **Stock Price Predictor** — TensorFlow, yfinance, Streamlit; predicted and visualized stock trends in an interactive UI.
- **Zomato Data Analysis** — Cleaned, preprocessed, visualized food review data using Pandas, NumPy, Matplotlib, Seaborn.
- **Movie Recommendation System** — Built collaborative filtering model using cosine similarity in scikit-learn.
- **Email spam detections** — Developed a spam classifier using TF-IDF Vectorization and Logistic Regression. Learned fundamentals of NLP, feature engineering, and text classification.

TECHNICAL STRENGTHS

Language	Python (DSA), C, SPSS
Visualization	Power BI, Matplotlib, Seaborn
Statistics	Hypothesis Testing, Regression
Databases	SQL (MySQL, PostgreSQL)
Big Data	Hadoop, Spark (Basic)
ML Frameworks	TensorFlow, PyTorch, Scikit-learn, Keras
Deep Learning	CNNs, RNNs, LangChain

CERTIFICATIONS

Data Structures and Algorithms in Python(Geeks For Geeks)-Ongoing
Learn Python Programming- Beginner to Master(Udemy)
Building Language Models on AWS (AWS)
Big Data Hadoop and Spark Developer Training(Simplilearn)
MongoDB Developer and Administrator(Simplilearn)