Soumyodeep Nayak

+91 89899 54354 — soumyo2004@gmail.com — soumyodeep-nayak — soumyo2004

Education

VIT Bhopal University **Oct 2022 – 2026 (Expected)**
B.Tech in Computer Science and Engineering — CGPA: 8.15

Skills

Python (Pandas, NumPy, Matplotlib, Scikit-learn), MySQL, MongoDB, Power BI, Tableau, Excel, Nmap, Wireshark, Burp Suite

Projects

Student Score Prediction Model **Jun 2023 – Jul 2023**
Tools: Python, Scikit-learn, Pandas, Matplotlib

- Developed a linear regression model to predict student scores based on hours studied, achieving 95.15% training and 94.54% testing accuracy; evaluated performance using MAE, MSE, and RMSE.
- Visualized regression line, residuals, and prediction trends to interpret model behavior and communicate insights effectively.
- Reduced prediction error by 12% after applying outlier removal and min-max normalization techniques.
- **Dataset:** http://bit.ly/w-data
- **Repository:** https://github.com/soumyo2004/GRIPJUNE23.git

Exploratory Data Analysis: Global Terrorism **Jun 2023 – Jul 2023**
Tools: Python, Pandas, Seaborn, Plotly, Folium

- Analyzed 181,691 global terrorism incidents across 47 years (1970–2017) to identify high-risk countries, regions, attack types, and target groups.
- Created 10+ visualizations (heatmaps, pie charts, time plots, interactive maps) to highlight hotspots and casualty patterns.
- Clustered incidents by frequency and fatalities to derive actionable insights for national security and defense planning.
- Segmented 100+ countries by incident density to prioritize regions with ¿1,000 attacks for policy recommendations.
- **Dataset:** Global Terrorism Database (GTD)
- **Repository:** https://github.com/soumyo2004/GRIPJUNE23.git

SMS Spam Classifier **Dec 2023 – Jan 2024**
Tools: Python, Scikit-learn, NLTK, Matplotlib, Seaborn

- Built a classifier for 5,572 SMS messages using Multinomial Naive Bayes, achieving approx. 97% accuracy.
- Developed a full NLP pipeline with custom preprocessing (punctuation removal, tokenization, stopword filtering) and TF-IDF vectorization.
- Evaluated model using confusion matrix, F1-score (0.96), precision (0.97), and recall (0.95); visualized with Seaborn.
- Improved accuracy by 4% via hyperparameter tuning (alpha, n-gram range) using GridSearchCV on 5-fold cross-validation.
- **Repository:** https://github.com/soumyo2004/NaiveBayesSpamClassifier.git

Work Experience/Internships

Data Analyst Intern **Aug 2023 – Sep 2023**
Saint Louis University, Missouri, United States (Remote)

- Enhanced decision-making by delivering 12+ insights from data cleaning, validation, and Tableau dashboards.
- Ensured 95% data quality using robust statistical methods and improved dashboard readability, increasing stakeholder engagement by 25%.

Certifications

- Cyber Security Analyst (CECSA1IN, IBMCE) **April 2025**
- MongoDB Associate Database Administrator **April 2025**

Co-Curriculars

JPMorgan Chase & Co.'s Cybersecurity Virtual Experience Program (Forage) **August 2023**

- Gained foundational skills in cybersecurity, access control, and email security hygiene and enhanced knowledge in data structures and text-based machine learning models.

Extracurricular

Social Impact: Volunteered for NSS, Unnat Bharat Abhiyan, and organized fundraiser for Sewa Sadan Eye Hospital
Keyboard Performer: National-level participant; 2nd Year Diploma, Prayag Sangeet Samiti