SWATI BADOLA

Delhi, India |swatibadola39@gmail.com | (+91) 76003-79535 | LinkedIn | GitHub | Portfolio

SUMMARY

Hands-on data science enthusiast with experience building and deploying machine learning and NLP applications. Skilled in developing end-to-end ML pipelines, deploying models on Heroku and AWS, and performing deep data analysis. Strong in Python, DSA, and cloud-integrated solutions. Passionate about solving real-world problems using ML.

SKILLS

Competencies: Linear Regression, Logistic Regression, TF-IDF, Sentiment Analysis, Feature Engineering, Exploratory Data Analysis (EDA), Tokenization, Named Entity Recognition (NER), Lemmatization, Model deployment, Input Validation, Data Visualization, Deep Learning
Programming Languages: Python, SQL
Libraries/Frameworks: Scikit-learn, TensorFlow, NLTK, Pandas, NumPy, Matplotlib, Seaborn, Plotly, Regex, BeautifulSoup, Flask
Tools/Platforms: Git, GitHub, Heroku, Flask, Google Collab, Kaggle, AWS, MongoDB
Databases: MongoDB, SQL
Cloud/Deployment: Heroku, AWS

PROJECTS

AI-Powered Sponsorship Detection System **Personal Project**

- Designed and implemented a multi-modal ML pipeline to detect sponsored content across social media posts and YouTube videos using NLP and pattern recognition techniques.
- Engineered named entity recognition (NER) and brand detection modules to accurately extract and classify promotional brand mentions in content.
- Collected and labeled a custom dataset by scraping public content, enabling supervised learning without reliance on platform APIs.
- Visualized insights via a user-friendly dashboard, enabling trend discovery in brand-influencer collaborations.

Student Score Predictor **Personal Project**

- Developed a linear regression model to predict student math scores based on study habits and demographics ($R^2 = 0.84$; MAE < 5), improving prediction accuracy for academic planners.
- Engineered input validation logic on the front end, reducing erroneous submissions by ~20%.
- Deployed model via Flask on Heroku to simulate production ML workflows. [Link]

Sentiment Analysis on Movie Reviews **Kaggle Dataset**

- Built a text classification pipeline using TF-IDF and logistic regression, achieving 90% accuracy on 8,000+ product reviews.
- Preprocessed text via lemmatization, stop word removal, and tokenization using NLTK to boost model reliability.
- Visualized sentiment polarity using word clouds and distribution plots to uncover behavioral insights. [GitHub]

Customer Satisfaction Analysis **Kaggle Dataset**

- Conducted exploratory data analysis on telecom user data to identify churn patterns and retention drivers.
- Utilized heatmaps, box plots, and segmented bar charts to present actionable KPIs to business stakeholders. [GitHub]

EDUCATION

Delhi Skills & Entrepreneurship University **Delhi, India**
Bachelors Of Technology |CGPA: 8.04 *Jan 2021 - 2025*
Courses: Artificial Intelligence & Machine Learning, Robotics Technology
Achievement: Advanced to Semi-finals of *DSEU Incubation 2023* for a tech startup pitch

CERTIFICATIONS

Machine Learning Summer Internship Program