AWANTIKA SRIVASTAVA

Associate Data Scientist | Machine Learning | Analytics

+91-8920482037 |sawantika81@gmail.com| LINKEDIN |Github

PROFILE SUMMARY

Associate Data Scientist with **2+ years of experience** in **data analysis, machine learning, and statistical modeling**, skilled at transforming large, complex datasets into **actionable business insights**. Strong foundation in **Python, SQL, statistics, and ML**, with experience supporting **data-driven decision making** in fast-paced environments.

CORE TECHNICAL SKILLS

- **Programming & Data Science:** Python (Numpy, Pandas, Scikit-learn), SQL, MySQL, Jupyter Notebook.
- **SQL & Data Management:** SQL (Joins, Subqueries, CTEs), Data Retrieval, Data Cleaning, Data Validation.
- **Statistics & Mathematics:** EDA, Statistical Modeling, Hypothesis Testing, Confidence Intervals, A/B Testing, Scenario Analysis.
- **Machine Learning:** Supervised & Unsupervised Learning, Regression, Classification, Clustering, Random Forest, Decision Trees, SVM, KNN, K-Means, XGBoost.
- **Tools & Frameworks:** PyTorch, TensorFlow, Hadoop (Exposure), GitHub.
- **Data Engineering & Processing:** Data Cleaning, Data Mining, Structured & Unstructured Data Processing, Automation.
- **Deep Learning :** Neural Networks, CNN, RNN, LSTM, Transformers (BERT), Transfer Learning, Model Fine-tuning, Computer Vision.
- **Data Visualization & Reporting:** EDA, Data Visualization, Business Reports, Insight Presentation.
- **Professional Skills:** Analytical Thinking, Problem Solving, Technical Communication, Independent & Collaborative Work.

EXPERIENCE

Data Scientist / Machine Learning Engineer | PPS International Pvt. Ltd.   **January 2024-Present**
- Worked with **large, complex internal and external datasets** to evaluate trends and support business and operational strategies.
- Performed **data preprocessing and cleansing** on structured and unstructured data to ensure high data quality.
- Built and evaluated **machine learning models** (classification and regression) for analytical and predictive use cases.
- Conducted **exploratory data analysis (EDA)** to detect anomalies, data inconsistencies, and performance drivers.
- Automated data collection and analysis workflows using **Python and SQL**.
- Developed **reports and visualizations** to communicate findings to technical and non-technical stakeholders.
- Assisted in assessing system changes, testing updates, and evaluating impacts on production analytics systems.
- Proposed **data-driven solutions and strategies** to address business challenges.

PROJECTS

Railway Driver Assistance System | Enterprise ML Project
- Designed and implemented an **end-to-end ML pipeline** for real-time detection of unsafe driver behavior using video and sensor data.
- Performed **large-scale data preprocessing, labeling validation, and exploratory data analysis (EDA)** to identify noise, bias, and edge cases.
- Built and evaluated **classification models** using deep learning–based feature extraction to detect risk patterns in real time.
- Optimized inference pipelines to achieve **20–25 FPS low-latency performance**, enabling near real-time decision support.
- Conducted **error analysis and model performance monitoring** to improve robustness across varying lighting and motion conditions.
- Translated model outputs into **actionable safety insights** to support operational decision-making and system deployment.
- Collaborated with engineering teams to ensure **deployment readiness and system scalability**.

NLP-Based Intelligent Chatbot (RAG-enabled)
- Developed an **NLP-based chatbot** to provide automated question answering over structured and unstructured datasets.
- Implemented **retrieval-augmented generation (RAG)** techniques to improve response accuracy and contextual relevance.
- Designed efficient **text preprocessing and chunking strategies** to enhance retrieval performance.
- Evaluated chatbot responses using **relevance and accuracy metrics**, iteratively refining prompts and retrieval logic.
- Deployed the solution using **Python-based APIs and lightweight web interfaces** for scalable user access.

YouTube Comments Sentiment Analysis (BERT-based) | link - https://youtube-ai-analyzer-ndzqo6r2mepjrtsdjmwaxl.streamlit.app/
- Analyzed **large volumes of user-generated text data** to uncover sentiment trends and engagement patterns.
- Fine-tuned **BERT-based transformer models** for multi-class sentiment classification tasks.
- Performed **data cleaning, tokenization, and feature engineering** to improve model performance.
- Evaluated models using **precision, recall, F1-score**, and confusion matrix analysis.
- Communicated sentiment insights through **clear summaries and visualizations** to demonstrate business impact.

CERTIFICATION

- IBM Data Science & AI Certification
- AWS Generative AI with Large Language Models
- OpenCV Computer Vision Certification

EDUCATION

IMS Engineering College, Ghaziabad   **September - 2020**
Bachelor of Technology (Electrical and electronics engineering)