Aman Kashyap

Jagdalpur, Chhattisgarh devaman0110@gmail.com 62641 17894 portfolio
aman-kashyap-3a4bb7237 amankas864

About

Enthusiastic and results-driven Data Scientist with a strong foundation in digital analytics, machine learning, and data-driven insights, eager to optimize customer experiences. Proficient in Python, SQL, and big data processing, with proven ability to develop impactful analytical solutions and support experimentation. Adept at transforming complex data into actionable recommendations to drive business growth and efficiencies in dynamic, cross-functional environments.

Education

SRM Institute of Science and Technology *Sept 2021 – May 2025*
B. Tech in Computer Science (Data Science)
Chennai, India

- GPA: 8.69

Experience

Data Science Intern *Bhopal, MP*
InfoBharatInterns *June 2025 – July 2025*

- **Developed and deployed a machine learning-based sentiment analysis model** that achieved an **87% accuracy** in classifying restaurant reviews as positive or negative, by leveraging Natural Language Processing (NLP) techniques including TF-IDF vectorization and Logistic Regression.
- **Performed comprehensive data preprocessing and feature engineering** on a dataset of **568,454 restaurant reviews**, optimizing data quality and model input, through tokenization, stop word removal, and stemming/lemmatization.
- **Built and integrated a user-friendly web interface using Streamlit** to enable real-time sentiment prediction for new reviews, **reducing manual review classification time by an estimated 70%** and enhancing the practical application of the model.
- **Applied robust model evaluation techniques** to iteratively refine and select the optimal algorithm (Logistic Regression) for sentiment classification, leading to an **5% improvement in precision** over baseline models.

Projects

Dionysus - Github Repository Tool *github link*

- Engineered a full-stack AI codebase intelligence platform using **Next.js, TypeScript, and Prisma**, capable of processing and generating semantic embeddings for **100+ file codebases** from GitHub repositories.
- Developed an AI-powered code understanding system with **Google Gemini embeddings and LangChain** to provide intelligent Q&A, and integrated **AssemblyAI** for real-time meeting transcription and automated summaries.
- Built a highly performant semantic search engine using **PostgreSQL with the pgvector extension**, enabling developers to find relevant code snippets across **1000+ files in milliseconds** with automated indexing.
- Implemented a scalable **credit-based billing system with Stripe**, and integrated team collaboration features for project sharing and access control, supporting a multi-user environment.

Sentiment Analysis using BERT *colab link*

- **Developed a sentiment analysis model leveraging BERT** to classify text data, **achieving 92% accuracy** by implementing advanced data preprocessing and fine-tuning the BERT model on a labeled

dataset.

- **Optimized model hyperparameters** through rigorous experimentation, **improving prediction precision by 5%** and showcasing strong skills in NLP, deep learning, and model evaluation.
- Tools Used: Python, Pytorch, Pandas, Numpy, Matplotlib, Seaborn, Transformers, Scikit-learn

Fuzzy Energy System *github link*

- **Designed and implemented a fuzzy logic-based energy management system** to **optimize energy distribution and decision-making** under uncertain conditions, by developing a robust fuzzy inference system using Python and `scikit-fuzzy`.
- **Enhanced system efficiency and stability**, potentially leading to an **estimated 15% reduction in energy waste**, through the application of custom-defined fuzzy rules and membership functions.
- **Modeled complex energy system dynamics and processed relevant energy data**, demonstrating expertise in control systems and handling imprecise or ambiguous data, utilizing Python and scientific libraries.
- **Contributed to a project focused on smart energy solutions**, gaining practical experience in applying advanced computational intelligence techniques to real-world engineering problems, by designing and testing system responses.

Certifications and Tests

The Complete Python Pro Bootcamp *Udemy* July 2024

- Completed a comprehensive 100-day bootcamp, mastering **Python fundamentals**, **object-oriented programming (OOP)**, and advanced concepts including error handling, APIs, and web development with **Flask**.
- Developed multiple real-world applications, including **web scrapers, GUI apps, and API-driven websites**, demonstrating robust problem-solving and software development skills.
- Acquired strong skills in **building and consuming REST APIs** and applied **data manipulation and analysis** techniques using Pandas, NumPy, and Matplotlib.
- Applied fundamental **statistical and machine learning concepts** using Scikit-Learn, including regression, t-tests, and residuals analysis.

Supervised Machine Learning *Coursera* Oct 2023

- **Achieved a score of 99.6%**, demonstrating in-depth knowledge and practical skills in foundational supervised machine learning algorithms, including **linear regression, logistic regression, decision trees, and support vector machines**.
- Applied techniques for **model training, evaluation, cross-validation, and hyperparameter tuning** to solve real-world regression and classification problems.
- Developed proficiency in **statistical inference and hypothesis testing**, including the application of concepts for evaluating model performance and significance.

Skills

Technical Skills: Python (NumPy, Pandas, Scikit-learn, NLTK), C, C++, SQL (Postgres SQL), Big Data (Kafka, Redis), Data Visualization (Grafana, Matplotlib), Cloud Platforms (AWS), DevOps Tools (Docker, Kubernetes, Jenkins, Terraform, Git), NoSQL (MongoDB).

Analytical Skills: Data Mining, Statistical Analysis, Hypothesis Testing, Regression, Classification, Exploratory Data Analysis.

Soft Skills: Analytical, Collaborative, Communicative, Adaptable, Problem-Solver.

Extra Curricular Activities

Karate, Cycling, Badminton, Chess, Basketball