Sayan Chakraborty

Mobile: +91 8334869162
Email: sayanchakraborty289@gmail.com
LinkedIn: linkedin.com/in/sayanchakraborty97
Location: Kolkata

Data science graduate with a strong background in quantitative analysis, artificial intelligence, and experimentation. Highly skilled in Python and R for data manipulation and visualization. Exceptional problem-solving skills and a keen eye for detail. Looking to apply my expertise in data-driven decision- making to contribute to meaningful projects. Committed to staying current with industry advancements and continuously improving my skill set.

Skills

- **Programming**: Python, R, Java, HTML, CSS
- **Machine Learning & Deep Learning**: Pandas, NumPy, SciKit Learn, TensorFlow
- **Web Framework:** Flask, Django
- **Data Visualization:** Tableau, Matplotlib, Seaborn, Power Bi
- **Databases**: MySQL, Oracle SQL
- **ETL& Big Data:** Hadoop, Data warehousing
- **AI Applications**: Natural Language Processing (NLP), BERT model, CNN, LSTM
- **Cloud Technologies**: AWS EC2, AWS S3, Azure Storage services, Azure Databricks, Azure MlOps
- **Tools**: VS Code, PyCharm, RStudio, Jupyter Notebook, MS Office

EDUCATION

MSc Data Science (Distinction) University of Essex, UK **Oct 2022 - Sep 2023**

Modules included: Data Science and Decision Making, Mathematical modelling and Applied Statistics, Database Systems, Deep Learning and ANN, Machine Learning, Natural Language Processing.

BTech Computer Science Asansol Engineering College, India **Aug 2016 - Jul 2020**

Modules included: Data Structure Algorithm, Python, Java, Databases, Networking.
CGPA:7.78/10

WORK EXPERIENCE

Data Scientist Associate, BlackCoffer **Oct 2023 - Present**

- Developed a custom Python web scraping script for client-specific data extraction needs, demonstrating expertise in Python and web scraping.
- Creating python script for organizing data for client promotional activity.
- Utilized advanced SQL queries for efficient data retrieval from databases, contributing to improved data-driven decision-making.
- Managed client interactions, ensuring clear communication and tailored solutions to meet specific requirements.

Assistant System Engineer, **Tata Consultancy Services** **Apr 2021 - Sep 2022**

- Developed and maintained JCL scripts for efficient execution of batch jobs, ensuring smooth data processing.
- Fixed software bugs promptly, achieving a bug resolution rate of 95% and ensuring optimal system performance.
- Resolved Java-related issues with an average turnaround time of 24 hours, minimizing downtime and improving system stability.
- Collaborated with cross-functional teams to implement new features, resulting in a 20% increase in application functionality.
- Monitored client software daily, identifying and addressing performance issues proactively.

Awards and Recognition

- Star team award for successfully completing the aws migration project.

Project Engineer, Wipro Technologies **Sep 2020 - Mar 2021**

- Leveraged Hadoop, Spark, and Hive to handle and analyze large datasets, resulting in efficient data processing and improved insights.
- Conducted data cleaning operations, enhancing data quality and accuracy by 30% and enabling more reliable analysis.
- Utilized SQL queries to extract relevant data from databases, supporting data-driven decision- making processes.
- Created comprehensive reports presenting key findings and insights, facilitating informed decision- making by stakeholders.

PROFESSIONAL DEVELOPMENT

Internship

AI/ML Development, Baavlibuch **Aug 2023 - Oct 2023**

- Worked on the development of a **chatbot for patient education**.
- Modularised the code for semantic search on BioOntology portal.
- Deployed code using **Django**.
- Did **syntactinc string matching using ngrams and naïve methods**.
- Ranked the different methods for retrieval from a graph dataset.

Projects:

Dissertation: Detecting Aggression in Social Media Post using Transfer learning and Deep Learning methods

- Applied text preprocessing techniques like **tokenization** and **feature extraction**.
- Established a baseline using **Multinomial Naive Bayes**.
- Advanced models included **CNN** and **LSTM** architectures.
- **Fine-tuned BERT (LLM) for detecting aggression based on specific dataset**.
- Project Link: https://github.com/sayan936/Social-Media-Aggression-Detection

Cricket Player Statistics Chatbot Development:

- Integrated **Streamlit Chat** for seamless messaging, enhancing user engagement and providing a conversational interface.
- Employed **HuggingFacePipeline with the model "google/flan-t5-base"** for natural language understanding, enabling accurate responses to complex cricket-related queries and **HuggingFaceEmbeddings with 'thenlper/gte-large'** model to generate semantic embeddings.
- Configured **FAISS (Facebook AI Similarity Search)** for fast and efficient similarity search in large datasets, improving response times and relevance of chatbot answers.
- Created custom **PromptTemplate for structured query handling**, ensuring the chatbot provides relevant and contextually correct information to user inquiries.
- Developed a **RetrievalQA** chain integrating the chatbot with a retrieval system, allowing for accurate and data-driven answers based on T20I cricket player statistics.
- Project Link: https://github.com/sayan936/cricbot

LLama Model Fine-Tuning for Enhanced Language Understanding:

- Fine-tuned the **LLama 2-7B** language model on the **OpenOrca dataset** to improve natural language processing capabilities.
- Employed advanced NLP tools and techniques, including **Transformers and BitsAndBytes**, for efficient model optimization.
- Implemented **PEFT (Parameter-Efficient Fine-Tuning)** strategies to enhance model adaptation with minimal parameter adjustments.
- Utilized **HuggingFace Hub** for effective dataset management and versioning.

Real-Time Data Processing System using Azure Databricks

- Developed a real-time data processing system leveraging **Azure Databricks** for streamlined analytics and machine learning workflows.
- Integrated Azure Databricks with **Azure Blob Storage** to ingest and store large volumes of data efficiently.
- Implemented **Spark** Structured Streaming in Databricks for real-time data processing, enhancing decision-making processes with timely data insights.
- Configured Databricks clusters to auto-scale according to workload demands, optimizing resource utilization and reducing operational costs.

Azure Cloud Data Management System:

- Designed and implemented a cloud-based data management system utilizing Microsoft Azure services including **AzureSQL Database, Azure Data Factory, and Azure Blob Storage** for efficient data storage, transformation, and retrieval.
- Developed **Azure Logic Apps** to automate workflows, enabling seamless data integration and synchronization between cloud services and on-premises databases.
- Configured **Azure Active Directory** for secure user authentication and authorization, ensuring data security and compliance with industry standards.

.

Language and Certification:

German A1: Goethe A1 exam certified.