ANEESH KRISHNA

aneeshkrishna4739@gmail.com | Linkedin | Github | Portfolio | +1 716 303 1623

EDUCATION

University at Buffalo, The State University of New York **Aug 2023 – Dec 2024**
Master of Science, Majors: Data Science Buffalo, NY
Relevant Coursework: Statistical Learning and Data Mining, Machine Learning, Deep Learning

Visvesvaraya Technological University **Aug 2016 – Aug 2020**
Bachelor of Engineering in Computer Science India
Relevant Coursework: Data Structures, Algorithms, Software Engineering, Database Management

SKILLS AND CERTIFICATIONS

- **Languages**: Python, SQL, R, MATLAB
- **Software:** AWS, GCP, Langchain, TensorFlow, Pytorch, scikit-learn, Keras, Hugging Face Transformers, Power BI
- **Frameworks**: XGBoost, LightGBM, Optuna, Pandas, Numpy, Plotly, SQL, Seaborn, Pinecone, Groq, Huggingface, MLops

EXPERIENCE

Cybersecurity Research Assistant **Aug 2024 – Present**
University at Buffalo Remote

- Developed an Intrusion Detection System (IDS) achieving 99.98% accuracy using a BERT-based NLP model and 98.55% accuracy with a Naive Bayes classifier, leveraging GPU clusters via SSH for model training and optimization.
- Curated the dataset for Advanced Persistent Threat (APT) detection by collecting and processing over 1 million rows of ransomware attack data using commercial simulation tools like Picus.
- Benchmarked multiple ML models for APT detection by tracking accuracy and runtime performance, using tmux for session management and MLflow for tracking experiments.

Machine Learning Engineer Intern **Aug 2024 – Dec 2024**
Advance2000 Buffalo, NY

- Delivered a RAG-powered Helpdesk Assistant that boosted ticket resolution speed of support team by 250% by leveraging past solutions to provide relevant response suggestions for new tickets.
- Enhanced model accuracy by 10% through advanced prompt engineering, including chain-of-thought, few-shot prompting, and metadata filtering, optimizing ticket solution relevance.
- Reduced helpdesk response time by 25% by implementing an LLM system using Hugging Face embeddings, Pinecone vector databases, and Llama 3.1 through Groq for real-time ticket analysis.
- Processed solutions for 42,000 open tickets, achieving 98% completion rate within a day by implementing a custom rate-limiting algorithm and parallel processing, optimizing for API token usage and operational efficiency.

Data Migration Analyst **Oct 2020 – Apr 2023**
Infosys Ltd Remote, India

- Designed and automated Power BI dashboards to track data migration progress across 90+ sites, using DAX for calculations and providing real-time analytics for data-driven decisions.
- Coordinated UAT for Power BI reports, enhancing reporting accuracy by 40% through advanced data visualization and stakeholder-driven improvements.
- Streamlined SharePoint site redesign using PowerShell, SQL scripts, and C#, .Net, and JavaScript, reducing manual intervention by 50% and improving deployment efficiency.
- Led a team of 5 engineers in automating data workflows and modernizing legacy SharePoint sites, boosting processing efficiency by 60% with SPFX, Azure DevOps, and optimized data pipelines.

PROJECTS

YouTube Q&A Bot | Hackathon (~24 hrs) - GitHub *Python, RAG, Flask, LLM*

- Developed a chatbot using Flask, YouTube API, and Pinecone to build a chatbot that answers questions based on YouTube video content and provides timestamps.
- Applied RAG by vectorizing transcripts with Hugging Face embeddings and using GROQ API for LLM-based responses.

Sports Statistics Generator | Published Website (~100 hrs) - Site / GitHub *Python, GenAI, Streamlit, FireStore*

- Engineered a website using Streamlit, Python, Firestore, and Gemini Pro LLM to generate cricket statistics plots based on user queries on a dataset of over 1 million records.
- Collected user queries and feedback on plot accuracy to automatically generate Pandas code, refining the model using Google AI Studio on Google Cloud Platform.

Patient Outcomes Predictor | Team Project (~60 hrs) - GitHub *Python, Optuna, Seaborn, scikit-learn*

- Cleaned and analyzed healthcare data using Polars and Seaborn to reduce dataset size by 90%, identifying trends in diseases and post-operative complications.
- Devised predictive models (Logistic Regression, SVC, Random Forest, Gradient Boosting) to estimate post-operative complications with 82% accuracy after Bayesian optimization.

Business Loan Classifier | Team Project (~40 hrs) - GitHub *Python, Seaborn, PostgreSQL*

- Leveraged PostgreSQL and Python for data preprocessing to maintain integrity and optimize ETL processes; performed EDA to analyze business sectors and default rates, achieving 93% accuracy in loan repayment predictions using Random Forest.