Avinash Pandey

aopandey@purdue.edu | (317) 998-3501 | Indianapolis, IN 46204 | avinashpandey.streamlit.app/ | github.com/Aopandey

Education

Bachelor of Science in Computer Science   Graduated: December 2024
Purdue University, IN
Minor: Mathematics   Cumulative GPA: 3.3
Awards and Honors: Dean's List (Spring 2023 – Fall 2024), International Jaguar Excellence Award, ALDPES Honors society

Student & Professional Organization

- **Team Leader**, Boiler Gold Rush (BGR)   April 2024 – August 2024
  - Facilitated the orientation of 800+ students, leading a team to promote academic success and community engagement
- **Vice President**, Computer Science Club (CS Club)   May 2023 – May 2024
  - Revitalized the club, organized hackathons attracting 200+ students, secured sponsorship, and increased engagement by 100%

Work Experience

Machine Learning Engineering Intern   July 2024 – November 2024
Aider Ventures   Indianapolis, IN

- Developed automated data pipelines using FAISS embeddings and ChromaDB to summarize over 2,600 research papers from ICML 2024, optimizing data handling and ensuring accurate information extraction from scraped source icml.cc
- Engineered scalable pipelines leveraging Gemini API to process and classify over 10,000 research papers, improving the categorization process of datasets and increasing system efficiency by 40% through data processing techniques
- Implemented metadata filtering algorithms in LangChain, improving precision of summarized research outputs by 20%
- Constructed an interactive dashboard for visualizing and analyzing key insights like the number of papers by university and top research areas, enabling non-technical professionals to get insightful trends across over 10,000 papers

Business Analyst Intern – Office of Technology Services   December 2023 – March 2024
Legislative Services Agency   Indianapolis, IN

- Collaborated with Software Developers and Business Analysts to support and troubleshoot internal software using tools like Tableau, SQL Server, and Power BI, achieving a 95% resolution rate
- Executed data verification processes through ETL pipelines, ensuring 98% accuracy for Indiana General Assembly datasets
- Facilitated strategic interactions and technical support with Indiana lawmakers and Legislative Services Agency staff, optimizing business processes through the development of tailored applications and problem resolution tools

Projects

Mutual Learning Algorithm for News Classification (Senior Capstone Project)   August 2024 – December 2024

- Developed a mutual learning algorithm for news classification, achieving 98% accuracy on the BBC news dataset
- Integrated Multinomial Naive Bayes, SVMs, and MLP Neural Networks with preprocessing steps tokenization & lemmatization
- Enhanced performance by 3.21% through a "student-teacher" framework for resource-constrained environments

AI-Driven Diabetes Prediction Pipeline (Machine Learning and Deep Learning)   November 2024

- Achieved 97% recall and 95% ROC AUC in predicting diabetes using a pipeline combining ML (Random Forest, SVM, Decision Tree) and deep learning (Feedforward Neural Networks with MLP) models on a dataset of 100,000 samples
- Addressed class imbalance with SMOTE and SMOTE-Tomek and optimized hyperparameters via GridSearchCV
- Identified HbA1c and blood glucose as key predictors and improved model performance by 15**%** using batch normalization and regularization, demonstrating scalable and interpretable healthcare AI solutions

Weather Station Data Hub (Python)   March 2024

- Developed a Flask-powered API delivering real-time weather data and analytics with 99% uptime
- Integrated Pandas for data processing and year-filtered insights to streamline data ingestion and delivery
- Showcased expertise in API development, data manipulation, and scalable web services

Counts of Bushels Project (SQL)   August 2023 - October 2023

- Designed a 5NF database in SQL Server '22 for Indiana's corn & soybean yield, streamlining storage for 100+ stakeholders
- Conducted advanced analytics with SQL queries, including yield comparisons and trend tracking
- Optimized queries to deliver actionable insights on agricultural production and resource allocation

Technical Skills

- **Programming and Data management:** Python, SQL, R, Java, C++, JavaScript, SQL Server, ETL Pipelines
- **Frameworks and Tools:** TensorFlow, PyTorch, Scikit-learn, Django, Pandas, NumPy, Tableau, Power BI, Git, GitHub
- **Cloud and Big Data:** AWS (S3, Lambda, EC2), Azure Data Lake, Hadoop, Spark, Docker, Dagster
- **Machine Learning and AI:** OpenAI API, Gemini API, LangChain, Hugging Face, FAISS, ChromaDB, ZenML, Haystack