

AVINASH PANDEY

aopandey24@gmail.com | 317-998-3501 | Indianapolis, IN | <https://www.linkedin.com/in/avinashopandey>

TECHNICAL SKILLS

Scripting Languages and Packages: Python, SQL, Java, Pandas, NumPy, SciPy, PySpark

Big Data and Cloud Technologies: AWS (EC2, S3, Lambda), SQL Server, Docker, Kubernetes, Dagster, Git, Github

ETL and Data Processing: Data Warehousing, ETL Pipelines, Data Orchestration, Streaming Data Processing, Data Governance

PROFESSIONAL EXPERIENCE

Aider Ventures

July 2024 – Nov 2024

Machine Learning Engineering Intern

Indianapolis, IN

- Designed and optimized automated ETL pipelines to process over 2,600 research papers, leveraging FAISS embeddings and ChromaDB to transform unstructured research text into structured data, improving data retrieval and enhancing analysis accuracy for downstream applications.
- Built a scalable pipeline integrating the Gemini API, enabling efficient classification and metadata extraction from 10,000+ research papers, improving content organization and research trend analysis by 40%.
- Implemented metadata filtering algorithms in LangChain, refining data enrichment processes to improve precision and usability of extracted insights by 20%.
- Created a data visualization framework to analyze trends across 10,000+ papers, providing structured insights into research patterns and author affiliations for non-technical professionals to anticipate potential startup opportunities.

Legislative Services Agency

Dec 2023 – Mar 2024

Business Analyst Intern

Indianapolis, IN

- Designed and managed ETL pipelines to validate and preprocess Indiana General Assembly datasets, achieving 98% data accuracy for legislative bill and its report.
- Developed and optimized SQL-based workflows, improving data extraction, transformation, and loading efficiency within SQL Server.
- Collaborated in a cross-functional team with Business Analysts and Software Developers to enhance business intelligence workflows, leveraging Tableau and Power BI for improved data reporting and analysis.
- Provided technical support to Indiana Lawmakers and troubleshooted for data integration issues, ensuring seamless access to structured legislative data sources.

PROJECTS

AI-Driven Diabetes Prediction Pipeline | Python, Pandas, NumPy, SciPy

Aug 2024 – Nov 2024

- Designed and optimized data preprocessing pipelines to clean and structure a 100,000-sample dataset, reducing processing time.
- Implemented data transformation workflows to handle class imbalance, utilizing SMOTE and SMOTE-Tomek for improved model performance.
- Developed a scalable data pipeline to streamline feature engineering, including interaction terms and transformations, improving model accuracy.

Counts of Bushels Project | SQL, SQL Server 2022, Data Warehousing

Aug 2023 – Oct 2023

- Designed a 5NF relational database in SQL Server 2022 for Indiana's corn and soybean yield, optimizing data storage for 100+ stakeholders.
- Developed efficient SQL-based ETL workflows to ingest, transform, and structure agricultural yield data for downstream analysis.
- Optimized query performance for large-scale yield comparisons and trend tracking, improving data retrieval efficiency.

EDUCATION

Purdue University

December 2024

Bachelors of Science in Computer Science, GPA: 3.3/4.0

Coursework: Database Systems, Data Science, Introduction to Data Analysis and Mining

LEADERSHIP & AWARDS

- Dean's List x5, International Jaguar Excellence Award, ALDPES Honors society
- Led orientation of 800+ students fostering academic success and community engagement as a Team Leader
- Secured Industry Sponsorship and organized hackathons for 200+ students as the Vice-President of the Computer Science club resulting in doubled club engagement.