Vandit Mehta

Email: contact.vanditmehta@gmail.com
GitHub LinkedIn
Mobile: +1 (438) 467 2203

About Me

Passionate Software Developer/Data Engineer dedicated to creating scalable systems and solving complex challenges, with a strong ability to self-learn and adapt to diverse technological environments.

Education

- **Masters of Applied Computer Science** Montreal, Canada
  Concordia Univeristy; GPA: 3.6/4.3 *Jan '23 - Dec '24*
- **Bachelors of Technology in Computer Engineering** Ahmedabad, India
  Indus Univeristy; GPA: 3.8/4 *July '18 - May '22*

Experience

- **Data Engineering Intern** Montreal, Canada
  ALDO GROUP *May '23 - Sep '23*
  - **Data Validation Framework Development:** Designed and implemented a robust **PySpark-based** data validation framework integrated with **Apache Airflow** for job scheduling and automation. Reduced processing time from 8 hours to 45 minutes (90% improvement) for 2 TB of weekly data. Enhanced QA productivity by 80% and achieved 99.9% accuracy across 200+ data pipelines.
  - **Schema Comparison and Data Integrity:** Conducted 50+ schema comparisons and verified row counts for datasets with over 10 million records. Mapped 200+ columns to identify discrepancies and generated detailed reports that ensured 100% detection and resolution of 500+ cases of missing or mismatched data.
  - **Data Quality Assurance and Optimization:** Improved QA processes by implementing 10+ automated validation checks and rules using **PySpark** and **Great Expectations**, reducing manual errors by 95%. Ensured the seamless operation of 25+ automated data pipelines, increasing system reliability by 30%.
  - **Performance Metrics and Reporting:** Collaborated with cross-functional teams using **Jira** to define 15+ KPIs for data quality and system performance. Achieved a 20% improvement in accuracy and a 25% boost in processing efficiency by continuously refining validation workflows and deploying improvements via **Git** and **Docker**.
- **Machine Learning Intern** Chennai, India
  NJS Infotech *Jan '22 - Apr '22*
  - **Self-Diagnosis Chat-Bot Design:** Developed a self-diagnosis chatbot using a Decision Tree classifier (sklearn), integrating Natural Language Processing (NLP) techniques for user query understanding. The system processed 50,000+ daily queries, providing users with potential diagnoses, severity levels, remedies, and nearby healthcare options.
  - **Impact During COVID-19:** Enabled over 5 million individuals in rural India to access critical health information during the second COVID-19 wave through a scalable, cloud-based platform. The solution handled over 10 TB of health-related data, providing real-time assistance and updates on COVID-19 symptoms, prevention, and nearby healthcare facilities.
  - **Technology Stack and Scalability:** Leveraged technologies such as Python, Flask, Apache Kafka for real-time data streaming, and deployed the chatbot on AWS, ensuring 99.9% uptime. The solution utilized a microservices architecture, enabling seamless scalability to handle increased traffic during health crises, while maintaining low-latency response times.

Projects

- **DataFusion: Unified Data Flow Engine:** [Tech Stack:- PySpark, ADLS Gen2, Databricks, ADF]
  - **Scalable Data Pipeline Engineering**: Engineered a scalable data pipeline using Azure Services to ingest, clean, and transform large datasets. This enabled seamless data flow across different layers (Bronze, Silver, and Gold) for optimized querying and reporting, ensuring high performance and data integrity.
  - **Data Aggregation and BI Optimization**: Optimized data aggregation and queries in Azure Synapse, creating views and stored procedures for efficient analysis. Integrated with Power BI to provide real-time business intelligence dashboards, delivering actionable insights to stakeholders for better decision-making. **Project Link:** Click Here
- **Pneumonia Detection using Chest XRays** [Tech Stack :- Pytorch, OpenCV, Scikit-Learn, TorchVision]
  - **CNN Models:** Played Implemented and fine-tuned CNN models (ResNet18, DenseNet121, InceptionV3) for pneumonia detection in chest X-ray images, achieving an accuracy of 12%, with transfer learning improving model performance on a dataset of 100,000+ images.
  - **Model Optimization and Hyperparameter Tuning:** Optimized model training using Stochastic Gradient Descent and data augmentation, fine-tuning hyperparameters. InceptionV3 achieved a performance improvement of 18%, with a 95% accuracy rate and strong generalization across 5 different datasets. **Project Link:** Click Here

Technical Skills

- **Core Technologies:** Python, Java, C, C++, PHP, SQL, HTML, CSS, JavaScript.
- **Databases:** MongoDB, PostgreSQL, MySQL, AWS (DynamoDB, RedShift, Athena), Azure (Cosmos DB, Synapse)
- **Frameworks/Libraries:** PySpark, Databricks, Hadoop, Kafka, Airflow, DBT, Flask, Django, Scikit-learn, PyTorch
- **DevOps & Tools:** Git, Docker, Jenkins, Ansible, Maven, Kubernetes, Terraform

Additional Activities

Teaching Assistant: Conducted engaging tutorials that enhanced student comprehension and participation, while assisting with assessments and providing constructive feedback to support academic growth.