AARUSHI APTE

Champaign, IL | (447)-902-1981 | aapte3@illinois.edu | LinkedIn

Dedicated graduate student pursuing a Master's degree in Information Management. Adept at leveraging cutting-edge technologies to organize, analyze, and derive valuable insights from data. Passionate about the intersection of technology and business, with a strong foundation in data management, information systems, and strategic decision-making. Committed to continuous learning and eager to apply academic knowledge to real-world challenges.

EDUCATION

University of Illinois at Urbana-Champaign, IL — Expected May 2024
Master of Science in Information Management — GPA: 3.83
Courses: Programming & Quality in Data Analytics, Statistical Model, Database Scaling, Business Intelligence

NMIMS University, Mumbai, India — May 2021
Bachelor of Technology in Information Technology — GPA: 3.11
Courses: Data Warehousing & Mining, Data Structures and Algorithms, Advanced Database Management Systems

TECHNICAL SKILLS

- **Languages:** Python, R, Javascript, Typescript
- **Libraries & Cloud Computing:** NumPy, Pandas, Matplotlib, NLTK, scikit-learn, PySpark, Amazon Web Services, S3, Lambda, Azure
- **Databases & Softwares:** Apache Spark, MySQL, Dynamodb, Tableau, Power BI, MS Office Suite, SAP, SharePoint, Hadoop
- **IDE & Version Control**: Visual Studio, PyCharm, Jupyter Notebook, Postman, Git, GitHub, Android Studio

PROFESSIONAL EXPERIENCE

Kohler Co. — Champaign, Illinois
Data Science Intern — May 2023 - Present

- Enabled **predictive** & **prescriptive** analytics by integrating the smart factory ecosystem with enterprise **Azure** platform
- Built an **automated** pipeline to ingest data from on premise databases into **Azure Data Lake** using **Azure Data Factory**
- Created an **ETL pipeline** to transform, normalise & store the data into the **Azure SQL data warehouse** using **Spark** & **Databricks**
- **Automated** a **3** year old manual process to generate reports and create presentations using **Numpy, Pandas** and **python-pptx**
- Developed a nightly **batch process** designed to handle data deposited into **Azure Blob Storage** & generate PowerBI **dashboards**

Quantiphi Analytics Solution — Mumbai, India
Framework Engineer — November 2021 - July 2022

- Reduced turnaround time by **30%** by assisting in development of an internal helpdesk **chatbot** utilizing **Amazon Web Services**
- Boosted customer satisfaction to **43%** by **automating** resolution of account related issues through **Angular, AWS Lambda**, and **DynamoDB**, resulting in reduced wait and response times
- Built a pipeline to transform raw conversation logs into structured data using **PySpark**, **Pandas** & **AWS Glue**
- Developed schemas using **AWS Glue crawler** and **SQL queries** to expose the data for the **PowerBI** dashboard, in **AWS Athena**

Framework Engineer Intern — July 2021 - November 2021

- Conducted **trend analysis** using **NumPy**, **Pandas** & **Matplotlib** on power consumption patterns to identify maintenance windows that would minimize downtime
- Led a team of 4 developers to build a **web application** called “SplitEx” to allow users to keep track of their expenses and split bills using **MongoDb**, **React**, **Node.js** & **Figma**

MSPRO Technologies — Thane, India
Software Engineer Intern — May 2019 - June 2019

- Achieved **24%** efficiency boost by designing a **database** for a rental agency in order to **automate** manual processes
- Used **UML diagrams** and **MySQL** to define the database **schema** and execute queries for ad-hoc reporting

PROJECTS

Monte Carlo Simulation to Derive Landing Distance for a Flight | Python, Matplotlib, Pandas, geographiclib, CSV

- Randomised variables to **predict** the probability of potential outcome and the landing distance required by the flight
- Use **Pandas** to derive and manipulate **dataframes** to get all accommodating airports, in the flight’s path for **hypothesis testing**
- Rejected the hypothesis of less than 10% variation in distances to airports, but confirmed a 15% reduction in viable landing airports under dynamic conditions

Sentiment Analysis on Amazon Reviews | Python, Pandas, Numpy, Matplotlib, Seaborn, NLTK

- Compared the **sentiment** of the review versus the rating of an Amazon product using the **VADER** and **Roberta** pretrained model
- Used **NLTK** to tokenize the input data and categorize the tokens under their appropriate tags
- Concluded Roberta outperformed VADER using **polarity scores** on neutral and negative reviews

University Admit Prediction | Python, Pandas, Numpy, Matplotlib, Seaborn, scikit-learn

- **Predicted** probability of getting an admit/reject from a university using **sklearn** and **feature scaling**
- Compared regression/classification ML algorithms, achieved R2 score of **0.81** for **Linear Regression** and accuracy of **95%** for **Gradient Boosting Classifier**