DHAIRYA JAYESH CHHEDA

📞 930-333-5591 ✉️ dhairyajayeshchheda@gmail.com in LinkedIn GitHub 🌐 Portfolio

Education

Indiana University Bloomington   **August 2023 – May 2025**
Master of Science in Data Science — GPA: 3.85/4   *Bloomington, IN*

Experience

Department of History, Indiana University Bloomington   **May 2024 – Present**
Data Research Assistant   *Bloomington, IN*
- Engineered a dynamic web scraping solution using Selenium to interact with webpage elements, extracting historical data from 5+ museums and integrating it into a relational database, reducing manual data collection efforts by 80%.
- Orchestrated a research initiative leveraging 1M+ data points to create Tableau dashboards, enabling data-driven reporting and presenting key performance indicators (KPIs) to stakeholders, increasing funding opportunities by $200,000.
- Designed and optimized a predictive machine learning (ML) model to forecast the locations of ancient artifacts, reducing false positives by 16%, improving artifact recovery rates by 30% and potentially saving $100,000 in research costs.

LTIMindtree   **June 2021 – July 2023**
Software Engineer   *Mumbai, MH*
- Resolved 10+ paginated report service tickets within JIRA, and conducted comprehensive testing of various SQL stored procedures and functions, ensuring efficient operation of reports across applications and reducing downtime by 40%.
- Automated ETL processes through the development of complex SQL scripts, enhancing data pipeline efficiency by 50%.
- Spearheaded a team in achieving 75% acceleration in project completion by optimizing feature engineering and following SDLC and Agile methodologies, minimizing project duration from 40 days to 10 days.
- Refactored ML Python notebooks into modular code, performed System Integration Testing (SIT) and User Acceptance Testing (UAT), and deployed applications via CI/CD pipelines for seamless end-to-end migration.

Hawkeye MedTech   **September 2019 – June 2020**
Artificial Intelligence Research Intern   *Columbia, MD*
- Collaborated with 3+ cross-functional teams to gather requirements, define project scopes, and ensure alignment with business objectives, fostering effective teamwork and project success.
- Conducted in-depth market research and analysis on various open-source AI expert systems, resulting in the identification of rule-based engines that guided strategic processes for patient triage based on urgency.
- Optimized the predictive accuracy of the patient recommendation system by 20% through ensemble learning and hyperparameter tuning, improving patient outcomes by 25% and increasing the company's service adoption rate by 35%, contributing to an estimated revenue growth of $150,000 annually.

Projects

Customer Churn Prediction | *Python, Softmax Regression, SVM, Random Forest, Ensemble Learning*   **April 2024**
- Trained different classifiers, performed hyperparameter tuning and achieved an accuracy of 78.16%.
- Enhanced model performance by applying ensemble learning techniques such as the voting, bagging and stacking methods, boosting accuracy to 86.73%, surpassing individual classifier results.

E-Commerce Sales Analysis | *Python, EDA, SQL, PowerBI*   **December 2023**
- Ensured data integrity by managing missing values and encoding categorical variables, improving data quality by 33%.
- Analyzed sales data using PowerBI and conducted time series analysis, identifying key growth factors and forecasting sales for the next 15 days, contributing to a 30% increase in annual sales growth.

Modeling Human Tendencies for Password Guessing | *LSTM, GRU, GAN*   **May 2021**
- Constructed LSTM and GRU models with three stacked recurrent layers followed by two densely connected layers totaling 4.8M+ parameters, generating human-like passwords that matched ~55% of the 14M+ passwords within 10^9 guesses.
- Improved model accuracy by integrating a GAN model, yielding a gain of 12% in accuracy and refining the guessing performance of the models.

Technical Skills

Languages: Python, R, SQL, HTML, CSS
Frameworks: Pandas, Numpy, Scikit-Learn, Matplotlib, Plotly, TensorFlow, Keras, NLTK, Streamlit, Selenium
Statistics: Inferential Statistics, Hypothesis Testing (A/B Testing), Regression Analysis, ANOVA, Time Series Forecasting
Data Science: Exploratory Data Analytics, Data Visualization, Feature Engineering, Supervised & Unsupervised Learning Techniques, Data Mining, Machine Learning Algorithms, Deep Learning (Transfer Learning, Fine-tuning), Natural Language Processing (Tokenization, Stemming, Word2Vec), Model Performance Evaluation, Predictive Modeling, Generative AI, LLMs
Database: Microsoft SQL, MySQL, PostgreSQL
Cloud Computing: GCP BigQuery, GCP Looker, AWS EC2, AWS S3, AWS SageMaker, AWS Redshift, AWS Athena
Tools/Platforms: VS Code, Jupyter Notebooks, PowerBI, Tableau, Alteryx, Qlik Sense, Jenkins, GitHub, JIRA, Excel