Lin Dai

7dailingy@gmail.com | Website

EDUCATION

George Mason University **Aug 2023 - May 2025 (Expected)**
Master of Science in Data Analytics Engineering, GPA: 3.72/4.0 *Fairfax, VA*
Relevant Courses: Big Data Analytics, Application of Data Mining, Visualization for Analytics, Artificial Intelligence, Database Systems, Operations Research

Michigan State University **Aug 2016 - May 2020**
Bachelor of Science in Computer Science *East Lansing, MI*

WORK EXPERIENCE

Walmart Global Tech **Dec 2021 - May 2023**
Software Engineer II *Reston, VA*

- Designed and developed supplier interface features including onboarding, program management, and supplier program details/categories via JavaScript, React, and GraphQL, speeding up direct-to-store deliveries by 80%.
- Redesigned the CI/CD pipeline by automating manual build validations and merging deployment stages, reducing deployment steps by 50% and increasing release frequency.
- Resolved 20+ user-reported defects, including critical UI bugs, performance bottlenecks, and data synchronization issues, achieving 80% code coverage.
- Created and maintained 20+ release notes with features and enhancements across multiple software versions to ensure development and deployment transparency with developers and stakeholders.
- Guided interns in conceptualizing, designing, and implementing various application features to meet their internship objectives.
- Developed Splunk queries and dashboards that optimized system monitoring, reducing mean time to resolution (MTTR) for production issues by 40%, from an average of 30 minutes to 18 minutes.
- Led business requirement discovery and data analysis sessions with stakeholders to design application APIs, and coordinated with 6 teams to onboard 8 clients, enhancing application service and adoption.
- Engaged in on-call rotations and provided timely responses to resolve 3+ critical issues and incidents.

PROJECTS

Image Clustering

- Developed K-Means clustering algorithm to determine clusters for 10,740 images of handwritten digits. Applied t-Distributed Stochastic Neighbor Embedding to perform image matrix dimension reduction while preserving local structure. Developed K-Means clustering algorithm class with functions such as assigning data points to clusters with Euclidean distances, computing within-cluster distances, computing cluster Sum of Squared Errors to evaluate cluster compactness, computing new cluster centroids, and computing Silhouette Coefficient to evaluate cluster quality. Implemented function to generate total Sum of Squared Errors vs. iteration plot to evaluate optimal number of clusters. Utilized hyperparameter tuning to experiment on the training model and determine the optimal model configuration.
- Language/Tool: Python, Scikit-learn, Pandas, NumPy, Matplotlib, Jupyter Notebook, Anaconda

Yelp Reviews Natural Language Processing

- Developed a natural language processing algorithm with k-nearest neighbors classifier to predict sentiment for 18,000 reviews. Performed data exploration and preprocessing by generating graphics for descriptive statistics and label distribution, and identifying missing data. Developed stemming text preprocessing to tokenize sentences, remove punctuation, remove numbers, remove word contraction, remove stop words, and stem the collection with Porter Stemmer. Applied Term Frequency-Inverse Document Frequency vectorizer to capture word importance in numerical format. Applied chi-squared test to perform feature importance analysis to eliminate noisy features and improve model efficiency. Applied Singular Value Decomposition to perform dimensionality reduction on sparse matrices to capture important information and improve model performance. Developed k-nearest neighbors algorithm to perform sentiment classification with Euclidean distance. Applied k-fold cross-validation with classification accuracy to evaluate model performance. Utilized hyperparameter tuning to experiment on the training model and determine the optimal model configuration.
- Language/Tool: Python, Scikit-learn, Pandas, NumPy, Matplotlib, Jupyter Notebook, Anaconda

SKILLS

Languages: JavaScript, HTML/CSS, Python, Java, TypeScript, SQL
Frameworks/Libraries: Scikit-learn, Pandas, NumPy, Matplotlib, React, Redux
Tools/Technologies: Git, Jupyter Notebook, Anaconda, Kubernetes, Docker, MongoDB, AWS (EC2, S3, Glue DataBrew), Jira, Confluence, Figma
Others: Agile, SDLC, CI/CD, TDD, BDD, API Design