

TECHNICAL SKILLS

- **LANGUAGES:** Python, R, SQL
- **TOOLS:** Jupyter, R Studio, PyCharm, Tableau, MySQL, PostgreSQL, JIRA, MS Excel
- **SKILLS:** Data Structures & Algorithms, Machine Learning, Pivot Tables, Data Modeling, Statistical Analysis, Data Mining, Data Cleaning, Data Wrangling, Data Visualization, Queries, AWS
- **LIBRARIES:** Pandas, Numpy, Scikit-learn, Spacy, Beautiful Soup, Selenium, Tidyverse, Tidymodels, Keras

EDUCATION

Pace University New York, NY
Seidenberg School of Computer Science and Information Systems January 2021- May 2023

Master of Science in Information Systems

Data Science Major

4.0 GPA

Relevant coursework: Intro to Coding(Python), Data Analysis & Visualization (Tableau, Python, R), Big Data Analytics, Web Mining (Python), DBMS(SQL)

Awarded Information Systems Departmental Award

Karnataka State Open University Mysore, India
Bachelor of Commerce June 2012- May 2015

Business Administration Major

First Class

Relevant coursework: Business Statistics

EXPERIENCE

Computational Intelligence Lab, Pace University New York, NY
Lab Assistant January 2023 – May 2023

Tools and Technologies: Python, R, SQL, MySQL, Jupyter Notebooks, ML, Analytics, GitHub, MS Excel

- Collaborated with the faculty and another lab assistant to identify and work on a data science project, to build a model to predict the class of the patients in a pelvic study, classifying them into 'normal' and 'abnormal' based on factors such as pelvic tilt, incidence etc.
- Performed data cleaning and exploratory data analysis to assess the data and identify patterns.
- Visualized data in both Python and R to showcase trends in an effective manner.
- Responsible for feature selection and application of K-Nearest Neighbors algorithm with a 77% accuracy.
- Deployed Random Forest Classifier algorithm to increase accuracy of prediction by 11%.
- Incorporated libraries in the projects, including Pandas, Scikit-learn, TensorFlow libraries to analyze data.
- Authored and curated content on a data science project for the lab portal, resulting in 10% increase in user engagement and knowledge dissemination.
- Demonstrated adept repository management on GitHub, including updating course materials, assignments, and other resources, facilitating 7% reduction in material update time, and enhancing accessibility to course materials.
- Shared insights and knowledge with peers, fostering a culture of continuous learning and staying up to date with the latest data science trends, leading to 12% increase in participation in lab activities.

Pai, Nayak & Associates Udupi, Karnataka, India
Data Analyst Aug 2015- Aug 2019

Tools and Technologies: SQL, MySQL, MS Excel, Data Modeling, Queries

- Translated business needs into data analysis, business intelligence data sources, and reporting solutions for different types of clients.

- Proficient in SQL for data extraction, manipulation, and querying from relational databases, ensuring efficient data processing and analysis.
- Utilized Python and R for advanced statistical analysis and predictive modeling of data-related tasks.
- Implemented robust data validation procedures, identifying and resolving inconsistencies in datasets to ensure reliable analysis outcomes.
- Collaborated with cross-functional teams to design and implement dynamic visualizations to effectively communicate valuable insights to technical and non-technical stakeholders.
- Collaborated with business units to understand requirements and provide customized data solutions, improving overall business efficiency.

ACADEMIC PROJECTS

Librivox Audiobook Views Prediction

Project GitHub: <https://github.com/dPai12/LibriVox-Audiobook-Views-Prediction>

- Teamed with 2 peers to scrape and clean data from LibriVox and implemented WhisperAI to convert audio data to text, enabling comprehensive analysis.
- Applied Natural Language Processing (NLP) for Named Entity Recognition(NER) to preprocess and analyze text data.
- Developed a Linear Regression model to predict views on Audiobooks over a period of 5 years with 63% accuracy.
- Implemented libraries including Pandas, Scikit-Learn, Beautiful Soup, nltk, Spacy, Whisper to perform comprehensive statistical analysis of data.
- Visualized key insights using Tableau, enhancing the understanding of trends and patterns in audiobook consumption.

Commonlit Readability

Project GitHub: <https://github.com/dPai12/CommonLit-Readability>

- Trained Machine Learning models such as Linear Regression and SVM, for predicting the readability of text passages across grades 3-12 in R.
- Employed comprehensive Natural Language Processing (NLP) techniques for robust data preprocessing and feature extraction.
- Achieved 48% accuracy in predicting readability scores and identified optimal models for different grade levels.
- Incorporated a range of libraries including tidyverse, tidymodels, dplyr, parsnip, keras, ggplot2, textrecipes.

Major Safety Events

Project GitHub: https://github.com/dPai12/Major_Safety_Events

- As part of a collaborative team of 4, conducted in-depth Exploratory Data Analysis (EDA) on Major Safety Events under the Federal Transit Authority (FTA) spanning from 2014 to 2021 using R.
- Leveraged powerful Data Visualization techniques and descriptive statistics in R to communicate trends and the factors contributing to these events as well as their resulting impacts.
- Deployed a suite of libraries, including tidyverse, dplyr, and ggplot2, to ensure comprehensive data analysis and insightful visual representation in R.

SIDE PROJECTS

Book Digest (In Progress)

Summary: This app will be used to sell, and exchange used books online.

- Implementing a robust recommendation system to provide personalized book suggestions to users based on their preferences, purchase history, and browsing behavior.
- Leveraging advanced algorithms and data analytics techniques to optimize the accuracy of recommendations.
- Collaborating closely with the lead Android engineer to seamlessly integrate the recommendation system into the Book Digest mobile application.
- GitHub link: <https://github.com/shashwathkamath/BookDigest>
- Tech Stack: Jetpack Compose, Kotlin, Firebase (for backend), TensorFlow, Neo4j

CERTIFICATIONS

AWS Certified Cloud Practitioner

Amazon Web Services (AWS) - 05/2024