

Mina Mehdinia

Oregon, US (Willing to relocate) | ☎ 503-308-0778 | ✉ Email | 🔗 LinkedIn | 📝 Medium 🗣 Blog

SUMMARY

Formally trained data scientist from an intensive four-year university program, with over 180 credits in core areas of data science including computer science, statistics, math, and machine learning. Graduated with honors. Proficient in developing, training, and evaluating machine learning and deep learning models. Adaptable and versatile, I'm poised to excel in both collaborative and independent settings. Seeking a challenging position in a dynamic organization to address real-world problems and drive business growth.

EDUCATION

Portland State University

BS in Data Science, Summa Cum Laude, GPA: 3.92/4

Portland, OR

2019 - June 2023

TECHNICAL SKILLS

- **Programming Languages:** Python, R, SQL, Bash, C++
- **Libraries:** Pandas, numpy, scikit-learn, PyTorch, Transformers, NLTK, openai, matplotlib, seaborn, plotly, ggplot2, dplyr, tidyR
- **Developer Tools:** Git, Jupyter, Colab, PyCharm, RStudio, VS Code, DataGrip, AWS SageMaker
- **ML Techniques:** Deep Learning, Natural Language Processing, Large language models, Generative AI, Convolutional Neural Networks, Linear regression, Support Vector Machine, Random Forest, Decision Tree, K-means Clustering
- **Data Science Techniques and Statistical Methods** Predictive modeling, Time Series Analysis, Hypothesis Testing, Regression Analysis, Descriptive Statistics, Inferential Statistics, Dimensionality Reduction, Data Cleaning and Preprocessing, Data Visualization, Correlation Analysis

WORK EXPERIENCE

Freelance Data Scientist / Large language Model

Summer 2023

- **Fine-tuned BERT:** Fine-tuned BERT using Hugging Face's transformers library for predicting product ratings on Amazon from customer reviews.
- **Prompt Engineering with GPT:** Employed OpenAI's GPT-3.5 to analyze and score product reviews, while smoothly embedding it into a Python environment. Experimented with zero shot and few shot prompting techniques.
- **Naive Bayes Classifier:** Developed a prediction model using the Naive Bayes classifier to estimate Amazon product ratings based on customer review content.

Machine Learning Research Assistant

Jan 2022 - June 2023

Portland State University

- **Object detection:** Designed an end-to-end deep learning system using PyTorch to detect and compute the total value of US coins in diverse lighting and backgrounds from smartphone images. Personally collected data via smartphone and annotated data using the LandingLens platform.
- **K-means Clustering & KNN:** Designed and developed K-means clustering and KNN regression models for food categorization and recommendation using Python and Scikit-learn. Built a K-means clustering model from scratch, compared its performance with Scikit-learn's version, and processed datasets of over 8,000 food items.

Data Science Intern

Summer 2022

Summer Institute in Bio-statistics and Data Science

UC Irvine, CA

- **Bio-Statistical Modeling:** Utilized modern bio-statistics methodologies to statistically model stress based on biophysical, contextual, and demographic features. Conducted comprehensive data wrangling, exploratory data analysis, and data cleaning using R and Python. Prioritized model interpretability and addressed collinearity between predictors through linear modeling techniques.

Data Analyst Research Assistant

Sep 2021 - June 2023

Portland State University

- **Dimensionality Reduction:** Executed dimensionality reduction techniques including principal component analysis and recursive feature elimination. Engineered predictive models employing linear and quadratic discriminant analysis, generalized linear models, and KNN to predict loan default prediction.
- **Exploratory Data Analysis:** Implemented extensive exploratory data analysis using Pandas and various plotting libraries in Python to provide some insights to our client about the effect of COVID on the healthcare of transgender women.
- **Time series:** Analyzed Klamath River at Keno and Miller data to assess the impact of weather and river flow on thermal stratification. Conducted statistical analysis and hypothesis testing to identify significant factors affecting thermal stratification. Implemented generalized additive models, random forest and kmeans clustering for analysis and recommended the predictive models and features to the client.
- **Statistical analysis:** Implemented web scraping in R to collect the data from the Cherry Blossom 10-mile running race from 1973-2022. Performed data cleaning and wrangling in R. Modeled the data using a Linear Mixed Effects (LME) model and found statistical relationship between age and physical fitness with respect to how fast people run.

RELEVANT COURSEWORK

Algorithms, Machine learning, Artificial intelligence, Intro to database management system, Data visualization in R, Data structures, Modern regression analysis, Statistical learning, Large-scale data algorithms, Data science consulting skills, Data science practicum, Computer science I & II, Intro to Unix, Calculus I - IV, Statistics I & II, Applied Linear Algebra I & II, Discrete Structures I & II.

ADDITIONAL INFORMATION

US Citizen