

Mina Mehdinia

minamehdinia213@gmail.com • (503) 308-0778
[LinkedIn URL](#) [Medium](#) [Github.io](#)
Chicago, Illinois (willing to relocate)

Data Scientist

Graduated Summa Cum Laude in Data Science with significant hands-on coding and project management experience from internships across various industries. Applied machine learning and deep learning models to real-world problems, enhancing decision-making processes and marketing strategies. Skilled in Python and R, with a proven track record in developing and optimizing algorithms for predictive analytics, customer segmentation, and revenue growth. Experienced in leading data-driven projects from conceptualization through implementation, both collaboratively and independently. Driven to leverage analytical skills and coding expertise to contribute to innovative solutions within a dynamic organization.

Education

Bachelor of Science in Data Science, (GPA 3.92) – Summa Cum Laude
Portland State University, Portland, OR

Technical Proficiencies

Languages:	Python, R, SQL
Tools:	Git, Jupyter, Hugging Face Hub, Colab, PyCharm, RStudio, AWS SageMaker
Libraries:	Pandas, numpy, scikit-learn, PyTorch, Transformers, openai, matplotlib, seaborn, plotly, ggplot2, dplyr, tidyr,
AI and ML Techniques:	Deep Learning, Natural Language Processing, Large language models, Generative AI, Convolutional Neural Networks, Support Vector Machine, Random Forest, Decision Tree, K-means Clustering
Statistical Techniques:	Predictive modeling, Time Series Analysis, Hypothesis Testing, Regression Analysis, Dimensional Reduction, Exploratory data analysis, Marketing Mix Modeling (MMM), Multi-Touch Attribution (MTA)

Career Experience

ChannelAttribution, Remote

Apr 2024 – Present

Data Science Intern

- Leveraged advanced analytics techniques to enhance digital marketing strategies, focusing on Multi-Touch Attribution (MTA), Marketing Mix Modeling (MMM), and A/B testing.
- Creating insightful data-driven tutorials for company products and presenting complex analytical findings to stakeholders.

- Enhanced and automated product rating predictions on Amazon by fine-tuning a 110M params BERT model using Hugging Face's transformers library, achieving 52% improvement over ChatGPT. Deployed the model to the Hugging Face Model Hub for broader accessibility ([Try it](#)).
- Analyzed embeddings from a fine-tuned BERT model using t-SNE and Plotly, identifying sentiment trends and anomaly detection in Amazon product reviews, which informed targeted model refinement strategies.
- Leveraged OpenAI's GPT-3.5 to elevate product review scoring and analysis, achieving a very low error (MAE of 0.56) in sentiment accuracy and led to innovations in zero-shot and few-shot prompting techniques.
- Spearheaded the development of a Naive Bayes classifier as a baseline model, achieving 67% prediction accuracy for Amazon product ratings based on customer reviews.

- Led a statistical research team which collaborated with UCI Medical Center researchers in performing statistical inference on a 250-person clinical trial involving ecological momentary assessment data on pregnant mothers and child health outcomes.
- Spearheaded bio-statistical modeling techniques such as GLM, odds, and Logistic Regression on biomedical and demographic data. Achieving a 68% correlation between stress and biophysical, contextual, and demographic features. Improved model interpretability by 40% through advanced linear modeling.

- Conceived and executed a deep learning system using PyTorch that accurately computed the total value of US coins under varied lighting and backgrounds achieving 97.5% precision.
- Led the development of K-means clustering and KNN regression models to recommend food items, achieving a 93.3% accuracy in food categorization recommendations
- Streamlined predictive modeling using dimensionality reduction, achieving 81.4% accuracy in loan default prediction by employing advanced algorithms.
- Designed and executed a web scraping strategy in R, capturing data from the Cherry Blossom 10-mile race (1973-2022). Streamlined data preparation in R and employed a Linear Mixed Effects (LME) model, uncovering a notable link between age and running speed influenced by physical fitness.
- Analyzed Klamath River's thermal stratification by examining weather and river flow data. Employed generalized additive models, random forest models, and K-means clustering to uncover patterns and establish a stratification threshold, guiding water resource management strategies.
- Examined the impact of the COVID19 pandemic on transgender women's healthcare experiences, with a focus on their financial health and gender affirming care using statistical analysis and methods such as Network Analysis and Graph Theory.