SARAH CHEN, PhD

san.francisco.ca@gmail.com | (415) 555-9876 | linkedin.com/in/sarahchen | San Francisco, CA

PROFESSIONAL SUMMARY

Data scientist with 5+ years of experience developing machine learning solutions in healthcare and finance. PhD in Computer Science with specialization in machine learning. Strong track record of translating business problems into analytical models with measurable business impact.

SKILLS

- **Programming Languages**: Python (Advanced), R (Intermediate), SQL (Advanced), Java (Basic)
- **Data Science Libraries**: NumPy, Pandas, scikit-learn, TensorFlow, PyTorch, Keras
- **Cloud & Big Data**: AWS (SageMaker, S3, EC2), GCP, Spark, Hadoop
- **Data Visualization**: Tableau, Matplotlib, Seaborn, Plotly
- **Databases**: PostgreSQL, MongoDB, Snowflake
- **Tools**: Git, Docker, Kubernetes, Airflow

PROFESSIONAL EXPERIENCE

SENIOR DATA SCIENTIST

HealthMetrics Inc., San Francisco, CA

April 2021 - Present

- Lead a team of 3 data scientists developing predictive models for patient readmission risk, reducing 30-day readmissions by 22%
- Designed and implemented an NLP pipeline for processing clinical notes, improving diagnosis coding efficiency by 35%
- Collaborated with product teams to integrate ML models into the company's healthcare analytics platform
- Developed a real-time anomaly detection system for monitoring patient vitals using TensorFlow and AWS SageMaker
- Presented findings and insights to C-level executives and healthcare providers

DATA SCIENTIST

FinancialEdge Analytics, San Francisco, CA

June 2018 - March 2021

- Built credit risk assessment models using gradient boosting algorithms, improving default prediction accuracy by 18%
- Created time series forecasting models for market trend analysis with Python and PyTorch
- Engineered features from unstructured financial data using NLP techniques
- Developed an automated ML pipeline for model training, validation, and deployment using Airflow and Docker
- Collaborated with engineering team to implement models in production systems

DATA SCIENCE INTERN

Novartis Pharmaceuticals, Cambridge, MA

May 2017 - August 2017

- Analyzed clinical trial data to identify patterns in patient responses to experimental treatments
- Assisted in developing statistical models to optimize clinical trial design
- Created interactive data visualizations for research team presentations

EDUCATION

PhD in COMPUTER SCIENCE

Stanford University, Stanford, CA

2014 - 2018

- Dissertation: "Deep Learning Approaches for Medical Image Segmentation and Classification"
- Published 5 papers in top-tier machine learning and healthcare informatics conferences
- Teaching Assistant for Machine Learning and Data Mining courses

MS in STATISTICS

University of California, Berkeley, Berkeley, CA

2012 - 2014

- GPA: 3.92/4.0
- Research Focus: Statistical Learning Theory

BS in MATHEMATICS

Massachusetts Institute of Technology, Cambridge, MA

2008 - 2012

- GPA: 3.85/4.0
- Minor: Computer Science

PUBLICATIONS & PATENTS

- Chen, S., et al. (2023). "Improving Medical Diagnosis with Attention-Based Deep Learning Models." *Conference on Neural Information Processing Systems (NeurIPS)*.
- Chen, S., et al. (2022). "Robust Time Series Forecasting for Financial Market Prediction." *International Conference on Machine Learning (ICML)*.
- Patent: "Method and System for Predicting Patient Outcomes Using Multi-modal Data Integration" (US Patent App. 16/324,892)

CERTIFICATIONS

- AWS Certified Machine Learning - Specialty
- Google Professional Data Engineer
- Deeplearning.ai TensorFlow Developer