SARAH JOHNSON

Data Scientist & Machine Learning Engineer

Email: sarah.johnson@email.com

Phone: +1-555-0456

Location: New York, NY

LinkedIn: linkedin.com/in/sarahjohnson

GitHub: github.com/sarahjohnson

PROFESSIONAL SUMMARY

Data Scientist with 4 years of experience in machine learning, deep learning, and data analytics. Expert in Python, R, and advanced statistical modeling. Passionate about solving complex business problems using AI and data-driven insights.

TECHNICAL SKILLS

Programming: Python, R, SQL, Scala, MATLAB

Machine Learning: TensorFlow, PyTorch, Keras, Scikit-learn, XGBoost

Data Analysis: Pandas, NumPy, SciPy, Statsmodels

Visualization: Matplotlib, Seaborn, Plotly, Tableau, Power BI

Big Data: Apache Spark, Hadoop, Hive, PySpark

Cloud Platforms: Azure ML, Google Cloud Platform (GCP), AWS SageMaker

Databases: SQL Server, PostgreSQL, MySQL, MongoDB

MLOps: MLflow, Kubeflow, Docker, Git

PROFESSIONAL EXPERIENCE

Senior Data Scientist | DataTech Solutions | Mar 2022 - Present

• Lead data scientist for predictive modeling projects serving Fortune 500 clients

• Developed deep learning models using PyTorch for image classification with 92% accuracy

• Built recommendation systems using collaborative filtering and neural networks

• Implemented automated ML pipelines on Azure ML reducing model deployment time by 60%

• Conducted A/B testing and statistical analysis to validate model improvements

• Mentored team of 3 junior data scientists on best practices and model development

Data Scientist | Analytics Pro Inc. | Jul 2020 - Feb 2022

• Created machine learning models for customer churn prediction using Random Forest and XGBoost

• Performed time series forecasting for sales prediction using ARIMA and LSTM networks

• Developed natural language processing (NLP) solutions for sentiment analysis

• Designed interactive dashboards in Tableau for executive-level reporting

- Worked with big data processing using Apache Spark and PySpark
- Collaborated with engineering team to deploy models into production

Junior Data Analyst | Research Labs | Jan 2019 - Jun 2020
- Conducted exploratory data analysis on large datasets using Python and R
- Created data visualizations and reports for stakeholder presentations
- Performed statistical hypothesis testing and regression analysis
- Cleaned and preprocessed data from multiple sources
- Assisted in building predictive models for business forecasting

EDUCATION

Master of Science in Data Science

Columbia University, New York | 2017 - 2019

Focus: Machine Learning, Statistical Learning, Deep Learning

Thesis: "Deep Learning Approaches for Time Series Forecasting"

Bachelor of Science in Statistics

University of Michigan | 2013 - 2017

Minor: Computer Science

GPA: 3.8/4.0

CERTIFICATIONS

- TensorFlow Developer Certificate (Google, 2021)
- Azure Data Scientist Associate (Microsoft, 2022)
- Professional Data Scientist Certification (IBM, 2020)
- Deep Learning Specialization (Coursera, 2019)

RESEARCH & PUBLICATIONS

- "Advanced Neural Networks for Predictive Analytics" - Published in Journal of Machine Learning Research, 2021
- "Optimization Techniques in Deep Learning" - Presented at AI Summit 2022
- Co-author of "Big Data Analytics with Python" - Technical Blog Series

PROJECTS

Fraud Detection System
- Built real-time fraud detection system using anomaly detection algorithms
- Processed millions of transactions using Apache Spark
- Achieved 95% precision in fraud identification reducing false positives by 30%

Medical Image Analysis

• Developed CNN-based model for disease diagnosis from medical images

• Used transfer learning with ResNet and VGG architectures

• Collaborated with healthcare professionals to validate model accuracy

Customer Segmentation Platform

• Created unsupervised learning solution for customer clustering

• Used K-means, DBSCAN, and hierarchical clustering techniques

• Delivered insights leading to 20% increase in marketing campaign effectiveness

AWARDS & RECOGNITION

• "Outstanding Data Scientist Award" - DataTech Solutions (2023)

• "Best Research Paper" - ML Conference (2021)

• Dean's List - Columbia University (2018, 2019)

LANGUAGES

English (Native)

French (Professional Working Proficiency)

Mandarin (Basic)

PROFESSIONAL MEMBERSHIPS

• Member, Association for Computing Machinery (ACM)

• Member, Women in Data Science (WiDS)

• Active contributor to Kaggle competitions