

Sean Cooke

(516) 413-1876

SeanCooke8892@gmail.com

[LinkedIn](#) | [GitHub](#) | [Personal Website](#)

Skills

- Programming Languages: Python (Keras, TensorFlow, Pandas, pytest, Poetry), Java, JavaScript
- Databases: Snowflake, Oracle SQL, MySQL
- DevOps and Cloud Technologies: Kubernetes, AWS, Docker, AWS, Git, Linux, Airflow, Tableau
- Project Management: Agile Scrum, Kanban, Jira, Confluence
- Machine Learning, Data Pipelines, Cloud Computing, Test Driven Development, Data Visualization

Experience

athenahealth *Boston, MA*

Machine Learning Engineer

June 2022-Present

- Created a neural network using Keras to categorize medical documents into 27 categories. Deployed model to production and achieves 93% accuracy on 15M documents per month
- Extracted unstructured patient attributes from medical documents using GenAI. Prompt achieves 90% accuracy on 15M documents per month
- Summarized medical documents using generative AI to allow physicians to decide which patient records to read in-full

Data Analyst

March 2020-June 2022

- [Coauthored published research](#) in top cardiology journal showing addition of Apple watch atrial fibrillation detection software did not increase AFib claims using Difference-in-Differences Estimation
- Showed doctors were 7x more likely to mark patient conditions as acute/chronic after swimlanes were added to patient chart by creating Tableau dashboard querying Snowflake tables in Airflow

Data Engineer

May 2018-March 2020

- Reduced daily extract failures from 677 to 5 in 1 month through Oracle SQL optimization by leveraging indexes and limiting cross-joins. Helps prevent EHR outages for hospitals by decreasing ETL job failures

University of Rochester Medical Center *Rochester, NY*

Software Engineer

June 2014-April 2018

- Architected, implemented, and supported UR Attendance, an attendance tracking website for 1k medical residents and fellows using PHP; allows physicians more time with patients by reducing admin time by 2 hours per meeting

Independent Projects

Titanic

<https://github.com/SeanCooke/Titanic>

- Feature Engineering and XGBoost classifier achieves 86% accuracy on survival from Titanic Data Set

Tumor Classification

www.github.com/SeanCooke/tumor-classification

- Random forest classifier achieves 96% accuracy on diagnosis from Breast Cancer (Wisconsin) Data Set

Education

University of Rochester *Rochester, NY*

August 2015-May 2017

- Master of Science in Computer Science

SUNY Geneseo *Geneseo, NY*

January 2011-May 2014

- Bachelor of Arts in Mathematics, Minor in Computer Science