Diane M. Napolitano

Somerville, NJ 08876

📞 (914) 400-8714 • ✉ dmnapolitano@gmail.com • in dinapolitano
○ dmnapolitano • 🌐 diane.sdf-us.org

Relevant Skills

Python, pandas, scikit-learn, SQL (PostgreSQL, Redshift, MySQL), Git (GitHub, GitLab, Bitbucket), MongoDB, Apache Spark, Amazon Web Services (AWS), Bayesian Inference and Prediction, Personalization, Time Series Forecasting, Data Engineering, HuggingFace, PyTorch

Experience

Data Science

The Washington Post   **New York, NY**
Senior Data Scientist / Applied Machine Learning Scientist   *August 2023 to Present*
Data Scientist on two teams: the Post Pulse team, focused on prediction of election outcomes and models to provide insights into the electorate; and the Personalization team, responsible for user-specific content recommendations.
○ Led the development of a weighted, stratified regression model based on Ecological Inference used during Primary elections to infer the share of a candidate's votes in a prior Primary that is going towards a candidate in the current Primary.
○ Co-developer of the Post's Presidential polling averages, a Bayesian random-walk model written in Python and NumPyro which aggregates state-level and national horse-race polls conducted prior to the 2024 election. Contributed code, data, and lead the development of the model's prior using on economic features Collaborated with members of the Polling and Data Engineering teams. Polling averages generated over 9 million page views and was the top-performing Politics story in 2024.
○ Contributed model improvements to, and led the data preparation efforts for our regression model with bootstrapping used to produce live, real-time predictions with uncertainty during the 2024 election.
○ Led the effort to refine the measurement of Personalization products in terms of both model evaluation, quantifying the business value of content recommendation, and "beyond accuracy" metrics to convert non-paying customers and retain subscribers. Performed an extensive literature review and causal analysis while collaborating with other Data Scientists and Product stakeholders.

Associated Press (AP)   **New York, NY**
Data Science Manager   *December 2020 to August 2023*
Data Scientist tasked with improving AP's search and elections tabulation platforms.
○ Developed time series regression models with uncertainty to forecast the number of votes received per combination of U.S. geographical location both prior to and during live elections. Developed REST API and coordinated with infrastructure team to deploy, maintain, and refresh models on AWS. Oversaw integrated team of junior data analysts and data engineers to collect and clean additional data sources (such as U.S. Census data) to improve model performance, resulting in a substantial reduction in MAPE.
○ Using HuggingFace and PyTorch, fine-tuned a fast DistilBERT model to identify named entities in keyword search queries and improve search result retrieval. Lead annotation task on real user queries to create a human-labeled data set used to perform model evaluations. Coordinated with the Search Platform developers to ensure smooth production deployment.
○ Led annotation task to assign relevancy scores to named entities within news stories, resulting in improvements in relevancy sort-order for end-users of the AP Newsroom platform.

Product Management

Refinitiv (Formerly Thomson Reuters Financial and Risk)   **New York, NY**
News Content and Metadata Product Manager   *March 2020 to December 2020*
Tasked with improving the Eikon, Workspace, and API customer experience through news availability and discoverability.
○ Developed a method to classify images as informational graphics using a logistic regression model.
○ Lead the effort to build Tableau dashboards to aggregate user behavior and external client feedback data, including search queries and stories read per news source, topic, and language.

Data and Machine Learning Engineering

Rutgers University · **Piscataway, NJ**

Research Programmer · *January 2019 to August 2019*

Application Developer in the Laboratory for Computer Science Research, supporting the Department of Computer Science.

○ Supervised student developers, contributed new functionality to automated grading system, and created tutorials on how to use the department's Hadoop cluster for the Data Science MS degree program.

Educational Testing Service (ETS) · **Princeton, NJ**

Research Engineer · *May 2011 to December 2018*

Lead Back-End Software and Data Engineer and Architect in the NLP and Speech group in the R&D division writing production code.

○ Led the development and re-architecting of *TextEvaluator*, ETS's reading complexity and genre analysis engine, incorporating best-in-class NLP feature extraction into object-oriented Python code base. Worked with scientists, clients, and front-end developers to create the website and the functionality they needed.

○ With assistance from junior engineers and scientists, responsible for design and development of entire back-end code base of *The Writing Mentor*, based on Apache Storm with code in Python, Java, and Perl.

○ Main contributor to *e-rater*, ETS' engine for automated essay scoring used on major standardized tests. Used Scrum processes with cross-functional teams of senior and junior engineers to deliver production code and NLP-based scoring models. Engine is an Apache Storm-based architecture written in Python, Perl, and Java.

○ Lead Data Engineer responsible for building and maintaining PostgreSQL databases for research and operational use, as well as performing custom extract-transform-load (ETL) procedures between our databases and the corporate Data Warehouse (Oracle). Databases stored over thirty million test-taker essays, short-answer, and spoken responses.

Stroz Friedberg LLC · **Valley Stream, NY**

Natural Language Processing Developer · *February 2009 to March 2010*

Responsible for integrating state-of-the-art statistical NLP into existing Electronic Discovery platform, including Named Entity Recognition and a custom filter for irrelevant documents.

Teaching

SUNY College at Old Westbury · **Old Westbury, NY**

Adjunct Instructor, Math and Computer Information Sciences Department · *September 2008 to May 2011*

Created lectures, assignments and exams, assigned grades, and designed one to three courses per semester at small (approximately 4,000 students) open-enrolment public college. Taught on introductory Java, C++, and electives in Artificial Intelligence and Operating Systems.

Volunteering

Peer Review, Association for Computational Linguistics (ACL)	*2016 - Present*
Peer Review, Journal of Open Source Software (JOSS)	*2024 - Present*
Statistical Consultant, Statistics Without Borders	*2025 - Present*

Education

State University of New York at Stony Brook · **Stony Brook, NY**

Master of Science, Computer Science · *2006 to 2008*

Thesis: *Software for Writing Assistance and Improvement for Advanced Learners of English*

Advisor: Amanda Stent

State University of New York at Binghamton · **Binghamton, NY**

Bachelor of Science, Computer Science, Minor in History · *2002 to 2006*