Diane M. Napolitano

28 Fairmont Avenue – Somerville, NJ 08876

📞 (914) 400-8714 ● ✉ dmnapolitano@gmail.com ● 🌐 http://diane.sdf-us.org

Professional Interests and Skills

Professional Interests: Natural Language Processing, Time Series Forecasting, Information Retrieval and Extraction, Computational Social Science, Educational Applications

Programming Languages: Python, SQL (PostgreSQL, Microsoft), Java

Other: *nix (MacOS, Linux), Git (including GitHub, GitLab), MongoDB, Amazon Web Services (AWS), Atlassian suite, Elasticsearch, Tableau

Experience

Data Science

The Washington Post   **New York, NY**
Senior Data Scientist, Elections Engineering   *August 2023 to Present*

Associated Press (AP)   **New York, NY**
Data Science Manager   *December 2020 to August 2023*

Data Scientist improving AP's search and elections tabulation platforms.

○ Developed Bayesian time-series regression models to predict the number of votes received per combination of U.S. geographical location and electoral race. 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.

○ 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.

○ Fine-tuned a fast DistilBERT model to identify named entities in keyword search queries and improve search result retrieval. Evaluated model on in-house annotated dataset, and coordinated with the Search Platform developers to ensure smooth production deployment.

Product Management

Refinitiv   **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.

Software 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. Projects included:

○ *TextEvaluator*. Incorporated 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.

○ *The Writing Mentor*. With assistance from junior engineers and scientists, responsible for design and development of entire back-end code base, documentation, and collaboration with external front-end developers.

○ *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.

Community Involvement

Member
Association for Computational Linguistics (ACL) *2015 to Present*

Reviewer, Computational Social Sciences Track
ACL-IJCNLP, ACL, AACL, EACL, EMNLP *2021 to Present*

Reviewer, Main Conference
PyData Global *2021*

Reviewer, NLP Applications Track
AACL-AFNLP, EMNLP, ACL, NAACL-HLT, ACL-IJCNLP, EACL *2020 to Present*

Reviewer, Social Media Track
NAACL *2019*

Program Committee
ACL's Innovative Use of NLP for Building Educational Applications (BEA) Workshop *2018 to Present*

Reviewer, System Demonstrations Program
EMNLP, ACL, NAACL-HLT, IJCAI, EACL *2016 to 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*