Sanya Bathla Taneja

Email: sbt12@pitt.edu

Website | LinkedIn | GitHub

EDUCATION

PhD Intelligent Systems

University of Pittsburgh | 2020-24 — Pittsburgh, PA

Major: Artificial Intelligence

MS Intelligent Systems

University of Pittsburgh | 2018-20 — Pittsburgh, PA

B.Tech. Computer Science and Engineering

Indira Gandhi Delhi Technical University for Women | 2014-18 — Delhi, India

EXPERIENCE

Graduate Student Researcher | *February 2021 – Present*

University of Pittsburgh, Intelligent Systems Program

- Discovery and generation of mechanistic hypotheses for pharmacokinetic natural product-drug interactions (NPDI) using knowledge graphs, ontologies, machine reading and embeddings.
- Standardization of and generation of signals to identify NPDIs from FAERS data and poison center reports to assess clinical relevance of NPDIs.
- Part of the informatics core of the NaPDI Center created by the National Institutes of Health National Center for Complementary and Integrative Health (NCCIH).

Graduate Student Researcher | *February 2020 – 2021*

University of Pittsburgh, Intelligent Systems Program

- Responsible for longitudinal EHR data extraction, cleaning, and primary technical development of analyses.
- Implemented and tested machine learning and case-control epidemiological analyses for Alzheimer's disease onset and risk factors.
- Assisted with development of knowledge graph using biomedical ontologies and machine reading to discover novel associations to prevent the onset of Alzheimer's disease. Funded by the Pitt Momentum Teaming Grant (2020).

Research Assistant | *September 2018 – February 2020*

University of Pittsburgh, School of Medicine

- Developed and designed natural language processing and machine learning pipeline for twitter surveillance of vaping at the Center for Research on Media, Technology and Health.
- Responsible for RITHM software framework maintenance, documentation, and upkeep of the GitHub repository. (https://github.com/CRMTH/RITHM).

- Assisted with data extraction and processing for real-time Twitter data mining for public health research and analysis using Python and resources at the Pittsburgh Supercomputing Center (PSC).

Summer Short-Term Trainee Program | *June – August 2019*
University of Pittsburgh, Department of Biomedical Informatics | Malawi, Africa

- Developed Bayesian networks and machine learning models with decision tree analysis to diagnose and manage childhood malaria in Malawi.
- Consulted experts at health centers and Global Health Informatics Institute in Malawi and UPMC Children's Hospital to design the study and implement it as Master's thesis.

Software Development Engineer (SDE) Intern | *February – July 2018*
Amazon India

- Developed backend API's for the Seller and Retail website using Java, Spring MVC, Coral, JavaScript and Handlebars. Involved in adding order cancellation details to the Seller dashboard to supplement the seller website.

PEER REVIEWED PUBLICATIONS

- Visweswaran S, Colditz JB, O'Halloran P, Han NR, **Taneja SB**, Welling J, Chu KH, Sidani JE, Primack BA, Machine Learning Classifiers for Twitter Surveillance of Vaping: Comparative Machine Learning Study, *J Med Internet Res 2020;22(8):e17478*, URL: https://www.jmir.org/2020/8/e17478, DOI: 10.2196/17478
- Sidani JE, Colditz J, Shensa A, Melcher E, **Taneja SB**, Hoffman BL, Davis E, James AE, Primack BA, Chu KH. Identifying Contexts and Dissemination Patterns Of Electronic Nicotine Delivery System (Ends) Misinformation On Social Media. *In Annals Of Behavioral Medicine 2020 May 1 (Vol. 54, Pp. S186-S186)*
- Abhishek, A., **Taneja, S. B.,** Malik, G., Anand, A., & Awekar, A., Fine-grained Entity Recognition with Reduced False Negatives and Large Type Coverage. *Presented at the Automated Knowledge Base Construction (AKBC) Conference, 2019*
- Gupta A, **Taneja SB**, Malik G, Vij S, Tayal DK, Jain A. SLANGZY: a fuzzy logic-based algorithm for English slang meaning selection. *Progress in Artificial Intelligence. 2019 Apr 1;8(1):111-21.*
- Thesis: **Taneja, Sanya Bathla.** Bayesian Networks for Diagnosing Childhood Malaria in Malawi. Master's Thesis, University of Pittsburgh, 2020. Available from: http://d-scholarship.pitt.edu/38993/.

PRESENTATIONS

- S. Malec, **S. Taneja**, K. Witonsky, C. Shaaban, H. Karim, A. Levine, S. Albert, P. Monro, R. Boyce. Modeling Alzheimer's Disease by Combining Knowledge Extracted from Biomedical Literature with Biomedical Ontologies. *Poster Presentation. AMIA Informatics Summit 2021.*
- Hoffman BL, Colditz JB, Sidani JE, Davis EM, **Taneja SB**, James AE, Primck BA, Morris A, Brink L, Lynch M, Rose JJ, Chu KH. Correlation Of Twitter Data To Reported Cases

Of E-cigarette Or Vaping Product Use-associated Lung Injury (EVALI). *Poster Presentation. 2020 American Thoracic Society International Conference. May 17, 2020. (Canceled due to COVID-19)*

- **Taneja SB.** What do you think of vaping? Machine learning methods for Twitter Stance Detection. *Presented at: Intelligent Systems Program AI Forum, University of Pittsburgh; November 8, 2019. http://isp.pitt.edu/node/1997.*
- **Taneja SB**, Douglas GP, Druzdzel MJ. Using Bayesian networks to diagnose childhood illness in low- and middle-income countries: case of malaria in Malawi. *Poster presented at: DBMI Annual Training Program Retreat, University of Pittsburgh; August 22, 2019.*

SKILLS AND INTERESTS

Skills and Interests: Machine Learning, Natural Language Processing, OMOP Common Data Model, ETL of Electronic Health Records (EHR) data, Clinical Decision Support, Bayesian Networks, Knowledge Graph, Knowledge Representation
Technologies: Python, R, C, C++, Git, JavaScript, HTML, MATLAB, SQL
Libraries: NLTK, Spacy, Pandas, Scikit-learn, Jupyter Lab, Keras, Networkx

ADDITIONAL EXPERIENCE

American Medical Informatics Association (AMIA)

- Translational Bioinformatics Year-in-Review team, AMIA Informatics Summit 2021
- Member of the AMIA student working group

COURSE PROJECTS

Foundations of Biomedical Informatics | Fall 2019
Utilizing clinical notes in electronic medical records (MIMIC III) to predict mortality risk in the ICU.

Natural Language Processing | Spring 2019 | https://github.com/sanyabt/NLP-CS2731
Feature analysis and multilabel, multiclass classification of emotions in short texts using Random Forest.

Machine Learning | Spring 2019 | https://github.com/sanyabt/ML-CS2750
Comparison of supervised machine learning models to predict patient no-shows in primary care hospitals.

Undergraduate Research Project | November 2017 – May 2018
Development of algorithm for English slang meaning selection from social media using fuzzy membership functions and natural language processing with Python.