Sanya Bathla Taneja

Email: sbt12@pitt.edu
Pittsburgh, PA, USA
Website | LinkedIn | GitHub

EDUCATION

University of Pittsburgh — **Pittsburgh, PA**
PhD Intelligent Systems | *2020 – 2023*
Major: Artificial Intelligence

University of Pittsburgh — **Pittsburgh, PA**
M.S. Intelligent Systems | *2018 – 2020*
GPA: 4.0

Indira Gandhi Delhi Technical University for Women — **Delhi, India**
B.Tech. Computer Science and Engineering | *2014 – 2018*

EXPERIENCE | PROJECTS

University of Pittsburgh, Intelligent Systems Program
Graduate Student Researcher | ***February 2020 – Present***

- Discovery of novel associations to prevent the onset of Alzheimer's disease using large-scale electronic health records.
- Funded by the Pitt Momentum Teaming Award (2020)

University of Pittsburgh, School of Medicine
Research Assistant | ***September 2018 – February 2020***

- Real-time Twitter data mining for public health research and analysis through natural language processing and machine learning techniques at the Center for Research on Media, Technology and Health using Python and resources at the Pittsburgh Supercomputing Center (PSC).
- Responsible for RITHM software framework maintenance, documentation, and upkeep of the GitHub repository. (https://github.com/CRMTH/RITHM).

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

- Developed Bayesian model for diagnosis of childhood illness in low- and middle-income countries at the Global Health Informatics Institute in Malawi.
- Site visits to health centers, village health posts, district and central hospital to observe pediatric healthcare and diagnosis workflow to enhance the model.

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

- 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.

Indira Gandhi Delhi Technical University for Women
Undergraduate Research Project | *November 2017 – May 2018*
Natural Language Processing with Python

- Conceptualized an algorithm for English slang meaning selection using fuzzy membership functions on parameters and slang definitions found on Urban Dictionary.
- Developed scripts for text mining of English slang from popular social media sites (Twitter, YouTube, Reddit), processing of data using NLTK and Python, and execution of the algorithm.

PUBLICATIONS

- Machine learning for Twitter surveillance of vaping. *In review, 2020*
- 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)*
- 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, S. B.**, Malik, G., Vij, S., Tayal, D. K., & Jain, A. (2019). SLANGZY: a fuzzy logic-based algorithm for English slang meaning selection. *Progress in Artificial Intelligence, 8(1), 111-121.*

SKILLS

Skills: Python, C, C++, Git, JavaScript, HTML, MATLAB, MySQL, Machine Learning, Natural Language Processing
Libraries: NLTK, Spacy, Pandas, Scikit-learn, Jupyter Notebook, Keras

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
Feature analysis and multilabel, multiclass classification of emotions in short texts using Random Forest. (https://github.com/sanyabt/NLP-CS2731)

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