Sanya Bathla Taneja

WORK EXPERIENCE | PROJECTS:

Research Assistant | School of Medicine, University of Pittsburgh
October 2018-Present

- Leveraging Twitter to Monitor Nicotine and Tobacco-Related Cancer Communication with real-time monitoring of Twitter health messages with PI Dr. Brian Primack at the Center for Research on Media, Technology and Health. (https://github.com/CRMTH/RITHM)
- Responsible for RITHM software framework maintenance, documentation, and upkeep of the Github repository.
- Analyzed real-time Twitter data for public health research involving e-cigarettes, vaping through natural language processing and machine learning techniques using Python and Bridges infrastructure at the Pittsburgh Supercomputing Center (PSC).

Software Development Engineer (SDE) Intern | Amazon India
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.
- Implemented and documented an internal interface used for easier troubleshooting of seller issues.

Undergraduate Research Project | IGDTUW
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.
- Implemented 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.
- Created web portal with slang word definitions, along with the calculated 'Slang Factor' to evaluate the relevance of each definition.

Summer Research Internship | Indian Institute of Technology, Guwahati
June-July 2017
Natural Language Processing with Python

- Developed 'End-to-End Fine-Grained Entity Recognition System' from Wikipedia and Freebase using Python.
- Involved in mining, processing and annotation of training and test sentence corpus from Wikipedia pages for entity recognition and classification into 118 categories.

Amazon Campus Mentorship Series | Amazon India
November 2016-April 2017

- Worked on Locker Capacity Management System using Django and MySQL. Interface developed to manage and book lockers to deliver packages when placing an order. (Link: https://github.com/sanyabt/acms-1)

PUBLICATIONS:

Gupta, A., Taneja, S.B., Malik, G. et al. Prog Artif Intell (2018).
"SLANGZY: a fuzzy logic-based algorithm for English slang meaning selection".
https://doi.org/10.1007/s13748-018-0159-3

CONTACT:

Email: sbt12@pitt.edu
Phone: +1 (412)-626-4025
Website | LinkedIn | GitHub

EDUCATION:

MS INTELLIGENT SYSTEMS | 2018-20
Intelligent Systems Program
University of Pittsburgh, PA
Cumulative GPA: 4.0

B. TECH COMPUTER SCIENCE ENGINEERING | 2014-18
Indira Gandhi Delhi Technical University for Women (IGDTUW)
Percentage: 85.6% (First Division with Distinction)

SKILLS:

Python, C, C++, Git
JavaScript, HTML, CSS, Django, Java, Spring
MySQL, DynamoDB
Proficient communication and oratory skills
(TOEFL iBT Score: 120)

LIBRARIES AND MORE:

NLTK, Spacy, Scikit-learn, Jupyter Notebook, Tweepy, Numpy, Pandas, Slurm

INTERESTS:

Natural Language Processing, Data Mining, Artificial Intelligence, Clinical Informatics

SELECTED COURSEWORK:

Foundations of Artificial Intelligence
Public and Clinical Health Informatics
Applied Statistics
Natural Language Processing
Machine Learning

EXTRA-CURRICULAR:

Member of Pitt Python Linguistics group (PyLing)
Talk on Twitter Data Mining at PyLing, November 2018

ACHIEVEMENTS:

Contribution to Python open source visualization and image processing library 'scikit-image' (https://github.com/sanyabt/scikit-image)
Kaggle submission for 'Bag of words meets bag of popcorns' challenge
Selected for Undergraduate Summer School in Indian Institute of Science, Bangalore, July 2016
Participation in workshops: Women Techmakers (Google India), PyData conference (Delhi)