

# Sanya Bathla Taneja

Email: [sbt12@pitt.edu](mailto:sbt12@pitt.edu)

Pittsburgh, PA, USA

[Website](#) | [LinkedIn](#) | [GitHub](#)

## INTEREST

Research in natural language processing, machine learning, data science and clinical informatics.

## EDUCATION

### University of Pittsburgh

Pittsburgh, PA

MS Intelligent Systems | *August 2018 – April 2020*

Major: Artificial Intelligence

Cumulative GPA: 4.0

### Indira Gandhi Delhi Technical University for Women

Delhi, India

B.Tech. Computer Science and Engineering | *August 2014 – May 2018*

## EXPERIENCE | PROJECTS

### 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 based on the Integrated Management of Childhood Illness (IMCI) protocol by WHO.
- Site visits to health centers, village health posts, district and central hospital to observe pediatric healthcare in Malawi, Africa.

### University of Pittsburgh, School of Medicine

#### Research Assistant | *October 2018 – Present*

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

### 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.

## **Indian Institute of Technology, Guwahati**

### **Summer Research Internship | 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 India**

### **Amazon Campus Mentorship Series | 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.

## **PUBLICATIONS**

---

- Abhishek, A., **Taneja, S. B.**, Malik, G., Anand, A., & Awekar, A. (2019). Fine-grained Entity Recognition with Reduced False Negatives and Large Type Coverage. *arXiv preprint arXiv:1904.13178*.
- 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**

---

**Technical Skills:** Python, C, C++, Git, JavaScript, HTML, CSS, MATLAB, MySQL, Django

**Libraries:** NLTK, Spacy, Pandas, Scikit-learn, Slurm, Jupyter Notebook

**Languages:** English (TOEFL iBT score: 120) and Hindi

## **SELECTED GRADUATE COURSEWORK**

---

Foundations of Artificial Intelligence | Public and Clinical Health Informatics |

Machine Learning | Applied Statistics | Natural Language Processing |

## **EXTRACURRICULAR**

---

- Summer internship and course on Principles of Global Health Informatics in Malawi, Africa

## **Sanya Bathla Taneja**

**Email:** [sbt12@pitt.edu](mailto:sbt12@pitt.edu)

Pittsburgh, PA, USA

[Website](#) | [LinkedIn](#) | [GitHub](#)

- Member of Pitt Python Linguistics Group (PyLing)