# Sanya Bathla Taneja

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


## EDUCATION

**University of Pittsburgh** **Pittsburgh, PA**
M.S. Intelligent Systems | *August 2018 – April 2020*
Major: Artificial Intelligence
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, School of Medicine**
**Research Assistant | *September 2018 – Present***

- Leveraging Twitter to Monitor Nicotine and Tobacco-Related Cancer Communication with real-time monitoring of Twitter health messages with Dr. Kar Hai Chu and Dr. Shyam Visweswaran at the Center for Research on Media, Technology and Health (CRMTH).
- Real-time Twitter data mining for public health research involving e-cigarettes and analysis through natural language processing and machine learning techniques including word embeddings, supervised models and deep learning 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) and data analysis for miscommunication on social media, social network analysis and user behavior.

**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. Presented to officials at the Ministry of Health and at the annual department retreat as a poster presentation in August 2019.
- 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.

**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.

## PUBLICATIONS

- Machine learning for Twitter surveillance of vaping. *In review, 2019*
- Correlation of Twitter Data to Reported Cases of E-cigarette or Vaping Product Use-associated Lung Injury (EVALI). *Abstract in review, 2019*
- 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

**Technical Skills:** Python, C, C++, Git, JavaScript, HTML, MATLAB, MySQL, Linux environment
**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)