

JENIYA TABASSUM

sites.google.com/site/jeniyatabassum | +1-614-620-5644 | jeniya.tabassum@gmail.com | linkedin.com/in/jeniyat | github.com/jeniyat

- ◇ 8+ years of hands-on experience in building machine learning system for large scale data.
- ◇ 5+ years of experience in developing end to end deep learning solutions for structured/unstructured data
- ◇ Proficient in PyTorch, Tensorflow, scikit-learn
- ◇ Published first author long-papers in ACL & EMNLP

EDUCATION

Ph.D. in Computer Science and Engineering 12/2020
The Ohio State University (OSU), Columbus, Ohio, USA
Thesis: Information Extraction From User Generated Noisy Texts ([dissertation](#))

B.S. in Computer Science and Engineering 04/2012
Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh
CGPA: 3.87/4.00 (top 5%)

TECHNICAL SKILLS

- ◇ ML and Visualization libraries: **PyTorch**, **Tensorflow**, **HuggingFace**, **Scikit**, Scipy, NLTK, Matplotlib
- ◇ Programming Languages: **Python**, **R**, Java, Scala, , Matlab
- ◇ Cloud Platform and DBMS: **AWS**, **SageMaker**, **MySQL**, GCP, Oracle, JSON, PySpark

PROFESSIONAL EXPERIENCE

Applied Scientist, Amazon Science 04/2022 - present

- ◇ Amazon Selection Monitoring
 - Created automation for catalog generation by developing ML algorithms which enabled savings of 0.5 human hours/day.
 - Tools: Python, PyTorch, SQL, Scikit, AWS

Software Development Engineer, Amazon AWS 08/2021 - 04/2022

- ◇ Amazon SageMaker Python SDK ([code](#))
 - Developed functionality to the open source python library to include training and deploying huggingface framework
 - Tools: Python, Bash, CDK
- ◇ Amazon SageMaker Example Notebooks ([code](#))
 - Incorporated end to end ML notebooks to the open sourced repository depicting the workflow of State of the Art machine learning models
 - Tools: Python, Typescript, CDK

PUBLICATIONS

- ◇ **Jeniya Tabassum**, Mounica Maddela, Wei Xu and Alan Ritter, "[Code and Named Entity Recognition in StackOverflow](#)," **ACL '20**.
- ◇ **Jeniya Tabassum**, Syndey Lee, Wei Xu and Alan Ritter, "[WNUT-2020 Task 1 Overview: Extracting Entities and Relations from Wet Lab Protocols](#)," **WNUT @ EMNLP '20**.
- ◇ **Jeniya Tabassum**, Alan Ritter and Wei Xu, "[Time Expression Resolution for Social Media Data](#)," **WinNLP @ ACL '17**.
- ◇ **Jeniya Tabassum**, Alan Ritter and Wei Xu, "[TweeTIME: Minimally Supervised Method for Recognizing and Normalizing Time Expressions in Twitter](#)," **EMNLP '16**.
- ◇ **Jeniya Tabassum** and Alan Ritter, "[Distant Supervision for Temporal Resolution](#)," **MASC-SLL '16**.
- ◇ Asif Salekin, **Jeniya Tabassum** and Masud Hasan, "[Extract and Rank Web Communities](#)," **WIMS '13**.
- ◇ **Jeniya Tabassum**, Himel Dev, Mohammed Eunus Ali and Md. Fahim Abdullah, "[Role of Social Media during Disaster in the Context of Savar Tragedy](#)," **WADM '13**.

RESEARCH EXPERIENCE

Graduate Research Assistant, OSU (Advisors: Prof. Wei Xu & Prof. Alan Ritter) 08/2014 - 12/2020

- ◇ Fine Grained Entity Extraction From Software Text ([web-demo](#) / [code](#) / [data](#) / [paper](#) / [talk](#))
 - **Lead a team of 4 annotators** to create the first software domain named-entity corpus with **15k+ StackOverflow sentences**
 - Proposed an **embedding level attention** for the transformer based NER model
 - Proposed model achieved **F1 Score of 78.41** with 21.6 increase over vanilla BERT ([current State of the Art](#))
 - Tools: Python, PyTorch, Huggingface, Javascript, Tornado, Brat
- ◇ Entity and Relation Extraction From Wet Lab Protocol ([code](#) / [data](#) / [paper](#))
 - **Lead a team of 3 annotators** to create an entity-relation corpus for the procedural texts from **700+ wet lab recipes**
 - Developed **neural ensemble models** for both tasks
 - Proposed model achieved **F1 Score of 76.84** for NER task and **F1 Score of 81.32** for RE task ([current State of the Art](#))

- Tools: Python, PyTorch Scikit, Brat
- ◇ Time Information Resolution From Tweets ([code](#) / [data](#) / [paper](#) / [talk](#))
 - Developed a **temporal tagger** to detect & and normalize tweet time expressions by utilizing the **distant supervision approach**
 - Developed a **date resolver** that can combine the numerical date features with word vectors via **bi-linear BiLSTM model**
 - Proposed model achieved **F1 Score of 68.12** with 17% increase over SUTIME ([current State of the Art](#))
 - Tools: Python, Keras, Tensorflow, Scala, Sklearn
- ◇ User Profile Mining From Twitter ([code](#) / [data](#))
 - Modeled the **spread of information through tweets**
 - Analyzed the tweets from **40M+ users** to evaluate whether the profile is controlled by human or bots
 - Tools: Python, Tweepy, Humanizr, Botometer
- ◇ Learning Semantics From Software Social Networks ([code](#) / [data](#))
 - Extracted proximity from the followers activity of **84M+ GitHub repositories**
 - Created **user embeddings** and **repository embeddings** from the text contents of the repository-user network
 - Utilized the proposed **repository embedding** to evaluate similarities in between repositories
 - Tools: Python, PyGithub, Numpy

Undergraduate Research Assistant, BUET (Advisors: Prof. Masud Hasan & Prof. Eunus Ali)

02/2010 - 06/2013

- ◇ Social Media on Disaster Response ([paper](#))
 - Explored the **impact of social media in solving disaster** related problem by analyzing the Facebook posts on the Savar Tragedy
 - Proposed an approach to **co-ordinate the relief distribution** by filtering out the repetitive post
 - Tools: Python, LIWC, R
- ◇ Web Community Extraction ([paper](#) / [talk](#))
 - Proposed a **novel extraction and ranking algorithm** for web communities
 - Demonstrated **improvement in auctions of a sponsored search market** by utilizing the proposed algorithm
 - Tools: Java, Matlab

INVITED TALKS

- ◇ **Information Extraction form User-generated Text**. Megagon AI, March 2021.
- ◇ **NLP on Noisy User-generated text - NER for StackOverflow**. Aggregate Intellect AISC, July 2020.
- ◇ **Temporal Normalization from Noisy Twitter Text**. Bangla-AI, September 2018.
- ◇ **Minimally Supervised Time Expression Resolution for Social Media Domain**. Georgetown University, April 2017.
- ◇ **Probabilistic Graphical Model with Latent Variables for Temporal Tagging**. Guest Lecturer for CSE 5535, OSU, March 2017.
- ◇ **Distant Supervision for Temporal Resolution**. Clippers Meeting, OSU, October 2016.
- ◇ **Minimally Supervised Temporal Recognizer and Normalizer**. AI seminar, OSU, September 2016.

PROFESSIONAL SERVICES

- ◇ Reviewer: ACL '19-'23, NAACL '18-'22, EMNLP '18-'21, AAAI '20, HCC '19.
- ◇ Program Committee: WinNLP '19-'22, NAACL-SRW '19, WNUT '16-21, MASC-SLL '16.
- ◇ Organizer, Shared Task @ WNUT '20
- ◇ Student Chair: [ACL Student Research Workshop](#) '18,
- ◇ Panel Member at WIE session, ICCIT '16
- ◇ Student Organizer: [NLP Speaker Series](#) (OSU) ['16 - '18]

TEACHING EXPERIENCE

Senior Lecturer, OSU, CSE

01/2021 - 06/2021

- ◇ Instructed the course on **"Introduction to AI (Intermediate Concepts)"** to a class of 120 students
- ◇ Supervised **25+ student projects**

Lecturer, OSU, CSE

01/2020 - 12/2020

- ◇ Instructed the course on **"Introduction to AI (Basic Concepts)"** to a class of 40 students
- ◇ Designed 4 **programming assignments** to evaluate the student understanding of AI concepts
- ◇ Collaborated with faculty supervisors to update the syllabus and create the course contents with current ML algorithms

Teaching Associate, OSU, CSE

08/2019 - 12/2019

- ◇ Graded programming assignments for an Advance AI class
- ◇ Supervisor: Prof. Eric Fosler-Lussier