

Mohna Chakraborty

☎ 716.748.5386 | ✉ mohnac@iastate.edu | <https://github.com/mohna0310> | 📍 Ames, Iowa

EDUCATION

Iowa State University

Ph.D in Computer Science

Aug 2020 – Expected 2024

Ames, IA, USA

- **Research Area:** Review Analysis, True label inference, Data Mining, Natural Language Processing, Machine Learning

State University of New York at Buffalo

M.S in Computer Science

Aug 2018 – July 2020

Buffalo, NY, USA

SKILLS

- **Languages:** Python, SQL, Java, Ruby
- **Tools:** Jupyter Notebook, Anaconda, AWS, Google Colab, Eclipse
- **Libraries and Frameworks:** NumPy, Pandas, scikit-learn, Matplotlib, Tensorflow, Keras, Pytorch, Dash, Scipy, Plotly, SpaCy, NLTK, VADER, TextBlob, Gensim, Beautiful Soup, Regex, igraph, NetworkX

RESEARCH EXPERIENCE

Iowa State University

Graduate Research

Aug 2020 – Aug 2022

- Implemented an architecture for joint extraction of aspect and opinion terms from reviews in a **weakly-supervised setting** to tackle the problems of **noise** and **bias** in data using BERT based double-layer span extraction framework.
- Worked on a fact verification framework using the concept of **multi-instance** and **multi-task learning** to handle the issues of **data imbalance** and **error propagation** between the modules.
- Implemented a framework that utilizes a **graph-based representation learning** that integrates control flow, math, and natural language information of pseudocode and source code to search for pseudocode from a codebase composed of millions of source code snippets by modeling the source code and pseudocode as Interprocedural Control Flow Graph (ICFG) in an unsupervised search setting.
- Working on implementing a joint framework by modeling several different tasks under **Aspect-Based Sentiment Analysis** by considering the correlation among them.
- Working on a **Bayesian approach** for optimal budget allocation to improve labeling accuracy with minimal human effort.

PUBLICATIONS

Conference

KDD, ESEC/FSE, RANLP, OpenReview, AAAI, TSE

- Open-Domain Aspect-Opinion Co-Mining with Double-Layer Span Extraction. **Mohna Chakraborty***, Adithya Kulkarni*, Qi Li (<https://doi.org/10.1145/3534678.3539386>)
SIGKDD Conference on Knowledge Discovery and Data Mining: KDD 2022
- Does reusing pre-trained NLP model propagate bugs?. **Mohna Chakraborty**
(<https://doi.org/10.1145/3468264.3473494>)
ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering: ESEC/FSE 2021
- Does local pruning offer task-specific models to learn effectively?. Abhishek Kumar Mishra*, **Mohna Chakraborty***
(<https://aclanthology.org/2021.ranlp-srw.17>)
Proceedings of the Student Research Workshop Associated with RANLP 2021

- [Re] Domain Generalization using Causal Matching. Richard D Jiles, **Mohna Chakraborty**, (<https://openreview.net/forum?id=r43elaGmhCY>)
ML Reproducibility Challenge 2021
- Boosting Fact Verification via Multi Instance Learning. Adithya Kulkarni, Nasim Sabetpour, **Mohna Chakraborty**, Qi Li, Submitted AAAI-2023
- Neural Search for Algorithm Implementation using Pseudo Code. **Mohna Chakraborty***, Adithya Kulkarni*, Yonas Sium*, Wei Le, Qi Li, Submitted TSE-2023

HONORS AND AWARDS

SIGKDD Travel Award

Aug 2022

SIGKDD Conference

Washington DC, USA

- * One among 46 students selected by KDD for Student Travel Award among all the applicants.

Poster Presentation

May 2022

Iowa State University

Ames, IA, USA

- * First place at the 6th Annual Research Day Competition

Grace Hopper Celebration

Sep 2022

Iowa State University

Ames, IA, USA

- * Selected to represent Iowa State University for the prestigious and competitive Grace Hopper Celebration

WORK EXPERIENCE

Epsilon Data Management, LLC

Chicago, IL, USA

Data Scientist Intern

May 2022 - Aug 2022

- * Current production system uses **Spark SQL** that does not support atomic operations like **upsert**, **delete**, etc., so for every updates the system does overwrite of the entire table resulting in higher resource consumption and time.
- * **Apache Hudi** is proposed as an alternative tool to solve issues like upsert and delete operations. We tested multiple workflows using Apache HUDI including parameter tuning to validate its effectiveness by testing real life scenarios with 1 billion data for up to 10% upsert operations, and the result shows 37% increase in run time compared with Spark SQL.

Data Scientist Intern

May 2021 - Aug 2021

- * Ability to accurately classify individual names plays a crucial role in the quality of the final product. Yet this ability is hampered due to heterogeneity in the collection and validation of data.
- * Current production methods validate the name data using **rule-based approaches**, limiting its ability to update or scale. Therefore, to alleviate this problem, we propose using **machine learning algorithms** on top of rule-based features and encoded features with 191 million data.
- * Based on the performance of all the classifiers, **Random Forest** performed achieved 91% F1 score with **oversampling** and **customized features** explaining the need of incorporating better features to help the learn model better and faster.

Delaware North

Buffalo, NY, USA

Data Analyst Intern

May 2019 - March 2020

- * Worked on training a **Logistic Regression** model that can be used to predict the passenger occupancy across the US Airport and used its prediction to train another Logistic Regression model that predicts the number of transaction counts and labor force needed during various days in restaurants at airport terminals.
- * Worked with **Beautiful Soup** to web scrape data like attendance, duration of the game, home and opponent team details, and other details for the games like NBA, NHL, NFL and MLB from their official website, and used the data to train a **predictive model** to understand the trend of the crowd for all these games.

Ericsson India Global Pvt. Ltd

Mumbai, MH, India

Junior Data Scientist

Sep 2015 - Dec 2017

- Two years of work experience as a Junior Data Scientist in an **agile environment** with hands-on experience in designing and implementing Machine Learning Algorithms.
- Utilized alarms to perform **Anomaly detection** that captures unusual site behavior of base stations via **supervised and unsupervised machine learning models**. The developed models significantly reduced the alarms by 30% . The models helped prevent the faults from happening through early predictions and proactive decisions.
- Developed a **regression model** to identify the cause of weak network connection by performing a descriptive analysis to gain insights into the dataset, summary statistics, and analyzing features impacting the target correlation among variables. The developed regression model achieved a 10% more precise prediction than the previous year.

REFERENCES

Dr. Qi Li

- **Title** : Assistant Professor
- **Department** : Computer Science
- **Email** : qli@iastate.edu
- **University** : Iowa State University