

# MRINAL TAK

Website: [mrinaltak.github.io](http://mrinaltak.github.io)  
+1 4132474610 ◇ [mtak@cs.umass.edu](mailto:mtak@cs.umass.edu)

## EDUCATION

### University of Massachusetts Amherst

Jan'21- Apr'22

MS in Computer Science with specialization in Data Science

Ongoing Courses: Intelligent Visual Computing, Data Science, Software Design

### Indian Institute of Technology Guwahati

2013-2017

B.Tech in Computer Science & Engineering

GPA: 8.72

Courses completed: Database Management System, Advanced Computer Networks, Data Mining,

Computer Vision using ML, Data structures, Operating System, Artificial Intelligence, Algorithms, Probability

## INDUSTRY EXPERIENCE

### Goldman Sachs (Senior Software Engineer)

Jan'19 - Jan'21

- Developed scalable distributed search engine for Goldman Sachs proprietary data, that runs over an Apache Hadoop YARN cluster.
- Developed the batch pipeline to index documents using Apache Hadoop MapReduce framework and store it into Apache HBase NoSQL database.
- Spearheaded redesign of Query Understanding Module and brought down the latency by 82%.
- Responsible for design and development of topic modeling to identify latent topics in financial documents, resulting in 12% improvement in search results

**Technologies Used:** Java 8, Apache Hadoop, JUnit, Apache HBase NoSQL, YARN, REST, IntelliJ, SVN, Google Protocol Buffers

### Samsung Research (Machine Learning Engineer)

Jun'17 - Jan'19

- Spearheaded the development of Anti-counterfeit Engine, by building a fact checking pipeline, which utilized Machine Reading Comprehension techniques.
- Delivered an intelligent Conversational Intelligence Module for smart calendar, using NLP techniques, to power Samsung's Voice Assistant.

**Technologies Used:** Java 8, JUnit, PyTorch, Python, IntelliJ, Github.

## PROJECTS

### Aspect based Sentiment Analysis using zero-shot-learning

2021- ongoing

*Guide: Prof. Andrew McCallum, CICS, University of Massachusetts Amherst*

- To solve the Aspect based Sentiment Analysis task using unlabeled data and making it domain agnostic.
- Finetuning RoBERTa Language Model using intermediate-task fine tuning, and measuring the performance with zero-shot prompts.

**Technologies Used:** PyTorch, HuggingFace, Python

### Entropy: Language Model Based Readability Metric

2017

- Designed a new fine-grained, computational measure of readability.
- Used KL divergence as a computational measure of readability of text, by stating predictability of a text, as determined by standard language models, is a viable metric of its readability.

### Monitoring production line performance to reduce failures

2017

- Posed the task of fault detection as a binary classification problem.
- Used biased sampling method and used sparse online classification algorithms on the numerical features.

## TECHNICAL SKILLS

### Programming Languages

Java, Python, C++

### Libraries & frameworks

Hadoop, HBase NoSQL, Tensorflow, PyTorch, Protobuf, REST, Github, JUnit, AWS

### Softwares

IntelliJ, MATLAB, Visual Studio, Android Studio