

# Sohaib Syed

Lincolnwood, IL 60712 • 773-876-7254 • sohaibsyed6@gmail.com • <https://www.linkedin.com/in/sohaibsyed6/>

---

## SUMMARY

Graduating in May 2024 with both a Bachelors in Computer Science and Masters in Data Science. Experienced in using ML techniques to analyze datasets and fine-tune NLP models. Seeking internship opportunities to apply Data Science skills.

---

## EDUCATION

ILLINOIS INSTITUTE OF TECHNOLOGY, CHICAGO, IL

Expected May 2024

### Bachelor's in Computer Science and Master's in Data Science

Relevant Courses: Applied Statistics, Data Mining, Data Prep & Analysis

- Bachelor's GPA: 3.9/4.0; Master's GPA: 3.8/4.0
  - Dean's List: Fall 2019, Fall 2020, Spring 2021, Fall 2021, Spring 2022, Fall 2022
  - Enrolled in Accelerated Master's Program to simultaneously obtain a Bachelors and Masters with an additional school year
- 

## EXPERIENCE

*Machine Learning Intern*

May 2022 - August 2022

**Motorola Solutions Inc., Applied Technology**

Schaumburg, IL

- Fine-tuned wav2vec2.0 for Voice Activity Detection tasks using linear and convolutional layers with PyTorch framework
  - Extracted stand-alone wav2vec2.0 300M parameter NLP speech model from Hugging Face
  - Developed Python scripts to train and test wav2vec2.0 model on degraded audio
  - Produced score files to classify every 20ms as speech or non-speech to allow for efficient performance evaluation
  - Presented research findings about wav2vec2.0 and how it outperforms in-house models to management
- 

## PROJECTS

*Incident Resolution Time Prediction Modeling*

December 2022 - December 2022

**Illinois Institute of Technology**

Chicago, IL

- Analyzed incident event log with 100k+ records from UCI ML repository to determine time to resolve a problem
- Cleaned 32 dimension dataset by removing missing values and reducing dimensions to 10 relevant characteristics
- Built 3 regression models and 3 random forest models each with different sizes and compare performance
- Concluded best regression model is unfit for dataset and random forest is better at predicting interval of incidence response

*Movie Recommender System*

November 2021 - November 2021

**Illinois Institute of Technology**

Chicago, IL

- Developed a model to recommend a user 5 movies based on 10 movies a user likes
- Utilized simple cosine similarity function to determine similarity score to recommend movies
- Manipulated datasets to score user profiles based on watched genres of movies

*Employee Database*

April 2021 - April 2021

**Illinois Institute of Technology**

Chicago, IL

- Programmed an application to communicate with a PostgreSQL database to keep and update employee information
  - Implemented a SQL script to create a database used to store employee data
  - Configured database to allow admin, manager, and employee users with different information privileges
  - Managed a team of 3 by organizing meetings and delegating tasks to meet project deadline
- 

## SKILLS AND CERTIFICATIONS

PROGRAMMING LANGUAGES: R, Python, Java, SQL

TECHNOLOGIES: PyTorch, TensorFlow, Git, BitBucket, DBeaver, PostgreSQL

CERTIFICATES: Created and deployed predictive model on Abacus.AI platform to earn Certificate of Achievement

INDEPENDENT COURSES: Python for Data Science by CoRise to improve data science skills in Python