# Sushruti Mishra

+1 (718) 313-8596 mishrasushruti99@gmail.com linkedin.com/in/sushruti-mishra https://github.com/mishrasushruti99

## EDUCATION

**New York University - Courant Institute of Mathematical Sciences , New York** *September 2021 - May 2023*
***Masters in Computer Science***
- Relevant Courses: AI, Natural Language Processing and Understanding, OS, DBMS, Algorithms, Data Science

**Delhi Technological University (DTU) Delhi, India** *July 2017 - June 2021*
***B.Tech, Computer Science (Department Rank 2 out of 127)***
- Relevant Courses: ML, Natural Language Processing, Deep learning, DBMS, Design and Analysis of Algorithms, OOPs

## SKILLS

**Programming Languages** : (Proficient) C++, C, Java (Intermediate) R, Python, Javascript, SQL
**Platforms, Databases, and Tools** : Linux, Docker, AWS, ReactJS, Flask, Git, Tableau, MySQL, Hive, Hadoop, PowerBI, Tableau, Spark
**Libraries** : Pytorch, Tensorflow, Keras, (NLTK, Pandas, Numpy, Sklearn, Matplotlib)

## WORK EXPERIENCE

**SEI Investments** | ***SWE - Machine Learning*** *July 2023 -*
- Developing LLM based chatbots

**Unicode** | ***Machine Learning Researcher*** *July 2023 -*
- AI-driven content moderation in youth-oriented online spaces (such as social media platforms/gaming/dating apps)

**Intel** | ***SDE - AI Intern*** *May 2022 - Aug 2022*
- Worked on a tool that facilitates code conversion, translation and migration (at scale) between different coding languages using transformers, GANs, and GNNs coupled with classical AI techniques.
- Worked on tool that helps in improving test coverage and measure test case sensitivity using Bayesian Sampling and other ML-based optimization techniques and further introduced test case prioritization protocol and ML algorithm for the same, decreasing the run time and test cases by 10X and 20X.

**Optum : United Healthcare Group** | ***SDE Intern*** *June 2020 - Aug 2020*
- Developed a scalable web app using Login-Dashboard environment with a REST API backend; reducing administrative mistakes by 80% and hastening risk analysis response by 20%
- Predicted patient's survivability and length of stay using Random forest and Gradient Boosting Regressor by analysing 37 parameters with a confidence of 87%
- Calculated risk factor over a linear classifier with user based features and acuity scores to reduce false positive rate to 4.7%

**Norwegian Geotechnical Institute** | ***SDE Intern*** *May 2018 - July 2018*
- Digitally mapped soil's organic carbon content by using Random Forest, Gradient Boosting Machine and Bayesian Additive Regression Trees on 22 sampling points and discovered from spatial prediction that Landsat 8 and Sentinel 2 data models explain 60-70% and 47-55% of variance respectively
- Implemented K-D Trees and Quadtrees to decrease time to find nearest neighbours and insertion of new spatial data points .
- Implemented spatial search algorithms like packed Hilbert R-tree which enables faster spatial queries on a millions of objects, support faster organization and processing of spatial data points.

## PROJECTS

**Style-mimicking text-generator** *Sep 2021 - Dec 2021*
- Implemented Markov chain models that read 38 Shakespearean plays and 150 poems to generate realistic Shakespearean text.
- Used GLOVE word embedding and LSTM to train a Neural Network to generate text mimicking style of the training text

**Online chess phishing detection** | ***Prof. Sanjay Singh*** *March 2020 - June 2020*
- Detected Cheating in Online Chess matches using CNN - LSTM with an accuracy of 89% on a self curated dataset of 90,000 games
- Conceptualised nomenclature based on player reaction time per move with a special emphasis on sacrifices and mating sequences to streamline detection and improve response time by 30%

**Analysing news articles using performance indicators** | ***Manish Singhal*** *May 2021 - June 2021*
- Analysed financial news articles with a Mining-based hierarchical classifier using 4 Performance Indicators (lagging, lagging reverse, leading, directionality) with positive and negative dictionaries for sentiment analysis.
- Classified the same by assigning a rating of good, bad and neutral with an accuracy of 93%, 90%, 86% respectively

**Coverage and efficacy of ART therapy for Pregnant women and infants** | ***Prof. Maria Unda*** *Sep 2022 - Dec 2022*
- Predicted the Mother-To-Child Transmission Ratio using Linear Regression in the years 2030 and 2045 and analyzed expenditure of countries between 2008 to 2015 using Clustering (HAC) and consequently predicted Spain's expenditure via Support vector regression for administering ART to pregnant women and kids. Data was collated from 5 different sources such as WHO, UNICEF etc.

**Multi Linguistic study on Hate speech using deep learning** | ***Prof. Anil Singh Parihar*** *Sept 2019 - Dec 2019*
- Devised a Multi linguistic aspect annotation using 4 psycholinguistic cues by parsing 17,000 tweets of 10 high user engagement topics and subsequently attained highest accuracy of 89% using CNN-LSTM(variants of MLP, LSTM, BiLSTM).

## PUBLICATIONS

**Review of Text Summarization in Indian Regional Languages** *ICCIN 2020*
**Evaluation of Factors Affecting Compressive Strength of Concrete using Machine Learning** *IEEE ACCTHPA 2020*
**Machine Learning Approach Based on Hybrid Features for Detection of Phishing URLs** *Confluence 2021*
**Hate Speech Detection Using Natural Language Processing: Applications and Challenges** *ICOEI 2021*

## ACHIEVEMENTS

**International Maths Olympiad** | ***Silver Medallist*** *2016*
**National Science Talent Search Examination Scholar** | ***Gold Medallist in Math, Science and English*** *2015*
**Grace Hopper Conference Scholarship** *2020, 2021*
**International Finalist CERN Webfest** *2020*
Ranked 7 globally in a pool of 400 contestants and ranked first in the "Disrupting Healthcare" section with our innovative solution.
**Nutanix International Women in Tech** *2020*
Was awarded a research grant of 10,000 USD for leveraging deep learning techniques to improve the efficiency and performance for cloud based systems for Nutanix.