Sushruti Mishra

+1 (718) 313-8596 https://mishrasushruti99.github.io 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 *3.7/4.0*
- 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) *9.02/10*
- 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, Kafka, GCP, Kubernetes. **Libraries** : Pytorch, Tensorflow, Keras, (NLTK, Pandas, Numpy, Sklearn, Matplotlib)

WORK EXPERIENCE

SEI Investments | ***SWE - Machine Learning*** *July 2023 -*
- Coordinated with configuration management team for development of a contextually aware chatbot built on transformers.

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%

FICCI — Machine learning engineer intern | ***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

University Delhi | **Research scholar** | ***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%

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.

Xceedance : Berkshire Hathaway *SDE Intern* *May 2019 - July 2019*
- Designed and developed a resume management system composed of a resume parser using NER and recommender system using Doc2Vec and Cosine Similarity; easing the Talent Acquisition team's burden in reading and summarising resumes for recruitment.

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

Vid-synchronize *Jan 2022 - Feb 2022*
- Vid-synchronize is a real-time online video synchronization platform with chat functionality which allows remote users to host, and watch videos synchronously in user created rooms.

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.

Polycystic ovary syndrome detection | ***Prof. Maria Unda*** *Sep 2022 - Dec 2022*
- Developed a machine learning model and using decision tree, Support Vector Machine, logistic regression, random forest and KNN techniques further supplemented XGBoost and Catboost to enhance their performance to predict whether or not a woman had PCOS disease or not. Self collected the data from various sources (WHO etc) and cleaned it.

Big Data Analytics for Amazon Customer Reviews *Sep 2022 - Dec 2022*
- Performed analysis using Hadoop MapReduce, chaining, secondary sorting, joins, binning, summarization and filtration patterns. Also made use of Apache Pig to perform some analysis along with Mahout for recommendations.

Concurrent traffic simulation *Sep 2019 - Dec 2019*
- Built thread-safe communication protocol for real time traffic simulation using mutexes, locks, and message queues

Forecasting Travel Claims using machine learning *Feb 2022 - May 2022*
- Forecasted travel claims using MLP (without address data imbalance, with SMOTE and SMOTE coupled with Adaboost) on 11 key parameters with an accuracy of 78

Driver Drowsiness *Sep 2020 - Dec 2020*
- Detected driver drowsiness using OpenCV for blink detection and applied statistical methods

Hate Speech detection | ***Prof. Anil Singh Parihar*** *Sep 2020 - Dec 2020*
- Developed and designed a hate speech detection system hosted on Flask and using Multinomial naive Bayes, Support Vector Machine, logistic regression, random forest and KNN techniques further supplemented by AdaBoost.

Fake review detection *Sep 2022 - Dec 2022*
- Detected fake reviews using SVM and multinomial naive Bayes were chosen as classifiers, and model tuning was done using two distinct vectorizers, Count Vectorizer and TF-IDF Vectorizers. Overall, all of the trained models had an accuracy rate of 80%, indicating that the vectorizers functioned admirably and that there are distinctions between false and actual reviews. Out of the two, the count vectorizer improved the models' performance more, and , LR performed the best, with an accuracy rate of 85% and a recall rate of 92%.

Phishing website detection *Feb 2020 - May 2020*
- Developed a machine learning model and using decision tree, Support Vector Machine, logistic regression, random forest and MLP techniques further supplemented XGBoost to enhance their performance to detect whether or not a site is a phishing site or not. Self collected the data from various sources (Phishtank, UNB etc) and cleaned it while extracting 3 key features from data.

Context aware Chatbot *Sep 2023*
- Developed a real time PDF intake chatbot using tensorflow which is contextually aware, which uses huggingface transformers and is deployed on streamlit. Also tried different architectures, such as seq2seq, distilBERT etc. The PDF ingestion just takes 0.1 seconds, which shall increase the knowledge transfer by a huge margin.

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 ML *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.