

SIDHARTH PUROHIT

+1 (347) 513-6761 • sp6365@nyu.edu • [LinkedIn](#) • [Github](#)

EDUCATION

New York University Tandon School of Engineering | *Master of Science in Computer Science* **Brooklyn, NY, USA** | May 2023
Relevant Coursework: Machine Learning, Cloud Computing, Design and Analysis of Algorithms

Kalinga Institute of Industrial Technology | *B.Tech. in Computer Science & Engineering* **Bhubaneswar, India** | May 2021
Relevant Coursework: Data Structures and Algorithms, Statistics and Linear Algebra, Database Systems, Machine Learning, Data Analytics, Data Mining and Warehousing, Discrete Mathematics, Industry Audit Management

SKILLS

Programming Languages: Python, SQL, C++, C, JavaScript, Java, R

Tools and Libraries: Amazon Web Services, Flask, NumPy, Pandas, Keras, PyTorch, TensorFlow, SciPy, scikit-learn, OpenCV, NLTK, Matplotlib, Git, Jupyter, Spark, Kafka, Hadoop, Django, Bootstrap, No SQL Databases, Azure

Data Science Techniques: MLOps, Data Wrangling, Data Warehousing, Database design, ETL, Data Visualization, Deep Learning, Natural Language Processing, Mathematical Modelling

EXPERIENCE

GENIE CETL | *Software Engineering Research Intern* **Leicester, UK** | July 2020 - Dec 2020

- Member of the Technical Staff on COVID-19 computational analysis project under the supervision of DNA fingerprinting lab, University of Leicester, UK.
- Responsible for end to end design of the “Cosine Comparison” feature, from design documentation to unit and integration testing.
- Implemented NLP aided web application using AWS Codepipeline, CloudFormation, flask, Java, and JSON scripts which resulted in reducing the manual workload by 40%.
- Used Selenium to access data from NCBI (National Center for Biotechnology Information) web database. The simulation resulted in an efficiency improvement by 65% on the entire project.
- Built the UI components using Flask, Bootstrap, and developed RESTful APIs for database management using Python.
- Deployed application over AWS S3 bucket & integrated with front-end using javascript backed AWS SDK and API Gateway.

Mitacs Globalink | *Graduate Research Intern* **Edmonton, Alberta, Canada** | May 2020 - Jul 2020

- Selected among the top **150 undergraduate students** around the globe for the Graduate Research Internship (GRI) program.
- Led a team of 5 in academic Human Computer Interaction backed research at MacEwan University & University of Alberta, Canada.
- The significance of the project is to aid instructors in the assessment of students’ assignments and exams which can be used in a manner similar to a scantron(OMR) which uses only visual input.
- Created a prototype for analyzing complex Calculus problems in Mathematics using OpenCV, flask, Convolution Neural Networks.
- The simulation resulted in an overall accuracy of over 85% using various supervised Machine learning algorithms.

ACADEMIC AND INDUSTRIAL PROJECTS

DNA Sequence analysis Engine | Python, NLP, AWS Lambda, Sagemaker, DynamoDB, S3, API Gateway **Nov 2021 - Dec 2021**

- Created a distributed load balanced web application for DNA sequence analysis of COVID-19 pandemic.
- Built and tested a NLP aided algorithm to display relative similarity between any two DNA sequences on real time data.
- Enabled WhatsApp APIs to implement appointment booking dashboard for COVID-19 detection. Link: [Forward HealthTechSystem](#)

Automatic Email Spam tagging system | Machine Learning, MLOps, AWS-Sagemaker, SES, Codepipeline **Nov 2021 - Dec 2021**

- Created a machine learning classification model on AWS-SageMaker to predict whether an email is spam or not.
- Implemented MLOps operations using AWS-CloudFormation and Codepipeline on the entire system. Link: [Spam tagging system](#)

Risk Profiling and Tolerance for Investor | Machine Learning, Computational Finance, Stats modelling **Nov 2021 - Dec 2021**

- Created a Machine learning model for Portfolio management process and also to quantify the behavioral biases of investors.
- Implemented Linear Regression, Lasso, Elastic Net, KNN, CART, SVR on real time data. Link: [Investor Risk Tolerance](#)

Alleviate: Coping mental health via Web Application | Flask, SQL, Node.js, NLP, AngularJS, Heroku **Oct 2019 - Apr 2020**

- Built a web dashboard using AngularJS and deployed a NLP backed ML model at back-end to help users deal with Mental health.
- Used real time emotions data from Kaggle to help users counter depression and maintain good mental health. Link: [Alleviate](#)

PUBLICATIONS

- Purohit, S., Satapathy, S.C., Sibi Chakkaravarthy, S. et al. Correlation-Based Analysis of COVID-19 Virus Genome Versus Other Fatal Virus Genomes. Arab J Sci Eng (2021). <https://doi.org/10.1007/s13369-021-05811-4>