

# Shivani Tyagi

Data Science Enthusiast | Author

[linkedin.com/in/shivani-tyagi-09](https://www.linkedin.com/in/shivani-tyagi-09) | [github.com/SHIVITG](https://github.com/SHIVITG) (Github) | [shivitg.github.io](https://shivitg.github.io)

DOB: 09/11/1994

E-mail: [shivani.tg94@yahoo.com](mailto:shivani.tg94@yahoo.com) Phone:  
8130772899

## EXPERIENCE & INTERNSHIPS

### Unigreate Infotech Inc. , Bengaluru, Karnataka, India— *Data Scientist*

FEB 2021 - PRESENT

#### Role & Responsibilities :

- Working on handling data extraction problems ( Working on both Python and Java development).
- Handling python development environment setup.
- Bash scripting for development sanity check.
- Involved in client discussion to analyse client requirements.
- Actively working on Computer Vision.

### Scanta Inc. , Gurugram, Haryana — *Machine Learning Engineer*

SEPT 2019 - FEB 2021

#### Role & Responsibilities :

- Worked on data dashboard generation by analysing virtual assistants requests and responses to detect anomalies in conversations and report malicious events.
- Setup the pipeline for basic NLP preprocessing like text cleaning, tokenization, generating bags of words, evaluating n-grams. Evaluate similarity between text using cosine similarity, and much more.
- Used Kafka nodes as message brokers, capsulating modules into docker containers.
- Used SQL Server as primary database and redis as in memory database for superfast data fetching.
- Experimented with Uber AI's Plug Play Language Model to induce personalities in text.
- Responsible for development of a paraphrasing tool using Transformers.
- Responsible for end to end engineering on NLP products pipelines from data mining, data cleaning , to modelling and deployment on cloud.
- Managing the intern group and guiding them throughout the internship duration.
- Major tech stack used - python3, hugging face tokenizers, redis, nltk, sklearn, Docker, flask,html and more.

### Centre for Development of Advanced Computing Cyber Security, Noida,Uttar Pradesh — *Internship* (DEC 2018 - JUN 2019)

- Internship on Cyber Security & ML Application under joint program of Microsoft and DSCI.

## EDUCATION

### Jaypee Institute of Information Technology, Noida, U.P— *M.TECH*

2017 - 2019

- Branch: Computer Science & Engineering with Machine Learning as specialization , Secured an overall score of 8.3 CGPA
- Published two research papers and a chapter on findings related to the Imbalanced Data Classification Problem in Machine Learning.

### RKGIT, Ghaziabad — *B.TECH*

2013 - 2017

- Branch: Computer Science & Engineering , Secured 75.4%

## PROJECTS

## - Object Detection Model

Trained a convolutional neural network model in order to detect labels for multiple objects from given labelled image data.

## - ML Engine

Conversational analysis on conversation datasets applied some NLP strategies to perform domain-specific tasks for conversational AI, majorly focusing on security aspects. And raise flags based on low priority or high priority by categorizing anomalies in text.

## - Text Data Preprocessing Pipeline (Natural Language Processing)

A pipeline to automate tasks like tokenization, lemmatization, and various other NLP preprocessing tasks by setting up a flexible sequential system for performing the preprocessing tasks that earlier need to be done manually one after another, based on requirements.

## - Text Similarity Model

Tracking all the aspects of similarity between two inputs.

## -Human Vs AI Model: for text differentiation

## -Security QA Model

Model to automate the security authentication system for Virtual Assistant Platform.

## -Data Augmentation Tool

## -Other Minor Projects

- Network Anomaly Detection using Machine Learning- LINUX, Wireshark & Python
- Proposed a new undersampling approach "NOL- Neighbors on Line"-R & Python
- Working with Resampling Techniques to handle imbalanced dataset-Python
- Performance Evaluation of ML algorithms for Credit Card Fraud Detection-Python
- Data Mining Project About Recognizing Human Activity-Python
- Website Sample For AICTE-Final Year BTech Project-Java

## PUBLISHED WORK

- Sampling Approaches for Imbalanced Data Classification Problem in Machine Learning . (ICRIC-2019)- Springer
- Computational Techniques for Real Time Credit Card Fraud Detection. Jan 2019 CNCS2019
- Performance Evaluation of Machine Learning Algorithms for Credit Card Fraud Detection. Confluence- 2019 IEEE
- Neighbors On Line (NoL): An Approach to Balance Skewed Datasets. Dec' 2019 ICIMMI-2019 Springer

## TECHNICAL SKILLS

- **Programming languages:** Python, SQL, R,Java, sqlite3
- **Other framework experience:** Numpy, Pandas, Flask, Scikit-learn, Keras, nltk, spacy, tokenizers
- **Deployment experience:** Docker, Redis, EC2, API-Gateway, AWS Lambda
- **Web Technologies:** HTML, XML and CSS
- **Cyber Security:** Networking, System Fundamentals of LINUX & WINDOWS, Cyber Forensic, Web Server & Application Security, Cryptography & Network Security, Auditing.
- **Visualization Tools:** TABLEAU and XMDVTool