

Siddhant Mahurkar

☎ (+91) 7038244499 | ✉ sid.mahurkar@outlook.com | 📍 sidmahurkar | 🌐 siddhantmahurkar | 📧 Siddhant Mahurkar

Education

Vellore Institute of Technology (VIT) Vellore

Vellore, India

B.TECH IN COMPUTER SCIENCE & ENGINEERING

July 2017 - Present

- **CGPA:** 9.05/10.00
- **Selected Coursework:** Multivariate Calculus; Linear Algebra; Differential & Difference Equations; Probability & Statistics; Natural Language Processing; Image Processing; Machine Learning; Artificial Intelligence; Discrete Mathematics & Graph Theory; Database Management Systems; Data Structures & Algorithms; Java Programming; Object Oriented Programming; Software Engineering.

Work Experience

FactEntry Data Solutions

Vellore, India

DATA SCIENCE INTERN

Jan. 2020 - Jun. 2020

- Worked on different Machine Learning based approaches for feature extraction from corporate bonds.
- Built a pipeline for automating the process of extracting First coupon date and Issue currency related to bonds.
- Implemented an AI Model for bonds in different languages and in text and OCR format to give required fields as output.
- **Supervised By:** Prof. Delhi Babu R (Dept. of CSE, VIT Vellore)

FindMind Analytics

VIT Vellore, India

MACHINE LEARNING INTERN

Feb. 2020 - May. 2020

- Studied and worked on elastic search engine using BERT for legal professionals to increase output accuracy and reduce the time of to get results.
- Built a convolutional neural network model to mask specific fields on ADHAR Identification card.
- Built a web scraper for Indian Tribunal sites for obtaining information on legal issues.
- **Supervised By:** Arpit Jain (COO, FindMind Analytics)

Verzeo

Bangalore, India

MACHINE LEARNING INTERN

Jun. 2019 - Jul. 2019

- Studied and summarized various machine learning algorithms and techniques to tackle real world problems.
- Implemented Text Summarizers using various deep learning and graph-based algorithms.
- Built an object detection model using YOLO algorithm.

Publications

LRG at SemEval-2020 Task 7: Assessing the Ability of BERT and Derivative Models to Perform Short-Edits based Humor Grading

Workshop Paper

SUBMITTED TO SEMEVAL-2020

2020

Siddhant Mahurkar, Rajaswa Patil

Attributional Analysis of Multi-Modal Fake News Detection Models

Conference Paper

PRESENTED AT IEEE BigMM-2020

2020

Shashank Madhusudhan, Siddhant Mahurkar, Suresh Kumar Nagarajan

Citta: A Lite Semantic Recommendation Framework for Digital Libraries

Conference Paper

PRESENTED AT NDLI-UNESCO KNOWLEDGE ENGINEERING FOR DIGITAL LIBRARIES-2019

2019

Rajaswa Patil, Siddhant Mahurkar

Community Service

NSS (National Service Scheme NGO)

VIT, Vellore

CORE MEMBER

Nov. 2018 - Jan. 2020

- Participated in volunteer work for the NSS NGO during my undergraduate.
- Carried in ventures such as food distribution to poor, tree plantation and awareness drives.
- Organized events to commemorate 20th anniversary of Kargil War's Heroes.

Honors & Awards

2019 **Best Poster Award**, Won the award at NDLI-UNESCO KEDL-19 for poster presentation on "AI in Digital Libraries" among 35 participants

IIT Delhi, India

2019 **Finalist**, "VIT Hack" Hackathon organised during graVITas 19, Technical Fest of VIT

VIT Vellore, India

Projects

Semantic Search Framework for Digital Libraries

NLP & IR

FRAMEWORK/WEB APPLICATION

Nov. 2019 - Dec. 2019

- Developed a semantic search framework for Digital Libraries with a BERT-based approach.
- Built a Python-Flask based web-app to deploy the model and try in real-time.
- **URL:** <https://github.com/sidmahurkar/citta-app>

Fake News Detection Using Transformer Based Model

NLP

FRAMEWORK

Sept. 2019 - Oct. 2019

- Developed a transformer based model using BERT to classify a piece of news as potentially fake or real.
- Data-set gathered from around 5 different sources to include various domains. (Sports, Celebrity, Political etc.)
- **URL:** <https://github.com/sidmahurkar/Fake-News-Detection>

Exploratory Data Analysis on Abrogation of Article 370 Tweets

NLP & IR

FRAMEWORK

Aug. 2019 - Sept. 2019

- Built a Twitter-API based model to extract tweets about revocation of Article 370 & 35A.
- Around 300K Tweets with #Article370 & #Article35A were collected during 5 Aug - 15 Aug 2019.
- Extracted tweets analyzed for sentiment, tweets/hour and other valuable insights.
- **URL:** <https://github.com/sidmahurkar/Sentiment-Analysis-on-Abrogation-of-Article-370>

Indoor Scenery Image Classification

CV

FRAMEWORK

Aug. 2019 - Sept. 2019

- Designed a CNN based framework to classify different indoor house images into 67 different classes.
- Used a fine-tuned ResNet-152 to achieve the highest accuracy for the task.
- **URL:** <https://github.com/sidmahurkar/Image-Classification-on-Indoor-Scenery-Images>

Extracurricular Activity

LRG (Language Research Group)

BITS, Goa

CORE MEMBER

Jun. 2019 - Present

- A research group focused on Natural Language Processing and Speech consisting of students and professors from various universities.
- We attended various conferences that allowed to interact with researchers and remain abreast with the latest technology.
- Published two papers in collaboration with other members.

Skills

Languages Python, R, C, C++, Java

Frameworks PyTorch, Keras, Tensorflow, SpaCy, Pandas, Flask, NLTK

Operating Systems Windows, Linux-Ubunu

Spoken Languages English, Hindi, Marathi

Research Interests

Social Media Visualising different social media platforms for insights about a public decision, Abuse detection.

Computer Vision Real-time Object Detection, Malady classification in healthcare.

Natural Language Processing Fake news detection, Gender bias in NLP, Sentiment analysis.