

Nikhil Singh

NLP Graduate Student @ UCSC

☎ +1 (408) 334 1228 | ✉ nsingh58@ucsc.edu | 🌐 nikhilbyte | in nikhil-singh-8235b728a | 🕒 100am | 📍 Santa Clara, CA, USA | nikhilbyte.github.io

Education

University of California, Santa Cruz

M.S. in Natural Language Processing

Santa Clara, California

Sep. 2023 - Mar. 2025

Manipal University Jaipur

B.Tech. in Information Technology

Jaipur, Rajasthan

Jul. 2017 - Aug. 2021

Technical Skills

Programming Python

Frameworks PyTorch, TensorFlow, JAX, Scikit-Learn

Typesetting Google Suite, Office, L^AT_EX, Libre

Languages Hindi (Native), English (Professional)

Work History

BetterPlace

Data Scientist Intern

Bangalore, KA, India

Nov 2022 - Apr 2023

- Developed a Background Verification Tool, scraping 65 million court records from the National Judicial Data Grid with Selenium, and securely storing them in AWS's S3 bucket.
- Managed Databricks clusters using PySpark for efficient storage and rapid retrieval of court records
- Significantly reduced retrieval time through smart NLP techniques, enhancing background verification efficiency
- Expertise in scraper development, data synthesis, SQL and AWS Python SDK for data management

IIT Bombay

Machine Learning Research Intern

Mumbai, MH, India

Jul. 2021 - Aug. 2022

- Developed a Post-OCR model for the 'Akshar Anveshini' project, effectively correcting errors in OCR'd Sanskrit texts.
- Experimented with various sequence-to-sequence machine learning models, including traditional RNN-based encoder-decoder and transformer-based text-to-text generation models, showcasing strong machine learning skills
- Proposed and implemented a novel encoding technique and byte-level tokenization method, surpassing performance benchmarks set by baseline models.
- Contributed to research findings that got accepted at EMNLP (Empirical Methods in Natural Language Processing), highlighting the impact of our work in the field.
- Collaborated closely with a renowned professor and two post-doctoral researchers, gaining valuable experience in project management and research methodologies.
- Played a crucial role in digitizing Sanskrit books and improving the accuracy of OCR models

The University of Edinburgh

Research Intern

Remote

Dec. 2020 - Jun. 2021

- Led "Safety & Sustainability in Micro-Mobility" project, creating a multi-criteria route algorithm for cyclists and pedestrians, balancing safety and efficiency.
- Implemented a novel algorithm to modify A* route planning, contributing to advancements in mapping services.
- Managed COVID-19 hotspot data, converting it into usable coordinates with Google Geocoding API, and developed the web application backend
- Published project research in MDPI's ISPRS International Journal of Geo-Information.
- Enhanced coordination and collaboration skills while working closely with front-end developers, gaining expertise in version control and web development.
- Developed expertise in data analytics and interpretation during the internship.

iHub IT Solutions

Data Science Intern

Kurnad, KA, India

Jun. 2020 - Aug. 2020

- Created a Conversational bot using RASA and Google DialogFlow.
- Embedded the chatbot into a Django based webapp

UniConverge Technologies Pvt. Ltd. and The IoT Academy

Machine Learning Intern

Noida, UP, India

Apr. 2020 - May. 2020

- Created a machine learning model that could differentiate between toxic and normal text in multiple languages.

Publications

- [1] Sumit Mishra, **Nikhil Singh**, Devanjan Bhattacharya, et al. “Application-Based COVID-19 Micro-Mobility Solution for Safe and Smart Navigation in Pandemics” *ISPRS, International Journal of Geo-Informatics* [[MDPI](#)]
- [2] Ayush Maheshwari, **Nikhil Singh**, Amrith Krishna, Ganesh Ramakrishnan “A Benchmark and Dataset for Post-OCR text correction in Sanskrit” *Findings of the Association for Computational Linguistics : EMNLP 2022* [[ACL Anthology](#)]
- [3] Shikha Mundra, **Nikhil Singh**, Namita Mittal “Fine-tune BERT to Classify Hate Speech in Hindi English Code-Mixed Text” *FIRE 2021 CEUR* [[CEUR WS](#)]
- [4] **Nikhil Singh** “niksss at SemEval-2022 Task 6: Are Traditionally Pre-Trained Contextual Embeddings Enough for Detecting Intended Sarcasm ?” *Proceedings of SemEval 2022 Workshop @ NAACL* [[ACL Anthology](#)]
- [5] **Nikhil Singh** “niksss at SemEval-2022 Task7:Transformers for Grading the Clarifications on Instructional Texts” *Proceedings of the SemEval 2022 Workshop @ NAACL* [[ACL Anthology](#)]
- [6] **Nikhil Singh** “niksss at Qur’an QA 2022: A Heavily Optimized BERT Based Model for Answering Questions from the Holy Qu’ran” *Proceedings of the OSACT 2022 Workshop @ LREC 2022* [[LREC Conferences](#)]
- [7] **Nikhil Singh** “Niksss At Hinglisheval: Language-Agnostic Bert-Based Contextual Embeddings With Catboost For Quality Evaluation Of The Low-Resource Synthetically Generated Code-Mixed Hinglish Text” *15th International Natural Language Generation Conference 2022* [[INLG](#)]
- [8] Sumit Mishra, Devanjan Bhattacharya, Atanshi Chaturvedi , **Nikhil Singh** “Assessing micro-mobility services in pandemics for studying 10-minutes cities concept in India using geospatial data analysis: an application” *IWCTS '22: Proceedings of the 15th ACM SIGSPATIAL International Workshop on Computational Transportation Science* [[ACM Digital Library](#)]

Awards and Honors

May. 2022

Contest: Special Prize in AI4Science Hackathon Hosted by CalTech [[Winners](#)][[Project](#)]

Online

Competitions

- 4th Position in Text Paraphrase Competition hosted by Syracuse University (CRI-COMP 2022). [[Leaderboard](#)]
- 1st Position in Sub-Task B of Quality Evaluation of the Low-Resource Synthetically Generated Code-Mixed Hinglish Text. [[Leaderboard](#)]
- 8th Position in Mozilla Luganda Automatic Speech Recognition . [[Leaderboard](#)]
- 4th Position In Enzyme Classification Competition hosted by Instadeep, London. [[Leaderboard](#)]
- 5th Position in WNUT-2020 Task 2 @ EMNLP 2020. [[Leaderboard](#)]