

# NIKHIL BALWANI

✉ nikhil.balwani@columbia.edu    nikhilbalwani    nikhilbalwani    nikhil.info   📞 +1 (646) 431 7195

## Education

### Columbia University

Aug 2022 – Dec 2023

*Master of Science in Computer Science*

- Relevant Courses: High-Performance ML, Natural Language Processing, Conversational AI

### Ahmedabad University

Sep 2016 – May 2020

*Bachelor of Technology in Information and Communication Technology with Summa Cum Laude*

CGPA: 3.60 / 4.00

- Relevant Courses: Data Structures and Algorithms, Software Engineering, Machine Learning, Artificial Intelligence
- Honors: Merit - Full Tuition Fee Waiver (2019, 2020), Scholastic Distinction with “Highest Excellence” (first rank holder)

## Research Interests

Deep Learning (DL), Natural Language Processing (NLP), Time Series, Statistical Machine Learning (ML)

## Experience

### Infocusp Innovations Private Limited (Client: Google X)

Jan 2021 – Jun 2022

*Machine Learning Engineer, Confidential Project*

Ahmedabad, India

- Wrote scalable big data scripts in Apache Beam on Cloud Dataflow runner to process 100 million+ samples at scale.
- Created a document similarity engine API based on SimHash and Multi-Indexed Hashing that can search 21 million documents in under 150 ms.
- Automated an end-to-end solution for evaluation of 50+ Transformer models on Google Cloud Virtual Machines.
- Managed the complete life cycle of 50+ Transformer (Deep Learning, Natural Language Processing) models - data preparation, model creation, and training on Google Cloud TPUs, and deployment of APIs on GCP Virtual Machines.
- Enhanced the accuracy of Transformer models by 30% points using a structure-aware attention mechanism for translation tasks. Exact match accuracy increased by 4%.

### Embibe - AI in Education

Dec 2019 – Jan 2021

*Data Scientist*

Bengaluru, India

- Implemented the organization's first Knowledge Tracing model called Bayesian Knowledge Tracing (BKT) based on an HMM (Statistical Machine Learning) - the backbone of two different APIs.
- Architected an end-to-end data pipeline for BKT models to automatically update concept mastery scores using Spark.
- Developed eGo - a simulation engine API based on pre-trained BKT models. The simulations of student behavior helped uncover corner cases and bugs in two upcoming products.
- Trained an LSTM model for concept mastery - which led to an AUC performance gain of 0.21 on the validation set.
- Implemented an in-depth, scalable, and reproducible analysis of 10 million+ student attempts for Test-on-Test student performance and concept mastery improvement.

*Intern, Data Science Lab*

Bengaluru, India


- Wrote an internal research article capturing 15 different Knowledge Tracing approaches in the literature - classified under Bayesian and Non-Bayesian techniques - which proved beneficial for new joiners for a comprehensive study.

### University of Liverpool - Ahmedabad University

Apr 2018 – Nov 2018

*Undergraduate Student Researcher, Project DST-UKIERI*

Ahmedabad, Gujarat

- Trained a customized LSTM network (Time Series Modeling) to leverage the temporal correlation in signal data in Cognitive Radio to improve the detection performance by 0.16. [Publication 

## Technical Skills

**Languages**: Python, Java, SQL, C, C++, R, HTML/CSS, L<sup>A</sup>T<sub>E</sub>X, JavaScript, Bash.

**Developer Tools**: Cider/Cider-V (G3), Cloud Shell, VS Code, Eclipse, Google Cloud Platform, Jupyter Notebooks, Git.

**Technologies/Frameworks**: Flume (G3), Critique (G3), TPUs (GCP), DataFlow (GCP), Compute-Engine (GCP), Tensorflow, NumPy, Flask, Tornado Web Server, RESTful APIs, Apache Beam, Spark, MATLAB.

**Certifications**: Machine Learning (Coursera), Deep Learning Specialization (Coursera).

## Projects

### MLify - Machine Learning from Scratch | Python, Numpy

- Bare-metal implementations of some common supervised classifiers (Feed-Forward Neural Network, Decision Tree, Random Forest) and unsupervised clustering techniques (K-means, Gaussian Mixture Model using EM).

### Face and Expression Recognition | TensorFlow, Keras, Pandas, OpenCV, NumPy, Python, Computer Vision

- A comparison of neural and non-neural network-based approaches for the problem of face and expression recognition.

## Leadership and Achievements

- Interviewed 10+ candidates for entry-level Software Engineering and Machine Learning roles (Infocusp).
- Delivered several talks: “Apache Beam Programming” (2021, Infocusp), “CLIP Loss” (2021, Infocusp), “Knowledge Tracing” (2019, Embibe), and “Machine Learning for Cognitive Radio” (2018, Ahmedabad University).