Nikhil Balwani

nikhil.balwani@columbia.edu nikhilbalwani nikhilbalwani nikhil.info

Education

Columbia University in the City of New York **Aug 2022 – Dec 2023**
Master of Science in Computer Science

Ahmedabad University **Sep 2016 – May 2020**
Bachelor of Technology in Information and Communication Technology with Summa Cum Laude *CGPA: 3.60 / 4.00*

Experience

Google X via Infocusp Innovations Private Limited **Jan 2021 – Jun 2022**
Machine Learning Engineer, Confidential Project *Ahmedabad, Gujarat*

- Responsible for the complete life cycle of **Transformer** models - **Data preparation**, **model creation and training** on **Google Cloud TPUs** and **deployment** of **APIs** on **GCP Compute Engine - Virtual Machines**.
- Improved on the baseline accuracy of **Transformer** models by **30% points** using a hybrid **Attention Mechanism** with relative position representations for translation task. **Exact match accuracy** increased by **4%**.
- Wrote scalable big-data scripts in **Apache Beam** on **Google Cloud Platform (GCP)** for automatic dataset curation. Fetched over **100 million+** training samples which led to an increase in translation performance.
- Reduced the manual effort involved in the evaluation of models by setting up an **automatic evaluation pipeline**.
- Developed a document similarity engine based on **SimHash** and **Multi-Indexed Hashing** that can search **21 million** documents in under **150 ms**.

Embibe, Indiavidual Learning Private Limited **Dec 2019 – Jan 2021**
Data Scientist *Bengaluru, Karnataka*

- Developed the organization's first **Knowledge Tracing** model called **Bayesian Knowledge Tracing (BKT)** based on **Hidden Markov Model (HMM)** - backbone of various downstream products.
- Trained an **LSTM** model for concept mastery - led to an **AUC performance gain of 0.21** on the validation set.
- Set-up an end-to-end **data pipeline** for **BKT** to automatically update concept mastery scores using **PySpark**.
- Worked on an **in-depth, scalable and reproducible** analysis of **10 million+ student attempts** for Test-on-Test student performance and concept mastery improvement. The report that was sent to the **National Testing Agency**, the organization which conducts national level entrance examinations - **JEE** and **NEET** - in India.
- Developed **eGo** - a **Simulation Engine** based on **pre-trained BKT models**. The simulations of student behavior helped uncover corner cases and bugs in **two upcoming products**.

Intern, Data Science Lab *Bengaluru, Karnataka*

- Conducted several talks on applications of **Knowledge Tracing** to familiarize the team with the literature.
- Wrote an internal **Survey Article** on different Knowledge Tracing approaches in the literature - classified under **Bayesian** and **Non-Bayesian** approaches - which proved beneficial for new joiners for a comprehensive study.

University of Liverpool - Ahmedabad University **Apr 2018 – Nov 2018**
Undergraduate Research Intern, Project DST-UKIERI *Ahmedabad, Gujarat*

- Project title: **"Non-parametric Smart Sensing Analytics based on Large Spectrum Data and Estimation of Channel Activity Statistics"** as part of **DST-UKIERI**, a UK-India Education and Research Initiative.
- Trained a customized **LSTM network** to leverage the **temporal correlation** in signal data in **Cognitive Radio** to improve the detection performance. For FM broadcasting, the detection probability increased by **0.16**.

Technical Skills

Languages: Python, Java, SQL, C, C++, R, HTML/CSS, LaTeX.
Developer Tools: VS Code, Eclipse, Google Cloud Platform, Jupyter Notebooks.
Technologies/Frameworks: Tensorflow, NumPy, Flask, Google Cloud TPUs, Google Computer Engine, Tornado Web Server, Git, RESTful APIs, Apache Beam.
Certifications: Machine Learning (Coursera), Deep Learning Specialization (Coursera).

Publications

- N. N. Balwani, D. K. Patel, B. Soni, and M. Lopez-Benitez **Long Short-Term memory-based spectrum sensing scheme for cognitive radio**, 2019 IEEE 30th Ann. Int. Sym. on PIMRC, Istanbul, Turkey, Sep. 2019.

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**).

Leadership and Achievements

- **Gold Medal** - Awarded scholastic distinction with "Highest Excellence" for 2020 batch - Ahmedabad University.
- **Scholarship** - Full tuition fee waiver, for being the first rank holder - 2019 and 2020 - Ahmedabad University.
- **Speaker** - Faculty Development Programme (2018) - "Advanced 5G Wireless Communications" - conducted a talk on applications of Artificial Neural Networks for Spectrum Sensing in Cognitive Radio.