Nihar Sanda

+91-8097283441 sanda.n@northeastern.edu nihar-sanda-07ab6118b koolgax99 koolgax99.github.io

EDUCATION

Northeastern University - Khoury College of Computer Sciences **2023 - 2025 (Expected)**
MS in Computer Science *Boston, MA, USA*

Indian Institute of Information Technology, Dharwad **2019 - 2023**
Bachelor of Technology, Computer Science and Engineering *Karnataka, India*

TECHNICAL SKILLS

Languages: Python, R, Java, HTML, CSS, C++, C, JavaScript, Typescript, OWL, RDF
Libraries: Numpy, PyTorch Matplotlib, Seaborn, Scikit, Pandas, Keras, TensorFlow, Dlib, SciPy
Technologies: Node, React, RESTful services, Flask, Django, Docker, Jenkins, CI/CD, R Shiny, AWS, GCP
Database: SQLite, PostgreSQL, MySQL, MS SQ, MongoDB, NoSQL, DynamoDB, Redis.

PROFESSIONAL EXPERIENCE

Indian Institute of Technology, Bombay **May 2023 – Present**
Research Associate *Mumbai, India*

- Scaling the **"Affect Aware Tutoring System Using Video Bots"** to learn **student engagement with adaptive feedback**, **improving accessibility** and **enhancing learning outcomes**.
- Demonstrating **leadership and project management skills** by overseeing the development of a high-performance **learning management system**, designed to handle concurrent usage by over **10,000 users**.
- Engaged in extensive research on **"Privacy Protection of Student Video Data in Diverse Learning Environments"** exploring innovative approaches to **safeguard student privacy** and confidentiality within various educational settings.

Patenti Technology Solutions **January 2023 - May 2023**
Software Engineering Intern *Bengaluru, India*

- **LookupTM Tool:** Creating a comprehensive IP infringement search tool by **utilizing OCR models** and employing **Django's Model-View-Template (MVT) architecture** to develop the web application with **MySQl database**.
- **Devops Operations:** Deployment and Maintenance of multiple **production and testing environments** on **AWS EC2, S3, and Redis**, utilizing **Github Actions** for **CI/CD automation.**
- **Backend Development:** Developed many backend modules and executed a **pipeline with LLMs for topic modeling and scientific term retrieval** in patent documents, **incorporating APIs** and **implementing RateLimiting** for seamless integration.

Google Summer of Code, PEcAn Project **2022 and 2023**
Student Intern *Bengaluru, India*

- Engineering the **APIs for the various PEcAn packages** of **data assimilation and meta-analysis** in the R language, contributing to its creation and enhancement.
- Enhanced the **authentication of the existing REST APIs** by incorporating robust **API Key authentication** and implementing efficient **rate-limiting features**.
- Leveraging **R Shiny** to create a **robust and user-friendly dashboard**, empowering users to generate dynamic **SDA (State Data Assimilation) and forecasting graphs** for various researchers around the world.

PROJECTS

Covid-19 Mortality Prediction | *Data Visualization, Machine Learning, Deep Learning* **May 2022**

- **Data analysis and processing:** Proficient in **cleaning, exploring, and transforming private hospital data** with **1 lakh data points** to better understand and extract information using **Pandas** and **Python**
- **Data Imbalance Mitigation:** Employed techniques such as SMOTE and undersampling to effectively tackle the challenge of imbalanced data.
- **Mortality Prediction:** Implemented **state-of-the-art machine learning models** to predict **COVID-19 patient mortality**, utilizing an ensemble approach and achieving an impressive **accuracy of 93%**

Protein Fold Recognition | *NLP, Transformers, Bio-Informatics* **August 2022**

- **Protein Fold Recognition** Implemented advanced **NLP techniques** to improve protein fold recognition for low similarity datasets such as **DD, EDD, TG, and SCOPe** baseline datasets encompassing diverse amino acid-based protein sequences and their corresponding folds.
- **Feature Extraction:** Extracted **features** by utilizing evolutionary **PSSM and HMM profiles** of protein sequences, and concatenating them with global **Convolutional** and **Skip Bi-gram features**.
- **Multiclass Classification:** Implemented **BERT and ESM by Meta transformer-based models** for classification and achieving an impressive accuracy **exceeding 93%** across all datasets, **surpassing the previous 85% accuracy.**

ACHIEVEMENTS AND LEADERSHIP

- Lead a team to the Grand Finals at Smart India Hackathon 2022 and also won many National Level hackathons
- 2 times Google Summer of Code (2022, 2023) Recipient PEcAn Project
- Open Source Contributor for Rucio(CERN), CircuitVerse, PEcAn Project with many accepted PRs.
- Founder and President at Velocity, Web Development Club of IIIT Dharwad • Captain of the University Tennis Team.

RESEARCH PAPERS AND PUBLICATIONS

- **Interestingness from COVID-19 Data: Ontology and Transformer-Based Methods** - ICON 2022, proceedings in ACL Anthology (SCI and SCOPUS Indexed)
- **Ontology-Based Semantic Data Interestingness Using BERT Models** - Taylor and Francis' Connection Science Journal (SCI and SCOPUS Indexed)