NIHAR SANDA

📞 +1 (617) 959-5688 ✉ sanda.n@northeastern.edu in niharsanda ○ 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*

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.

EXPERIENCE

Indian Institute of Technology, Bombay   **May 2023 – Sep 2023**
Research Associate   *Mumbai, India*
- Developed an innovative tutoring system, **"Affect Aware Tutoring System Using Video Bots"**, which includes a learning management system capturing click-stream log data and facial data to predict the **user's affect state** in real-time using an optimized transformer-based deep learning model trained on the DAiSEE dataset for 300 hours.
- 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   **Jan 2023 - May 2023**
Software Engineering Intern   *Bengaluru, India*
- Created 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**.
- Deployed and maintained multiple **production and testing environments** on **AWS EC2, S3, and Redis**, utilizing **Github Actions** for CI/CD automation.
- Developed numerous 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*
- Leveraged **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.
- Created the **REST APIs for the various PEcAn packages** of **data assimilation and meta-analysis** in the R language.
- Enhanced the **authentication of the existing REST APIs** by incorporating robust **API Key authentication** and implementing efficient **rate-limiting features**.

PROJECTS

Covid-19 Mortality Prediction | *Data Visualization, Machine Learning, Deep Learning*   **May 2022**
- Leveraged various data **cleaning, pre-processing methods and transformed private hospital data** with **1 lakh data points** to better understand and extract information using **Pandas** and **Python**
- Employed techniques such as **SMOTE and undersampling** to effectively tackle the challenge of imbalanced data.
- Implemented **SOTA machine learning and deep 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**
- Implemented advanced **NLP techniques** to improve protein fold recognition for low similarity baseline datasets such as **DD, EDD, TG, and SCOPe**, encompassing diverse amino acid-based **protein sequences** and their **corresponding folds**.
- Extracted **features** by utilizing evolutionary **PSSM and HMM profiles** of protein sequences, and concatenating them with global **Convolutional** and **Skip Bi-gram features**.
- 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

- Director's Gold Medal for the Best Outgoing Student at IIIT Dharwad for 2023 Batch
- 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 at 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

- **An Effective Framework for the Prediction of Protein Folds using Natural Language Processing and Evolutionary Features** - IEEE/ACM Transactions on Computational Biology and Bioinformatics
- **Interestingness from COVID-19 Data: Ontology and Transformer-Based Methods** - Proceedings in ACL Anthology
- **Ontology-Based Semantic Data Interestingness Using BERT Models** - Taylor and Francis' Connection Science Journal