

AMRUIT SAHOO

+91-63-7170-0699

amruit2k@gmail.com

linkedin.com/in/amruit-sahoo

github.com/AmruitW'@23

EDUCATION:

Indian Institute of Technology (IIT), Kharagpur, India

Interdisciplinary Dual Degree Programme (IDDP) 5 year

Bachelor of Technology - Electrical Engineering &

Master of Technology - Artificial Intelligence, Machine Learning & Applications

July 2019 - Present

CGPA: 8.39/10

SAI International School, Bhubaneswar, India

Class 12, Central Board of Secondary Education (CBSE)

July 2018 - June 2019

Percentage: 92.4%

DAV Public School Chandrasekharpur, Bhubaneswar, India

Class 10, Central Board of Secondary Education (CBSE)

July 2016 - June 2017

CGPA: 10/10

INTERNSHIP:

CarDekho, Girnarsoft

ML/AI Summer Intern

May 2022 - July 2022

Remote

- Used **Flask** to build a **parser** for extracting important information from motor policies of various companies
- Used **pdftotext** and **tabula.py** as text extraction softwares, to extract important key-value pairs in the raw PDFs
- Used **MongoDB (compass)** and **MySQL (phpMyAdmin)** as databases to store raw pdfs and extracted information
- Used **Google Vision API** as an OCR to parse JPEG documents for registration cards and Aadhar cards
- Tried replicating Google's **Document AI** paper, to automate the process of extracting important information from documents

RESEARCH EXPERIENCE:

CNeRG, IIT Kharagpur

under supervision of Prof. Animesh Mukherjee

Jun 2022 - Present

Kharagpur, West Bengal

- Built a **crawler** using **Selenium** to scrape user query data points containing useful information, from **Ubuntu Launchpad** site
- Worked on **retrieval of duplicate questions** on the AskUbuntu platform, utilizing both **text & network** based features
- Implemented **SBERT & Node2Vec** based MLP model with **Bi-Encoder** architecture trained on **AskUbuntu triplet dataset**
- Successfully **replicated** results of past papers on duplicate question detection: **DUPE & DupPredictor**
- Performed **literature review** of papers in domain of Deep Learning applied on Software Engineering & Code related tasks
- Worked on **generating answers** for new questions posted on **AskUbuntu** site, by **prompting Bloom & OpenAI API**
- Engineered **prompts** using top related questions retrieved by a **GAT + PageRank** model trained on **AskUbuntu** dataset

PUBLICATIONS:

- Rima Hazra, Debanjan Saha, **Amruit Sahoo**, Somnath Banerjee and Animesh Mukherjee. Duplicate Question Retrieval and Confirmation Time Prediction in Software Communities. *Communicated to SIGIR 2023.*

PROJECTS:

Learning Equality - Curriculum Recommendations | *Design Lab project under Prof. Plaban Bhowmick* **Feb 2023 - April 2023**

- Streamlined the process of **matching educational content with topics** in a curriculum, using **Kolibri** library as the dataset
- Implemented a **multi-lingual SBERT** model to generate the **initial embeddings** for topics and contents
- Implemented a **Node2Vec** based model to incorporate the structural information available in the **context tree of topics**
- Trained a **Bipartite Graph Neural Network** to learn the **final embeddings** for topics and contents

Social Media Engagement Prediction for News posts | *Course project under Prof. Adway Mitra*

Oct 2022 - Nov 2022

- Analyzed different news posts from various pages of news articles and used them to generate **engagement predictions**
- Observed an **LDA model** with 7 topics and a **coherence score of 0.606** as a suitable topic model
- Trained different models on the processed dataset to finalize a model with **40.9% improvement** over naive predictions

Predicting Future Population Distribution | *Course project under Prof. Adway Mitra*

Feb 2022 - April 2022

- Extracted **population data of India**, from the World Bank site, for years **1960 to 2021**, over several age groups and genders
- Implemented **LSTM** model to forecast the change in population distribution of India till 2050, for all age groups and genders
- Observed and reported the analytical results of the data, which can be helpful to find out solutions and **tackle the increasing population problems** over the years

HACKATHONS & COMPETITIONS:

- Winner of Source Code Finals**, a nation-wide competition organized by Kshitij, IIT Kharagpur.
- Obtained Global rank of **712 out of 11,000** contestants in Google Kistart Round D Coding competition.

AWARDS & ACHIEVEMENTS:

- Secured **All India Rank of 1030 (within top 0.1%)** among **1.4 million candidates** in JEE Advanced 2019 (**IIT-JEE**).
- Qualified National Standard Examination in **Physics** (2018) and National Standard Examination **Astronomy** (2017).
- Awarded **scholarship** for clearing the final round of National Talent Search Examination (**NTSE**) 2017, out of more than **1 million candidates (within top 0.1%)**.

SKILLS SUMMARY:

- Languages:** Python, C++, HTML **Frameworks:** Pytorch, Tensorflow, Keras, Flask
- Libraries:** Pandas, Numpy, Matplotlib, PyG, Selenium, pdftotext, tabula.py, SpaCy, Sci-kit learn, OpenCV, C++STL
- Software/Tools:** Git/Github, VSCode, Jupyter Notebook, Overleaf

COURSEWORK INFORMATION:

- University credit courses:**
Programming and Data Structures, Machine Learning Foundation & Applications, Artificial Intelligence Foundations & Applications, Deep Learning Foundations & Applications, Dependable & Secure AI/ML, Graphical and Generative Models, Linear Algebra for AI & ML, Mathematics-1, Mathematics-2, Probability & Stochastic Processes, English for communication.
- Coursera certification courses:** Machine Learning Specialization, DeepLearning Specialization (offered by DeepLearning.AI).

EXTRA CURRICULAR ACTIVITIES:

- Part of Gold winning Sketching & Cartooning team in Inter-Hall General Championship Social & Cultural Cup.
- Part of Gold winning team of Azad Hall of Residence in Inter-Hall Illumination.
- Part of Basketball & Western Musicals team of Azad Hall of Residence in Inter-Hall General Championship Cup.