# SARNAB BANERJEE  |  21MF10054

## MANUFACTURING SCIENCE & ENGG.(B.Tech 4Y)

## EDUCATION

| Year | Degree/Exam | Institute | CGPA/Marks |
|---|---|---|---|
| 2025 | B.TECH | IIT Kharagpur | 8.36 / 10 |
| 2021 | ISC | Julien Day School, Ganganagar | 99.5% |
| 2019 | ICSE | Julien Day School, Ganganagar | 97.6% |

## INTERNSHIPS

### Research Intern | IIM Udaipur | Prof. Samadrita Bhattacharya                [May 2023 - Present]
*Objective: Application of Machine Learning on Reviews of Local Businesses to predict their trajectory*
• Utilized **fuzzy-wuzzy** text similarity to classify **3** specific business domains from a vast pool of diverse business reviews
• Conducted **eWOM** analysis on these domains by preprocessing and analyzing the reviews of such businesses using **NLTK**
• Employed state-of-the-art models like **RoBERTa** and **SentiStrength** to analyze the changing **sentiments** of customer reviews
• Implemented a topic modeling pipeline using **BERTopic** and **LDA** to classify reviews into clusters and identify crucial themes

### Research Intern | Stanford University | Prof. Pascal Geldsetzer                [May 2023 - Present]
*Objective: Use of Machine learning to monitor health in low and middle-income countries*
• Performed **Exploratory Data Analysis** to visualize the data via **scatter plots** and **heatmap** by **matplotlib** and identify outliers
• Preprocessed the dataset by imputing the **NaN** values with **mean** and **mode**, and **one-hot encoding** the categorical features
• Used an ensemble of **XGBoost**, **CATBoost** and **Random Forest Regressor** models to predict values of **6** independent labels
• Improved the value of **RMSE** of test data by **14.94%** by tuning the parameters of the ensemble models using **GridSearchCV**

## PROJECTS

### AlgoZenith Summer Hackathon | DSA Search Engine                [May 2023]
• Scraped **3000+** algorithmic questions from three coding sites using **Selenium** and performed data preprocessing using **NLTK**
• Implemented **TF-IDF** algorithm and computed **Inverted Index** of the corpus by identifying keywords and counting frequencies
• Built a **Flask** app, displaying the top **20** similar questions with decreasing TF-IDF scores using a spawned Python process

### Open IIT Data Analytics Competition | Gender Bias Detection in Text                [Aug'2022 - Sept'2022]
• Classified individual sentences using a unique **Depth First Search** approach into known biased (0) and unbiased (1) labels
• Performed **Text Preprocessing** and **Text Cleaning** while keeping **pronouns** intact to identify **three** different kinds of biases
• Employed various text vectorization techniques and used an **ensemble** approach for text and sentence pair classification
• Improved **validation accuracy** of SVM by **11.63%** from **0.43** to **0.91** through hyperparameter tuning using **GridSearchCV**

### Pokemon Classification using ResNet-50 | Self Project                [May 2022]
• Implemented **Data Preprocessing** and **Data Augmentation** on a small dataset of Pokemon Images consisting of **4** labels
• Utilized the **ResNet-50** as a base model for feature extraction from images and done **Fine-Tuning** on its **last** layer
• Achieved a **validation accuracy** of **0.9643** and **F1 - score** of **0.89** by using an **Artificial Neural Network** as the FC layer

## AWARDS AND ACHIEVEMENTS

• Attained a Rank among the **top 1%** in Joint Entrance Examination Advanced 2021 among **0.25 million** shortlisted candidates
• Stood among the **top 1%** by achieving a score of **99.10 percentile** in JEE Mains 2021 among **1.1 million** applicants
• Secured a Rank of **652** in West Bengal Joint Entrance Examination 2021 among **1 Lakh** applicants from all over West Bengal
• Achieved a peak rating of **1649** by securing a rank of **447** in **Codechef Starters 88 Division 3** among **25,000 contestants**
• Awarded the **Highest Academic Excellence award** at Julien Day School, Ganganagar from class **IX** to **XII [2017-2021]**
• Secured a state rank of **140** in the **International Informatics Olympiad 2016**, organized by Silverzone Foundation, New Delhi

## COURSEWORK INFORMATION

**Computer Science**: Programming and Data Structures (Theory and Lab) | Machine Learning - Foundations and Applications
**Math and Statistical:** Linear Algebra | Real and Complex Analysis | Probability and Statistics | Advanced Calculus | Economics
**MOOCs**: DSA in C++ | Intermediate Machine Learning With Python | Deep Learning With MatLab | Convolutional Neural Networks | Geospatial Analysis | Natural Language Processing (CS224N) | AZ Web Development Bootcamp

## SKILLS AND EXPERTISE

**Programming Languages**: C | C++ | Java | Python | Matlab | HTML | CSS | JavaScript
**Libraries**: Numpy | Pandas | Matplotlib | Seaborn | Scikit-learn | Tensorflow | Keras | OpenCV | STL | Flask | Selenium | NLTK
**Softwares and Tools**: Matlab | Git | Github | VSCode | FreeCAD | Jupyter | Google Colaboratory

## EXTRA CURRICULAR ACTIVITIES

• National Cadet Corps (**NCC**) - Volunteered as a cadet in the National Cadet Corps - 1 Bengal EME COY NCC, IIT Kharagpur
• **Mentored** first-year students by clarifying their doubts in various academic and non-academic areas
• Active Member of Bengali Technology Dramatic Society - **Druheen**, IIT Kharagpur, participated in 2 onstage productions
• Part of the **Rangoli** Team of the Rajendra Prasad Hall of Residence in Inter-Hall Gymkhana Championship 2022-2023