

# ARNAB BANERJEE

M.Sc. Big Data Analytics Student  
RKMVERI, Belur Math, West Bengal, India

@ arnabbanerjee6704@gmail.com

arnab-banerjee-1740662a4

the-blue-panther

9123733834

Portfo-



## PROJECTS

- Sentiment Analysis on Product Reviews**  
ML | Classification [\[View here\]](#) July 2025
  - Built an NLP model to classify customer product reviews into positive, negative, or neutral sentiments.
  - Applied comprehensive text preprocessing (tokenization, stemming/lemmatization), TF-IDF vectorization, and trained classification algorithms (e.g., Logistic Regression, SVM).
- Customer Churn Prediction Model**  
ML | Classification [\[View here\]](#) July 2025
  - Developed a machine learning model to predict customer churn for a telecom company.
  - Focused on robust feature engineering, effective handling of imbalanced datasets, and thorough model evaluation using metrics like Precision, Recall, F1-score, and ROC-AUC.
- Exploratory Data Analysis (EDA) on COVID-19 Data**  
Data Analysis | Visualization [\[View here\]](#) June 2025
  - Conducted in-depth EDA on global COVID-19 dataset, focusing on trends in cases, deaths, and vaccination rates.
  - Performed data cleaning, handled missing values, and generated insightful visualizations (Matplotlib, Seaborn) to uncover epidemiological patterns.
- Speech-to-Text Script (Benglish NLP)**  
NLP | Fine-tuning [\[View here\]](#) April - May 2025
  - Fine-tuned the Whisper model for accurate transcription of Bengali-English code-switched speech.
  - Involved extensive dataset creation and multilingual fine-tuning techniques using PyTorch and Hugging Face.
- Wine Quality Classifier**  
ML | Classification [\[View here\]](#) Nov 2024
  - Developed an ML model to classify red wine quality, achieving 89.5% accuracy with Gradient Boosting.
  - Utilized IQR filtering and robust feature selection, evaluating performance with ROC-AUC.

## COURSEWORK

- Data Structures and Algorithms
- Java & Hadoop
- Machine Learning
- Linear Algebra
- Probability and Stochastic Process
- Statistics-I
- Universal Human Values

## EXPERIENCE

### Summer Intern

Dept. of Computer Science, Ramakrishna Mission Vivekananda Educational and Research Institute

June 2025 - Present

- Developing a post-synthesis voice conversion framework for Sanskrit audio, enabling voice transformation without altering the existing TTS pipeline or original voice dataset.
- Implementing cross-lingual voice transfer, utilizing an English voice dataset to generate natural-sounding Sanskrit speech, while preserving linguistic content and adopting the target speaker's identity and fluent Sanskrit prosody.

| Advisor- Swami Punyeshwarananda (Vikas Mj)

## EDUCATION

- Ramakrishna Mission Vivekananda Educational and Research Institute, Howrah**  
M.Sc. in Big Data Analytics  
2024 - Present
- Dakshin Dinajpur University, Balurghat**  
M.Sc. in Mathematics  
2021 - 2023 CGPA: 8.72
- New Alipore College, University of Calcutta**  
B.Sc. in Mathematics  
2017 - 2021 Percentage: 78.63%
- Behala Aryya Vidyamandir (HS), WBCHSE**  
Higher Secondary (10+2) | PCMB  
2015 - 2017 Score: 82%
- Behala Aryya Vidyamandir (HS), WBBSE**  
Secondary (10)  
2009 - 2015 Score: 89.86%

## TECHNICAL SKILLS

- Programming Languages:** Python, C++, Java, R,  $\text{\LaTeX}$
- Frameworks:** Pytorch, scikit-learn, Seaborn, PySpark, Hadoop, NumPy, Pandas, Matplotlib, Manim
- Tools:** Git/Github, Docker, MS Office
- Operating System:** Windows, Linux (Ubuntu)

## ACTIVITY

- Co-founder & Initiator, Clava Mathematica (2019)**
  - Initiated and co-founded "Clava Mathematica," a mathematics club at New Alipore College.

## HOBBY

- Anime, Football, Playing Guitar Cajon
- Thriller literature, Sitcoms, Fantasy fiction