# Sudam Kumar Paul

📞 +91-9933407223
✉ 2002sudam@gmail.com

in LinkedIn Profile
⭘ GitHub Profile

## EDUCATION

- **Ramkrishna Mission Vivekananda Educational and Research Institute** *2024-26*
  *M.Sc. in Big Data Analytics (Data Science)* — Ongoing(CGPA: 7.17 till 1st Sem)

- **Barasat Govt. College** *2020-23*
  *B.Sc.(H) in Mathematics from West Bengal State University, Barasat, West Bengal* — Percentage: 87.45

- **Habra High School(H.S.)** *2020*
  *West Bengal Council of Higher Secondary Education, West Bengal* — Percentage: 96.4

## ONGOING PROJECTS

- **Comparative Analysis of Asset Pricing Models: CAPM, APT and Fama-French Factors** *June 2025 to Now*
  *Implementing and comparing traditional asset pricing models for portfolio performance analysis.*

  – Tools & technologies used: Python (pandas, numpy, matplotlib, scikit-learn), Statsmodels: Linear Regression, Multivariate Regression

  – Performing a comparative analysis of CAPM, APT, and Fama-French models using regression techniques to evaluate their effectiveness in explaining asset returns.

- **GauGan: Conditional Image Generation Using Gaussian-GAN** ⭘ *Project Repository*
  *GaussianGAN from scratch in TensorFlow/Keras using custom Generator, Discriminator and Encoder.* May 2025 to Now

  – Tools & technologies used: Python, TensorFlow, Keras, Google Colab, NumPy, Matplotlib, GANs, VAEs, SPADE (Spatially-Adaptive Normalization)

  – Implementing a conditional GAN model based on SPADE and VAE principles (GauGAN) to generate photorealistic images from semantic segmentation maps using the PASCAL_VOC_2012 and Facades datasets.)

## COMPLETED PROJECTS

- **Vision Guided Robotic Manipulation** ⭘ *Project Repository*
  *A Computer-Vision Integrated Robotic Arm for Real World Object Interaction* Feb-April 2025

  – Tools & technologies used: OpenCV, NumPy, Python, ESP32, Servo Motors, IoT, Inverse Kinematics.
  – Developed real-time object interation system based on computer vision

- **Implementation of ANN from Scratch** ⭘ *Project Repository*
  *Two-Layer Fully Connected Neural Network on MNIST dataset from scratch* March-April 2025

  – Tools & technologies used: Python, Numpy, Matplotlib, Custom-built Neural Network, Cross-Entropy Loss & Softmax activation function, Gradient Descent & Momentum.

  – Designed and implemented a two-layer fully connected neural network from scratch using NumPy for handwritten digit classification on the MNIST dataset with sigmoid and softmax layers, trained via cross-entropy loss and gradient descent variants, achieving 96.92% train accuracy and 94.52% test accuracy with custom forward/backward propagations and manual parameter initialization.

- **Wildfire Prediction using H2O** ⭘ *Project Repository*
  *Distributed Machine Learning using the H2O framework with Random Forest Classifier* March-May 2025

  – Tools & technologies used: Java-11, Python 3.12, H2O.ai Framework (h2o 3.46.07), H2OFrame, Numpy, Pandas, Matplotlib, H2ORandomForestEstimator, H2O Cluster, Linux Environment.

  – The project involved data cleaning, data preprocessing, feature engineering, exploratory analysis and training a Random Forest model(30 trees, max depth 20) to predict the "Confidence" level of wildfire detection (Low, Nominal, High) using H2O's distributed architecture, achieving an AUC of 0.96 and RMSE of 0.18 on "viirs-snpp_2023_all_countries" dataset.

- **Unsupervised Learning using Py-Spark** ⭘ *Project Repository*
  *Implementation of K_Means_Clustering in Distributed way using Py-Spark* March-May 2025

- Tools & technologies used: Java-21, Python 3.10, Pyspark 3.3.2, Numpy, Pandas, Matplotlib, VectorAssembler, Spark MLlib(KMeans), ClusteringEvaluator, Spark Cluster, Linux Environment, SparkSession.

- Developed a distributed K-Means clustering pipeline using PySpark on the Iris dataset, including feature engineering, scaling, and optimal cluster selection using Silhouette scores. Visualized cluster results with 3D scatter plots and evaluated model performance using PySpark's ClusteringEvaluator.

- **Evaluating Regression Models for Cost of Living Index Prediction**  Project Repository

  *Prediction of Cost_of_Living_Index in Several Countries using Various Regression_techniques*  August-September 2024

  - Tools & technologies used: Google Colab, Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn(for ML models: Linear Regression, Ridge, Lasso, ElasticNet, SGD), PolynomialFeatures, StandardScaler.

  - A comparative study of various regression techniques, including linear regression, polynomial regression, gradient descent methods and regularization techniques to predict the Cost of Living Index and analyze their performance and computational efficiency on Cost_of_Living_Index_2024 dataset.

## Technical Skills and Interests

- **Languages**: Python , R (Basic) , LaTeX
- **Data Structure and Algorithms**: Array , Stack , Queue , Linked List , Tree , Graph , Hashing , Priority Queue , Sorting (Mergesort, Quicksort, Heapsort etc.) , Searching
- **Data Science and ML Tools**: Numpy , Pandas , Matplotlib , Scikit-learn , OpenCV
- **DL Frameworks**: Keras , TensorFlow , PyTorch
- **Tools**: Google Colab , Jupyter Notebook , VS Code , ChatGPT , Git
- **Image Processing, Computer Vision**
- **Distributed Data Processing Tools and Graph Database**: Py-Spark , Neo4j
- **Operating Systems**: Windows and Linux

## Key Courses Taken

- **Data Structure and Algorithms using Python**
- **Artificial Intelligence and Machine Learning :** Machine Learning , Computer Vision , Deep Learning and it's application in NLP.
- **Big Data Technologies :** Apache Spark , Graph Database- Neo4j
- **Mathematics for Computing :** Linear Algebra and Matrix Computation , Probability and Basic Statistics , Advanced Statistical Methods , Time Series and Survival Ananlysis

## Achievements

- **Successfully Completed a Course of DSA using Python in NPTEL**  *Jan - March 2025*

  *Programming, Data Structures And Algorithms Using Python*

## Experience

- **IASc-INSA-NASI Summer Research Intern**  *June-July 2025*

  *Research Intern in Financial Economics*  Indian Institute of Technology, Kharagpur

  - Working on Comparative Analysis of Asset Pricing Models: CAPM, APT and Fama-French Factors.

## Positions of Responsibility

- **Organizing Team Member,** Perceptron 2025 Puzzle Group (Tech Fest at RKMVERI)  *Jan, 2025*