Subhasish Saha Data Scientist

subhasishsaha007@gmail.com

https://github.com/Subhasish-Saha

https://www.linkedin.com/in/subhasish-saha-7994bb209/

C-4 ,Mandir Park, Nabanagar, Birati, 700051 Kolkata, India

30 Aug 1998

8276887228

SKILLS

- Python
- Machine Learning
- Deep Learning
- Reinforcement Learning
- Computer Vision
- NLP
- Data Manipulation
- Data Analytics
- Data Mining
- Data Wrangling
- Tensorflow
- Docker

SOFT SKILLS

- Problem Solving
- Teamwork
- Communication
- Attention to detail
- Adaptability
- Time Management
- Creativity

EXPERIENCE

Junior Data Scientist, Cloudcraftz solutions pvt ltd — Jul 2022 – present | Kolkata, India

Options Backtesting Engine:

- Developed a **high-frequency trading strategy backtesting engine** to analyze and optimize trading strategies for maximizing profits.
- Designed and implemented the portfolio module for real-time trade storage and updates.
- Utilized the **Black-Scholes** method to generate **synthetic option prices** for unavailable market data.
- Employed **classical machine learning** techniques to enhance the strategy, including **supervised** and **unsupervised methods**.
- Created a **classification model** to **predict trading signals** based on implied volatilities and realized volatilities.
- Employed **hierarchical** and **K-means clustering** for trend analysis and trade quantity control.
- Achieved a **25% improvement in PnL** over the baseline strategy using supervised techniques and a **34% improvement** using unsupervised methods.
- Proficient in data visualization, chart creation, and report preparation.

Portfolio Allocation using Machine Learning:

- Worked on a project for **portfolio allocation** based on risk-taking capacity and financial goals.
- Leveraged Classical **Machine Learning for future price prediction of stock time series data**, resulting in accurate price forecasts.
- Explored **reinforcement learning** and **deep learning** approaches for portfolio optimization.
- Employed portfolio optimization techniques, including **max sharpe** and **min vol** methods.

Data Science Intern, Indian Statistical Institute, Kolkata — May 2022 – Jul 2022 | Kolkata, India

Incidental Scene Text Detection:

- Successfully developed an efficient model for accurate **detection of textual regions** within images.
- Leveraged the **ICDAR 2015** and **ICDAR 2013** datasets for training and validation.
- Implemented **rectangular bounding boxes** for precise text **region localization**.
- Utilized **ResNet-50** architecture to build the model, resulting in robust and accurate text detection.
- Incorporated innovative techniques and ideas from various research papers into the project, such as **Vision Transformers**, **EAST**, and other cutting-edge models.

Data Science Intern, Indian Institute of Technology, Patna — Dec 2021 – Apr 2022 | Patna, India

Cyberbully Detection on Multi-modal Indian Languages:

- Developed a model that predicts the **sentiment and bully class of tweets/sentences** in various **Indian languages**.
- Collected a dataset by **scraping 6436 random tweets** from Twitter.
- Compiled a **Hindi-English corpus** based on the collected tweets.
- Designed and built a model utilizing **BERT** and **BiLSTM** architectures, achieving excellent accuracy in classifying sentences.

PERSONAL PROJECTS

Notebook Renderer Library

- Developed a Python library hosted on PyPI for rendering and playing YouTube videos within Jupyter notebooks.
- Implemented modular coding practices for maintainability and scalability.
- Integrated GitHub testing for automated testing of releases.

Student Performance Indicator

- Conducted data collection, exploratory data analysis (EDA), and feature engineering.
- Developed a predictive model for students' test scores considering variables like Gender, Ethnicity, Parental Level of Education, Lunch, and Test Preparation Course.

Flight Price Prediction

- Implemented a model for predicting flight fares between source and destination based on date, time of arrival and departure.
- Applied machine learning techniques, regression, exploratory data analysis (EDA), feature engineering, and model selection.

EDUCATION

M Sc. in Data Science, University of Kalyani — 2020 – 2022 | Kalyani, India
CGPA - 9.1

B Sc. (Honours) in Mathematics, University of Calcutta — 2016 – 2019 | Kolkata, India
(53.56 %)

COURSES

Machine Learning, Coursera — 2022

Introduction To TensorFlow For Artificial Intelligence, Machine Learning, And Deep Learning, Coursera — 2021

LANGUAGES

- Bengali
- English
- Hindi