Deepanjan Saha

Data Scientist

Contact

Address
Hyderabad, India 500019

Phone
8887824591

E-mail
deepsaha20@gmail.com

LinkedIn
https://www.linkedin.com/in/deepanjan-saha-india/

Github
https://github.com/DeepanjanSaha-INDIA

Skills

Machine learning

Advanced

Deep Learning

Upper intermediate

Natural Langueage Processing

Upper intermediate

An ML enthusiast who did Post Graduation in Business Analytics from IIM Calcutta, ISI Kolkata and IIT Kharagpur. Currently working as a Data Scientist in Providence with primary work on Deep learning, Statistical testing, NLP, Time Series and LLMs. Prior to my post-graduation was working as a software engineer in Fidelity Investments, where I got exposure to financial market. Part of robotics team during my B.Tech from NIT Allahabad. Always learning new skills and perfecting existing ones.

Work History

2023-06 - Current

Data Scientist
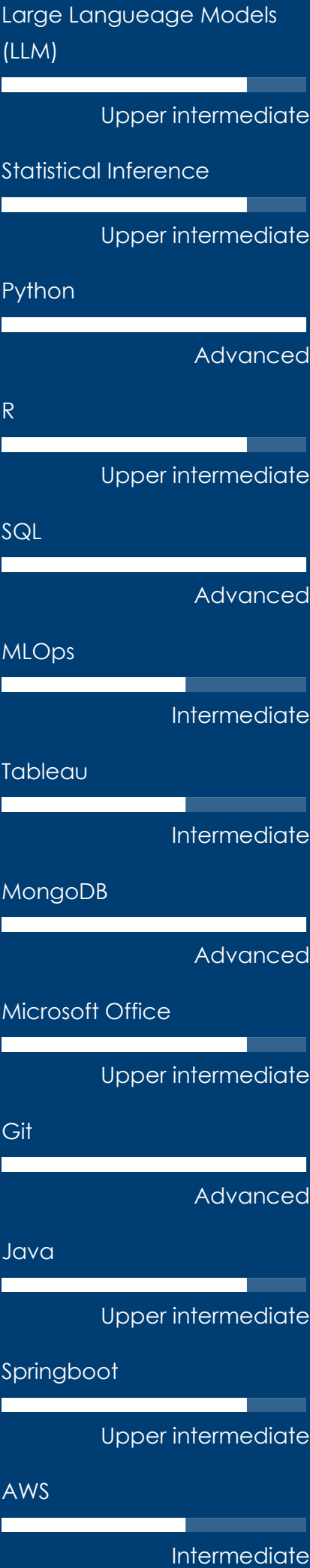
Providence India, Hyderabad

- Worked with stakeholders to develop **metric extraction pipeline** from text data using NLP techniques.
- Use Deep learning, NLP and Data Structures concepts to build **custom LLMs** from scratch for internal business use case.
- Utilized advanced querying, visualization and analytics tools to analyze and process complex data sets.

2019-08 - 2021-10

Software Engineer

Fidelity Investments, Bangalore

- Built scalable applications in Agile setup, using Microservice architecture & Spring frameworks
- Worked on E2E delivery: Requirement gathering, API development, testing, deployment & documentation
- Developed **Retirement Healthcare Cost API**, to estimate customers post-retirement healthcare cost
- Enhanced **Savings Optimization API**, to get optimal redistribution of funds to maximize tax benefits
- Added **Role Based View** feature to existing Production environment API monitoring platform
- Developed new KPI dashboard using **QlikSense**, to track platform's API traffic for management team
- Developed and deployed automated mailing system using **Qlik NPrinting**, for weekly KPI reports
- Received **Luminary Award** under technical

Skills (sidebar)

Large Langueage Models (LLM)
Upper intermediate

Statistical Inference
Upper intermediate

Python
Advanced

R
Upper intermediate

SQL
Advanced

MLOps
Intermediate

Tableau
Intermediate

MongoDB
Advanced

Microsoft Office
Upper intermediate

Git
Advanced

Java
Upper intermediate

Springboot
Upper intermediate

AWS
Intermediate

---

excellence category for exceptional performance in 2020
- Collaborated with other engineers to improve existing automated testing projects.
- Implemented **MongoDB Aggregation pipeline** to pre-process and store API **Splunk** log data to database.

Education

2021-10 - 2023-05

Post Graduate Diploma in Business Analytics: Business Analytics

Indian Institute of Management - Calcutta
Completed Post Graduate Diploma in Business Analytics, which is a tri-institute program by **IIM Calcutta**, **IIT Kharagpur** and **ISI Kolkata**.

2015-07 - 2019-05

Bachelor of Technology: Electronics And Communication

Motilal Nehru National Institute of Technology - Allahabad
- Graduated with Electronics and Communication degree with overall CGPA of **9.03**.
- Member of robotics team, organized Techfest event named circuit of the day, and workshop lead in Prosang Robotics fest.
- Built collage's first quadcopter drone from scratch.

2013-05 - 2014-05

12th Class: PCM + Computer Science

Kendriya Vidyalaya OLF - Dehradun
Passed Class 12th with overall **90.33%**

2011-05 - 2012-05

10th Class

Jaswant Modern Sr. Sec. School - Dehradun
Passed Class 10th with overall CGPA of **9.8** out of 10.

Accomplishments

- Ranked among Top 10 percentile in class of 60 students of 7th PGDBA Batch of 2021-23 (2023)
- Awarded merit-based scholarship for securing 3rd highest GPA in 1st semester of PGDBA at ISI

Kolkata (2022)

- Co-Authored an article on Blockchain and its application in Digital Identity for PGDBA Magazine (2023)
- Ranked 1st in Fidlatica 6.0, Propensity cut-off prediction competition by SCMHRD Pune (2022)
- Ranked 1st in Cypher 3.0, Data Visualization and analysis competition by IIM Raipur (2022)
- Ranked 3rd in Trilytics (2600+ teams), Market prediction & launch strategy case by PGDBA (2022)
- Ranked 7th in Kaggle data science competition to predict survival rate for hospitalised patients (2022)
- Top 8 (Finalist) out of 574 teams in Abhigyata 2.0, Business case competition, by IIM Amritsar (2022)
- Top 8 (Finalist) out of 571 teams in Vridhi: Social Impact Case competition, by IIM Lucknow (2022)
- Ranked 2nd in Electromania, Hardware game console development Event in IIT Kanpur (2018)
- Awarded PPO as Software Engineer by Fidelity Investment after 2-months UG Internship (2018)
- Ranked 1st in Embetronix, a Maze solving bot building Event organized by IIT Kharagpur (2017)
- Ranked 1st in Measure It 3D, an Innovative product development Event in IIT Kanpur (2017)
- Secured 3rd State rank in International Olympiad of Science and overall Olympiad rank of 613 (2009)

Certifications

2020-05	Deep Learning: Convolutional Neural Network in Python, Udemy
2020-05	A Complete Guide to Tensorflow 2.0 using Keras API, Udemy
2020-06	MongoDB - The Complete Developers Guide 2020, Udemy
2020-09	The Data Science Course 2020: Complete Data Science Bootcamp, Udemy
2022-03	Winter School for Deep Learning, ISI Kolkata

Projects

Forecasting Air Quality Data (Time Series)

- Predicted PM2.5 levels on 5 years' Air quality index data. Performed ADF test for stationarity check of data
- Plotted ACF/PACF to identify model order. Modelled AR, MA & ARMA process to fit the univariate series
- Achieved best AIC score of 11.14 on 13-day test window. Tested goodness of fit using Ljung Box statistics

Patient Survival Prediction (Classification)

- Utilized 64K medical records to build a Classification model for predicting survival rate of patients
- Performed EDA; Over-sampling to handle class imbalance & KNN Imputer to impute missing data
- Built voting classifier using Random Forest, XGBoost, CatBoost & LGBM achieved ROC-AUC score of 0.89

Complaint Classification (Deep Learning)

- Built text classification model, trained on 1M+ textual complaints data to identify financial products
- Performed under-sampling of data & Applied Bag of words, TFIDF, pre-trained GloVe word embedding, with LSTM & also leveraged transfer learning using pre-trained BERT model to achieve F1 score of 0.78

Metric Extraction (Deep Learning)

- Built custom Named Entity Recognition model to extract multiple medical entities from textual data. Fine-tuned BERT transformer model to obtain overall F1 score of 0.8628