Tamoghna Saha

DATA SCIENTIST

Kolkata, West Bengal, India
tamoghnasaha.22@gmail.com

LinkedIn
GitHub
kaggle

Data Scientist with 4+ years of experience in implementing Machine Learning and Deep Learning models in the domain of Natural Language Processing and Computer Vision, possessing strong analytical and design principal skills, and passionate to explore state-of-the-art algorithms.

EXPERIENCE

Ford Motor Company, Chennai — *Data Scientist*

NOVEMBER 2020 - PRESENT

Working with AI Advancement Center team under Global Data Insight & Analytics division, specializing in NLP-based use cases and creating modules for Ford's internal NLP pipeline.

Mu Sigma, Bangalore — *Decision Scientist*

OCTOBER 2017 - NOVEMBER 2020

Worked in the R&D team on creating various re-usable modules on ML, assisted in delivering solutions to multiple client-facing teams across industries such as aviation, retail, and conducted Python and various Deep Learning sessions across the company.

EDUCATION

Institute of Engineering and Management, Kolkata — *Bachelor of Technology*

2013 - 2017

Stream: Electronics and Communication Engineering
DGPA: 9/10
Board: Maulana Abul Kalam Azad University of Technology (M.A.K.A.U.T.)

Patha Bhavan, Kolkata — *Higher Secondary Certificate*

2011 - 2013

Stream: Science
Additional Subject: Statistics
Percentage: 83.6
Board: West Bengal Council of Higher Secondary Education (W.B.C.H.S.E)

Patha Bhavan, Kolkata — *Secondary School Certificate*

2009 - 2011

Percentage: 82
Board: West Bengal Board of Secondary Education (W.B.B.S.E.)

SKILLS

Language: Python, SQL

Framework: HuggingFace, spaCy, PyTorch, RayTune, Keras, scikit-learn, torchserve, Dash, Streamlit, Plotly

Platform: Jupyter, VS Code, Docker, Poetry, Google Cloud Platform

AWARDS/ACHIEVEMENTS

Top 3 in Ford India IoT Hackathon 2021: Vibration Anomaly Analysis

Recognition Award in Ford for taking sessions and seminars on NLP

Spot Awards in Mu Sigma:
1. Developing demo of a web app that showcases object detection and OCR in less than 2 months (2018)
2. Best contributor to Mu Sigma Training program in Python course (2020)
3. Recognized developer for creating re-usable modules and handling client threads simultaneously (2020)

Student Performance Award in college (2017)

Smart Glove: 1st prize in Engineering Model Competition in NEN Kolkata (2015)

Former Governor of West Bengal Mr. Gopal Krishna Gandhi appreciated and took my painting (2008)

LANGUAGES

English, Bengali, Hindi

Ford

PII MASKING

Developed a robust NER pipeline using HuggingFace and deploying in GCP to mask PII data from customer query records in the Customer Relationship Center to ensure data privacy before passing it to downstream tasks such as sentiment analysis and topic modeling. Additionally, removed signatures using spaCy as a part of preprocessing. Assisted in developing beta UI of the entire NLP pipeline using Dash.

RESUME PARSER

Implemented a technique to segment various sections of a resume and identify the best match on a search engine using spaCy and HuggingFace and created the beta search interface using Streamlit.

Mu Sigma

MACHINE LEARNING MODULES

Developed re-usable code modules known as Foundation Brick of various Machine Learning estimators such as EDA, Regression, Classification, and Time Series with interactive visualizations in Jupyter notebook to accelerate generating insights by the delivery teams.

OPTICAL CHARACTER RECOGNITION

Implemented Deep Learning model for a major multinational energy corporation, which distinguishes areas of handwritten and machine-printed texts and extracts texts from Engineering Blueprint's images.

NATURAL LANGUAGE QUERY

As a part of R&D, NLQ was implemented using a Dash-based Web Application where the user can upload a dataset and enter a query as natural question. This question is translated to its corresponding SQL query using BERT and based on the queried output, a visualization was generated. It benefits a non-technical person to draw basic analysis without coding.

HUMAN TRACKING ANALYTICS

For a US Airline Company, various Deep Learning models were explored to track humans and luggage and get their respective counts while boarding the plane. It was implemented to estimate the overhead bin capacity and check if it has reached the threshold to allow the remaining passengers to put their carry-in luggage inside beforehand to reduce turn-around time.

ENTITY RESOLUTION IN FRAUD ANALYTICS

For an Australian Insurance Company, a Deep Learning model was implemented to identify duplicate profiles based on the customer's demographics and resulted in a 60% reduction in profiles by performing de-duplication.

CLOUD CERTIFICATIONS
1. Google Cloud Platform Big Data and Machine Learning Fundamentals
2. Smart Analytics, Machine Learning, and AI on GCP

CO-CURRICULUM AND HOBBIES
5 years course in Art and Painting from Ramakrishna Mission Institute of Culture, Golpark, Kolkata
Played Table Tennis for 4+ years, learned swimming for 1+ year, solving sudoku