Sourav Mandal *Data Scientist*

Vill.:Rajbagan, P.O.: Chakdaha, Dist.: Nadia, State: West Bengal, Pin: 741222

+91 8618241725

Indian

Sourav9827

mandal.srv1998@gmail.com

27/11/1998

sourav-mandal-390064210

Portfolio

PROFILE

As a highly motivated and skilled data scientist with a postgraduate degree in Mathematics, I possess a solid understanding of various data science concepts and tools, including Python, SQL, Machine Learning, Deep Learning, NLP, and PowerBI. With hands-on experience in multiple data science projects and internships, I have developed a strong problem-solving and analytical mindset. Additionally, my past experience in teaching mathematics and mentoring a team of interns for a backend operation internship showcases my excellent communication and leadership skills. As a result, I am confident in my ability to leverage my skills and experience to contribute to any data-driven organization's success.

SKILLS

Python (*Pandas, Numpy, Matplotlib, Seaborn, Flask, Plotly, Dash, Sklearn etc.*)

Data Analytics (*MySql, Tableau, PowerBI, Snowflake, Pentaho etc.*)

Databases (*SQL and Mongo DB.*)

MS-Office (*Advanced Excel, Power point, Word, Access.*)

Team Leader (*I lead a group of interns under me, monitor their work and as a team we solve various task in Highway Delite.*)

Machine Learning (*Linear and Logistic Regression, SVM, Xgboost, Random Forest, Naive Bayes etc.*)

Deep learning (*ANN, CNN and RNN*)

NLP (*NLTK, Machine Learning, RNN, LSTM, GRU, Self-Attention model, Transforms, Chatbot etc.*)

Big Data (*Hadoop, Kafka and Spark*)

Cloud Services (*AWS, Azure, Heroku, Render etc.*)

MLOps (*Github Action, CI/CD pipelines, Deployment Techniques In AWS,AZURE, Dockers And Kubernetes, MLflow, DVC.*)

R Software (*ggplot2, dplyr, plotly kniter, data.table, mlr3 etc.*)

CERTIFICATES

Data Analytics Consulting Virtual Internship(KPMG) (*Data Quality Assessment Data Insights Data Insights and Presentation May 8th, 2021*)

Data@ANZ Program (*Exploratory Data Analysis Predictive Analytics May 13th, 2021*)

Full Stack Data Science (*inuron.ai*)

Python for Data Science and Machine Learning Bootcamp (*udemy*)

Introduction to Machine Learning (*Coursera*)

Data Science foundations (*Great Learning*)

PROFESSIONAL EXPERIENCE

Teacher(Mathematics), *Self Employed*
2017 – present | Chakdaha, India

- Freelancing Teacher of Mathematics and Computer Science
- Creating Presentations and Learning designs for everyday classes.
- Good understanding of Concepts and Application in real life situations.
- Noticeable improvements in Student's Academic results.

Mentor of NH Mapping, *Highway Delite*
02/2022 – 11/2022 | Work from home, India

- Building database for highway related data.
- Train and monitor the work of new interns.
- Research and create documents for tourism data
- Verify and upload the highway information in the backend of the website

Data Analyst Intern, *The Sparks Foundation*
08/2021 – 09/2021 | Work from home, India

- Prediction using Supervised ML
- Prediction using Unsupervised ML
- Exploratory Data Analysis - Retail
- Exploratory Data Analysis - Terrorism
- Prediction using Decision Tree Algorithm

Data Science Intern, *ineuron.ai*
01/2023 – 02/2023 | Work from home, India

- **Project Title**: News Article sorting
- **Objective:** To classify news articles into predefined categories like Sports, Technology, Entertainment, etc. using deep learning methods.
- **Tools:** Python, Google's BERT model, Azure App Services, Flask, HTML/CSS , Pandas, NumPy, Scikit-learn, TensorFlow, and Keras.
- **Final Output:** Achieved 97.96% accuracy on test set, deployed model on Azure App Services using Flask web app for predictions
- **Roles & responsibilities:**
  - Pre-processing the data to make it suitable for the model
  - Fine-tuning the BERT model for News Category Classification
  - Developing the Flask application for the model
  - Deploying the model on Azure app services
  - Testing and debugging the application
  - Documentation of the project

PROJECTS

Tweets Sentiment Analysis,
Data analysis and machine learning using Python and NLTK
10/2022 – 11/2022

- **Objective:** Analyze the sentiment of tweets and classify them as positive, negative, or neutral using machine learning
- **Tools:** Python, Scikit-Learn, SciPy, NLTK, Pandas, Heroku, Falsk, HTML/CSS.
- **Final Output:** Achieved 85% accuracy on test set, deployed model on Heroku as a web app for predictions
- **Roles & responsibilities:** Sole contributor, responsible for all aspects of the project including data collection, preprocessing, model training, and deployment. Managed the entire project from start to finish.

Mushroom Classification,
Develop an accurate and reliable model for classifying mushrooms.
09/2022 – 12/2022

- **Objective:** Classify mushrooms into edible or poisonous categories using machine learning

EDUCATION

M.Sc Mathematics(Pure), *University Of Kalyani*
08/2019 – 08/2021 | Kalyani, India

Diploma in Information Technology,
Nehru Yuva Computer Shiksha Kendra
02/2021 – 02/2022 | Chakdaha, India

Bachelor in Education (B.Ed), *W.B.U.T.T.E.P.A*
09/2021 – present | Santipur, India

B.Sc (Hons.) Mathematics, *University Of Kalyani*
07/2016 – 07/2019 | Kalyani, India

Full Stack Data Science program, *ineuron.ai*
03/2022 – present
Data Science and Analytics, Machine Learning, Deep Learning and Big data basics.

LANGUAGES

English **Hindi** **Bengali**

COURSES

Learning Analytics Tools, *NPTEL*
07/2021 – 10/2021

Probability for Computer Science, *NPTEL*
07/2021 – 10/2021

Introduction to Machine Learning, *NPTEL*
07/2021 – 09/2021

Data Analytics with Python, *NPTEL*
01/2021 – 04/2021

Data Science for Engineers, *NPTEL*
01/2021 – 03/2021

Programming, Data Structures And Algorithms Using Python, *NPTEL*
09/2020 – 11/2020

Introduction to R Software, *NPTEL*
09/2020 – 11/2020

Python for Data Science, *NPTEL*
09/2020 – 10/2020

DECLARATION

I hereby declare that all the information furnished above is correct to the best of my belief. I am responsible for the authenticity of all the information.

Sourav Mandal
Chakdaha, 24/02/2022

- **Tools:** Python, scikit-learn, pandas, SciPy, NumPy, Flask, Heroku
- **Data**: Mushroom dataset with over 8,000 records and 22 features
- **Preprocessing**: Data cleaning, encoding categorical variables, splitting into training and test sets
- **Model training:** Comparison of various algorithms and selection of best performing algorithm for classification
- **Final output:** Achieved 99% accuracy on test set, deployed on Heroku as a Flask web app for predictions
- **Roles & responsibilities:** Sole contributor, responsible for data collection, preprocessing, model training, deployment, and project management

Insurance Fraud Detection, *Develop and implement a fraud detection system for an insurance company.*
07/2022 – 08/2022

- **Problem Statement/Objective:** To develop a machine learning model that can accurately detect fraud in insurance claims to minimize losses and improve claim processing efficiency.
- **Tools:** Python, Jupyter Notebook, Scikit-learn, Pandas, NumPy, Flask, Heroku, HTML/CSS.
- **Final Output:**
  - A machine learning model with accuracy of 77.2% in detecting fraudulent claims that can predict the likelihood of a claim being fraudulent based on various features such as claim amount, policy holder history, and claim type.
  - The model was deployed on Heroku as a web application that can be used to submit insurance claims and receive an instant fraud prediction.
- **Roles & Responsibilities:** Sole developer and project manager. Conducted data analysis and preprocessing, trained and evaluated machine learning models, and deployed the final model on Heroku.
  -

California Housing Price Estimation, *Developed a machine learning model to predict housing prices in California using data from the California Census*
06/2022 – 07/2022

- **Objective:**To develop a machine learning model that can predict the median housing price in any district in California based on various factors such as population, median income, etc.
- **Tools:** Python, Scikit-learn, Jupyter notebook, Pandas, Matplotlib, seaborn, Flask, MLops, Docker Heroku, Git, GitHub Actions
- **Final Output**
  - Created a machine learning model that achieved an R squared value of 0.87 on the test set, demonstrating strong predictive power
  - Developed a Flask application on Heroku following MLOps practices for continuous integration and monitoring.
- **Roles & Responsibilities:** As a solo project, I was responsible for every aspect of the project, including data exploration, data cleaning, model building, MLOps practices implementation, and deployment on Heroku.

Algerian Forest fire prediction, *Created a machine learning model that can accurately predict the likelihood of future forest fires in Algeria*
06/2022 – 06/2022

- **Objective:** To develop a machine learning model that can predict the occurrence of forest fires in Algerialikelihood of future forest fires using the data.
- **Tools:** Python, Scikit-learn, Jupyter notebook, Pandas, Matplotlib, seaborn, Flask, heroku, HTML/CSS.
- **Data:** The project used a dataset obtained from Algerian Forest Services, which contains historical data on forest fires in Algeria from 2000 to 2019. The dataset consists of various features such as temperature, humidity, wind speed, and rain, among others.
- **Final Output:** The developed machine learning model can predict the occurrence of forest fires in Algeria with an accuracy of 98.76%. The model was integrated into a Flask app and deployed on Heroku, enabling users to input real-time data and get predictions.
- **Roles & Responsibilities:**Sole contributor, responsible for all aspects of the project including data collection, preprocessing, model training, and deployment. Managed the entire project from start to finish.