

SHAONI MUKHERJEE

• Data Scientist •

ABOUT ME

A professional data scientist with background in research in Machine Learning. Highly analytical with a knack for analysis, math, statistics. Proactive in fetching information from various sources and analyze it for better understanding. Highly accurate in analyzing and interpreting large datasets, developing new forecasting models.

EDUCATION

2020- 2021 • Liverpool John Moores University Master's with Merit in Data Science and Machine Learning

2019 - 2020 • International Institute of Information Technology Graduated with CGPA 3
PGDDS in Data Science

EXPERIENCE

2021 Ineuron

Intern

- Tasked to perform Market Basket analysis on E-Commerce.
- Objective of the case study was centered around predicting the product rating a user would rate after purchasing the product.
- Use of various ML model such as Logistic Regression, Random Forest Classifier, LSTM was used to find the best working model. Finally, applying the Apriori algorithm to find the frequently bought items in the dataset.
- Attained a score of above 90% in predicting the rating also was able to draw valuable insights from the dataset.

2021 Sparks Foundation

Intern

The main aim of the project was to create an end to end website.

- The objective of the project was to create an end to end application to predict the price of flight. Data used is available at kaggle.
- The Machine Learning model was created using Linear Regression, Decision Tree Regressor, KNN. To fine tune the model Grid Search CV and Randomized Search was used.
- The model was evaluated using RMSE and R Square. The Random Forest Model was selected as the model achieved a RMSE score of 2015.6 and R square of 82%.
- This was Flask web app which was deployed using Heroku.

Github link:-https://github.com/ShاونiMukherjee/Model_Deployment

web app link:-<https://flight-price-prediction-api.herokuapp.com/>

SKILLS

Python



Flask



SQL, HQL



Machine Learning



Deep Learning



Exploratory Data



Analysis

LIBRARIES

Scikit Learn, NLTK, Tensorflow,
Pandas, Numpy, Keras, Matplotlib,
Matplotlib, Flask,
Seaborn, OpenCV, Spark
Plotly, MLlib, Theano

CONTACT ME AT

📍 Bangalore, 560075

✉️ shaomukherjee@gmail.com

🌐 <https://www.linkedin.com/in/shaoni-mukherjee/>

💻 <https://shaonimukherjee.github.io/shaoni.github.io/>

RESEARCH

The novel research tested the hypothesis that the prices in Airbnb listings in New York City is influenced more by physical location than other features. Further, the aim was to find out post covid effect on the Airbnb Properties Pricing. An example of the result indicated that \$210 will add to the price if the property is located in East Village. Also, \$85 for an extra bathroom. Different Machine Learning model was built and bench marking was done to find out the best one. Using different ML model, the research evaluates more than the physical location to predict the prices.

<https://github.com/ShاونiMukherjee/Airbnb-Price-Prediction>

END TO END PROJECTS

MACHINE LEARNING

Bangalore Home Prices Prediction: Data Science Regression Project to find the prices of the properties in Bangalore.

- The initial model was built using Linear Regression with sklearn module. Further, the model was tuned using different ML model such as Lasso and Decision Tree.
- The model attained a score of above 86%. Further, a python flask server that uses the saved model to serve http requests was created.
- The ML model was deployed to production using amazon AWS EC2 instance.

Link:-https://github.com/ShاونiMukherjee/Bangalore_price_deployment

DEEP LEARNING

Plant Village Dataset: Plant Disease Detection using CNN.

- The dataset was a collection of different vegetables healthy and unhealthy leaves images.
- Analysis of 54,306 images of plant leaves was attempted to predict the crop-disease. Deep Convolutional Neural Network was applied for the classification problem.
- The overall accuracy obtained varied between 85.53% to 98%.
- The model was deployed using Heroku.
- However, further improvements are going on to make a mobile app using tensorflow serving, FastAPI, React JS, React Native.

Link:-

https://github.com/ShاونiMukherjee/Plant_Village_Disease_Classifications_1
