Nishant Sinha

STUDENT

Flat No.502, Tower16, The Close North, Nirvana Country, Sec.50, Gurgaon, India

(+91) 9958868674 | nishantsinha00@gmail.com | https://github.com/nishantsinha00 | https://www.linkedin.com/in/nishant-sinha-201885191/

Summary

A creative and hardworking fresher passionate in the field of Artificial Intelligence with experience in solving business problems and have executed various projects in Computer Vision and NLP. An analytical thinker with the ability to come up with solutions for complex problems.

Education

Indian Institute of Technology Kharagpur   *Kharagpur, West Bengal*
DUAL DEGREE PROGRAM IN CIVIL ENGINEERING/ARTIFICIAL INTELLIGENCE MACHINE LEARNING AND APPLICATIONS   *jun. 2019- jun 2024*

RESEARCH/INDUSTRIAL EXPERIENCE

Anicca Data Science Solutions   *Bengaluru, Karnataka*
ML ENGINEER INTERN   *April. 2022 - Jun. 2022*

- Project 1: Built an image classifier after applying various image classification models like VGG-16, ResNet50, Inceptionv3, and EfficientNet and finally came up with a custom
- Project 2: Researched various Time Series Forecasting, Machine Learning, and Deep Learning algorithms to predict the stock price of Fortune 500 companies for 15 days into the future. Also contributed to the development of the web application using the Flask API.

Feynn Labs   *Remote*
MACHINE LEARNING INTERN   *Dec. 2021 - Feb. 2022*

- Ideated ML/AI-based solutions for small/medium businesses and helped them with their sales, business operations, marketing, and more.
- Analyzed the Ed-Tech Market in India using segmentation analysis and came up with a feasible strategy to enter the market.
- Using various clustering techniques came up with 4 segments that an Ed-Tech startup should focus on while entering the market.

IIT Kharagpur   *Kharagpur, West Bengal*
DEEP LEARNING RESEARCH INTERN   *Dec. 2021 - Feb. 2022*

- Ideated various problems in the field of Agriculture and how Artificial Intelligence can provide solutions to those respective problems.
- Acquired satellite data and processed the satellite image data using Geo-technical software (QGIS) to get tabular data consisting of various extracted properties of soil as features which were used as labels for model training.
- Applied multitask 2d-CNN model architecture on the satellite image data set to predict soil organic carbon, bulk density, clay content, coarse fragments, sand, silt, and nitrogen of the soil.

Awards/Achievements

- Vice-captain of MMM hall Data Analytics Team which placed 5th amongst 21 halls of residence in the IIT inter-hall data analytics competition.

AREAS OF EXPERTISE

Data Visualizations, Clustering, Classification, Quantitative Analysis, Predictive Analysis, Web Scraping, ML/DL Algorithms, Time Series Analysis, Time Series Forecasting, Computer Vision (Object Detection and Image segmentation), NLP, Generative models, FastAPI, Flask, Tensorflow/PyTorch.

Self-Projects

SENTIMENT ANALYSIS

- Data Preprocessing: Applied various Over-Sampling techniques like Random Over Sampling and SMOTE due to class imbalance problem between Positive, Neutral, and Negative Sentiments. Also applied various Text Preprocessing techniques like Bag of Words, Tfidf vectorization and also applied Word2Vec models for text preprocessing. Sentiment Prediction.
- Sentiment Prediction/Models: Used various Machine Learning models like LogisticRegression, MultinomialNB, RandomForest and XgBoost and performed hyperparameter tuning with RandomSearch and GridSearch. Also used Deep Learning models like Artificial Neural Network and LSTM and achieved the best validation accuracy of 90.3 percent.
- Topic Modelling: Performed topic modeling on the data set using various techniques like LDA and NMF and concluded that 'love' and 'great' were the most popular terms in the amazon reviews which explains the class imbalance problem (highly abundant positive classes) of the dataset.

IMAGE CAPTIONING

- Used the Flicker8k_Dataset dataset for images and the Flickr_8k_text dataset for captions.
- Used the pre-trained model called **Xception** which is already trained with large datasets to extract the important features from these models and these extracted features were used as input data for model training.
- Applied various text preprocessing techniques on the captions textual data and finally converted the textual dataset to a numerically vectorized dataset by applying **Tokenization** followed by **Padding** on each sequence to ensure that all the sequence vectors are of the same length.
- Used the extracted feature vector representations of the image and the numerically vectorized text sequences to create a Data Generator which was useful for model training.
- Created an **LSTM** model from scratch and used the Data Generator to train the model. The model was trained on **Categorical cross-entropy** loss.
- Trained the model for 3 epochs and got a training loss of **3.67** and a validation loss of **4.13**.