

Syed Yousuf

Software Engineer and Aspiring Data Scientist

yousufsyed900@gmail.com | 0336-1280433 | <https://github.com/yousuf9292> |
<https://www.linkedin.com/in/syed-yousuf-701454178/> |
<https://yousufport.herokuapp.com>

Education

Sir Syed University of Engineering and Technology | (BSc) Software Engineering

2017 – 2020

Built a Deep Learning model to assist Doctors in diagnosing disease. **Image dataset.** Malaria cells, pneumonia and brain tumor. **Tabular dataset.** Diabetes, heart disease, kidney disease and breast cancer. Model predictions are saved to a database for further medical analysis. Used Dash and Flask to create a simple to use web application.

National College | Pre-Engineering

2014 - 2016

St. Patrick's High School | Computer Science

2004 – 2014

Certification

Coursera Python and Statistics for Financial Analysis. Completed July 2020

<https://www.coursera.org/account/accomplishments/verify/KMBHZ3MUQ96U>.

Skills

- Coding and libraries: Python, NumPy, Pandas, Scrapy
- Machine Learning: Scikit-Learn, TensorFlow, Keras
- Data visualization: Matplotlib, Seaborn, Plotly
- Web applications: Dash, Flask
- Database: MySQL

Data Science Projects

- Freelance work for a client. Built a CNN (Convolutional Neural Network) for a binary prediction. The model was predicting real face images vs face made by generative adversarial network the project was of a student test accuracy was 75 percentage.
- Freelance work for a client. Built a multipage Dashboard using Dash Plotly and for data processing Pandas framework was used.
- Built a CNN model on various image datasets including face expression, pneumonia, malaria. Productionalized the model as a web application.
- Freelance project that required accessing data from online sources. Built a web scraper to automate the data extraction using Python ScraPy.
- Performed sentiment text classification using a RNN (Recurrent Neural Network) on various public datasets. Productionalized the model as a web application.

Data Science Experiments

I have primarily used Kaggle datasets to hone my data analytical skills. This was supplemented by freelance client work conducted during term holiday.

EDA (Exploratory Data Analysis). Performed EDA on a wide range of datasets including: Coronavirus, hotel booking, sales, diamond, and heart disease. Use Matplotlib, Seaborn and Plotly for data visualization. Created an interactive dashboard in Tableau:
https://public.tableau.com/profile/syed.yousuf5438#!/vizhome/TitanicBook_16014606357110/TitanicStory.

Neural Networks. Used CNN, LSTM, and Autoencoder models. Used GPU cloud services like Google Colab.

Time series. Aware of temporal train-test split of time-series data to avoid data leakage. Experimented with Neural Network LSTM architecture models.

Natural Language Processing. Improved model by removing stop words and punctuation.

Imbalanced data. Worked with imbalance data sets and applied resampling using the SMOTE library.

Machine Learning algorithms. Very familiar with classification and regression problems. Used K-Means to cluster unlabeled data for a customer segmentation problem.

Hyper parameter tuning. Used grid search on smaller data sets and random search on large data sets.

Machine Learning pipelines. Taken advantage of scikit-learn pipeline library to automate data preprocessing, model selection, k-fold cross-validation and hyper parameter tuning.