Fahad Mehfooz

Data Scientist | Delhi

✉ fahad.mehfoooz@gmail.com

☐ fahadmehfooz.github.io/Fahad_Portfolio/

○ github.com/fahadmehfooz

☐ +91-8285588545

in linkedin.com/in/fahadmehfooz

WORK EXPERIENCE

Software Engineer
Tata Consultancy Services Ltd.

11/2020 - 02/2022　　　　　　　　　　　　　　*Noida, India*

– Working extensively on databases. Handling monthly runs of critical projects. Optimized scripts by using better SQL hints to cut down on script runtime by 70%.

– Working on data preparation for the machine learning models. Also working on data analysis to find the relationship between large datasets.

Junior Machine Learning Engineer
Omdena

04/2021 - 07/2021　　　　　　　　　　　　　　*Remote*

– Developed a regression model to predict the **ROI** of a building when the existing lighting system is upgraded to a new cost-efficient lighting system. ☐

– Filling missing values in features by KNN imputation. Predicting ROI for the buildings using a **random forest regressor.**

– Performed hyperparameter tuning to get the best parameter values. Achieved an accuracy of **82%** on the test data and an r2 score of **0.86**.

Python Instructor
Cuemath Online Classes

12/2020 - 04/2021　　　　　　　　　　　　　　*Remote*

– Led a classroom of students from class 8th to 12th on the basics of python programming.

Data Engineering Intern
Convosight

03/2020 - 04/2020　　　　　　　　　　　　　　*Delhi, India*

– Worked on data warehousing, wrote scripts for mass extraction of data, and updated redshift and other warehouses.

PROJECTS

Health Insurance Cross Sell Prediction (11/2021 - 12/2021) ☐

– Developed a binary classification model with **Gradient Boosting** as a baseline model to predict if existing policyholders in an insurance company would be receptive to buy vehicle insurance.

– **Target encoding** to handle features with large number of categories and one hot encoding the remaining categories.

– Class imbalance was observed, for every sample of 'customer is interested', 7.15 samples of 'not interested' are available. Used **SMOTE** boosting to oversample the minority class.

– Achieved an improved accuracy of **84% ROC-AUC** score on the test set with hyper-tuned LGBM classifier applied on PCA data do as to remove overfitting.

– Applied **SHAP** and **ELI5** plots such as dependence plots to understand feature interactions and summary plots to identify the important factors responsible for customer involvement.

Ted Talk Views Prediction (10/2021 - 11/2021) ☐

– Developed a regression model using **Gradient Boosting** algorithm to estimate the number of views on TED video with a **MAPE** score of **38%** on the test set.

– Used **Google's** pre-trained **Word2Vec** model to vectorize 'Transcript' feature in the dataset which held the maximum amount of textual information.

– Implemented **IQR** method to filter out outliers from numerical features and tuned the hyper-parameters of the model using **GridSearchCV.**

– Applied **SHAP** to understand important features influencing views taking XGBOOST as baseline model.

Malaria Classification (09/2021 - 10/2021) ☐

– Developed an image classification model with **pretrained VGG19** on Malaria patient images to identify presence of malaria in the patient.

– Performed data augmentation using techniques like: **rotation, translation, zooming, cropping, and flipping.**

– Achieved an improved accuracy of **94%** with **pretrained VGG19** on test data over baseline model developed using **standard CNN's.**

EDUCATION

Bachelor Of Technology
Guru Gobind Singh Indraprastha University

08/2016 - 10/2020

77 %

Senior Secondary (XII)
St. George's School

04/2015

86 %

TECH STACK

Languages
Python, C++, C, Java, R.

ML Frameworks
pandas, NumPy, Matplotlib, Seaborn, Plotly, scikit-learn, Keras, PyTorch, NLTK.

Databases
MySQL, PostgreSQL, PL/SQL.

ACHIEVEMENTS

Kaggle Master (10/2020 - Present) ☐
I rank among the top 75. *Current Rank: 75/ 2,00,900.*

Paper presentation (07/2020)
Presented paper titled, 'Medical Chatbot for Novel COVID-19', at ICT4SD Conference, at Goa.

PUBLICATIONS

Research Paper
Medical Chatbot for Novel COVID-19 ☐
12/2020
ICT Analysis and Applications: Springer

Review Paper
In what scenarios are machine learning models safe to apply? ☐
03/2019
Journal of Emerging Technologies and Innovative Research

RELEVANT COURSEWORK

AlmaBetter: Full Stack Data Science Specialization (07/2021 - Present)

Udacity Nanodegree: Data Science with Python (05/2020 - Present) ☐

IBM Data Science Professional (06/2020 - Present) ☐