

# ALI FARID KHAN

*Data Science Intern*

## CONTACT



+91 93925 90178



alifaridkhan313@gmail.com



Virar(W), Mumbai-401303

## SOCIAL PROFILES

[GitHub](#)

[Kaggle](#)

## EDUCATION

### University of Mumbai

2022 - Present

B. Sc. Data Science

### Mount Carmel High School

2020 - 2022

High School (PCMB)

## SKILLS

- Python, R
- Data Preprocessing (EDA, FE, FS)
- ML (Supervised & Unsupervised)
- Regression & Classification Algorithms
- Matplotlib, Seaborn
- Hyperparameter Tuning
- DL (ANN, CNN, RNN, LSTM, GRU, etc)
- Scikit-Learn, TensorFlow, Keras, PyTorch, etc
- NLP
- Git & GitHub
- Databases (MySQL, NoSQL(MongoDB), PostgreSQL)

## PROFILE

Hardworking and enthusiastic in the world of Data Science. Seeking to utilize my strong foundation in statistical modeling, programming languages (such as Python (with frameworks) and R), and data visualization (EDA) to effectively serve your company in an internship position. Dedicated and committed to becoming a dependable and valuable team member, leveraging my skills to drive business growth and inform data-driven decisions.

## PROJECTS

### TITANIC DATASET

I worked on Titanic dataset, performed Data Preprocessing (EDA, FE, FS). Used various classification algorithms. XGBoost and Random Forest was performing best and giving almost same accuracy and With 80% train and test accuracy.

### CIFAR-10 DATASET

CIFAR-10 dataset is famous for image detection prediction of 10 classes in deep learning. Used simple basic and advance ANN and CNN. It was my biggest project as i worked on it for 100+ hours. It took runtime of 9907.6s on GPU P100.

### MNIST DATASET

Fashion MNIST is again a famous image detection dataset of jackets, shoes, shirts and flops, etc in deep learning. I used Simple RNN, LSTM-RNN, GRU in this and LSTM RNN was performing the best. Used GPU Tesla T4 x2, it took a runtime of 1794.8s.

### MALL CUSTOMER DATASET

This notebook is on unsupervised learning algorithms upon dataset of mall that clusters the data in 4 clusters and it also contains PCA algorithm which is used for handling multiple features aka The curse of dimensionality.

### GOOGLE STOCK DATASET

In this notebook i had played with Simple RNN, GRU and LSTM and compared their train loss and valid loss. With the help of predicted and actual GRU and LSTM was performing best. GRU was better among all with no fluctuations.