

# FARAZ AHMAD KHAN

M.Tech, Industrial & Management Engineering  
Indian Institute of Technology Kanpur

Email: [farazdob14@gmail.com](mailto:farazdob14@gmail.com) | [farazamd@iitk.ac.in](mailto:farazamd@iitk.ac.in)  
Contact No.: 8527253218  
LinkedIn: [linkedin.com/in/faraz-khan-a6404a18b](https://www.linkedin.com/in/faraz-khan-a6404a18b)

## ACADEMIC DETAILS

YEAR	QUALIFICATION	EDUCATIONAL INSTITUTE	PERCENTAGE
2019-21	M.Tech, Industrial & Management Engineering	Indian Institute of Technology Kanpur	7.81 CPI*
2013-17	B.Tech, Mechanical Engineering	Jamia Millia Islamia, New Delhi	7.72 CPI
2013	Class XII (CBSE)	Central Academy, Lucknow	79.80 %
2011	Class X (CBSE)	Central Academy, Lucknow	8.80 CPI

## ACADEMIC PROJECTS AT IITK

### Data Mining and Knowledge Discovery

#### Movie Review Sentiment Analysis | NLP

(Aug'19 to Nov'19)

- To **Classify Movie Reviews** into 5 classes of sentiments, the dataset was having **156,061** tuples
- Performed **Stemming**, text cleaning, used **Up-sampling**, made Word Cloud, used **Bag of Words, tf-idf** and **n-grams**
- Accuracy for various models: **Decision Tree** = 58%, **Naïve Bayes** = 60%, **KNN** = 62%, **Logistic Reg.** = 64% and **Random Forest** = 64%
- Used Hyper-parameter tuning and voting **Ensemble** to achieve final test accuracy = 65%

### Statistical Modelling for Business Analytics

#### Identify Best Subset of Predictors that Effect Student Test Score | Regression

(Jan'20 to Feb'20)

- To develop a **Multivariate linear regression** model using best subset of variables out of 30 variables to predict test score
- Performed **One Hot Encoding**, tested **Heteroscedasticity** using **BP-test** and Residual plots, Multicollinearity using **VIF**
- Used **Forward Attribute Selection** to finalise an 8 variable model and achieved **Adjusted R<sup>2</sup>** value of 0.29

#### Room Occupancy Detection using Binary Classification | 0-1 Classification

(Mar'20 to Apr'20)

- Developed a binary classifier using **Logistic Regression** model with **20,560** data points
- Performed scaling, used **L1 regularization** and used **p-value @ 95%** significance level for Variable selection
- Results: Accuracy = 92%, **Precision** = 78%, Recall = 93%, **F1 score** = 0.85 and AUC under **ROC curve** = 0.93

#### Fashion Apparel Classifier using CNN | Classification

(June'20 to July'20)

- To Classify **Images** of fashion apparels into 10 classes using data having **70,000 gray scale images** pixel information
- Trained **Convolution Neural Network** with **Keras** as **CONV-POOL-DROP-CONV-POOL-FC-FC** architecture
- Results: Accuracy = 91%, improved it by 0.05% by applying **Dropout** and increasing kernels

#### Time Series Forecast using ARIMA/ARMA model | Time Series

(June'20 to July'20)

- To **Forecast Number of Female Births** and **Number of Inventory Items** based on past data
- Used **Dickey-Fuller test, ACF, PACF** plots, **AIC** value in auto ARIMA **grid search** to get AR, I and MA parameters
- Finalised **ARIMA(1,1,1)** model for non-stationary(Inventory data) and **ARMA(2,2)** model for stationary(Female births) data
- Results: **RMSFE** = 7 for stationary data which is **17% of mean** and 7789 for non-stationary data which is **0.59% of mean**

## INTERNSHIP

### Marketing and Data Science Intern at Harvesting

(Apr'20 to June 20)

#### Objective: Twitter **data scraping**, **user tagging** based on location and **user matchmaking**

- Proposed, brainstormed and implemented a marketing strategy for **Business outreach with zero investment**
- Finalised and Implemented the proposal to connect customers and sellers using a **periodically triggered twitter bot**
- Developed code to **scrape tweets**, followers, location information and **tag users** with automated greeting messages
- Learnings: **Team work, Conflict resolution**, used libraries: NLTK, re, pandas, tweepy, twitter API, twitterscraper, query\_tweets

## COURSEWORK AND SKILLS

Data Mining and Knowledge Discovery	Statistical Modelling for Business Analytics	Operation Research for Management
Probability and Statistics	Analytics in Transport and Telecom	Advanced Decision Models
Operations Management	Introduction to computing-JAVA	Business Process Management

Python (Numpy, Pandas, Sklearn, Matplotlib) | SQL | **Machine Learning** | Statistical Analysis | Natural Language Processing | **Deep Learning**

## POSITIONS OF RESPONSIBILITY AND ACHIEVEMENTS

- Teaching Assistant for **130 students** for the Data Mining and Knowledge Discovery course, IME Department, **IIT Kanpur**
- Orientation Team Member, **Counselling Service**, IIT Kanpur, responsible for managing hall allocation to new admissions
- Secured All India Rank 1500 in GATE 2019 (**Mechanical**) out of 167376 total candidates