Hassan Ali Khan

239 Kearny Avenue, Kearny, New Jersey, 07032 | +1 (917) 674 7848 | hak27@njit.edu



# EDUCATION

New Jersey Institute of Technology Newark

**Master of Science in Data Science** Sep 2018-present

Lahore University of Management Science (LUMS) Lahore

**Bachelor of Science in Computer Science** Aug 2014-May 2018

Relevant Courses

Machine Learning, Deep Learning, Applied Statistics, Data Mining, Data Analytics, R programming, Advanced Programming, Software Engineering, Computer Vision, Databases, Operating Systems

# SKILLS

**Programming Languages:** C, C++, Python, MATLAB, HASKELL, Java, Socket Programming, SQL, R

**Data**: Data Analytics, Data Wrangling, Statistical Inference, Data Modeling (Decision Tree, KMeans Clustering, SVM, Gaussian Naïve Bayes, Neural Network), Machine Learning (Supervised and Unsupervised Learning), Deep Learning, Data Visualization, Data Reporting, Data Engineering,Natural Language Processing, Data Mapping, A/B Testing

**Technical**: scikit-learn (sklearn), Pandas, NumPy, TensorFlow, Pytorch, SQL, MapReduce, GitHub, GitLab, Tableau

# EXPERIENCE

**Lahore University of Management Sciences** Lahore

**Undergraduate Research Assistant:** Aug 2017-May 2018

**Senior Year Project :** ’YouTube Kids Explicit and Forged Content Detection’

Successfully detected explicit and/or inappropriate content as well forged content in YouTube Kids videos using a deep learning model in Keras and TensorFlow Python libraries that takes in video features such as its individual frames as input. Used skicit, sklearn, NumPy, matplotlib and pandas Python libraries for data exploration, data cleaning and data visualization. Achieved 95% accuracy.

**Educative Inc.** Lahore

**Machine Learning Internship** May 2018-August 2018

Successfully developed a machine learning algorithm using convoluted neural networks (CNN) to recommend courses to the customers that are best suited to their needs using python libraries such as Keras, TensorFlow, matplotlib and

Pandas.

# PROJECTS

* **TensorFlow:** TensorFlow is an open source software library for numerical computation using data flow graphs; primarily used for training deep learning models. Worked on APIs and performance for training models on Tensor Processing Units (TPU).
* **PLAsticc Astronomy on Kaggle:** Performed data cleaning and exploration on the astronomy dataset on Kaggle and tested using neural networks for time series classification. Used skicit, sklearn, numpy, seaborn and pandas Python libraries for data exploration, data cleaning and data visualization.
* **K-means clustering:** Implemented K-Means clustering (PAM) from scratch and used the RFM (Recency, Frequency and Monetary) model to select attributes to cluster customers into different group.
* **Support Vector Machines:** Used sklearn library to implement support vector machines and feature selection to test on a simulated dataset of single nucleotide polymorphism (SNP) genotype data containing 29623 SNPs (total features) and achieved accuracy of 80%.
* **Generative Adversarial Networks (GAN):** Generated fake faces of celebrities to fool the discriminator into classifying a fake image as real image. Created a discriminator and a generator function using TensorFlow python library
* **Recommender System:** Developed a music and movie recommender system using collaborative filtering on public datasets