# MOHAMMAD YUSUF ALI

**GitHub | LinkedIn | myusufali2015@gmail.com | 8800977296 | Noida**

I am a Lead Data Scientist having over 4.5 years of experience in creating intelligent products and solutions leveraging Machine Learning (ML), Deep Learning (DL), Natural Language Processing (NLP) and Computer Vision (CV). I have been involved in complete project cycles - Problem Definition, Research and Discovery, Experiments, Architecture Design, Implementation and Deployment.

## WORK EXPERIENCE

### NEWGEN SOFTWARE TECHNOLOGIES LTD. | LEAD DATA SCIENTIST
**June 2017 – Present | Noida, India**

- **ID card data extraction product –** Worked with a team to create an Identity card data extraction solution leveraging YOLOv3 object detection techniques along with highly configured OCR engine and image processing techniques using OpenCV to increase extraction accuracy.
- **Risk Analytics solution -** Risky/Personal information from documents is identified using custom built named entity recognition. Custom LSTM based architecture is designed along with Custom Embeddings to get more than 5 percent F1 score improvement on state of art spacy and nltk.
- **Document Classification product -** Developed and deployed a solution for classification of structured and unstructured image data using a combination of different Machine Learning models like Convolutional Neural Networks (Layout based Classification), LSTM or SVM (Text based Classification) and Object Detection Model (Region based Classification).
- **Model Training Tool Backend -** Designed and integrated a backend server using Flask, CherryPy and REST APIs to support a UI where a user can train their Deep learning models. Data and Model pipelines are created to store user-uploaded data and trained-models in MongoDB databases.
- **Liveness Detection -** Designed and deployed a semi-active liveness detection algorithm using a webcam based on the distortions that occur on a human face based on the distance of the user from the webcam.

## PROJECTS

### SENTIMENT ANALYSIS ON MOVIE REVIEWS | GitHub | Kaggle
- Dataset: IMDB 50K Movie Review Dataset
- Model: BERT Tranfer Learning along with Bidirectional LSTMs
- Accuracy: Above 97% on Test Set
- Libraries: Tensorflow, HuggingFace Transformers and Pandas

### STOCK ANALYSIS ON REDDIT | GitHub
- Dataset: Real World Data collected through Reddit API from r/investing
- Model: Named Entity Recognition along with Sentiment Analysis
- Libraries: Spacy, Flair and Pandas

## CERTIFICATIONS

- **Deep Learning Specialization from DeepLearning.AI**
- **Natural Language Processing with Transformers in Python**
- **Master SQL for Data Science**

## EDUCATION

### JAMIA MILLIA ISLAMIA
B. TECH COMPUTER ENGG.
9.60 CGPA (July 2013 – May 2017)

### Dr. Virendra Swarup Memorial Public School, Kanpur
ICSE
10th: 88.14% (2010)
12th: 93.17% (2012)

## SKILLS

- Constructing Predictive Models
- Deep Learning
- Computer Vision
- Natural Language Processing
- Data Extraction
- Python
- R
- Keras
- Tensorflow
- Image Processing (OpenCV)
- Statistics
- Data Visualization (Power BI)
- Databases (SQL, MongoDB)
- REST APIs

## ACHIEVEMENTS

- Awarded Rising Star (Annual Award for Freshers) in Newgen Software Technologies in July 2018.
- Won the Cloudburst 1.0 hackathon at Newgen with a team of 6 in March 2019.
- Awarded Pat-on-Back (Monthly Award) in Newgen Software Technologies for August 2020.
- Awarded the Standing Ovation Award for July 2021 at Newgen.