

Mohit Shukla

9643455396 | <https://mohitshukla.live> | iammohitshukla9@gmail.com | [linkedin.com/in/mohit](https://www.linkedin.com/in/mohit) | github.com/Mohit

EDUCATION

Galgotias College of Engineering and Technology

Master of computer Application

Machine Learning Certification by Stanford University(Coursera)

Greater Noida —IN

Aug 2017 – Oct 2020

Maharaja Agrasen College , Delhi University

Bachelor of Science (Physcial Science)

New Delhi —IN

Aug 2014 – May 2017

TECHNICAL SKILLS

Languages:Python(Numpy, Pandas, Scikit-learn, Seaborn, Keras), SQL(Postgres), HTML, R(Beginner)

Data science: A/B Testing, Data science pipeline(cleaning, wrangling, visualization, modeling, interpretation)
Statistics, Hypothesis testing, OOPs, Tableau

Analytics:Exploratory Data Analysis, Visualization, Machine Learning, NLP, Deep Learning(intermedite), Hadoop

Cloud :Linux, Git, Flask, Google Cloud Platform, Heroku

PROJECTS

Image Captioning using Deep Learning | *Python, visualization, CNN, Git*

Jan 2021 – Present

- An image classification model based on CNN Algorithm and Transfer learning Algorithm (InceptionNet) was trained on the Flickr 8K Dataset for which the captions were serialized with the name of the image in a new file.
- Text preprocessing was applied on the captions and tokens were generated from the word. Padding of the tokens was created to deploy it on the LSTM model.

Netflix Movie Recommendation System | *Python,Pandas, Matplotlib,Git*

July 2020 – Sep 2020

- Built a Recommender System using item-item similarity and collaborative filtering based models.
- Used 300k data points and 500 tags and Performed Preprocessing on text data using NLP.
- I created a sparse matrix from data frame and found Global average of all movie ratings
- The most similar movies was found using similarity matrix and Matrix Factorization Techniques were used.
- Used content based filtering with combination of KNN and XGBoost regressor to predict ratings of movies.
- Got final RMSE score of 1.0675.

Quora Question Pair Similarity | *Python, NLP, Fuzzywuzzy, Seaborn*

Jan 2020 – Mar 2020

- Identify which questions asked on Quora are duplicates of questions that have already been asked.
- Binary-class classification problem, Used approx 300k question pairs as a data
- I also explored Advanced Feature Extraction (NLP and Fuzzy Features) with hyperparameter tuning. I also generated a WordCloud Generation.
- Reduced test log-loss from 0.88 to 0.36. Best Performing model is XGBoost.

Stackoverflow Tag Prediction | *Python, NLP, Machine Learning, Git*

Oct 2019 – Dec 2019

- Suggest the tags based on the content that was there in the question posted on Stackoverflow.
- Used 300k data points and 500 tags and Performed Preprocessing on text data using NLP.
- Used TF-IDF to featurize data and applied multi-class classification techniques such as One-Vs-Rest SGD with Logistic Regression and Linear SVM
- Achieved better Micro F1 score with One-Vs-Rest Linear SVM.

EXTRACURRICULAR ACTIVITIES

- Participated in a 12-week program by Google Cloud where I learned extensively about Google Cloud Platform and implemented cloud solutions via Quiklabs.
- Campus Ambassador role is an internship where I represented a Scaler Academy in Galgotias College of Engineering and Technology.
- Developed Salary predicting using Linear Regression and deploying it on Heroku.
- Participated in One month Data Science internship by The Sparks Foundation