



Aftab Anjum

Computer Engineer

Contact

 +9203078898056

 Anjumaftab.cs@gmail.com



About me

I am currently working as NLP Engineer at scenarios. My research interest involves: Natural Language processing, Evolution computation and its application in mathematical modeling.

EXPERIENCE

09/2020 – Present

NLP Engineer, Senarios | Lahore, Pakistan

- Performed preliminary data analysis using descriptive statistics and handled anomalies
- Involved in creating various regression and classification algorithms by using decision trees, NB, KNN, RF, SVM, regression models, LSTM, Transformer based models (Bert, Roberta, NERD), neural networks, clustering algorithms
- Developed solutions for an American-based company for Topic Extraction, Sentiment Analysis, text analytics, Name entity extraction, Binary, multi class and multi-label text classification, and Speech biometric.
- Worked on MLOPS, created and designed the architecture for models to be hosted on AWS cloud (AWS) and provide model scalability according to the user need.
- AWS services I have worked with and used till now: Sagemaker, EC2, Lambda, API Gateway, Dynamo DB, S3, Cloud Watch, SQS, IAM, Cron Jobs, Amazon Elastic Inference, Amazon Comprehend.

02/2018 – 01/2020

Research Assistance, University of Jinan | Shandong Jinan, China

- Worked and assisted on different project of natural language processing included text summarization, text classification, dimensionality reduction
- Worked on series of evolutionary Projects which are related to Genetic programming (GP, GEP, MEP, CGP and NEEP) and also worked on series of optimization algorithms (GA, PSO, DE and CMAES)

- Supervised the undergraduate student on research projects based on Genetic programming, NLP and Neural network
- Conducting literature review on NLP and Evolutionary Algorithms, provide summarization and in-depth institution to my supervisor and team members

04/2016 – 08/2016

Teaching Assistant, University of Lahore | Lahore, Pakistan

- Teach students the basic skills of C++
- Conducted labs and assisted students in semester projects
- Grad student Assignment and semester Quiz

EDUCATION

Graduated 06/2020

Master of Engineering in Computer Science and Technology

University of Jinan (UJN) | Shandong Jinan, China

Coursework includes: advance Engineering Mathematics, Machine learning Deep Learning and Evolutionary Computing and image processing

Graduated 04/2017

Master in Computer Science

University of Lahore (UOL) | Lahore, Pakistan

Coursework includes: Object Oriented Programming, Data Structures, Computer Networking, Database Management, Software Engineering, Theory of Automata, Analysis of Algorithm, and Artificial Intelligence

Graduated 08/2014

Bachelor of Science

University of Punjab (PU) | Lahore, Pakistan

Coursework includes: Physics, Chemistry, Mathematics and English

PUBLICATIONS

- **Aftab Anjum**, Fengyang Sun, Lin Wang, Jeff Orchard. "A Novel Neural Network-Based Symbolic Regression Method: Neuro-Encoded Expression Programming". 28th International Conference on Artificial Neural Networks, ICANN, 2019, 373-386. [\[Paper\]](#)
- **Aftab Anjum**, Mazharul Islam, Lin Wang. "Gene Permutation: A new Probabilistic Genetic Operator for Improving Multi Expression Programming". In 2019 IEEE Symposium Series on Computational Intelligence (SSCI), 2019, 3139-3146. [\[Paper\]](#)

- Mazharul Islam, **Aftab Anjum**, Tanveer Ahsan, Lin Wang. "Dimensionality Reduction for Sentiment Classification using Machine Learning Classifiers".

In 2019 IEEE Symposium Series on Computational Intelligence (SSCI), 2019, 3097-3103,

(Published as co-first author). [[Paper](#)]

Aftab Anjum, Mazharul Islam, Lin Wang. "Dimensionality Reduction for Sentiment Classification: Evolving for the Most Prominent and Separable Features". ArXiv preprint [[Paper](#)]

- **Aftab Anjum**, Mazharul Islam, Lin Wang. "Dimensionality Reduction for Sentiment Classification: Evolving for the Most Prominent and Separable Features".

Journal of Applied Soft Computing, 2020. [**Under Review**] [[Paper](#)]

SKILLS

- **Python:** Data wrangling, Exploratory data analysis, debugging using anaconda (Spyder and jupyter Notebook) and Matlab
- **Statistics/Machine Learning:** Statistical Analysis, linear/ logistic Regressor, Clustering, handling outlier and anomalies, Feature reduction, Feature selection, Bagging and boosting algorithms
- **Deep Learning Skills:** Clustering Models, Quantitative Analysis, Pytorch, SkLearn, Keras and OpenCV
- **Data mining:** SQL, MongoDB, S3, pandas, Seaborn, Matplotlib, Scipy, NLTK.
- **MLOPS:** Sagemaker, Lambda, API Gateway, Dynamo DB, S3, SQS, Coron jobs, EC2, IAM and Cloud watch

PROJECTS

- **IMDb Movie Reviews Sentiment Analysis:** To widen my area of expertise I have worked on a movie review data set for sentiment analysis. I tweak a Transformer based model (ROBERTA) by transfer learning and end to end learning. The model tries to classify the reviews based on their semantic meaning. I have achieved 86% on this data set.

Designed Model: ROBERTA

([Project detail and code](#))

- **Multi Class Text Classification:**

The project is about multi class text classification problem. After performing the data pre-processing and data cleaning I applied different machine learning and deep learning techniques to classify the tweet of the user according to their semantic information. I also optimized the hyper-parameters of RF by using Randomized Search CV.

Designed Model: Naive Bayes, Ada boost (AB), Gradient Boosting Classifier (GBC), Random Forest (RF), Deep Neural Network, SMOTE (handling class data imbalance)

[\(Project detail and code\)](#)

- **Named Entity Recognition for GMB Dataset:**

The project is about Named-entity recognition which locate and classify named entities mentioned in unstructured text into pre-defined categories. After performing some data analysis then I trained Bi-directional LSTM and CRF against two embedding Glove Embedding, forward and backward Flair embeddings.

Designed Model: Bi-directional LSTM and CRF

[\(Project detail and code\)](#)

- **Dimensionality reduction for sentiment classification:**

In this project we proposed a new method entitled as Sentiment Term Presence Count (SentiTPC) and Sentiment Term Presence Ratio (SentiTPR), These techniques reject the features by considering term presence difference and ratio of the distribution distinction for SentiTPC and SentiTPR. Respectively. In order to determine the effectiveness of this method I used 4 different classifiers (LR, SVM, NB, RF) for five high dimensional textual data set.

- **The Social Dilemma Tweets Text Classification:**

The project is about raw text classification. After the Data wrangling and some Exploratory data analysis then I applied Bidirectional LSTM techniques to classify the tweet correctly. I also perform few data imbalance techniques in-order to handle class data imbalance.

Designed Model: Glove pre-trained Embedding, Bidirectional LSTM, SMOTE (handling class data imbalance)

[\(Project detail and code\)](#)

My all-other **Designed projects** and their code can be seen on my [GitHub](#)

Online Courses & Certification

- **Python 3 Programming**

University of Michigan, Coursera 'Specialization', 2020 ([See credential](#))

- **Mathematics for Machine Learning: Linear Algebra**

Imperial College London, Coursera, 2020 ([See credential](#))