

Arushi Sharma

☎ (+91) 9205858274 | ✉ arushi.net@outlook.com | 📱 [arushi-08](#) | 💻 [arushi--sharma](#)

Experience

Optum Global Advantage

Hyderabad, India

ASSOCIATE DATA SCIENTIST · BACKEND PYTHON PROGRAMMER · AI & ML TEAM

Aug. 2020 - Present

- Developed and maintained Document Processing Pipelines for a Medical Document Indexing Site in Python using Celery, Flask and Pandas.
- Built a Text Similarity feature to find the similar pages between two documents using Cosine and Levenshtein Similarity metrics.
- Built a Date Search feature using Regex to extract various date format types from documents.
- Designed Flask APIs to analyze daily Data Inventory Metrics, Document Processing Time, and Disk Storage.

IHS Markit Ltd.

Gurugram, India

DATA TRANSFORMATION INTERN · ENGINEERING AND PRODUCT DESIGN TEAM

Jun. 2019 - Aug. 2019

- Developed a Web Crawling Automation tool to navigate through websites and extract information, using Selenium Frameworks of Python. This decreased data extraction time by 22%.
- Programmed data files to a set format, using NumPy operations on Pandas DataFrames, and built a GUI using Tkinter Python package.
- Deployed it as a full-fledged desktop application, using cx_Freeze.

Projects

Feature Enhanced Capsule Networks for Robust Automatic Essay Scoring.

PUBLISHED IN ECML PKDD 2021: MACHINE LEARNING AND KNOWLEDGE DISCOVERY IN DATABASES. APPLIED DATA SCIENCE

2021

- Proposed a novel architecture with two key components: BERT enhanced CapsNet model and the Feature Extraction model.
- Boosts the overall average Quadratic Weighted Kappa metric to 81%.
- The proposed work successfully scores most of the adversarial text lower.

Ceasing hate with MoH: Hate Speech Detection in Hindi-English Code-Switched Language

ACCEPTED IN INFORMATION PROCESSING AND MANAGEMENT JOURNAL

2021

- Classified social media tweets as not offensive, abusive and hate inducing.
- Extracted features from tweets: TF-IDF Word level vectors, Word Level N-gram Vectors, and employed Machine Learning Classifiers.
- Developed a Language Transliteration Mapper and finetuned mBERT and MuRIL classifier models.
- Compared the Empirical results of all the methods and reported scores higher than state-of-the-art precision, recall and F1 scores.

English-Hindi Neural Machine Translation-LSTM Seq2Seq and ConvS2S.

PUBLISHED IN ICCSP'20 IEEE CONFERENCE

2020

- Cleaned and tokenized text using Python NLP libraries: NLTK library and regex.
- Developed the language translator in Python using ConvSeq2Seq and LSTM Seq2Seq deep learning models.
- Compared these models on the basis of BLEU metric score, training time and hardware utilization.

Kaggle Problems

SENTIMENTAL ANALYSIS ON SOCIAL MEDIA TEXTS

2019

- Performed text preprocessing using NLTK and regex and employed ML, CNN and LSTM models on the preprocessed text.
- Evaluated performance on Accuracy, Precision, Recall and F1 metrics.

ICU REQUIREMENT PREDICTION

2019

- Analyzed Kaggle dataset to find the pertinent independent factors for ICU Requirement (y value).
- Reported best accuracy scores with Logistic Regression model and built Flask APIs rendering HTML templates.

Education

Delhi Technological University

New Delhi, India

B.TECH. IN ELECTRONICS AND COMMUNICATION ENGINEERING

Aug. 2016 - June 2020

- 7.91/10 CGPA · Smart India Hackathon Winner · TEDxDTU 2019 Corporate Head · Dance Society Council Member

Indian Language School

Lagos, Nigeria

HIGH SCHOOL DIPLOMA · CENTRAL BOARD OF SECONDARY EDUCATION

Mar. 2016

- 92.5% in Class 12 · 10 CGPA in Class 10 · Awarded Sportsperson of the Year.

Skills

Programming Python · Java · Bash · Git **Frameworks:** Pandas · NumPy · Flask · Celery

ML & NLP Scikit-Learn · Pytorch · NLTK · Keras

DevOps OpenShift · Jenkins · Docker **Database:** MongoDB · Redis **Testing:** Selenium

Certifications

National University of Singapore Big Data Analytics with Artificial Neural Networks

Hewlett Packard Enterprise Hands On Training in Big Data and HADOOP

Coursera Machine Learning: Regression and Classification