Arushi Sharma

☎ (+91) 9205858274 | ✉ arushi.net@outlook.com | arushi-08 | in arushi--sharma

Experience

Optum Global Advantage   *Hyderabad, India*
ASSOCIATE DATA SCIENTIST · BACKEND PYTHON PROGRAMMER · AI & ML TEAM   *Aug. 2021 - Present*

- Developed Document Processing Pipelines for a Medical Document Indexing Site in Python using Celery, Flask and Pandas.
- Built a Text Similarity Feature to find the delta between two documents using Cosine and Levenshtein Similarity metrics.
- Designed Flask APIs to evaluate Data Inventory Metrics and Daily Reporting.
- Reviewed feature enhancements and Flask APIs.

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 program decreased data extraction time, of the firm, 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 PAPER 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 used Machine Learning Classifiers: Naive Bayes, Logistic Regression, SVM Classifier, Random Forest Classifier, XGBoost Classifier, AdaBoost Classifier and Gradient Boosted Decision Trees.
- Developed a Language Mapper and used Tensorflow's BERT classifier model.
- 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 processed 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.

Sentimental Analysis on Social Media Texts

KAGGLE PROBLEM   *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.
- Designed a GUI for the model and made it a user-friendly Sentiment Detection Application using Tkinter.

Education

Delhi Technological University   *New Delhi, India*
B.TECH. IN ELECTRONICS AND COMMUNICATION ENGINEERING   *Aug. 2016 - Till Date*

- 7.91/10 CGPA · Smart India Hackathon Winner

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 · NumPy · Pandas · Flask · Celery · Java · Bash
ML & NLP	Scikit-Learn · Pytorch · NLTK · Keras
Web Crawling	Request · Selenium · BeautifulSoup
Tools	OpenShift · MongoEngine · Docker · Jenkins

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