Sumon Kanti Dey

https://github.com/SumonKantiDey
https://www.kaggle.com/sumon23
https://www.linkedin.com/in/sumon-kanti-dey-96321b10b
sumonkantidey23@gmail.com | +8801879862842

EDUCATION

NOAKHALI SCIENCE AND TECHNOLOGY UNIVERSITY | B.Sc. in Computer Science
Passed Dec 2018 | Noakhali,Bangladesh • GPA: 3.32 / 4.00
Research Interests: Natural Language Processing, Machine Learning, Deep Learning, Information Retrieval, Sensemaking.
Related Courses: Data Structure and Algorithm, Digital Image Processing, Artificial Intelligence and Neural Networks, Software Engineering and Information System Design, Database Management, Andrew NG's Machine Learning on Coursera,Linear algebra.

EXPERIENCE

LEADBOOK.COM | Engineer Scientist
Apr 2020 - Present | Singapore

- Work on the back-end data acquisition and data pipelining. Improve data reliability, efficiency and quality.
- Build large scale Apache Kafka based email generation and validation system which will generate email pattern from user and company domain and recommend best email pattern to customer.
- Design, refactor and developed backend APIs using Flask and AWS.
- Developed email classification module which will take an email and classified it as a business or private email.
- Design rule based information retrieval tool which will crawl company postal address from company website meta data.

HIFI DIGITAL LTD | Python Developer
Mar 2019 - Present | Banani, Dhaka

- Developed large scale information Retrieval tool using Selenium and Beautiful Soup.
- Bootstrapping production ready machine learning models.
- Present information using data visualization techniques like Tableau, Plotly.
- Data Engineering and pipe-lining the data in python.
- Processing, cleansing, and verifying the integrity of data used for analysis.
- Design, development, refactor and deploy backend API in Django and AWS.

TECHNICAL SKILLS

Languages : Python, C/C++, Java, JavaScript, D3.js, SQL, Prolog, Swift, LaTeX.
Frameworks/Packages : BeautifulSoup, Selenium, Git, Tableau, Flask, Django, Apache Kafka.
Libraries and Tools: Scikit-learn, Pytorch, Keras, NLTK, Matplotlib, Jupyter, Spyder, Tweepy, Pandas, AWS.
Operating System : Windows, Linux.

PROJECTS

JIGSAW MULTILINGUAL TOXIC COMMENT CLASSIFICATION [github]

- The target of this competition is to identify toxicity in online conversations.
- Used RoBERTa XLM pre-trained models and different architectures on top of XLM-Roberta.
- The small-scale model finished at 151th place out of 1621 teams.

QUORA INSINCERE QUESTIONS CLASSIFICATION [github]

- Applied 5 folds stack attention based Bi-GRU with different embedding (Glove, Word2Vec, FastText).
- Build a web app to see real time sincere vs insincere question classification.

BENGALI.AI HANDWRITTEN GRAPHEME CLASSIFICATION [github]

- Multi class classification challenge where need to classify three constituent elements from an image: grapheme root, vowel diacritics, and consonant diacritics.
- 64x64 pixels with one single channel image fit into a MultiOutput CNN model to classify three class from the image. Build a Flask WebApp which will help to classify real time Grapheme with three constituent elements.

RNN AND CNN NETWORKS FOR SARCASM DETECTION VIA AUXILIARY AND EMOTION FUSION | bachelor thesis

- Collected 32k tweets with sarcastic and non-sarcastic marker and developed an annotated sarcasm dataset.

- Proposed an improved context, contrast and semantic modeling network for sarcasm detection.
- Proposed a noble fusion methodology (Handcrafted text features for stylistic pattern extraction and Deepmoji pretrained vectors for emotion mapping for short text) with convolution and recurrent neural networks.

REAL TIME ROAD CONDITION MEASUREMENT SYSTEM

- Designed a system can detect Crack, Pothole, Faded severity from road image.
- Implemented Faster R-CNN InceptionV2 model pretrained on COCO large-scale object detection dataset.
- Used Django Rest framework and AWS for deploying machine learning model.

APTOS 2019 DIABETIC RETINOPATHY DETECTION TO STOP BLINDNESS

- Used Pre-trained EfficientNet with fine-tuning for classification task.
- The small-scale model finished at 199th place out of 2931 teams.

IRON ORE FUTURE PRICE PREDICTION ON MULTIVARIATE TIME SERIES DATA [GITHUB]

- Implemented Deep Learning and time series (ARIMA, SARIMA, VAR, VARMAX ,Facebook Prophet) model to forecast the future price of Iron Ore.
- In backend used Django Rest framework and AWS for deploying the forecasting model.

TOXIC COMMENT CLASSIFICATION CHALLENGE ON WIKI DATA

- Six types of of toxicity detection from wiki conversation text using Logistic Regression with TF-IDF words and char n-grams. Accuracy 0.9754.
- Implemented simple (Bi-GRU+LSTM+CNN+Pooling) with Pre-trained FastText Word Embeddings which was outperformed previous one model. Accuracy 0.9828.

SELECTED PUBLICATIONS

- Md. Rezaul Karim, Sumon Kanti Dey , and Bharathi Raja Chakravarthi, "DeepHateExplainer : Explainable Hate Speech De-tection for Under-resourced Bengali Language", under review at IEEE Access, December 2020. [arXiv:2012.14353v1]

ACHIEVEMENTS AND COMPETITIVE PROGRAMMING

I have joined several national and regional programming contest and solved almost 700 online judge problems in Lightoj,UvaOJ, Codeforces, Codechef, Spoj. I have completed Andrew NG's machine learning Course on Coursera.

- Kaggle Competition Expert (Ranked 3471 among 144,292 kagglers in the world). 2020
- Earn a Certificate for completing Andrew NG's machine learning course on Coursera. 2018
- 24th place at SUB Inter University Programming Contest. 2017
- 17th place at DIU ACM ICPC Dhaka Region Worlds finals Warm-up. 2016
- 1st place at inter University programming Contest. 2016
- Honorable mention at ACM ICPC Asia Dhaka Regional Contest. 2016
- 1st place at CSTE carnival'15 Programming Contest,NSTU. 2015
- Top 1% in Noakhali S&T University Entrance Exam. Ranked 54 among thousands participants. 2014

SOCIETIES

- Served as the General Secretary at departmental computing club. Help to grow up a community of 150+ people. 2017-2018
- Organized 8+ programming competitions, 1 boot-camp, 6+ workshops and 10+ seminars.
- Chief Judge at Intra NSTU PI Day Programming Contest(26 teams). 2018
- Organizer and Mentor at Junior Programming Boot-camp goicpc18 [45 participants]. 2017
- Workshop Trainer at CSTE Club Competitive Programming. 2016-2017
- Judge and Problemsetter at NSTU Intra CSTE programming Contest [30 teams]. 2017

TALKS

- Introduction to Machine learning and Kaggle. April 2020
- Choice of a Career on becoming a Potential Computer Science Grad at Noakhali S&T University. Oct 2019

WRITINGS

HTTPS://SUMONKANTIDEY.GITHUB.IO