Kaushik Amar Das

Email: kaushikamardas@gmail.com
Website: https://cozek.github.io/
Github: www.github.com/cozek
Linkedin: www.linkedin.com/in/kaushikamardas

RESEARCH WORKS

Ensemble of ELECTRA for Profiling Fake News Spreaders *May - June 2020*
Created an ensemble of 15 ELECTRA models for the task of identifying Possible Fake News Spreaders on Twitter in PAN at CLEF 2020. Our approach scored an accuracy of 0.70 and 0.69 on the English and Spanish test set respectively.
Paper: https://pan.webis.de/downloads/publications/papers/das_2020.pdf
Code: https://github.com/cozek/profiling-fake-news-spreaders

Checkpoint Ensemble of Transformers for Hate Speech Classification *Feb - May 2020*
Used Checkpoint Ensembling to create ensembles of transformers for building hate speech detection systems in OffensEval 2020 at SemEval'20. We showed that checkpoint ensembling can improve the performance of these systems. The transformers used in these systems were GPT2, RoBERTa and DistilRoBERTa.
Paper: https://www.aclweb.org/anthology/2020.semeval-1.267.pdf
Code: https://github.com/cozek/OffensEval2020-code

Multimodal Classification of Internet Memes *May 2020*
Analysed two approaches for the internet meme classification challenge of Memotion Analysis at SemEval'20. The first approach combined features extracted from DistilRoBERTa and EfficientNet, while the second approach used only EfficientNet. Our classifier achieved 9'th Rank for the Humor Classification task of the challenge.
Paper: https://www.aclweb.org/anthology/2020.semeval-1.152.pdf
Code: https://github.com/cozek/memotion2020-code

Multilingual Automated Aggression Identification *Feb - May 2020*
Joint project with Arup Baruah where we worked on multilingual automated aggression Identification in TRAC-2 at LREC'20. I built the RoBERTa based models for this work. My XLM-RoBERTa model ranked 2nd out of 10 in the Misogynistic Aggression Identification subtask for the Hindi language.
Paper: https://www.aclweb.org/anthology/2020.trac-1.12.pdf
Code: https://github.com/cozek/trac2020_submission

Transfer-Learning for Detection and Classification of Hate Speech *Aug - October 2019*
Built Stacked BiLSTM and CNN based models for detecting hate speech in Hindi, English and German languages at HASOC 2019 at FIRE'19. We used transfer learning to analyse whether a model trained to do the binary classification of hate speech can be fine-tuned for fine-grained classification of the same.
Paper: http://ceur-ws.org/Vol-2517/T3-19.pdf
Code: https://github.com/cozek/hasoc-2019-falsepostive

Hate Speech Detection In Social Media – Masters Thesis *Aug 2019 - June 2020*
Under the supervision of Dr Ferdous Ahmed Barbhuiya, IIITG
The thesis studies detection of hate speech in social media. The thesis compiles relevant literature in the domain in addition to exploring techniques for detecting hate speech in social media. The thesis also works on the detection of online aggression and differentiating hate speech from other sentiments in internet memes.
Thesis and code: https://github.com/cozek/hate-speech-detection-social-media

ACHIEVEMENTS

MTech Batch Topper, IIITG for the class of 2018-2020

Ranked 2nd out of 10 in Misogynistic Aggression Identification subtask in the Hindi language using XLM - RoBERTa in the second workshop on Trolling, Aggression and Cyberbullying - 2020.

Ranked 9th out of 29 in Humor Classification task of SemEval 2020 Task 8, Memotion Analysis, using EfficientNet.

EXPERIENCE

Technology R&D Specialist - Internship *January 2021 - July 2021*
Accenture Technology Labs, Bangalore, India

P.hD. Scholar *August 2020 - Present*
Indian Institute of Information Technology, Guwahati

Summer Internship *July 2017*
Indian Oil Corporation Limited, Guwahati Refinery

OTHER PROJECTS

Smart Garbage Management and Monitoring System *March-May 2019*
Class Project under Instructor Asst. Prof. Rakesh Matam, IIIT Guwahati
Developed a smart garbage management system that would monitor the status of the disposal sites and alert the concerned authorities when the sites needed emptying in real-time.
Technology Used: Python, Raspberry Pi, AWS EC2, AWS SNS, AWS IoT Core

Twitter Sentimental Analysis using Apache Spark and Kafka *April - May 2019*
Class Project under Instructor Asst. Prof. Dip Sankar Banerjee, IIIT Guwahati
Mined tweets with certain hashtags from Twitter and pipelined it through Apache Kafka to Apache Spark's streaming API and analyzed their sentiment.
Technology Used: Docker, Apache Spark, Apache Kafka, Apache Toree, Scala

Result Portal *Oct - Nov 2018*
Class Project under Instructor Prof. Gautam Barua, IIIT Guwahati
Developed an online portal where the semester results of the students would be uploaded and managed.
Technology Used: XAMPP, PHP, MySQL, HTML, CSS, Bootstrap.

TECHNICAL SUMMARY

Programming	Proficient: Python Amateur: C++, Java Previous Experience: Scala, MATLAB, PHP, C
Markdown	HTML, CSS, Bootstrap
Environments	Windows, Linux
Software Tools	Jupyter Notebook, VSCode, Docker
Frameworks	Numpy, Pandas, Scikit-Learn, NLTK, Keras, PyTorch, Spark, Apache, Flask

EDUCATION

MTech in Computer Science and Engineering *CGPA: 9.60 / 10*
Indian Institute of Information Technology, Guwahati *July 2018 - June 2020*

BTech in Computer Science and Engineering *CGPA: 7.04 / 10*
Assam Science and Technology University, India *July 2014 - June 2018*

CERTIFICATIONS

Deep Learning Specialization *June 2019*
by deeplearning.ai on *Coursera*

Machine Learning *May 2019*
by Stanford University on *Coursera*

Cloud Computing Certification *Aug - Sep 2018*
by IIT Kharagpur on *NPTEL*

REFERENCES

Dr. Ferdous Ahmed Barbhuiya, Associate Dean (R&D), IIITG. ferdous@iiitg.ac.in

Dr. Kaustuv Nag, Assistant Professor (CSE), IIITG. kaustuv.nag@gmail.com