Amey Hengle

Pune, India | Professional Email | Personal Email | Linkedin | Github | Personal Website

RESEARCH INTERESTS

Deep Learning, Natural Language Processing, Multilingual NLP, Computational Social Science, and Dialogue Systems.

EDUCATION

Savitribai Phule Pune University (PVG's COET) — Pune, India
Bachelor's of Engineering in Computer Science, CGPA: 8.38/10 — 2016 – 2020

PUBLICATIONS (* DENOTES EQUAL CONTRIBUTION)

1. **Combining Context-Free and Contextualized Representations for Arabic Sarcasm Detection and Sentiment Identification.** [paper] ***Shared Task Runner up!***
Amey Hengle, Atharva Kshirsagar, Shaily Desai and Manisha Marathe.
EACL 2021 *Workshop on Arabic Natural Language Processing (WANLP).*

2. **Cluster Analysis of Online Mental Health Discourse using Topic-Infused Deep Contextualized Representations.** [paper]
Atharva Kulkarni, **Amey Hengle**, Pradnya Kulkarni, and Manisha Marathe.
EACL 2021 *Workshop on Health Text Mining and Information Analysis (LOUHI).*

3. **An Attention Ensemble Approach for Efficient Text Classification of Indian Languages.** [paper] ***Shared Task Winner!***
Atharva Kulkarni, **Amey Hengle**, and Rutuja Udyawar.
The 17^{th} *International Conference on Natural Language Processing (ICON* 2020*).*

4. **Smart Cap: A Deep Learning and IoT Based Assistant for the Visually Impaired.** [paper]
Amey Hengle, Atharva Kulkarni, Nachiket Bavadekar, Niraj Kulkarni, and Rutuja Udyawar.
The third IEEE International Conference on Smart Systems and Inventive Technology (ICSSIT 2020*).*

EXPERIENCE

Machine Learning Engineer — Aug 2021 – Present
***Skit.ai** [company website]* — *New York, USA*

- Currently working on the end-to-end design, implementation and deployment of SKIT's voicebot product.
- My primary responsibilities include improving the voicebot's NLU, NER, and Distress-Detection capabilities by implementing multilingual, multimodal (audio and text) PLMs.
- I am also working on projects like label-noise detection, unsupervised intent discovery, and domain-biasing our in-house speech-to-text software.

Data Science Engineer — April 2021 – July 2021
***Twimbit** [company website]* — *Noida, India*

- Lead the design and development of Twimbit's unsupervised topic discovery, semantic search similarity, and kubernetes-integration projects.
- Improved overall topic discovery by 35% after implementing a hybrid PLM, which combined feature vectors obtained from an LDA model with contextual embeddings from XLM-Roberta.

Research Colaborator — May 2020 – Present
***Cognitive and Behavioural Neuroscience Lab, IIT Bombay** [lab website]* — *Mumbai, India*

- Currently working on two research problems in the area of computational social science and clinical psychology.

- The first project entails the linguistic analysis and classification of Depression-Anxiety comorbid posts from Reddit.

- As part of the second project, I am working on explainable deep neural networks for depression classification from social-media posts.

Research Assistant Aug 2020 – Feb 2021
PVG's College of Engineering and Technology *Pune, India*

- Conducted research on discourse themes mining from online mental health communities.
- Proposed a unique data representation technique of *Topic-infused Deep Contextualized Representations*, which combines contextual embeddings from Pre-trained Language Models with topical information from LDA.
- The embeddings were generated using a deftly crafted autoencoder model that captured both common and complementary information from the source embeddings.
- Employed UMAP for dimensionality reduction and HDBSCAN to draw out prominent clusters.
- Creating a novel and comprehensive dataset of PTSD related posts on Reddit from 2015-2020.
- Designed a multi-input, multi-task, transformer based model for EACL WASSA 2021 Shared Task on empathy and distress score prediction ***(Shared Task Winner)***.

Research Intern (ML and NLP) Aug 2020 – Dec 2020
Optimum Data Analytics [company website] *Pune, India*

- Involved in the end-to-end R&D of BuddyBot, a stress-relieving chatbot for mental health patients.
- Improved the chatbot's topic and dialogue act classification pipeline using attention-based models.
- Devised deep learning based solutions for hate speech detection on social media.
- Developed an attention ensemble CNN-BiLSTM model for the ICON 2020 TechDoFication Shared Task ***(Shared Task Winner)***.

Bachelor's Thesis Intern Aug 2019 – April 2020
Optimum Data Analytics [company website] *Pune, India*

- Worked on their flagship venture, 'Bindu Smart Cap' – a multimodal AI agent designed to assist visually impaired people. [video demo]
- Implemented image captioning using attention-based encoder-decoder model, face recognition based on dlib's face recognition, and OCR using Google Vision.
- Implemented face recognition using opencv and python. Deployed the model on an ubuntu web server using Flask, Javascript and Ajax.

Application Developer Intern June 2019 – Aug 2019
Schlumberger [link] *Noida, India*

- Developed an application software for automating the data pipelines in SAP using REST, Postman, and TkInter. Deployed existing API servers on Google Apigee using Javascript and REST.
- Represented the Schlumberger Cloud-For-Customer (C4C) team at the Schlumberger's Global Hackathon challenge 2019.

SELECTED PROJECTS

A Hybrid Transformer-based Model for Irony Detection in Arabic [code]

- Developed a multi-channel hybrid model to detect sarcasm in Arabic Tweets. The system was a part of our team SPPU_AASM's submission in the ArSarcasm Shared Task-2021.
- The model combines word representations generated using AraBERT, a language specific transformer-based model, with static word vectors trained on Arabic Twitter corpus.
- Our model outperformed multiple baseline models for the given task, securing a 2nd rank amongst 34 teams in the sarcasm detection subtask. [scoreboard]

Attention Ensemble approach for Marathi Text Classification [code]

- Developed a Hybrid CNN-BiLSTM Attention Ensemble model for the task of coarse-grained automatic technical domain identification of short texts in the Marathi Language.

- Our system ranked 1st for the TechDoFication Shared Task organized at ICON 2020. [link]

Mental Health Information Clustering using Meta-embeddings
- Employed a denoise-autoencoder model to generate meta-embeddings from contextualized sentence representations (RoBERTa) and topic models (LDA), to identify latent themes pertaining to mental health discussion groups on Reddit.
- Performed clustering using HDBSCAN and dimensionality reduction using UMAP.

BuddyBot: A Chatbot System for Stress Detection
- Implemented a retrieval-based conversational agent using DialogFlow, Flask and Firebase
- Integrated the response-framework of DialogFlow with a custom built sentiment analysis model, helping to compute the polarity of a conversation in real time.
- Designed a cloud-based logging system to store user conversations. [Demo]

Dynamic Sea Route Optimization [code]
- Developed a graph-based strategy to connect all the lat-long coordinates in a shipping lane.
- Designed Algorithms for finding the distance-based optimal sea route using Depth First Search (DFS), Dynamic Programming (DP), and Beam Search.

Nautical Calculations [PyPI Package]
- Nautical-calculations is a first-of-its-kind python library that implements the theoretical geo-spatial calculations such as bearing angle, rhumb line and great-circle distance in python.
- The library also supports custom-user functions such as getting the midpoints coordinate or equidistant points on a rhumb line or a great circle.

TECHNICAL SKILLS

Programming Languages: Python, Java, C/C++, SQL.
Frameworks: Pytorch, Tensorflow, Huggingface, Keras, Scikit-learn, OpenCV.
Tools: Git, Latex.

RELEVANT COURSES

- **Machine Learning:** Artificial Intelligence and Robotics, Machine Learning, Data Analytics, Data Mining and Warehousing.
- **Computer Science:** Data Structures and Algorithms, Object Oriented Programming, Design & Analysis of Algorithms, Database Management Systems, Computer Networks, Theory of Computation, Cloud Computing, Soft Computing and Optimization Algorithms, High Performance Computing.
- **Mathematics:** Linear Algebra, Differential and Integral Calculus, Differential Equations, Probability and Statistics, Discrete Mathematics.

SELECTED ACHIEVEMENTS AND AWARDS

- Winner of the **TechDoFication Shared Task** organized at **ICON 2020**.
- Runners up at the **EACL WANLP 2021 Shared Task**.
- Finalist at the **ZS Prize Competition** (link) from amongst 33,000 entries and won a cash prize of 2 lakhs rupees for the project Smart Cap.
- Bagged the 2^{nd} Prize in the **ASPIRE 2020**, a national level project competition organized by Computer Society of India (CSI) for Bachelor's Thesis Project.