Tejas Vaidhya

Indian Institute of Technology Kharagpur

✉ iamtejasvaidhya@gmail.com   ⚙ tejasvaidhyadev   ☐ +91 9340004079   🌐 Homepage

Interests

Causal Inference | Natural Language Processing | Computer Vision | Deep Learning | Machine Learning

Publications

- *Vaidhya T. et al* , "Domain Specific BERT Representation for Named Entity Recognition of lab protocol", Wnut workshop proceedings of **EMNLP 2020** [**Paper** | **Code**].
- Ojsav K., Adarsh K., *Vaidhya T.*, "Hostility Detection in Hindi leveraging Pre-Trained Language Models", Constraint workshop proceedings of **AAAI 2021** proceedings [**Paper** | **Code**].
- Kaushal A., *Vaidhya T.*, "Leveraging Event Specific and Chunk Span information to Extract COVID Events from Tweets", Wnut workshop proceedings of **EMNLP 2020** [**Paper** | **Code**].
- Vaidhya T. et al, "ArP-Gen:Architectural Plan Generator", Under review at **ICCV 2021** [**Paper** | **Demo**].
- Jin Z, Peng Z, *Vaidhya T.* et al, "Causal Direction in Data Matters: Implications of Causal and Anticausal Learning in NLP", Under review at *EMNLP 2021* [**Paper** | **Code**].
- Jin Z, Kügelgen J, *Vaidhya T.* et al, "Mining the Cause of Political Decision-Making from Social Media: A Case Study of COVID-19 Policies across the US States", Under review at *EMNLP 2021* [**Paper** | **Code**].

Education

Bachelor of Architecture (minor in Mathematics and computing)   **Graduating 2022**
Indian Institute of Technology Kharagpur, West Bengal, India   *Top 25 % of class*

Research Experience

Institute for Machine Learning , ETH Zurich
Advisor: Prof. Mrinmaya Sachan and Zhijing Jin   *Project-Link* | Mar 2021 - Present

- Implemented the experiment demonstrating the independent causal mechanisms in the machine translation task and in Semi-Supervised learning.
- Implemented training Algorithm for learning production rules of 9 language from semantic treebanks .
- Implemented SOTA Transformer based sentiment analysis classifier for COVID-19 related policies in US with accuracy of 74 %. Co-author two paper submitted at EMNLP 2021 during the first two and half months of internship.

Google SoC, JuliaLang
Advisor: Aviksen Gupta   *Project-Link* | May 2020 - Aug 2020

- Developed complete framework for traditional Language model and implemented Maximum Likelihood Estimation, Lidstone, Laplace, Witten Bell interpolation and other algorithm's APIs for users.
- Re-implemented SentencePiece unigram encoder in Julia and Sentencepiece tokenizer wrapper for Julia users.
- Contributed complete framework for ALBERT and 2+ Transformer based Language model from scratch in TextAnalysis (Natural language Processing Library in Julia ecosystem), their Documentation and tutorials.

ATLAS experiment at CERN, CUHK
Advisor: Prof. Tom Cheng   *Project-Link* | Mar 2020 - May 2020

- Setup the ATLAS environment in a Singularity container over a GPU cluster, after dealing with several data handling and machine learning frameworks natively developed at CERN such as Scientific Linux for CERN, ROOT, LWTNN etc.
- Reduced the training time of existing algorithms – recurrent neural networks based ensembles – five folds by efficient data handling through caching, effective parallelization and multi-GPU distributed training.
- Developed an attention based pipeline to model correlations between particle energy-momentum measures from different time-stamps. Created a two-stream transformer architecture to separately model Track and Cluster measures.

Uniworks Design, Hyderabad
Advisor: Arun Chaubey   *Project-Link* | *Demo* | Dec 2020 - Mar 2021

- Developed a novel end-to-end architectural plan generator to generate plan without the involvement of Architect.
- Empirically demonstrated the importance of the Genetic algorithm as an efficient starting point for plan generation and Conditional GANs to learn room semantics.
- Contributed a new spatial mapping task and the Spatial mapping dataset in computer vision community.

Centre for Artificial Intelligence, IIT Kharagpur

Advisor: Prof. Sudeshna Sarkar   *Aug 2019 - Jan 2020*

- Developed the bidirectional auto encoder-based model which uses the information extracted from context with out-of-reference knowledge to get the softmax probabilities over all the possible answers to a question.
- Extracted the edges based relations from commonsense corporas like ConceptNet and NELL.
- Improved the accuracy of Bi-LSTM based baseline model by 3% by incorporating the edge based relations

Independent Research Projects

- Formulated system for Extracting COVID-19 Events from Twitter ranking 1 st on the leader-board for the EMNLP 2020 Noisy Text Workshop Shared Task 3; Delivered Oral Presentation at EMNLP 2020 Workshop.
- Proposed and experimented an approach for Named Entity Recognition over lab protocols using Domain- specific Bert; Published at EMNLP 2020 noisy text workshop.
- Secured the 2 runner up position in the intra-university Data Analysis at Indian Institute of Technology Kharagpur.

Open Source Projects

Julia Ecosystem   *Project-Link | Jan 2020 - Present*

- Released checkpoints for pretrained weights in BSON format as desired by Julia and provide APIs for conversion.
- Author and maintainer of GoogleDrive.jl, Julia's official package to automate file download from Google-Drive.
- Maintainer of TextModel.jl and TextAnalysis.jl, Julia package for text pre-processing and analysis.

Kinship Relationship Prediction   *Project-Link | May 2019 – Jul 2019*

- Developed a Siamese convolutional neural network-based model for feature extraction and binary classification.
- Tuned the hyperparameters to achieve more than 85% (then SoTA) accuracy on the Face in the Wild dataset.

Energy Efficient Design   *Project-Link | Sep 2019 – Nov 2019*

- Developed a 3D model of Residential Complex for Visiting Researchers with Revit and rendered it with Lumion.
- Performed Energy Analysis and Wind Analysis on the building design through all stages, from the earliest conceptual phase through detailed design, to ensure constant working towards the most energy-efficient building possible.

Awards and Achievements

- Among top 0.12% of the 0.2 million candidates in the Joint Entrance Examination (Advanced) 2017.
- Among top 0.07% of the 1.2 million candidates in the Joint Entrance Examination (Mains) 2017.
- 1st position at W-NUT, shared task-3, EMNLP 2020 on Extracting COVID entities from Twitter.
- Secured 2nd runner up in Advertisement Design for Azad Hall of residence in 2019-2020.
- Secured 2nd runner up position in Inter Hall Data Analysis competition, 2019 for Azad Hall of Residence.
- Cleared Architecture Aptitude Test, 2017 for admission in Architecture Department at IITs.
- Selected for attending Eastern European Machine Learning Summer School 2021.

Relevant Coursework and Technical Skills

Skills: python, C/C++, Julia, TensorFlow, Pytorch/ OpenCV, Keras, Flux, HTML, CSS, Git, Shell, Unix, Latex, Photoshop, Illustrator, Microsoft Office.

Course work: Natural Language Processing with Deep Learning, Probability and Statistics, Discrete Mathematics, Linear Algebra, Deep Learning, Programming and Data Structure (*L*), Machine Learning, Stochastic Process, Design and Analysis of Algorithms (*L*), Operating Systems, NLP

'L' marked course have a laboratory component with them

Positions of Responsibility

- **Department Representative, CDC IIT Kharagpur**: Part of a team of 30 students, responsible for planning and Execution of the placement of 3000+ students at IIT Kharagpur.
- **Senior member of ShARE, IIT Kharagpur**: Member of a Global Student Think Tank with 620+ members from over 30 universities across 12 countries. Taught a workshop on Data Analysis to 100+ 1st year students.
- **Vice captain of Inter Hall Advertisement Design and Short film making**: Lead 15 members team in the Ad-Design and Short Film Making in Technology General Championship 2020 at the IIT Kharagpur.
- **Mentor at Google Summer of Code and Julia Summer of Code**: Mentored the following projects: "Plug and Play Language Models in Julia" and " XLNet - A Generalized Autoregressive Pretraining Model for JuliaText".