

RAJ RATN PRANESH

+1 8149968381 | [Google Scholar](#) | [E-mail](#) | [LinkedIn](#) | [GitHub](#) | [Website](#)

EDUCATION

Penn State University

Master's in Computer Science

Aug. 2021 – April 2023

Birla Institute of Technology, Mesra

Bachelor of Engineering in Computer Science

Overall GPA 8.5/10.0 (top 3%)

Aug. 2017 – April 2021

RESEARCH EXPERIENCE

AI/Data Science Research Intern

J.P. Morgan & Chase

June 2022 – Ongoing

New York, USA

- Developing parser independent document/webpage ingestion pipeline for Question-Answering task
- Built hybrid multi-level webpage ranking algorithm based on BM25 and PLM (neural re-ranking) model which boosted the MRR score by 4.21% over the single ranking model
- Designing and implementing Detectron and Tesseract based layout parser for extracting structured text from webpages followed by QA task

Research Assistant

Penn State University

August 2021 – May 2022

Pennsylvania, USA

- Worked on the extraction of textual features and generating reprehensibility score for research literature's claims.
- Developed model agnostic meta-learning algorithm for robust resource multilingual question-answering systems

AI Engineer

Marsview.AI

June 2021 – July 2021

Bangalore, India

- Developed transformer and Elastic search based Machine Reading Comprehension ML pipeline for extracting answers out of long context inputs
- Designed End-to-End Automatic Speech Recognition system with online real-time decoding for transcribing audio inputs

Keutzer Lab Research Assistant

University of California, Berkeley

May. 2020 – November 2020

Remote

- Trained a BERT model from scratch using Digital Corpus of Sanskrit dataset Collected a large unsegmented-segmented sentence pair in SLP1 format.
- Developed transformer-based encoder decoder BERT model for Sanskrit word segmentation generation task and fine-tuned collected dataset

AI-ML-NLP Lab Research Intern

Indian Institute of Technology, Patna

December 2019 – February 2020

Bihar, India

- Developed a deep learning multi-model framework for a protein-protein interaction model using protein sequences and 3D structure
- Implemented and fine-tuned various visual and linguistic models for extracting multi-modal features and experimented with multiple feature fusion techniques; intra-model (self-attention, relation-attention, transformer-attention) and cross-model (Factorized Bilinear Pooling, Weighted average, Attention concatenation)

CNRS-INRIA Lab Research Assistant

Université Grenoble Alpes, LIG-CNRS

May 2019 – July 2019

Grenoble, France

- Developed a human-guided data exploration approach for efficiently extracting knowledge from disaster dataset (micro-blogs) based on a given user-generated query
- Developed a novel statistical (term co-occurrence term frequency) and semantic (thesaurus, ontology, word embedding) based hybrid query processing and generation model with human-in-loop feedback mechanism
- Proposed and implemented a modified query expanding algorithm along with disaster class-specific vocabulary; resulted in an optimized information extraction system with 2.51% faster output delivery

PUBLICATIONS

- [1] **CMTA: COVID-19 Misinformation Multilingual Analysis on Twitter**
Raj Ratn Pranesh, Mehrdad Farokhenajd, Ambesh Shekhar, Genoveva [Paper]
59th ACL, 2021
- [2] **A Conglomerate of Multiple OCR Table Detection and Extraction**
Smita Pallavi, Raj Ratn Pranesh, Sumit Kumar [Paper]
22nd International Conference on Document Analysis and Recognition, 2020
- [3] **Automated Medical Assistance: Attention Based Consultation System**
Raj Ratn Pranesh, Ambesh Shekhar, Sumit Kumar [Paper]
NeurIPS, 2020 MLPH: Machine Learning in Public Health Workshop)
- [4] **MemeSem: A Multi-modal Framework for Sentimental Analysis of Meme via Transfer Learning**
Raj Ratn Pranesh, Ambesh Shekhar [Paper]
37th International Conference on Machine Learning (ICML), 2020 (4th Lifelong Learning Workshop)
- [5] **QuesBELM: A BERT based Ensemble Language Model for Natural Questions**
Raj Ratn Pranesh, Ambesh Shekhar, Smita Pallavi [Paper]
5th IEEE ICCCS (International Conference)
- [6] **Analysis of Resource-efficient Predictive Models for Natural Language Processing**
Raj Ratn Pranesh, Ambesh Shekhar [Paper]
EMNLP, 2020 (SustainLP Workshop)
- [7] **Biomedical Network Link Prediction using Neural Network Graph Embedding**
Sumit Kumar, Raj Ratn Pranesh, Ambesh Shekhar [Paper]
ACM CoDS-COMAD, 2020 (Young Researchers' Symposium)
- [8] **Classifying Micro-text Document Datasets: Application to Query Expansion of Crisis-Related Tweets**
Mehrdad Farokhen, Raj Ratn Pranesh, Javier A. Espinosa-Oviedo [Paper]
ICSOC, 2020 (STRAPS 2020 Workshop)
- [9] **A Fine-Grained Analysis of Misinformation in COVID-19 Tweets**
Sumit Kumar, Raj Ratn Pranesh, Kathleen M. Carle, [Paper]
Computational and Mathematical Organization Theory, Springer
- [10] **CLPLM: Character Level Pretrained Language Model for Extracting Support Phrases for Sentiment Labels**
Raj Ratn Pranesh, Sumit Kumar, Ambesh Shekhar [Paper]
17th International Conference on Natural Language Processing, 2021 (Short Paper)

RESEARCH PROJECTS

- M2L2QA: Meta-Learning for Multilingual Question-Answering System** July 2021 – Ongoing
 - We propose a meta-learning based multilingual COVID-19 QA system. Utilizing a model-agnostic training algorithm in which a prior to the conventional gradient update, a meta-learning update is done to make the multilingual QA model understand and learn the semantic relation between the passage and the questions present in a resource deficient language using the semantic knowledge of a high resource language model. With parallel training of a student-teacher model on multiple languages QA data, we directed the model to understand the text underlying contextual knowledge and generalize well for the resource-less languages.
- Improving Neural Text Summarization using Knowledge Graphs** September 2020 – December 2020
 - Developed a method for extractive text summarization using auto-regressive transformers. For better learning procedure we adopt the knowledge graph method to convert our textual data to more informative text and unsupervised training methods for wide use. We feed the informative text to our pre-trained generative model to summarize the text more properly and infer on generating a proper summary. The model is able to summarize input text into adequate information and is capable of performing several Natural Language Processing(NLP) tasks.
 - Paper is available [Here](#)
- RMHIDD A Reddit Mental Health Intervention Dialogue Dataset** August 2020 – September 2020
 - Created RMHIDD, a new dialogue corpus for automated mental health intervention. The dataset consists of over 200K Reddit posts collected from 18 different sub-Reddit groups with each post consisting of sequential conversation between the users. On this dataset, we also trained various models for dialogue generation task, namely- 'BART' and 'DialogPT'.

- Dataset is available [Here](#)

M2D A Multi-modal Framework for Automatic Medical Diagnosis

July 2020 – August 2020

- Developed a multimodal deep learning framework for automatic medical condition diagnosis via transfer learning. M2D leverages acoustic and textual features extracted from the audio utterance and the corresponding transcription describing a patient's medical symptoms. Our model utilizes ResNet-34 to learn audio feature via log mel-spectrogram and BioBERT language model to learn textual feature. We conducted a comparative performance analysis of M2D with baseline models based on textual or acoustic feature.
- Paper is available [Here](#)

Multi-modal Speech Emotion Analysis

August 2019 – October 2019

- This project involves predicting the speaker's emotion based on annotated speech and text data. Used IEMOCAP and MELD dataset with BERT for extracting the linguistic feature and raw audio log mel-spectrograms is generated using Librosa and supplied into visual model for speech feature extraction from the audio.
- Code is available [Here](#)

TECHNICAL SKILLS

Languages: Java, Python, C/C++, Ocaml, SQL, HTML/CSS, \LaTeX

Scientific Libraries: TensorFlow, Keras, PyTorch Sklearn, Pandas, Numpy, NLTK

Developer Tools: Git, Google Cloud Platform, VS Code, Google Colab, Terminal, Android Studio

TEACHING EXPERIENCE

Teaching Assistant: AI and Expert System

August 2019 – November 2019

- Preparing assignments quizzes, lecture slides and mentoring student project groups

Teaching Assistant: Database Management Systems

January 2019 – April 2019

- Supervised lab programming sessions, setting and grading questions for exams

OPEN-SOURCE CONTRIBUTION

Contributed to HuggingFace pretrained models

August 2020 – October 2020

- Trained BERT-base and RoBERTa language model from scratch using Sanskrit DCS corpus in IAST and SLP1 format. Models were published. [Link](#)

Google Code-In Mentor-TensorFlow

November 2019 — January 2020

- Selected as mentor by Paige Bailey for the TensorFlow organization to supervise and guide young students in the field of Deep Learning and Open Source Community through TensorFlow.

Microsoft Repository's Bug Fix

August 2019 — October 2019

- Spearheaded the development of Microsoft's Mt-DNN repository. Apex and PyTorch bug fix. [Link](#)

ACHIEVEMENTS AND AWARDS

- Recipient of visiting scholar fellowship at Sorbonne University, Paris, France. Fully funded research internship under the supervision of Dr Prof. Antoine Mine at APR Team of LIP-6 Lab(CNRS).
- Received ICML 2020 registration grant as a support for attending and presenting my work at the ICML 2020
- Received EMNLP 2020 diversity and inclusion grant as a support for attending and presenting my work at the EMNLP 2020
- Received NIPS 2020 registration grant as a support for attending and presenting my work at the NIPS 2020
- Virtually presented our work "COVID-19 question-answering exploration system" at the LIG-CNRS workshop on databases and information systems. Slides available [here](#).
- Recognized on the university's website as young research scholar on the recommendation of Institute Director.
- Received a special mention for excellence in research and abroad internship in the yearly university magazine
- Start-Up idea accepted for presentation at the annual start-up summit "National Seminar on Pragmatic Role of Technological Innovation in Start-Up NPTIS-2020" organized by the government of India. Abstract available [here](#).
- Developed and deployed Google assistant service for Google console that was used worldwide for promoting E-Learning.
- Selected as Microsoft Student Partner for the year 2019-2020. Conducted seminar and workshop for creating awareness about latest skills and technology.
- Actively participated in competitive coding contest and data science competitions. Six start coder on HackaRank platform and Kaggle competitions contributor. Profile [link](#).