Sanjeev Kumar

snjev421@gmail.com @snjev310 snjev310 sanjeev snjev310
sanjeev@cse.iitb.ac.in

Research Interests

- Focuses on investigating the domain of extremely low-resource Indian languages; aims to explore the transferability of knowledge from existing state-of-the-art language models trained on extensive data to develop tools for extremely low-resource languages.

Education

2022 – 2027 **Ph.D., Department of Computer Science & Engineering,**Indian Institute of Technology Bombay **(IITB)**, Mumbai-76, Maharastra, India
Thesis: Shallow NLP for Extremely Low Resource Indian Languages.
Advisors: Preethi Jyothi, Pushpak Bhattacharyya
Recipient of TCS Research Fellowship Cycle-17 (2023-2027).
CGPA: 7.93/10

2018 – 2021 **M.Tech., Statistical Computing (Data Science)**, School of Computer & Systems Sciences, Jawaharlal Nehru University **JNU**, New Delhi-67, India
Dissertation: Spatial Analysis Based Classification of Audio Data.
Advisor: Sonajharia Minz
CGPA: 7.03/9

2014 – 2018 **B.Tech. Information Technology**, School of Computer Engineering, **K**alinga **I**nstitute of **I**ndustrial **T**echnology **(KIIT)** University, Bhubaneswar-24, Odisha, India
BTP: Spearheaded the creation of an innovative platform connecting users for buying, selling, and borrowing diverse items, promoting resourcefulness and reducing waste.
CGPA: 8.33/10

Research Publication

1. S. Kumar, P. Jyothi, and P. Bhattacharyya, "Part-of-speech tagging for extremely low-resource Indian languages," in *Findings of the Association for Computational Linguistics ACL 2024*, L.-W. Ku, A. Martins, and V. Srikumar, Eds., Bangkok, Thailand and virtual meeting: Association for Computational Linguistics, Aug. 2024, pp. 14 422–14 431. URL: https://aclanthology.org/2024.findings-acl.857.

Employment History

Nov 2021-April-2022 **Senior Engineer-1** LG Soft India, HAIL Lab, Greater Noida, UP, India.
May 2021-Oct-2022 **AI/ML Engineer Trainee** LTTS, Bengaluru, Karnataka, India.
Apr 2019-July-2019 **BI Analyst Intern** Limeroad, Gurugram, Haryana, India.
Jan 2018-Apr-2018 **Web Development Intern** bringmyfood.in, Bhubaneswar, Odisha, India.

Skills

Programming Languages: C, Python, Java, SQL, LaTeX
Tools and Libraries : NLTK, Numpy, Pandas, Matplotlib, openCV, PyTorch, HuggingFace

Selected Projects

- **A Hybrid Word2vec Approach for Capturing Local and Global Semantic Similarity:** Gathered textual data encompassing a vocabulary of 1000 words, then trained a word embedding model to obtain embeddings for 10 animal and 10 bird words. Evaluated the accuracy by comparing it with Google's word2vec model.
- **Neural Attention Model for Question Answering:** Constructed a question-answering model on the Stanford Squad v2 dataset, utilizing an attention-based RNN encoder-decoder architecture for effective information retrieval and comprehension.
- **Measuring Semantic Consistency in Generated Image Caption:** Develop an advanced model that utilizes attention mechanisms to carefully examine the coherence between generated captions and their respective images, enhancing the dependability and informativeness of image descriptions.
- **Natural Language Inferencing (NLI) for Hindi:** Constructed a classification model on the IndicXNLI dataset, utilizing an attention-based architecture that assesses if a premise entails, negates, or is neutral towards the hypothesis statement.[1]
- **Beyond Words: Recognizing Emotions from Speech:** Constructed a classification model based on LSTM and evaluated its performance on the Acted Emotion Dynamic Database (AESDD). Analyzing audio samples representing five basic emotions (Anger, Happy, Fear, Sad, and Disgust), Our novel LSTM-based model surpassed baseline performance.[2]
- **Development and Demonstration of a Portable Device for Rapid Detection of Epileptics Discharges:** This project, led by the Principal Scientific Advisor of the Government of India and AIIMS New Delhi, involves the development of a system designed to detect elliptic discharges from the brain. We created an ML model to enable the early detection of these elliptical discharges.
- **Prediction of Hepatic Venous Pressure Gradient (HVPG) value from Liver CT Images:** With the goal of streamlining HVPG assessment and offering a more accessible diagnostic tool, our ILBS New Delhi project builds a classification-based model for fatty liver CT scans. This innovative approach promises to transform healthcare by simplifying HVPG measurement and potentially impacting countless patients.

Coursework

PhD Coursework
- Speech and Natural Language Processing and the Web, Foundations of Machine Learning, Deep Learning for Natural Language Processing, Automatic Speech Recognition, Critical Thinking for the Digital Age

MTech Coursework
- Optimization Techniques, Bigdata System, Probability & Stochastic Process, Statistical Inference & Multivariate Techniques, Bigdata Algorithm, Information Theory, Machine Learning, Probabilistic Graphical Model, Geospatial Informatics, Computer Vision

Awards and Achievements

2023-27 — Awarded **TCS Research Scholar Fellowship cycle-17** for PhD. **(48 Lakhs INR)**

2019 & 2020 — **UGC NET**, for the eligibility of Assistant Professor qualified in Dec 2019 & Nov 2020.

2018-2020 — **Merit-cum-Mean** Scholarship for M.Tech Study

[1] https://github.com/snjev310/cs-626-project/
[2] https://github.com/snjev310/cs-725-project

Conference/Symposium Attended

- Oct-2024 — Attended **Nvidia AI Summit** at Mumbai, India, from Oct 23-25, 2024.
- Aug-2024 — The 62^{nd} Annual Meeting of the Association for Computational Linguistics (**ACL'24**) Bangkok, Thailand, August 11–16, 2024
- Jul-2024 — Selected to attend 14^{th} Lisbon Machine Learning School (**LxMLS'24**) at Lisbon, Portugal from July 11-17, 2024.
- Dec-2023 — IndoML-2023 at IIT Bombay.
- 20th International Conference on Natural Language Processing **(ICON)**-2023 at Goa University sponsored by CFILT lab IIT Bombay.
- Feb-2023 — ACM-India Annual event at OSIT Bhopal sponsored by IIT Bombay.

Miscellaneous Experience

Reviewing

- COLING (Reviewer, 2025). IEEE EDITS (Reviewer, 2025)

Presentations

- Dec-2023 — Invited as speaker for a seminar titled "Decoding Big Data: Unveiling Insights and Applications for the Modern Era" at Pillai College of Engineering Navi Mumbai.

Teaching Assistant

- Apr'22-July'22 — **CS152+CS154**: Abstractions & Paradigms for Programming
- Mar'23-June'23 — **CS101**: Computer Programming and Utilization
- Aug'23-Dec'23 — **CS335+CS337**: Artificial Intelligence and Machine Learning.

Volunteer Experience

- Dec -2023 — Volunteer at IndoML-2023 at IIT Bombay.
- 2019-21 — Training and Placement Co-ordinator at Jawaharlal Nehru University Placement Cell.
- Councillor at Jawaharlal Nehru University Student Union.
- 2017-18 — KiiTFest Event Co-ordinator at KIIT Student Activity Center.

References

Prof Preethi Jyothi
Professor
Department of Computer Science & Engineering
IIT Bombay, Mumbai-76, India.
pjyothi@cse.iitb.ac.in

Prof Pushpak Bhattacharyya
Professor
Department of Computer Science & Engineering
IIT Bombay, Mumbai-76, India.
pb@cse.iitb.ac.in

Prof Sonajharia Minz
Professor
School of Computer & Systems Sciences
JNU, New Delhi-67, India.
sonaminz@mail.jnu.ac.in
sona.minz@gmail.com