

SAYALI GHODEKAR

MS CS applicant for Fall 2021

sayalighodekar26@gmail.com ♦ linkedin.com/in/sayali-ghodekar26

EDUCATION

Bachelor of Computer Engineering

June 2015 - June 2019

Pune Institute of Computer Technology, Savitribai Phule Pune University

Overall GPA: **9.09 (First Class with Distinction)**

Relevant Coursework: Machine Learning, Soft Computing and Optimization, Engineering Mathematics, Data Mining and Analysis, Cloud Computing, Embedded Systems and Internet of Things.

EXPERIENCE

DeepAffects, AI powered Voice Intelligence Startup

Jan 2020 - Present

Data Scientist

Mumbai, India

- Responsible for research and analysis of NLU and speech intelligence, including deep neural network techniques to transform speech and extract insights from text, audio and video of multi-speaker conversations.
- Development and scaling of DeepAffect's abstractive summarization API for video conference data using the future n-gram prediction model. Roughly processed over 1M+ minutes of audio into summaries.
- Developed a text segmentation algorithm to improve topic units for multi-turn dialogue data.
- Built DeepAffect's conversation analytics and metrics including dialogue act tagging, intent classification, question and answering systems using PyTorch.

Center for Indian Language Technologies, IIT Bombay

June 2019 - Dec 2019

Research Intern under the guidance of Prof. Pushpak Bhattacharyya and Prof. Malhar Kulkarni

- Designed experiments for True Cognate Detection across 10 Indian language pairs using mono-lingual word embeddings and Siamese deep neural networks. (Best F-score **0.82**)
- Developed the novel Textual History Analysis Tool to capture the historical evolution of texts through various temporal stages and digitized manuscripts using computational phylogenetics.
- Experimented on EyeLink Eye-Tracking System to utilize gaze features for cognate identification task. Demonstrated an improvement in scores using gaze features over crosslingual embeddings (work under review at EACL).
- Investigations into the cross-lingual embeddings and transfer for low-resource NLP.

SusthitSoft Technologies

June 2018 - December 2018

Trainee Associate

Pune, India

- Implementation of an OCR based tool for Devanagari script to digitize Sanskrit manuscripts.
- Conducted an extensive survey of the challenges faced during the recognition of Devanagari scripts due to the diacritics and complex writing system.

PUBLICATIONS

- Diptesh Kanojia*, Sai Sravan Munukutla*, **Sayali Ghodekar**^{*1}, Pushpak Bhattacharyya "Keep Your Dimensions on a Leash: True Cognate Detection using Siamese Deep Neural Network", 7th ACM IKDD CoDS and 25th COMAD, Young Researcher's Symposium, 2020 [link]

^{1*} denotes equal contribution

- Diptesh Kanojia, Malhar Kulkarni, Pushpak Bhattacharyya, **Sayali Ghodekar**, Eivind Kahrs "An Introduction to the Textual History Tool", Proceedings of the 6th International Sanskrit Computational Linguistics Symposium. [link]
- Diptesh Kanojia, Malhar Kulkarni, **Sayali Ghodekar**, Pushpak Bhattacharyya. "Strategies of Effective Digitization of Commentaries and Sub-commentaries: Towards the Construction of Textual History"

NOTABLE PROJECTS

Optical Character Recognition for Marathi Text with Braille Conversion Unit

Pune Institute of Computer Technology

June 2018 - April 2019

Prof. Hemlata Channe

- Proposed and implemented an automated system that extracts Marathi text from images and further converts text in the Braille format using CNN-BiLSTM based model (F-score: 0.82).
- Constructed a dataset of ~12k Marathi word images with their annotated labels helpful for image processing and character recognition tasks.

Ensemble Learning for Astronomical Data

August 2018 - October 2018

- Experiments to predict the photometric redshifts of distant galaxies for the Sloan Digital Sky Survey Data using machine learning methods.
- Demonstrated an improved accuracy on the galaxy classification task using ensemble classifiers.

CONFERENCE PRESENTATION

Poster and paper presentation at *Young Researcher's Symposium 2020*, 7th ACM IKDD CoDS and 25th COMAD, "Keep Your Dimensions on a Leash: True Cognate Detection using Siamese Deep Neural Network"

TECHNICAL STRENGTHS AND OTHER SKILLS

Programming Languages	Python, Java, C/C++, R, Shell Scripting
Deep Learning Frameworks	PyTorch, Keras, TensorFlow
Database Technologies	MySQL, MongoDB, PostgreSQL
Web Technologies	Javascript, PHP, Flask API, VueJS
Languages	English, Hindi, Marathi, Japanese (limited proficiency)

CERTIFICATIONS

- Japanese Language Proficiency Test (N5 Level) by The Japan Foundation and Japan Educational Exchange and Services. (July 2019)
- Certificate Course in Japanese by The Department of Foreign Languages, University of Pune. (**89%**)(April 2019)
- Data-driven Astronomy by The University of Sydney on Coursera. (October 2018)

EXTRACURRICULARS

- Volunteer at **Tathapi** for Women and Health Resource awareness. Gathered resources from non-profit organizations all over Maharashtra for the Helpline.
- Successfully organised **TEDxPICT 2017** under the Branding and Operations team. Contributed to TedxPICT blog under the Content Development team.