SAYAR GHOSH ROY

📞 +91-7980704678　　@ sayar.1997@gmail.com　　⌗ sayarghoshroy　　🔗 Sayar's Work　　in sayarghoshroy　　✉ 53/24 R.N Guha Road, Kolkata 700028

EDUCATION

International Institute of Information Technology, Hyderabad
📅 Jun 2017 – Apr 2022

- Bachelor of Technology (Honours) in Computer Science
- Master of Science (by Research) in Computational Linguistics

CGPA: 9.73/10

RELEVANT RESPONSIBILITIES

System Administrator
📅 Jan 2020 - Present

- Managing servers + other resources shared by Centre for Innovation and Entrepreneurship (CIE) & Information Retrieval and Extraction Lab

Lecturer at 'Intro to Deep Learning Workshop'
📅 May 2020

- Conducted four sessions for prospective candidates joining the Language Technologies Research Centre (LTRC), IIIT-H
- Created teaching material can be accessed here

Research Mentor
📅 Jun 2020 - Present

- Mentoring four senior students in their research work in the areas of Text Coherence Analysis, Natural Language Generation from Tabular Data, and domain specific Text Summarization

HONORS AND AWARDS

★ Winners, Constraint Shared Task at AAAI 2021: Secured First Position for Adverse Content Detection in Hindi among 45 participant teams

★ Winners, LaySumm & LongSumm at EMNLP 2020: Achieved SOTA results on two text summarization tasks with ROUGE-1 scores of 0.46 and 0.50

★ Recipient of Dean's List Academic Award, IIIT-H, 2017 - 2020

★ Endorsed by IIIT-H for the S.N Bose Scholars Program, 2018

★ Selected for INSPIRE Scholarship, 2016 by Dept. of Science & Technology, Govt. of India

★ Secured 1st School Rank, 8th State Rank in the Indian School Certificate (ISC) with an aggregate of 97.25% & a perfect score in Computer Science

★ Award for General Proficiency, 2015, Salt Lake Point School, Kolkata

★ Student of the Year 2014, St. Mary's Orphanage and Day School, Kolkata

PUBLICATIONS

Summaformers @ LaySumm 20, LongSumm 20

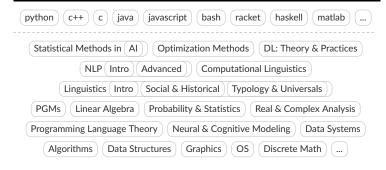*S Ghosh Roy, N Pinnaparaju, R Jain, M Gupta and V Varma*

Proceedings of the First Workshop on Scholarly Document Processing at EMNLP 2020, November 19, 2020, Online | url ● slides ● poster

Leveraging Multilingual Transformers for Hate Speech Detection

S Ghosh Roy, U Narayan, T Raha, Z Abid and V Varma

To Appear: HASOC Identification in Indo-European Languages at Forum of Information Retrieval Evaluation (FIRE), 2020 | pdf

SKILLS AND COURSEWORK

python　c++　c　java　javascript　bash　racket　haskell　matlab　…

Statistical Methods in AI　Optimization Methods　DL: Theory & Practices
NLP Intro Advanced　Computational Linguistics
Linguistics Intro　Social & Historical　Typology & Universals
PGMs　Linear Algebra　Probability & Statistics　Real & Complex Analysis
Programming Language Theory　Neural & Cognitive Modeling　Data Systems
Algorithms　Data Structures　Graphics　OS　Discrete Math　…

EXPERIENCE

Applied Research Engineer Intern
Apple
📅 Sep 2020 – Present　　📍 Remote

- Part of Apple Maps' Web-Discovery team
- Applying state-of-the-art Natural Language Processing techniques to complete specific business objectives
- Curated multiple datasets and built several Machine Learning models
- Improved relevance classification precision by over 22 percent

Undergraduate Researcher
Information Retrieval and Extraction Lab
📅 Aug 2019 – Present　　📍 Hyderabad, India

- Guided by Prof. Vasudeva Varma and Dr. Manish Gupta
- Designed an acclaimed architecture for generating blog like posts from Scientific Research Papers
- Formulated and built Hate Speech detection and classification systems for multiple languages
- Created a system for predicting sentence-wise relevance of online news articles in answering Microsoft Bing search queries

Software Engineering Intern
Digitalee
📅 May – Jul 2019　　📍 Hyderabad, India

- Built a pipeline for generating summaries of executive briefings as part of a three member team
- Implemented a PDF to text converter adhering to the structure of large executive briefing documents
- Created an end-to-end architecture for accepting PDF documents & producing text summaries

Software Engineering Intern
Theodor
📅 Sep – Nov 2018　　📍 Hyderabad, India

- Built a system facilitating conversion of hand-filled forms into a digital format
- Utilized CTPNs for text detection and CRNNs for text recognition
- Implemented a classifier for form fields such that scanned images of forms can be converted into JSON
- Created a web app using Django for users to upload scanned forms, receive digitized information, and have a dashboard of all activities

SELECTED PROJECTS

Multilingual Hate Speech Detection | python ● pytorch
- Detection and fine-grained classification of Hateful and Offensive Twitter posts in English, German and Hindi through Transformer models

Optimization and Learning | python ● scikit-learn
- From scratch implementations of various optimization techniques such as Adam Optimizer, Nesterov's Accelerated GD, Polyak's Learn Rate, Newton's Method, Rprop, and Quickprop: Tested out on classification & regression tasks

Acronym Sense Disambiguator | python ● nltk
- Identifies acronyms in a text file and disambiguates possible expansions, where a guided disambiguator module runs queries on an indexed Wikipedia dump

Image classification formulations | python ● scikit-learn ● TensorFlow
- Performance of various deep learning architectures for face recognition and miscellaneous image classification tasks: insights and analysis

COVID-19 modeling for India | python ● scipy
- Modeling the Coronavirus Statistics for India using the Susceptible-Infected-Recovered (SIR) Model

Manifold Learning Mechanisms | python ● scikit-learn
- Illustration of various Spectral Clustering Techniques + Manifold Visualization schemes such as MDS, LLE, & Isomap; with applications in Computer Vision

→ Do check out my work here.