

# SOUMIL MANDAL

+1(469)2260388 ◇ soumil.mandal@utdallas.edu ◇ soumandal.github.io ◇ linkedin.com/in/soumandal

## EDUCATION

---

### Master of Science (M.S) in Computer Science

May 2021

The University of Texas at Dallas, Richardson, USA

Volunteering for Intelligence Analytics Society (IAS) club

### Bachelor of Technology (B.Tech) in Computer Science & Engineering

June 2019

SRM Institute of Science & Technology, Chennai, India

GPA 3.52/4

Volunteered for SRM WebArch & SRM UAV student clubs

## SKILLS

---

### Programming Languages

Python, C++, Java, JavaScript, C

### Libraries

Tensorflow, Keras, PyTorch, Numpy, Scipy, NLTK, Spacy, Pandas

### Technologies

MongoDB, MySQL, PostgreSQL, Firebase, Docker, Git

## EXPERIENCE

---

### Research Assistant | Intelligent Software and Security Lab, UTD

Sep 2019 - Present

- Developing adversarial attack strategies for Natural Language Processing (NLP) systems.

### NLP Research Intern | Jadavpur University, Kolkata

Dec 2018 - Jan 2019

- Engineered a system for translating code-mixed to a monolingual instance using language-based partitioning and char seq2seq with attention architecture.
- Programmed a back-transliteration system using char seq2seq architecture.

### NLP Research Intern | Indian Institute of Technology (BHU), Varanasi

June 2018 - July 2018

- Developed a code-mixed language tagger which takes context information using Bi-LSTM-CRF architecture.
- Experimented on solving out-of-vocab (OOV) problem in NMT using cognate mapping.

### NLP Research Intern | Jadavpur University, Kolkata

Dec 2017 - Jan 2018

- Developed a code-mixed language tagger using character and phonetic encoding with LSTM.
- Carried out a comparative study on the translation of simple sentences by NMT and SMT systems.

### NLP Research Intern | Jadavpur University, Kolkata

May 2017 - July 2017

- Prepared data for the shared task on sentiment analysis for Indian languages (SAIL 17).
- Designed a system for automated collection of code-mixed data from Twitter using Twitter4j.
- Developed a binary and ternary sentiment classifier for code-mixed data using multinomial-Naive-Bayes.

## PROJECTS

---

**Insincere Question Classifier** - Created a classification model which tags whether a question is rhetorical, hate speech, hypothetical, sexually explicit, sincere, or other type (FIRE-CIQ 19' 1<sup>st</sup> place) Aug 2019

**AI for Legal Assistance** - Built a rudimentary information retrieval system which ranks statutes and precedents based on relevance, given a case as query using text embeddings (FIRE-AILA 19' 4<sup>th</sup> place) Aug 2019

**Normalization of Transliterated Words** - Developed a Bengali-English transliterated word normalizer using char seq2seq model and Levenshtein distance (paper accepted at W-NUT 2018, EMNLP). May 2018

## TECHNICAL ARTICLES

---

Sainik Kumar Mahata, **Mandal, Soumil**, Dipankar Das and Sivaji Bandyopadhyay. "Code-Mixed to Monolingual Translation Framework." - *Forum for Information Retrieval Evaluation (FIRE)*, Kolkata, India, 2019

**Mandal, Soumil**, and Sankalp Sanand. "Strategies for Language Identification in Code-Mixed Low Resource Languages." - *15th International Conference on Natural Language Processing, Student Paper Competition, Patiala, India, 2018*

Sainik Kumar Mahata, **Soumil Mandal**, Dipankar Das and Sivaji Banyopadhyay. "SMT vs NMT: A Case Study on Hindi & Bengali Simple Sentences." - *15th International Conference on Natural Language Processing, Patiala, India, 2018*

**Mandal, Soumil**, and Anil Kumar Singh. "Language Identification in Code-Mixed Data using Multichannel Neural Networks and Context Capture." - *The 4th Workshop on Noisy User-Generated Text, Hong Kong, 2018*

**Mandal, Soumil**, and Karthick Nanmaran. "Normalization of Transliterated Words in Code-Mixed Data Using Seq2Seq Model & Levenshtein Distance." - *The 4th Workshop on Noisy User-Generated Text, Hong Kong, 2018*

**Mandal, Soumil**, Sourya Dipta Das, and Dipankar Das. "Language Identification of Bengali-English Code-Mixed data using Character & Phonetic based LSTM Models." - *15th International Conference on Computing and Information Technology, Bangkok, Thailand, 2018*

**Mandal, Soumil**, Sainik Kumar Mahata, and Dipankar Das. "Preparing Bengali-English Code-Mixed Corpus for Sentiment Analysis of Indian Languages." - *The 13th Workshop on Asian Language Resources, Miyazaki, Japan, 2018*

**Mandal, Soumil**, and Dipankar Das. "Analyzing Roles of Classifiers and Code-Mixed factors for Sentiment Identification." - *Research in Computing Science, ISSN 1870-4069, 2018*