[github.com/gitlost-murali](github.com/gitlost-murali)
[Musings of Murali – Blog]
[Murali Manohar – Medium]

# KONDRAGUNTA MURALI MANOHAR

+31-625596220
kmanoharmurali@gmail.com

## EMPLOYMENT

**ML Engineer** **AskUI** **Jun 2022 - Aug 2022**

- Proposed and maintained
  - Unsupervised classification approach for a module (Neural Vector Match)
  - Post processing for OCR: Spelling correction module (Seq2Seq)
- Trained to push production ready code. Well versed in code reviews, agile methodology.

**Data Scientist** **Gramener** **Jun 2019 - July 2021**

**NLP Division, AI Labs** **Jun 2019 - July 2021**

- Had setup Text Platform - NLP Division in AI Labs, which currently serves :

- Smart Contract Analysis Tool to identify critical clauses and key entities in legal documents.
- Financial Journey Step Identification to map customer complaints to different journey steps.
- Polarity app to classify the live comments from VoC (Voice of Customers).

- Worked on Patient De-identification. Involves entity recognition and anonymization based on risk-score.

**AI Labs** **Jun 2019 – July 2021**

- EmoBGM - An emotion classifier to identify the emotion of a scene using BGM.
- As a part of Microsoft's AI for Earth initiative, I collaborated with:

- CameraTraps team to contribute to their project of identifying animal species on a large scale.
- Researchers from UMass & Cornell on scaling up their project on Bird Migration.

**Research Intern** **LTRC, IIIT Hyderabad** **Dec 2018 - May 2019**

- Worked on *Robust Initialization and Fine-tuning Methods for Unsupervised NMT*.
- Supervisor: Dr. Manish Shrivastava

**Summer Research Intern** **LTRC, IIIT Hyderabad** **May 2018 - July 2018**

- Worked on *Shallow Parser in a Low Resource Setting*, accepted in CICLing 2019.
- Supervisor: Prof. Dipti Misra Sharma

**Developer Intern** **Geek Online Ventures Pvt. Ltd** **May 2017 - July 2017**

- Built a Desktop Application to extract details of the employee such as Desktop screenshots, actual working hours, files uploaded, and a web application for the Project Manager(Or equivalent) to view the data.
- Contributed to the foundations of the Chatbot Development Team by creating a prototype.

## RESEARCH

**Unsupervised NMT** **Dr. Manish Shrivastava & Prof. Sudip** **Dec 2018 - May 2019**

- Hypothesized and showed that an additional step of Lexical substitution using Cross-lingual embeddings can converge the models faster.
- Proved that fine-tuning NMT models trained in Unsupervised Setting perform better and complement the learning from under-resourced settings.

**Determining Relevance in VQA** **Prof. Sudip Sanyal** **Aug 2018 - Dec 2018**

- Experimented whether the inclusion of Visual Concepts, Natural Questions generated from the image, and premises from text impart auxiliary information to the existing Visual Question Answering systems.

**Under Resource Shallow Parser** **Prof. Dipti Misra Sharma** **May 2018 - July 2018**

- We have tried the feasibility of Transfer learning, Multi-Task learning using Pre-Trained Neural(LSTM Networks) Language Model in NLP and worked on various approaches for improving sequence labeling task, and achieved encouraging accuracies.

**Data Engineering & Word Sense Disambiguation (WSD)** **Dr. Satyendr Singh** **Jan 2017- Mar 2018**

- Created Sense Annotated Corpus for Indian Languages, which involved cleaning and parsing raw corpora to build datasets for the native language speakers to annotate.
- Did an empirical analysis of how morphological variants, POS Tags, stemming, stopwords, and syntactico-semantic relations (karaka relations) would affect the objective of WSD.

## Education

**Europe** — **University of Groningen & Basque Country** — **2021-Present**

- **Masters** in Computational Linguistics. Recipient of **Erasmus Mundus Scholarship**. Y1 in the University of Basque Country and Y2 in the University of Groningen.

**Gurgaon, India** — **BML Munjal University** — **2015-2019**

- Bachelor of Technology in Computer Science & Engineering. CGPA: [ 9.28 / 10 ]
- Undergraduate Coursework: Data Structures; Algorithms; Operating Systems; Databases; Information Retrieval; Comp. Architecture; Discrete Mathematics; Machine Learning;

## Projects

**MEME Retriever** — Poster — **May 2022 - June 2022**

- Given a situation (text), relevant meme images are retrieved. Used CLIP, Sentence Transformers, etc.

**Vaccine stance prediction**

- Predict a twitter user's stance based on his tweet, connections and metadata. Used RoBERTa in unison with graph-based, metadata features as input to a classic ML model.

**Patient De-identification** — **Aug 2020 - June 2021**

- Built a product for de-identifying patient's information in Clinical Summary Report & Health records of a major Pharmaceutical company.
- Followed Data-Flywheel concept in creating a loop of annotating data with user's feedback (active learning). Used Data Augmentation techniques (Pattern-Exploitation techniques, template-based generation) to handle data scarcity.

**Large-Scale Species Identification** — **Mar 2020 - Jul 2020**

- Worked with Microsoft's CameraTraps team on identifying animal species on a large scale. Used EfficientDet model architecture and deployed using Azure.

**Detecting Bird Migration** — **Nov 2019 - Mar 2020**

- A Microsoft Initiative: Worked with researchers from UMass & Cornell on scaling up their project on Bird Migration.
- Converted existing MATLAB model MISTNET to Python(PyTorch) for detecting birds from the NEXRAD radar scans. Scaled the model using AzureBatch for processing 200 million scans.

**Text Platform** — **Aug 2019 - Nov 2019**

- Smart Contract Analysis Tool to identify critical clauses and key entities in legal documents. Implemented using a pipeline of LSTMs, SpaCy, and Label-Studio.
- Financial Journey Step Identification to map customer complaints to different journey steps like Opening, Fraud, Closing, etc. Used Transfer Learning to tackle smaller datasets.
- Polarity app to classify the live comments from VoC (Voice of Customers). Leveraged OpenAI's unsupervised sentiment neuron model for polarity recognition.

**EmoBGM** — **June 2019 – Aug 2019**

- Collaborated with a media network to find what ensemble of emotions lead to a hit episode using BGM.
- Extracted BGM from vocals and generated spectrograms to train a CNN model for BGM identification.

## Languages and Technologies

- Python, Javascript, Java, HTML/CSS, jQuery, C++
- PyTorch, Keras, Numpy, Pandas, Scikit-learn, Flask.
- Git, Docker, Azure, AWS, GCE

## Notable Mentions

3rd place NLP Hackathon — Reference — Alibaba Hackathon - 2018
Promising Star Award — Gramener - 2019
Star Innovation Award — Gramener - 2020