Kartheek Akella

GitHub, Website

Email: sukruthkartheek@gmail.com

EDUCATION

- **Birla Institute of Technology and Science**   Hyderabad, TS
  Bachelor of Engineering in Computer Science   *Aug. 2016 - May. 2020*

PUBLICATIONS

- **Kartheek Akella**, Sai Himal Allu, Sridhar Suresh Ragupathi, Aman Singhal, Zeeshan Khan, CV Jawahar, and Vinay P. Namboodiri. Exploring Pair-Wise NMT for Indian Languages. *In Proc. ICON 2020, Patna, India*

- Sai Praveen Kadiyala, **Kartheek Akella**, Truong Huu Tram. Program Behavior Analysis and Clustering using Performance Counters. *In Proc. DYNAMICS, ACSAC 2020, Austin, Texas* (**Equal Contribution**)

PATENTS

- **Internal Patent of A*STAR**: *Program Behavior Analyser using Performance Counters.* Sai Praveen Kadiyala, **Kartheek Akella**, Truong Huu Tram.

RESEARCH EXPERIENCE

- **NMT for Low-Resource Languages (On-going)**   IIIT Hyderabad
  Aug 2020 - Present   *Prof. CV Jawahar & Prof. Vinay Namboodiri*
  - **Contribution**: Working on improving performance of translation for low-resource languages.
- **Detection of Malware Behaviour**   Institute of Infocomm Research, A-Star, Singapore.
  Feb 2020 - Aug 2020   *Dr. Sai Praveen Kadiyala*
  - **Contribution**: Worked on detection of malware and analysing behaviour of malware in windows 10 environment using Machine Learning.
- **Named Entity Recognition in Telugu sentences**   BITS Pilani
  May 2019 - Dec 2019   *Prof. N L Bhanumurthy*
  - **Contribution**: Attempted to use **Attention** mechanism to extract the contextual information of a word to improve the performance of NER task over the FIRE dataset. [code]
- **Detection of RPL Attacks in LLNs**   BITS Pilani
  Jan 2019 - Dec 2019   *Prof. G Geetha Kumari*
  - **Contribution**: Developed models for detection of topology based RPL (Protocol designed specifically for Low-powered Lossy Networks) attacks, specifically **Worst Parent Attack**.

PROJECTS

- **Workshop On Asian Translation (WAT) 2020**: Participated in the Indic Multilingual Evaluation Task (Team Name: **cvit**). Worked on improving the multi-lingual translation performance of 3 languages, particularly Odiya, Telugu, and Tamil, using a novel method and already present techniques like **bert-nmt**, fine-tuning, and similar language regularisation.

- **Sentiment Analysis**: Implementation of popular deep learning algorithms for sentiment analysis in Pytorch-Lightning. Aim of the repository is **re-usability** and implementing new architectures and train easily. [code]

- **smol**: A Pure Python deep learning library built entirely on Numpy in the smallest number of lines possible for more transparency and easier to learn for beginners. [code]

- **Plagiarism Checker**: A Python source code plagiarism checker. Dataset is created by collecting assignment submissions of students and manually identifying similar codes. Extracted features of the code, which are of various types like simple string level features, statistical measures, etc. implemented a **Complement Naive Bayes** classifier as the dataset is highly imbalanced. Planning to extend this to detect plagiarism across multiple files and find better features for this task. (**WORK IN PROGRESS**) [code]

- **Statistical Analysis of B-School Questionnaire Responses**: A Study on Effect of various facilities on Student Satisfaction in B-Schools. Analysis from raw Pilot questionnaire survey data received from several management schools in and around Hyderabad affiliated to Osmania University. [code] [report]

- **Search Engine**: Information Retrieval system that searches lyrics of **262,000+** songs. Each song is represented as a vector of tf-idf scores in the semantic space of the entire vocabulary. In the query time, a similar tf-idf representation of the query is generated, and songs are searched for similarity using the cosine distance similarity. [code]

- **Movie Recommender System**: A comparative study of several recommender system algorithms including **Collaborative Filtering, SVD, CUR** and **Latent Factor Model**. The general idea is to reduce the dimensions of a very sparse dataset and recommend based on the movie-movie similarities. [code]

- **Statistical Machine Translation**: Translation from English to Spanish is done by implementing the IBM models based on word-based translation and developed a model based on the idea of **phrase-based translation**.

- **Fraud Detection**: Detection of fraudulent transactions using the **Outlier Detection** algorithms, **DBSCAN** (Density-based spatial clustering) and **LOF** (Location Outlier Factor). The mentioned methods calculate the density of points in the surroundings and identify a point/transaction as an outlier/fraudulent when the ratio is very low.

OTHER WORK EXPERIENCE

- **BITS Pilani** — Hyderabad, TS, India.
  Teaching Assistant — *May 2019 - Present*
  - **Courses**: Regression, Feature Engineering, Deep Learning

- **FastCrave** — Hyderabad, TS, India.
  Co-Founder — *Dec 2018 - Present*
  - **Aim**: India's first standardised restaurant chain on highways.

- **National Textile Corporation** — Coimbatore, TN, India.
  Summer Intern — *May 2018 - July 2018*
  - **IT Department**: Developed a website that helps in improving productivity by scheduling reported issues.

- **Department of Technical Arts (BITS)** — Hyderabad, TS, India.
  Android App Developer — *Jan 2017 - May 2019*
  - **Pearl 2018**: Official app for Pearl '18, served about 800+ active users during the fest period. [code]
  - **Fest Manager**: Android app framework for the development of a general-purpose app for the fests in BITS. [code]

AWARDS AND DISTINCTIONS

- **Workshop on Asian Translation - Indic Translation Challenge (2020)**:
  Team Name: **cvit**
  - Odiya (En To Od) - 9.85 (2nd) [result]
  - Malayalam (En to Ml) - 5.40 (2nd) [result]
  - Telugu (En to Te) - 5.12 (3rd) [result]
  - Tamil (En to Ta) - 4.53 (3rd) [result]

- **Indian National Astronomy Olympiad, 2013 and 2015**: Top 1% (among 100,000+)
- **Indian National Junior Science Olympiad, 2013**: Top 1% (among 100,000+)
- **Google Moonbots 2012**: Top 30 in 147 teams world wide   [team blog (inactive)]