Ketan Kumar Todi

https://www.linkedin.com/in/ketan-kumar-todi/
https://todiketan.github.io

Email : ktodi@andrew.cmu.edu
Mobile : +1-412-616-6134
Google Scholar

Education

Carnegie Mellon University, School of Computer Science — Pittsburgh, PA
Master of Science in Intelligent Information Systems — *Dec 2022*
Courses (Ongoing): Advanced Natural Language Processing, Intro To Machine Learning, Search Engines

Manipal Institute of Technology — Manipal, India
Bachelor of Technology in Computer Science and Engineering; GPA: 9.36/10.0 (Top 5%) — *Jul 2019*

Experience

Carnegie Mellon University — Pittsburgh, PA
Graduate Research Assistant - Prof. Graham Neubig — *Aug 2021 – Present*

- Improving performance on multi-hop QA datasets by enhancing the second stage retrieval process.

Taiger — Singapore
Research Engineer — *Jun 2019 – July 2021*

- **Semi-automatic Ontology Generation**: Developed an active-learning based trainable NER, relation extraction and relation classification model cutting down data requirements by upto 40% for ontology generation from domain specific documents; awarded as the best performer in the R&D team of about 20 employees for this project.
- **Information Extraction from legal documents**: Designed an CRF based stacking ensemble model for extracting named entities from legal documents, where ensemble model had a 3% gain over the best performing individual model.
- **Information retrieval for QA systems**: Developed a universal-sentence-encoding and BM-25 based paragraph retrieval system for documents with nested structures, like company policies, bank agreements thereby augmenting the dowstream QA performance.

Nanyang Technological University — Singapore
Research Assistant - Prof. TA Nguyen Binh Duong — *Jan 2019 – May 2019*

- Created a decentralized Air Traffic Management POC system using Reinforcement Learning and blockchain (Hyperledger Fabric) to assist the workforce in scheduling aircrafts.
- Concluded that global reward and actions determined using action-value algorithm by smart contract for different agents decreased total system delay more as compared to local optimization and local reward calculation.
- Reduced running time from 50 seconds to 4 seconds and distributed the entire system across multiple machines.

International Institute of Information Technology, Hyderabad (IIIT-H) — Hyderabad, India
Research Intern - Prof. Manish Shrivastava — *May 2018 – June 2018*

- Devised a cross-sentence attention based neural network for Semantic Textual Similarity task, achieving a PCC score of 0.76 as compared to the state-of-the-art PCC score of 0.81 at the time of internship.

Research Intern - Prof. Dipti Misra Sharma — *Dec 2017 – Jan 2018*

- Developed part of speech tagging (POS) module for an Indian Vernacular language, named Kannada using Machine Learning (mainly CRF and SVM) and LSTM, achieving state-of-the-art results of 92% F1 score, which was 6% higher as compared to the previous results, with more POS classes.

Publications

- **Knowledge Graph Generation with Deep Active Learning** : Abhishek Pradhan*, Ketan Kumar Todi*, Anbarasan Selvarasu, Atish Sanyal. IJCNN, Glasgow, UK, 2020
- **Decentralizing Air Traffic Flow Management with Blockchain-based Reinforcement Learning**: Ta Duong, Ketan Kumar Todi, Umang Chaudhary, Hong-Linh Truong. INDIN, Finland, 2019
- **Building a Kannada POS Tagger Using Machine Learning and Neural Network Models**: Ketan Kumar Todi*, Pruthwik Mishra*, Dipti Misra Sharma. CICLING, Vietnam, 2018
- **Sentiment Classification of review Data Using Sentence Significance Score Optimization**: Ketan Kumar Todi, Muralikrishna SN, Ashwath Rao B. IJDATS Journal

Programming Skills

- **Languages**: : C++, Python, Java, LATEX
- **ML Libraries**: : Pytorch, Spacy, Pandas, NLTK, scikit-learn