

Nickil Maveli

CONTACT INFORMATION

86/3 Spring Gardens
Edinburgh EH8 8HX
United Kingdom

Email: n.maveli@sms.ed.ac.uk
Website: <https://nickilmaveli.com>
Phone: +44 7342 792335

RESEARCH INTERESTS

I am broadly interested in Natural Language Processing, Machine Learning, and Information Retrieval. My recent research has focused on multilingual unsupervised syntactic parsing, semi-supervised learning, interpretability and robustness of machine learning algorithms, and understanding social inequality and bias in language.

EDUCATION

The University of Edinburgh, Scotland, UK 2020–2021
M.Sc. by Research, Informatics (Distinction)

Research Unit: Institute for Language, Cognition and Computation (ILCC)

Research Area: Unsupervised Parsing, Knowledge Graph Completion

Advisor: Prof. [Shay Cohen](#)

Committee: Frank Keller (internal), Ankur Parikh (external)

CMR Institute of Technology, Bangalore, India 2011–2015
Bachelor of Engineering in Electronics and Communication
G.P.A.: 7.72/10

PUBLICATIONS

Preprints

2. Wagging the Entity Tail: Using Coupled Tensor Decomposition to Improve Rare Entity Knowledge Graph Completion.
Esma Balkır, **Nickil Maveli**, Shay Cohen
Under review at ARR
1. Co-training an Unsupervised Constituency Parser with Weak Supervision [PDF]
Nickil Maveli, Shay Cohen
Under review at ACL 2022

Workshop Publications

1. EdinburghNLP at WNUT-2020 Task 2: Leveraging Transformers with Generalized Augmentation for Identifying Informativeness in COVID-19 Tweets [PDF]
Nickil Maveli
Proceedings of the 2020 EMNLP Workshop W-NUT

Theses

1. Co-training a Weakly Supervised Constituency Parser of Natural Language [PDF]
Nickil Maveli
Master's thesis

WORK EXPERIENCE

Cookpad Inc., Bristol, UK Oct. 2021–Present
Applied Researcher - NLP

- Operationalized an active learning based multilingual meal-classification algorithm for recipes to improve the mean reciprocal rank by 45% compared to existing methods.

Niki.ai, Bangalore, India May 2018–Sept. 2020
Senior Software Developer - NLP

- Developed a real-time pause detection algorithm on speech signals to identify end-of-speech conditions, which was later integrated on-device. Up to an SNR of 8 dB, the algorithm was able to correctly segment with an accuracy of 91% on different lengths of speech.
- Designed a probabilistic model capable of generating a binary outcome and intends to classify if an utterance is valid parse or not at every stage in the conversation. Engineered 63 statistical features from parse tree capable of achieving an accuracy of 97% on the holdout test set.
- Developed a bot that could respond to about 75 FAQs in regional languages and English through a Pooled GRU in conjunction with the FastText model. Customer Satisfaction scores increased by 35%, leading to better assistance at every dialogue stage.
- Built a character-level RNN model to classify utterances as ‘Hinglish’ (blend of Hindi and English) words with minimum lexicon support in combination with a dictionary lookup.
- Implemented an algorithm formed on Likelihood and Reliability scores to automatically generate contextual synonyms at n-gram context and augment concept-grammar data to avoid manual intervention.

Sociograph Solutions, Bangalore, India**Sept. 2016–Feb. 2018***Data Scientist - Research and Development*

- Built a deep learning time series model capable of generating weekly/monthly sales forecasts of products at SKU level with only a few thousands of supervised training data.
- Devised a robust machine learning model to predict and optimize various e-commerce metrics like sales conversion rate, customer lifetime value, average order value, shopping cart abandonment rate for retailers that led to 30% decrease in the drop off rate.
- Created an image search engine capable of recommending similar images based on segmentation as a selective search to increase the MAP by 45% compared to the available benchmark.

Hotify.ai, Bangalore, India**Feb. 2016–May 2016***Junior Data Scientist*

- Aggregated tweets from pre-selected news sources to estimate trending hashtags.
- Created Actionable Insights by analyzing how select Twitter users through A/B testing react to real-time events by organizing frequencies into temporal buckets to observe the distribution of tweets over time.
- Prepared training data of close to 10K samples by carefully examining RSS feeds of news sources and rated each source using factors such as true reach, amplification, and network impact.

INTERNSHIPS**Google Summer of Code (GSoC), Remote****May 2021–Aug. 2021***Student Developer*

Organization: Red Hen Lab and FrameNet Brazil

Project: Detecting Joint Meaning Construal by Language and Gesture [\[Details\]](#) [\[Slides\]](#)Mentor: Prof. [Tiago Torrent](#)**RESEARCH
ASSISTANTSHIP****University of Edinburgh, Remote****Aug. 2021–Oct. 2021***Research Assistant*

Organization: Usher Institute (University of Edinburgh) and Institute of Health Informatics (UCL)

Project: Clinical Natural Language Processing for automated medical coding [\[Details\]](#)Mentor: Prof. [Honghan Wu](#) and Dr. [Hang Dong](#)**TECHNICAL
SKILLS****Specialties**

Natural Language Processing, Machine Learning, Deep Learning, Reinforcement Learning, Recommender Systems, Time-series Forecasting

LanguagesPython, R, Matlab, Javascript, SQL, HTML/CSS, \LaTeX **Tools and
Frameworks**

Pandas, NumPy, SciPy, Matplotlib, Scikit-Learn, NLTK, Keras, PyTorch, TensorFlow, PySpark, Flask, Django, Docker, Kubernetes, AWS, Jira, Git, Elastic Stack, MongoDB

ACHIEVEMENTS	Selected to participate in the Lisbon Machine Learning School (LxMLS 2021).	2021
	Secured Top 2% in the Jigsaw Multilingual Toxic Comment Classification Challenge hosted on Kaggle; Won a silver medal. [Details]	2020
	Finished 2 nd out of 479 competitors in the Quality Forecasting in Cement Manufacturing Challenge hosted on CrowdAnalytix; Won \$2500 as prize money; Winner representing India. [Details]	2019
	Secured Top 3% in the Quora Insincere Questions Classification Challenge hosted on Kaggle; Won a silver medal. [Details]	2019
	My team was awarded 2 nd and 1 st place in the Niki Annual Hackathon for the years 2019 and 2020 respectively and won cash prizes.	2019/20
	Represented niki.ai to volunteer at the NITI Aayog Indic NLP Workshop for tackling challenges in regional languages concerning NLP. [Details]	2018
	Secured Top 0.85% in the Cold Start Energy Forecasting Challenge hosted on DrivenData. [Details]	2018
	Placed 6 th out of 2310 competitors in the Predicting Poverty Challenge hosted on DrivenData. [Details]	2018
	Finished 2 nd out of 750 competitors in the Predicting How Points End in Tennis Challenge hosted on CrowdAnalytix; Won \$2500 as prize money; Featured in Tennis Australia's official press release . [Details] [Media Coverage]	2018
	Placed 3 rd out of 2975 competitors in the Click Prediction Challenge hosted on AnalyticsVidhya; Awarded a cash prize of ₹25000. [Details]	2017
	Finished 4 th out of 404 competitors in the Funding Successful Projects Challenge organized by HackerEarth. [Details]	2017
	Among Top Stack Overflow Users for tags – Python, Pandas, NumPy, Matplotlib with an overall 25000+ reputation score and more than 4 million post views worldwide. [Details]	2016–present
INVITED TALKS	1. University of Oxford, IMCC seminar [Details]	Dec. 2021
SERVICE	Reviewer:	
	ACL ARR, AmericasNLP	2021
	W-NUT@EMNLP (Shared Task 2), Scipy (Tutorials, Talk & Poster Presentations)	2020
	Volunteer:	
	ACL, ICML, NAACL, ICLR, EACL	2021
	EMNLP	2020
	Respondent:	
	Semantic Scholar Diary Study project conducted by the Allen Institute for AI	2020
DISCUSSANT	Discussant:	
	Selected at UAI 2021 to discuss about a Explainability/Fairness related paper	2021