

# Nickil Maveli

---

CONTACT INFORMATION	10 Crichton Street Newington, Edinburgh EH8 9AB United Kingdom	Email: <a href="mailto:n.maveli@sms.ed.ac.uk">n.maveli@sms.ed.ac.uk</a> Website: <a href="https://nickilmaveli.com">https://nickilmaveli.com</a>
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	<b>The University of Edinburgh</b> , Scotland, UK <i>M.Sc. by Research, Informatics</i> Research Unit: Institute for Language, Cognition and Computation (ILCC) Research Area: Unsupervised Parsing, Knowledge Graph Completion Advisor: Prof. Shay Cohen	2020–2021
	<b>CMR Institute of Technology</b> , Bangalore, India <i>Bachelor of Engineering in Electronics and Communication</i> G.P.A.: 7.58/10	2011–2015
PUBLICATIONS	Authors who contributed equally are indicated using a superscript ‘dagger’ symbol, i.e. †, while joint (co-) supervision is denoted using a superscript ‘double dagger’, i.e. ‡. All publications, except ‘Preprints’, underwent the peer-review process.	
	<b>Workshop Publications</b>	
	W1. EdinburghNLP at WNUT-2020 Task 2: Leveraging Transformers with Generalized Augmentation for Identifying Informativeness in COVID-19 Tweets [PDF] <b>Nickil Maveli</b> <i>Proceedings of the 2020 EMNLP Workshop W-NUT</i>	
WORK EXPERIENCE	<b>Niki.ai</b> , Bangalore, India <i>Senior Software Developer - NLP</i>	May 2018–Sept. 2020
	<ul style="list-style-type: none"><li>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.</li><li>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.</li><li>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.</li><li>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.</li><li>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.</li></ul>	
	<b>Sociograph Solutions</b> , Bangalore, India <i>Data Scientist - Research and Development</i>	Sept. 2016–Feb. 2018
	<ul style="list-style-type: none"><li>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.</li></ul>	

- 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**  
*Junior Data Scientist*

Feb. 2016–May 2016

- 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.

#### TECHNICAL SKILLS

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

**Languages** Python, R, Matlab, Javascript, SQL, HTML/CSS, L<sup>A</sup>T<sub>E</sub>X

**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

Secured Top 2% in the Jigsaw Multilingual Toxic Comment Classification Challenge hosted on Kaggle; Won a silver medal. [[Details](#)]

2020

Finished 2<sup>nd</sup> out of 479 competitors in the Quality Forecasting in Cement Manufacturing Challenge hosted on CrowdAnalytix; Won \$2,500 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

Selected by [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<sup>th</sup> out of 2310 competitors in the Predicting Poverty Challenge hosted on DrivenData. [[Details](#)]

2018

Finished 2<sup>nd</sup> 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<sup>rd</sup> out of 2975 competitors in the Click Prediction Challenge hosted on AnalyticsVidhya; Awarded a cash prize of ₹25000. [[Details](#)]

2017

Finished 4<sup>th</sup> 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

Ranked under 2000 worldwide out of ≈1 million users on Sphere Online Judge (SPOJ). [[Details](#)]

2015–present

My team was awarded 2<sup>nd</sup> and 1<sup>st</sup> place in the Niki Annual Hackathon for the years 2019 and 2020 respectively and won cash prizes.

2019/20

**SERVICE****Reviewer:**

AmericasNLP

2021

W-NUT@EMNLP (Shared Task 2), Scipy (Tutorials, Talk &amp; Poster Presentations)

2020

**Volunteer:**

NAACL, ICLR, EACL (Social events)

2021

EMNLP (Lead Zoom)

2020

**Respondent:**

Semantic Scholar Diary Study project conducted by the Allen Institute for AI

2020

**SPOKEN LANGUAGES**

Hindi      Native

Malayalam      Limited Working

Kannada      Limited Working

English      Full Professional

**REFERENCES**

Available upon request.