# Nickil Maveli

Email: n.maveli@sms.ed.ac.uk
Website: https://nickil21.github.io

## 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, understanding social inequality and bias in language.

## Education

- **The University of Edinburgh** — Scotland, UK
  *MSc by Research in Informatics - Institute for Language, Cognition and Computation (ILCC)* — *Sept. 2020 – Present*
  Advisor: Prof. Shay Cohen
- **CMR Institute of Technology** — Bangalore, India
  *Bachelor of Engineering in Electronics and Communication; CGPA: 7.58/10* — *Aug. 2011 – July. 2015*

## Publications

### Preprints

* - equal contribution

1. EdinburghNLP at WNUT-2020 Task 2: Leveraging Transformers with Generalized Augmentation for Identifying Informativeness in COVID-19 Tweets
   **Nickil Maveli**
   *arXiv:2009.06375 (under review), 2020.*

## Work Experience

- **Niki.ai** — Bangalore, India
  *Senior Software Developer - NLP* — *May 2018 - Sept. 2020*
  - **Voice Activity Detection**: Developed a real-time pause detection algorithm on speech signals to identify end-of-speech condition which was later integrated on-device. Upto an SNR of 8 dB, the algorithm was able to correctly segment with an accuracy of 91% on different lengths of speech.
  - **Parse Classification**: 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.
  - **FAQ Bot**: Developed a bot that could provide responses for about 75 FAQ's in regional languages as well as English through a Pooled GRU in conjunction with FastText model. Customer Satisfaction scores increased by 35% leading to better assistance at every dialogue stage.
  - **Code Mixing**: 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. This gave an F score of 82% on the holdout test set and helped in identifying such users for better targeted campaigns.
  - **Discovery Platform**: Implemented an algorithm formed on Likelihood and Reliability scores to automatically generate contextual synonyms at n-gram context and augment concept-grammar data so that manual intervention can be avoided. Perplexity metric was minimized upon several iterations.
- **Sociograph Solutions** — Bangalore, India
  *Data Scientist - Research and Development* — *Sept. 2016 - Feb. 2018*
  - **Timeseries Forecast**: Built a deep learning time series model capable of generating weekly/monthly sales forecasts of products at SKU level with only few thousands of supervised training data. The RMSE metrics were converging after several iterations.
  - **Opportunity Analysis**: 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 drop off rate.
  - **Image Recommender**: Created an image search engine capable of recommending similar images based on segmentation as a selective search to increase the mean average precision by 45% compared to the benchmark that was available.
- **Hotify.ai** — Bangalore, India
  *Junior Data Scientist* — *Feb. 2016 - May 2016*

- **Twitter Mining**: Aggregated tweets from pre-selected news sources to estimate trending hashtags. Naive Bayes algorithm was shown to be most effective when combining both content and contextual data, successfully classifying tweets as containing a trend or not with an accuracy of 75% - 85%.
- **Data Visualization**: 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.
- **Data Consolidation**: 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.

## Skills

- **Specialties**: Machine Learning, Deep Learning, Reinforcement Learning, Natural Language Processing, Recommender Systems, Time-series Forecasting
- **Languages**: Python, R, Matlab, Javascript, SQL, HTML/CSS, LaTeX
- **Tools/Frameworks**: Pandas, NumPy, SciPy, Matplotlib, Scikit-Learn, PySpark, NLTK, OpenCV, Keras, PyTorch, TensorFlow, Flask, Django, Docker, Kubernetes, AWS, Jira, Git, Elastic Stack, MongoDB

## Achievements

- **Jigsaw Multilingual Toxic Comment Classification**: Secured Top 2% hosted on Kaggle; Won a silver medal. [Details]
- **Quality Forecasting in Cement Manufacturing**: Finished 2nd out of 479 competitors hosted on CrowdAnalytix; Won $2,500 as prize money; Winner representing India. [Details]
- **Quora Insincere Questions Classification Challenge**: Secured Top 3% hosted on Kaggle; Won a silver medal. [Details]
- **NITI Aayog Indic NLP Workshop**: Selected by niki.ai to volunteer for tackling challenges in regional languages concerning NLP. [Details]
- **Cold Start Energy Forecasting**: Secured Top 0.85% conducted on DrivenData. [Details]
- **Predicting Poverty**: Placed 6th out of 2310 competitors organized by DrivenData. [Details]
- **Predicting How Points End in Tennis**: Finished 2nd out of 750 competitors hosted on CrowdAnalytix; Won $2500 as prize money; **Featured in Tennis Australia's official press release**. [Details] [Media Coverage]
- **Click Prediction**: Placed 3rd out of 2975 competitors conducted on AnalyticsVidhya; Awarded a cash prize of ₹25000. [Details]
- **Funding Successful Projects**: Finished 4th out of 404 competitors organized by HackerEarth. [Details]
- **Stack Overflow Ninja**: Among top Stack Overflow users for tags - Python, Pandas, Numpy, Matplotlib with an overall 20000+ reputation score and more than 2 million post views worldwide. [Details]
- **Sphere Online Judge**: Ranked under 2000 worldwide out of 0.3 million users. [Details]
- **Niki Hackathon**: My team was awarded 2nd and 1st place in the Niki Annual Hackathon for the years 2019 and 2020 respectively and won cash prizes.

## Activities

- **Reviewer**: Scipy 2020 (Tutorials, Talk and Poster Presentations), W-NUT 2020 (Shared Task 2)
- **Volunteer**: EMNLP 2020 (Lead Volunteer).
- **Respondent**: Semantic Scholar Diary Study project conducted by AI2.

## References

Available upon request.