

Rohan Kshirsagar

CONTACT INFORMATION	+1 732 857 4254	rohan.kshirsagar@columbia.edu
EDUCATION	Columbia University , NY NY <i>MS Computer Science, Track: NLP & Machine Learning</i> GPA 4.0	2016 – current
	Duke University , Durham NC <i>BSE Computer Engineering, BS Computer Science</i>	2009 – 2013
EXPERIENCE	Research Scientist at Koko	December 2016 – current
	<ul style="list-style-type: none">• Trained text classification models using LSTMs, Global Pooling, and Mixture of Experts for detecting crisis in long texts. Beat IBM Watson Classification Service by AUC score by 5 points on test data.• Built error analysis and visualization tools for classification task in Python and Jupyter notebook	
	Research Assistant at Columbia University	August 2016 – December 2016
	<ul style="list-style-type: none">• Developed Recurrent Neural Network models in Keras for Target Extraction in Targeted Sentiment Analysis for Arabic Text• Beat the current state of the art CRF results in Target prediction using a combination of forward and bidirectional LSTMs on a variety of Arabic morphological schemes. Presentation here.	
	Senior Software Engineer at Bloomberg LP NLP Group	July 2013 – August 2016
	<ul style="list-style-type: none">• Technical owner of the News Extraction and Headline Generation Product. Interfaces with product stakeholders, plans and implements enhancements for UI, Engine, and Metrics.• Developed Structural Analysis (inferring semantic parent-child relationships in tables) within Table Understanding & Extraction Project. Created a tagging methodology for structural analysis, and yielded .93 F1 Score using a linear model• Conducted novel research on Network Monitoring ASIC Chips. Implemented a n-gram language model using regular expressions to see a 3x gain on smaller sized language models. Reversed Engineered <code>langid.py</code> to implement Language Identification on the chip to see a 6x performance gain. Research published as a short paper in ACL 2015	
SIDE PROJECTS	Trump vs Clinton Authorship Prediction and Discourse Analysis	Github
	<ul style="list-style-type: none">• Gathered a balance corpus of speech transcripts from Trump and Clinton of around 7,000 sentences• Trained several binary classifiers on a bag of lexical and part-of-speech ngrams to achieve .89 accuracy on hold out set. Wrote a Medium post that analyzes the informative features• Deployed model on heroku and built a web app to allow anyone to analyze the Nominees' discourse and predict authorship. Link: nomination-analysis.herokuapp.com	
	Columbia Graduate Student Affairs FAQ Bot	Presentation and Demo
	<ul style="list-style-type: none">• Worked with Columbia Graduate Student Affairs to bound the question-answering domain and gather question/answer dataset• Developed technology to train a bot automatically from an annotated question-answer dataset. Utilized word2vec, Watson, nltk, and scikit-learn to extract entities and intents from incoming queries.	
CONFERENCE PAPERS	Kenneth Heafield, Rohan Kshirsagar , and Santiago Barona. Language Identification and Modeling in Specialized Hardware. ACL, Beijing, China, July, 2015.	
LANGUAGES	C/C++, Python, bash, Javascript, SQL	