

Dhruv Kumar

(+1) 226-9783751 • ddhruvkr@gmail.com • Github • Website • LinkedIn

EDUCATION

University of Waterloo, Ontario, Canada

Sep 2018- Jul 2020

M. Math.(Thesis) in Computer Science

GPA: 86.75/100

Thesis: Iterative Edit-based Unsupervised Sentence Simplification

Coursework: Deep Learning, Trust Modeling and Online Social Networks, Blockchains: Foundations and Applications, Privacy and Fairness in Data Science

Indian Institute of Information Technology, Allahabad, U.P., India

Jul 2012- Jun 2016

B. Tech.(Hons) in Information Technology

GPA: 8.68 / 10.00 (9.07 - Last 2 years)

Thesis: Compressed Knowledge transfer via Factorization Models in Recommender Systems

Relevant Coursework: Information Retrieval, Artificial Intelligence, Natural Language Processing, Cognitive Process Modelling, Optimization Techniques, Data Mining, Probability & Statistics, Mathematics(I, II, III).

PUBLICATIONS

- O. Vechtomova, G. Sahu, **D. Kumar**, "Generation of lyrics lines conditioned on music audio clips", *In Workshop on NLP for Music and Audio (NLP4MusA) at ISMIR 2020*
- **D. Kumar**, L. Mou, L. Golab, O. Vechtomova, "Iterative Edit-based Unsupervised Sentence Simplification", *In Proceedings of ACL 2020 - Long Paper*
- R. Cohen, R. Agarwal, **D. Kumar**, A. Parmentier, T. H. Leung, "Sensitivity to risk profiles of users when developing AI systems", *33rd Canadian Conference on Artificial Intelligence (Canadian AI 2020)*
- R. Agarwal*, **D. Kumar***, L. Golab, S. Keshav, "Consentio: Managing Consent to Data Access using Permissioned Blockchains", *IEEE International Conference on Blockchain and Cryptocurrency (ICBC) 2020*
- **D. Kumar**, R. Cohen, L. Golab, "Online abuse detection: the value of preprocessing and neural attention models", in *10th workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA) at NAACL-HLT 2019*
- N. Kushwaha, S. Mehrotra, R. Kalia, **D. Kumar**, O. P. Vyas, "Inclusion of Semantic and Time-Variant Information Using Matrix Factorization Approach for Implicit Rating of Last.Fm Dataset," in *Springer Arabian Journal of Science and Engineering*, May 2016.

RESEARCH EXPERIENCE

Borealis AI, Toronto, Canada,

Machine Learning Research Intern

Sep 2020-Present

- "*Neural semantic parsing*": Working on neural semantic parsing models for text-to-SQL. More specifically, exploring if explicit schema linking and regularization techniques can help improve cross-domain generalizability and performance of a state-of-the-art semantic parser (RAT-SQL).
[Supervisor: Dr. Yanshuai Cao]

University of Waterloo, Waterloo, Canada

- *Unsupervised sentence simplification*

Jul 2019-Jul 2020

Designed an unsupervised algorithm building on edit-based text generation techniques for sentence simplification. The model is highly adaptable and can achieve SARI and BLEU scores of **30.44** and **26.47** respectively on the Newsela dataset.

[Supervisors: Prof. Lili Mou, Prof. Lukasz Golab, Prof. Olga Vechtomova]

- *Music audio conditioned lyrics generation*

Sep 2019-Apr 2020

Designed bimodal neural network models based on variational autoencoders (VAEs) to generate lines of lyrics for instrumental piece of music.

[Supervisor: Prof. Olga Vechtomova]

- *Anonymize personal attributes in human generated text*

Oct 2019-Dec 2019

As part of the course "Privacy and Fairness in Data Science", worked with the PASTEL and Yelp datasets to build an unsupervised neural model for multi-attribute style transfer. The model uses a combination of denoising, cycle consistency and classification losses.

[Supervisor: Prof. Xi He]

	<ul style="list-style-type: none"> ▪ <i>Consent Management System</i> Jan 2019-Aug 2019 Designed Consentio, a permissioned blockchain-based consent management system (CMS) that can handle up to 6000 transactions per second. [Supervisors: <i>Prof. Lukasz Golab, Prof. Srinivasan Keshav</i>] ▪ <i>Online abuse detection in Social Networks</i> Oct 2018-Feb 2019 Proposed a co-attention based neural model to detect implicit aggressiveness and incivility in online social networks. The model achieves a F1 score of 82.41, 77.75 and 76.07 for the minority abuse class on the Wikipedia toxicity/attack/aggression datasets respectively. [Supervisors: <i>Prof. Robin Cohen, Prof. Lukasz Golab</i>] ▪ <i>Attention-based Text classification</i> Dec 2018-Jan 2019 Implemented various state of the art deep learning models (e.g. Co-attention, Self-attention, Hierarchical attention) for text classification. The models were inspired by research conducted for different tasks in NLP. ▪ <i>Span-based Question Answering</i> Oct 2018-Dec 2018 As part of the "Deep Learning" course, implemented the Bi-Directional Attention Flow (BiDAF) model from scratch over the SQuAD1.1 dataset. [Supervisors: <i>Prof. Ali Ghodsi</i>]
	<p>Universität Paderborn, Paderborn, Germany Jan 2016-Jun 2016</p> <ul style="list-style-type: none"> ▪ Bachelor's Thesis: "Compressed Knowledge transfer via Factorization Models in Recommender Systems". Developed an algorithm to incorporate metadata in Factorization Machines, achieving the RMSE value of 0.836 as compared to 0.853 when using a Joint Matrix Factorization method on the MovieLens 1M dataset. [Supervisors: Dr. Artus Krohn-Grimberghe, Prof. Ratna Sanyal]
	<p>Indian Institute of Information Technology, Allahabad, India Jul 2015-Dec 2015</p> <ul style="list-style-type: none"> ▪ "Analysis of Time-Aware and Semantic Feature Based Music Recommender System". Proposed a Joint Matrix Factorization algorithm for Music Recommendation that utilizes geographical and time-based tagging information of artists, in addition to implicit user feedback (user clicks) to provide better recommendations. [Supervisor: Dr. O.P.Vyas]
TEACHING ASSISTANT EXPERIENCE	<ul style="list-style-type: none"> ▪ CS 115: Introduction to Computer Science, Fall 18, University of Waterloo. ▪ CS 230: Introduction to Computers and Computer Systems, Winter 19, Spring 20, University of Waterloo. ▪ CS 231: Algorithmic Problem Solving, Spring 19, University of Waterloo. ▪ CS 241: Foundations of Sequential Programs, Fall 19, University of Waterloo. ▪ CS 489/698: Topics in Computer Science, Neural Networks, Winter 20, University of Waterloo.
SERVICE	<ul style="list-style-type: none"> ▪ Program Committee / Reviewer: NeurIPS2019 AISG (AI for Social Good) Workshop ▪ Program Committee / Organizer: GEM: Natural Language Generation, Evaluation, and Metrics- ACL 2021
OTHER EXPERIENCE	<p>Arcesium(DE Shaw Group), Hyderabad, India Jul 2016-May 2018 Software Engineer, Fund and Investor Accounting</p> <ul style="list-style-type: none"> ▪ Enhanced the post-trade automation platform for funds operated by J.P. Morgan and D.E. Shaw. <p>Citigroup, Pune, India Software Engineering Intern, Equities May 2015-Jul 2015</p> <ul style="list-style-type: none"> ▪ Designed the first prototype of the Trading Controls application. Declined the full-time offer.
SKILLS	Python, Pytorch, Java, C/C++, SQL, Flask
ACHIEVEMENTS & EXTRA-CURRICULAR	<ul style="list-style-type: none"> ▪ Received International Masters Student Award and University of Waterloo Entrance Scholarship for graduate studies. ▪ Got Graduate Studies Research Travel Assistantship to attend the NAACL-HLT Conference, 2019. ▪ Accepted in the <i>Deep and Reinforcement Learning Summer School</i>, 2019 (less than 25% acceptance rate). ▪ Stood in the top 0.5% in the All India Engineering Entrance Examination 2012. ▪ Worked as the Events Head of the annual cultural cum technical festival <i>Effervescence</i> 2014. ▪ Represented the college band as a drummer.