

# Vivek Gupta

*Ph.D Applicant in Machine Learning*

CONTACT INFORMATION	Microsoft Research India #9, Lavelle Road Bangalore, India - 560001	website keviv9@gmail.com +91-8765696260
RESEARCH INTERESTS	Distributional Semantics, Extreme Classification, Resource Constraint Machine Learning	
CURRENT POSITION	<b>Microsoft Research India</b> <i>Research Fellow</i>	(June 2016 - Present) Advisors: Dr. Praneeth Netrapalli & Dr. Prateek Jain
EDUCATION	<b>Indian Institute of Technology Kanpur</b> M. Tech with Honors in Computer Science and Engineering CGPA: 9.3/10 <b>Indian Institute of Technology Kanpur</b> B. Tech with Honors in Computer Science and Engineering CGPA: 7.5/10	(July 2015 - May 2016)  (July 2011 - July 2015)
ACCEPTED PUBLICATIONS	<b>Gupta, V.</b> , Mekala, D.,Paranjape, B., Karnick, H. " <i>Sparse Composite Document Vectors using soft clustering over distributional representations</i> " , <b>EMNLP 2017</b> (Long Oral) [ArXiv] <b>Gupta, V.</b> , Karnick, H.,Bansal, A., Jhala, P. " <i>Product Classification in E-Commerce using Distributional Semantics</i> " , <b>COLING 2016</b> (Long Poster) [Paper][Poster] <b>Gupta, V.</b> , Mittal, S.,Bhaumik, S., Roy, R. " <i>Assisting Humans to Achieve Optimal Sleep by Changing Ambient Temperature</i> ", <b>BIBM 2016, BHI 2016 &amp; HI-DS 2016</b> workshop [Paper] [PPT]	
UNDER REVIEW PUBLICATIONS	Wadbude, R., <b>Gupta, V.</b> , Mekala, D., Karnick, H. " <i>User Bias Removal in Review Score Prediction</i> ", Under review in <b>DAB 2017</b> (Workshop Paper) [ArXiv][Poster] Dohare, S., Karnick, H., <b>Gupta, V.</b> , " <i>Exploring a full-fledged pipeline for Text Summarization using Abstract Meaning Representation</i> " Under review in <b>IJCNLP 2017</b> (Long Paper) [ArXiv] Wadbude, R., <b>Gupta, V.</b> , Rai, P., Natararjan, N.,Karnick, H. " <i>Leveraging Distributional Semantics for Multi-Label Learning</i> ", Under review in <b>NIPS 2017</b> Mahajan, D., <b>Gupta, V.</b> , Keerthi, S.,Sundararjan, S. " <i>Efficient Estimation of Generalization Error and Bias-Variance Components of Ensembles</i> ", Under review in <b>ICDM 2017</b> (Long Paper)	
MASTER THESIS	<b>Product Categorization in E-Commerce using Distributional Semantics</b> <i>Advisor: Prof. Harish Karnick (IIT Kanpur) &amp; Pradhuman Jhala (Flipkart.com)</i> [Thesis][PPT] <ul style="list-style-type: none"><li>Proposed a new distributional semantics representation for document which outperform other representation methods</li><li>Proposed a novel two-level ensemble approach utilizing (with respect to the taxonomy tree) a path-wise, node-wise and depth-wise classifiers for product classification</li><li>Achieve state of art results on various standard and novel proposed evaluation metrics</li></ul>	
RESEARCH EXPERIENCE	<b>Resource Constraint Anomaly Detection</b> <i>Advisor: Dr. Prateek Jain and Dr. Praneeth Netrapalli (Microsoft Research India)</i> <ul style="list-style-type: none"><li>Working on extension of ProtoNN (Resource Constraint Machine Learning) for semi-supervised setting</li><li>Working with company <i>magma international</i> on predictive maintenance, investigating time series features, novel evaluation metrics for anomaly detection</li></ul>	(June 2016 - Present)

**Leveraging Distributional Semantics for Multi-Label Learning** (June 2016 - May 2017)  
*Advisor: Dr. Nagarajan Natarajan (Microsoft Research India) Dr. Piyush Rai (IIT Kanpur), Dr. Harish Karnick (IIT Kanpur)*

- Design extreme multi-label learning algorithms using distributional semantics with more efficient training procedures
- Extended the approach to naturally incorporate other sources of side-information, in particular the label-label co-occurrence matrix

**Text Summarization using Abstract Meaning Representation** (June 2016 - May 2017)  
*Advisor: Dr. Harish Karnick (IIT Kanpur)*

- Explored a novel full-fledged pipeline for text summarization with an intermediate step of Abstract Meaning Representation (AMR)
- Proposed method achieves state-of the-art results compared to the other text summarization routines based on AMR
- Pointed out some significant problems in the existing evaluation methods making them unsuitable for evaluating summary quality

**Text Categorization using Sparse Composite Document Vector** (June 2016 - May 2017)  
*Advisor: Dr. Harish Karnick (IIT Kanpur)*

- Proposed a novel topic-based document representation which outperforms state-of-the-art models in multi-class and multi-label classification tasks
- Showed that fuzzy GMM clustering on word-vectors lead to more coherent topic than LDA and can be used to detect Polysemic words
- Embeddings provide a robust estimation of the query and document language models, thus improving the MAP of language model based retrieval systems

**Bayes Optimal Classification for Hierarchy** (June 2016 - Present)  
*Advisor: Dr. Purushottam Kar (IIT Kanpur) & Prof. Harish Karnick(IIT Kanpur)*

- Extended the consistency of hierarchical classification algorithm on asymmetric tree distance loss using calibrated surrogates
- Showed under reasonable assumption over hierarchy that the Bayes optimal classification for this asymmetric loss can be found in  $O(\log(n))$

**Generalization vs Optimization for Deep Learning** (June 2016 - Present)  
*Advisor: Dr. Praneeth Netrapalli, Microsoft Research India*

- Working on more efficient techniques (tail-averaging) for non-convex optimization for training deep neural nets
- Investigating the broad question of whether depth of neural nets help in generalization or optimization

RESEARCH  
INTERNSHIP

**Microsoft Research India, Bangalore** (May 2016 - Jul 2016)  
*Dr. Sundararajan Sellamanickam*

Estimation of generalization error for ensembles

- Worked on efficient Estimation of generalization error for ensembles using normality assumption on classification scores
- Worked on efficient prediction of accuracy, ensemble parameters, bias and variance of generalization errors using minimal number of ensembles

**Flipkart Internet Pvt. Ltd., Bangalore** (May 2015 - July 2015)  
*Pradhuman Jhala*

Web Scale Product Classification

- Developed a machine learned model for deep hierarchal classification of flipkart catalog hierarchy using multilevel classifiers and ensembles methods
- Model outperformed there naive based classifier with 89% with limited features for 900 classes with faster prediction time of 100x

	<b>Samsung R&amp;D Institute, Bangalore</b> Mobile and Healthcare Solution Y2014	(May 2014 - July 2014) <i>Sandip Bhaumik &amp; Raj Roy</i>
	<ul style="list-style-type: none"> <li>Did an innovation in S-Health Wearable technology on sleep applications by automatically change ambient temperature by using recurrent feedback signal from S-Watch</li> </ul>	
DEVELOPMENT INTERNSHIP	<b>Flipkart Internet Pvt. Ltd., Bangalore</b> Extracting Attribute Value Pairs	(Dec 2015 - Jan 2015) <i>Pradhuman Jhala</i>
	<ul style="list-style-type: none"> <li>Developed a machine learned multilabel multiclass model for extraction of attribute values from product apparel images</li> <li>Successfully extracted attribute values collar ,crew &amp; vneck ,full sleeve &amp; half sleeve for shirts {pointy, non-pointy} for sandals,{baseball, basketball, bracket &amp; ballet} for sport shoes with pair wise accuracy of 74%</li> </ul>	
	<b>Synopsys. Inc., Bangalore</b> Trend Generation for Dali Coverage	(May 2013 - July 2013) <i>Yogesh Pandey</i>
	<ul style="list-style-type: none"> <li>Created interactive visualization of Scenario coverage trend for functional verification project <i>Dali</i> using data visualizations technique</li> <li>Coded Object Oriented Filter pattern which parse json file and perform analysis on selected properties</li> </ul>	
SCHOLASTIC ACHIEVEMENTS	Selected among the top 7 finalist team in <b>Ericson's Innovation Awards, India</b> in 2016 Selected in top 100 teams all over India in <b>Internet for all challenge</b> by Unisys in 2015 <b>Best poster presentation</b> for Internship works at Flipkart Internet Private Ltd. in 2015 <b>Best project award</b> for project <i>Distributed Classroom</i> in Topic in Distributed System in 2015 Secured <b>All India Rank 183</b> in <b>IIT-JEE 2011</b> amongst more than 500,000 candidates Qualified for <b>Kishore Vaigyanik Protsahan Yojana (KVPY) Scholarship</b> in 2011 Selected among the top 1% of students in India, <b>Indian National Physics Olympiads</b> in 2011 Selected among the top 1% of students in India, <b>Indian National Chemistry Olympiads</b> in 2011 Selected among the top 1% of students in State, <b>Indian National Physics Olympiads</b> in 2010	
TRAVEL GRANTS	Student Volunteer for <i>EMNLP 2017</i> Conference in Copenhagen, Denmark Scholarship for attending <i>EMNLP 2017</i> by Microsoft Research India, Bangalore. Scholarship for attending <i>COLING 2016</i> by Flipkart Internet Pvt. Ltd., India. Scholarship for attending <i>Workshop on Brain, Computation and Learning, IISC, Bangalore</i> Scholarship for attending <i>IFCAM NMI Summer School, IISC, Bangalore</i>	
TALKS & SEMINARS	Text Categorization using Sparse Composite Document Vectors <i>CFILT Group, IIT Bombay</i>	(June 2017) <b>[PPT]</b>
	Research as a Student <i>Think Research Club, VIT, Bombay (Invited Talk)</i>	(June 2017) <b>[PPT]</b>
	Text Classification using Sparse Composite Document Vectors <i>YAML, Microsoft Research India, Bangalore</i>	(March 2017) <b>[PPT]</b>
	Natural Language Processing in E-Commerce, A Case Study <i>Botathon Event Forge Accelerator, Coimbatore</i>	(Feb 2017) <b>[PPT]</b>
	Speeding up Stochastic Gradient <i>RF Talk, Microsoft Research India, Bangalore</i>	(Nov 2016) <b>[PPT]</b>
	Product Classification using Distributional Semantics <i>Microsoft Research India, Bangalore and IBM Research Lab, India</i>	(August 2016) <b>[PPT]</b>

	Ocular, Vision for Visually Impaired <i>Erricson Innovation Award, IIT Delhi</i>	(May 2016) [PPT] [PIC]
	Parametric R Tree, Indexing of Moving Object <i>Indexing and Searching in Large Databases Seminar</i>	(Dec 2015) [PPT]
	Deep Classification in Large Scale Web Hierarchy <i>MLRD, SIGML IIT Kanpur and Flipkart Internet Pvt. Ltd.</i>	(August 2015) [PPT]
PROFESSIONAL SERVICE	<p><b>External Reviewer</b> for EMNLP 2017: Served as a reviewer for long and short papers for Empirical Methods in Natural Language Processing 2017</p> <p><b>Coordinator</b> Initiated and managed Special Interest Group in Machine Learning at Computer Science and Engineering Department, IIT Kanpur. Organize regular meetups for discussions and talks on topics in Machine Learning and related fields.</p> <p><b>Teaching Assistant</b> for MLT 2016 - Machine Learning Tool and Technique: Mentored a group of 30 M-Tech students part of an introductory course on Machine Learning. Set up a labeling software for project work, resulting in a new dataset.</p> <p><b>Student Secretary</b> in Promotion of Work Experience and Research PoWER, Office of Dean of Research and Development IIT Kanpur</p> <p><b>Student Mentor</b> in Alumni Contact Program (ACA) under Office of Dean of Research and Development IIT Kanpur</p>	
STUDENT MENTORSHIP	<p>Dheeraj Mekala, IIT Kanpur (Jointly with Prof. Harish Karnick and Prof. Puruhottam Kar) on Bayes Optimal Classifier for Hierarchical Classification [Report]</p> <p>Rahula Wadbude, IIT Kanpur (Jointly with Prof. Harish Karnick, Nagararjan Natarajan, Prof. Piyush Rai) on Leveraging Distributional Semantics for Multi-Label Learning (Submitted NIPS) [Report]</p> <p>Dheeraj Mekala, IIT Kanpur (Jointly with Prof. Harish Karnick) on Sparse Composite Document Vectors (EMNLP 2017) [Report]</p> <p>Rishabh Gupta and Sanjari Srivasatva, IIT Kanpur (Jointly with Prof. Harish Karnick) on Topic Modeling using Independent Component Analysis [Report]</p> <p>Rahul Wadbude and Dheeraj Mekala, IIT Kanpur (Jointly with Prof. Harish Karnick) on User Bias removal in Review Score Prediction (Submitted DAB 2017) [Report]</p> <p>Avikalp Srivastava, IIT Kharagpur (Jointly with Dr. Sundararajan Sellamanickam) on Interpretable and Efficient Word and Document embedding [Report]</p> <p>Shibhansh Dohare, IIT Kanpur (Jointly with Prof. Harish Karnick) on Text Summarization using Abstract Meaning Representation [Report]</p>	
OTHER PROJECTS	<p><b>S-Voice Commerce Dialog Learning</b> (Jan 2016 - Apr 2016) <i>Advisor: Prof. Harish Karnick (IIT Kanpur) &amp; Praveen Reddy (Samsung Research)</i></p> <ul style="list-style-type: none"> <li>• Worked on dialog state tracking system and problem of attribute value extraction for Samsung S-Voice Commerce</li> <li>• Referring to literature for state space models (DSTC Challenge), dependency based wordvec and sequence learning DNN models (LSTM, RNN)</li> </ul> <p><b>Independence &amp; Randomization in Deep Neural Net</b> (Jan 2016 - Apr 2016) <i>Advisor: Prof. Harish Karnick(IIT Kanpur) &amp; Gaurush Hiranandani (Adobe Research)</i></p> <ul style="list-style-type: none"> <li>• Worked on theoretic proof of <i>Improved Classification and Reconstruction using Independence and Randomization</i> in Deep Neural Networks DICTA 2015</li> <li>• Evaluated proposed non-linear dimensionality reduction techniques by appropriately scoring topology preservation and information gain perspective</li> </ul> <p><b>Introducing Diversity Priors in Latent Variable Model</b> (Jan 2016 - Apr 2016) <i>Advisor: Prof. Piyush Rai (IIT Kanpur)</i> [Report]</p>	

- Worked on Diversity Priors Regularizes for Latent Variable Models like Probabilistic PCA and Non-Negative Matrix Factorization
- Introducing log-det divergence, determinantal point process and pair-wise Mutual angular regularizes in standard probabilistic models

**Large Scale Multi Kernel Learning** (July 2015 - Dec 2015)  
*Advisor: Prof. Harish Karnick(IIT Kanpur)* [Report][PPT]

- Showed that Multi Kernel Learning work better than single kernel independent data separability and distribution in space
- Extended toolbox for multi-modal data using different kernels for data attributes from different heterogeneous sources

**Forest Cover Type Prediction** (Dec 2014 - Apr 2015)  
*Advisor: Prof. Harish Karnick(IIT Kanpur)* [Report][PPT]

- Predict forest type of forest from 7 classes using 54 attributes ( kaggle data ) and was ranked 42 with 82% accuracy on kaggle.
- Used classification algorithms like Random Forests, Extra-Trees Classifier, Auto-encoders, Gradient Boosting Model, K -Nearest Neighbor, Label propagation, Support Vector Machine etc.

**Location Based Approximate Query** (Dec 2014 - Apr 2015)  
*Advisor: Prof. Arnab Bhattacharya (IIT Kanpur)* [Report]

- Modified a Filter Effective Hybrid tree (FEH) for indexing approx geo-location queries using filter tree based on distance metric like Euclidean & Manhattan
- Result out perform in search time by significant margin for R\* tree using line extra space. Also developed a Graphical User Interface using Google Map API for demo

**Distributed Classroom (PANGO)** (Aug 2014 - Dec 2014)  
*Advisor: Prof. Ratan K Ghosh(IIT Kanpur)* [Report][PPT][Video][JED-I]

- Developed a server-less Distributed Application for File Sharing, Digital Whiteboard, Online Tests, Live Lecture Slideshare, Video and Document Tagging and automatic Lecture Scribing
- Implemented various distributed protocols like token based leader election, mutual exclusion joining, deadlock prevention, novel distributed file sharing and non-buffering syncing protocol

**Image Search Engine-PicBook** (Aug 2014 - Dec 2014)  
*Advisor: Prof. TV. Prabhakar(IIT Kanpur)* [Report][PPT][Video]

- Developed a collaborative social Image search engine (PICBOOK) using Neo4j graph database which capture collaborative clicks
- Software support automating tagging, re-ranking, graph search queries like *Photos of my friend friends, my photos, female friend photos*, text search, uploads, friends addition

**Image Retrieval & Re-Ranking** (Aug 2014 - Nov 2014)  
*Advisor: Prof. Harish Karnick* [Report]

- Work on a new approach to Image retrieval and ranking using latent semantic indexing and cross view learning. Project was co-mentored by Yahoo India R&D lab
- Learned a latent space for single representation of cross views (text-image) from Latent Semantic Analysis features for text semantic & SIFT, HOG and RGB for image semantics

## EXTRA CURRICULARS

Active Volunteer for teaching underprivileged student at Prayas IIT Kanpur for 2 years

C Rank National Cadet Corps is the Indian military cadet corps.

Attended Microsoft Research India Summer Workshop on Artificial Social Intelligence 2017

Attended Microsoft Research India Summer School on Internet of Things 2016