# Sanchit Agarwal

CONTACT INFORMATION

5819 Elwood Street Apt#6
Pittsburgh, PA 15232, USA

*+1-412-499-2271*
*sagarwa1@cs.cmu.edu*

EDUCATION

**Carnegie Mellon University, School of Computer Science, Pittsburgh**   December 2016
*Master of Science in Intelligent Information Systems*   **4.04/4.00**
**Indian School of Mines (ISM), Dhanbad**   May 2013
*Integrated Master of Technology in Mathematics and Computing*   **9.37/10.00**

TECHNICAL SKILLS

**Programming:** C, C++, Java, Python, MATLAB, Postgres, Android development, LATEX
**Technologies:** Eclipse, Git, SVN, Perforce, Lucene, PySpark, Hadoop, Scikit-learn, Stanford CoreNLP

WORK EXPERIENCE

**Amazon Inc., Cambridge, USA**   May 2016 – August 2016
*Machine Learning Scientist Intern*
Worked with Alexa science team to investigate the problem of miss-recognition of product names during voice shopping on Amazon Echo. Built a statistical re-ranker to extract the "best" product name starting with n-best ASR hypothesis. The re-ranker showed 11% relative decrease in product name mismatch.

**Samsung India Electronics, Noida, India**   June 2013 – July 2015
*Lead Engineer*
Worked on the app team of Samsung's intelligent personal assistant, *SVoice* as Android developer.

PROJECTS

**Classifying real and fake text**   April 2016
*Carnegie Mellon University, Pittsburgh, USA*
Built a classifier to segregate real Broadcast news articles from trigram-generated fake articles. Used features such as n-grams, tf-idf, article length and perplexity w.r.t 3-gram and 4-gram models. SVM with linear kernel provided the best accuracy.

**Re-ranking machine generated questions**   Jan 2016 – April 2016
*Carnegie Mellon University, Pittsburgh, USA*   *Advisor*: Prof Teruko Mitamura
Built a generic question re-ranking module using Learning-To-Rank (LETOR) approach that can serve as last component of a question generation system. Experimented with length based, linguistic and language model based features. Used open source software SVM$^{\text{rank}}$ for training the system.

**Question Generation and Answering System**   Jan 2016 – April 2016
*Carnegie Mellon University, Pittsburgh, USA*
Worked in a team of 4 to build a system that can produce coherent, grammatical and fluent questions and answers. Used Wikipedia articles as dataset. The system was written in python using NLTK and Stanford CoreNLP tools and provides a command-line interface. Our team won the best project award.

**Text based Information Retrieval**   Sept 2015 – Dec 2015
*Carnegie Mellon University, Pittsburgh, USA*
Built Ranked Boolean, Okapi BM25 and Indri retrieval models in Java using Lucene search engine library. Implemented several query operators and added query expansion and learning-to-rank capabilities. Conducted experiments to analyze the performance of the search engine on ClueWeb09 dataset using metrics like Mean Average Precision and Precision@N.

**Analysis of fMRI data using Machine Learning**   Nov 2015 – Dec 2015
*Carnegie Mellon University, Pittsburgh, USA*
Performed multi-class kernelized SVM classification to categorize the activity being performed by test subjects using their fMRI scans. Used PCA with ridge regression to predict missing voxels in fMRI scans of test subjects. Used holdout testing and cross-validation to test the model accuracy and prevent any over-fitting. Implementation was done in python using scikit-learn library.

**Predicting star ratings from customer reviews**   Nov 2015
*Carnegie Mellon University, Pittsburgh, USA*
Built a machine learning system from scratch to preprocess data, extract features (bag of words), and learn a multi-class logistic regression classifier. Implemented stochastic gradient descent (SGD) and batched SGD to train the logistic regression model. Implementation was done in python and experiments were run on a dataset from Yelp with 1 million reviews.

**Movie Recommender System** <span style="float:right">Oct 2015</span>
*Carnegie Mellon University, Pittsburgh, USA*
Developed, in MATLAB, a collaborative filtering based recommender system for movies using a subset of Netflix Prize dataset. Implemented both memory-based and model-based approaches for rating prediction. Tried clustering techniques to address cold-start problem.

**Fuzzy Transform and its Applications** <span style="float:right">May 2012 – May 2013</span>
*Indian School of Mines, Dhanbad, India*
Worked on Fuzzy Transform for discrete and continuous functions in one and two variables. Implemented in MATLAB, an approach to solve ordinary differential equations using fuzzy transform. Investigated the approach by comparing it to the exact and Euler's solutions. Also looked into image compression and reconstruction using fuzzy transform.

SELECTED
COURSEWORK

*Graduate:* Machine Learning, Search Engines, Machine Learning for Text Mining, Natural Language Processing, Language and Statistics

*Undergraduate:* Computer Organization, Data Structures, Algorithms Design & Analysis, Operating Systems, Software Engineering, Theory of Computation, DBMS, Probability and Statistics, Statistical Inference, Optimization Techniques, Mathematical Modeling & Calculus of Variations, Stochastic Process, Abstract Algebra, Linear Algebra, Numerical Methods, Fuzzy Set Theory and Applications

HONOURS AND
AWARDS

Awarded the **Best Student Shield** by **The President of India** at the convocation in May 2014

Selected twice for the University of Auckland International Summer Scholarship in 2011 and 2012

Topper of the college for the academic year 2009-2010

Recipient of the INSPIRE scholarship from the Department of Science and Technology, Govt. of India from 2008-2012

Placed in the National top 1.5% in the IIT-JEE 2008

EXTRA
CURRICULAR
ACTIVITIES

Won *Best Stock Price Data Hack* (Sponsored by FINRA) at PennApps XII, Sept. 2015

Placement Representative, Integrated M.Tech Mathematics & Computing, Batch 2k13

Hostel Prefect, *Amber Hostel* (3rd yr. hostel) for the academic session 2010-2011

Selected among top 150 Indian students to attend Level-0 session of Mathematical Training and Talent Search (MTTS), at Sambalpur University, Orissa from June 2009 - July 2009
(MTTS is a summer school organized by the National Board of Higher Mathematics (NBHM))

*Hobbies*: Reading