

Siddha Ganju

+1 (412) 789 6469 • sganju1@andrew.cmu.edu • <https://www.linkedin.com/in/sidgan> • <http://sidgan.me>

EDUCATION

2016	CARNEGIE MELLON UNIVERSITY, School of Computer Science Master of Computational Data Science
2015	NATIONAL INSTITUTE OF TECHNOLOGY, HAMIRPUR, (NITH) INDIA Bachelor of Technology in Computer Science & Engineering CGPA: 8.08/10, RANK: 1/103 Awarded Ambuja Cement Scholarship
2011	DAV PUBLIC SCHOOL, NEW SHIMLA, INDIA Scored 91.8%, Class XII, Central Board of Secondary Education, Delhi, India Rank: 1st in school, Top 1% in Himachal Pradesh state Awarded Amar Ujala Pratibha Samman for First rank

INTERESTS

Data Science, Deep Learning, Machine Learning, Natural Language Processing, Information Retrieval, Artificial Intelligence, Data Visualization

HONORS

2016	Accepted Speaker at Strata+Hadoop World, San Jose, USA Talk title, 'Atom Smashing using Machine Learning at CERN'
2016	Member, Open Leadership Cohort, Mozilla Science Lab
2015	Team 'OpenCosmics' session, Open Research Accelerator at MozFest, London
2015	Grace Hopper Conference Scholar , Texas, USA
2015	Winner, Best Innovative Outreach, CERN WebFest
2015	Firefox Student Ambassador, NITH
2014	Winner, Grace Hopper Conference Hackathon, Bangalore, India
2014	Finalist, New York University International Hackathon, Abu Dhabi
2013	Winner, India Scholarship Award, Institution of Engineering and Technology (IET) <i>Out of 5000 participants</i>
2013	Student Women Representative, Community Volunteers Conference, IET, Sri Lanka
2013	National Finalist, Nokia Do Good Hackathon, India

SKILLS

LANGUAGES	Python, C, Octave, R, C++, JAVA, L ^A T _E X
LIBRARIES	scikit-learn, pandas, NLTK, Weka, Theano, Caffe, PyLearn2, PyML, Mahout, Mallet
FRAMEWORKS	Hadoop, .NET, Spark, Azure, Amazon Web Services

EXTRA CURRICULAR ACTIVITIES

2014	Scored grade 'A' under Martin Stoelen, ShanghAI Lectures. Developed a solution for ' Swiss Robots with Adaptive Morphology '
2014	Attended International Summer School on Machine Learning Algorithms & Data Analytics, Thapar University, Patiala, India
2011-14	Chief English Editor, 'Srijan' Institute magazine, NITH
2011-14	Convener, 'Pixonoids' Graphics & Web team, NITH
2012-14	Design Head, Society for Promotion of Indian Classical Music, NITH
2011-14	Executive Member, Computer Science Engineering Community, NITH
2011-14	Student Secretary, Girls Hostel, NITH
2011	Volunteer, Literacy Mission, for educating economically challenged children

PROJECTS

- 2015 | **Collaborative Filtering with Word2Vec to gain sentiments from movie reviews to add emotional dimensions to movie recommendations. Mentor: Prof. William Cohen, CMU**
Use Word2Vec to analyze movie reviews and gain insight into the emotions that led to a movie being rated as 5 or 1. Helps to answer questions as 'What's a good romantic movie?'
- 2015 | **Evaluation of Apache Spark as Analytics framework for CERN's Big Data Analytics infrastructure, CERN Openlab Intern, Mentor: Valentin Kuznetsov, CMS**
Used Apache Spark to streamline different predictive prototypes by gathering information from CMS, ran predictive models and proposed datasets which will become popular over time. Evaluated quality of individual models, performed component analysis and selected best predictive model for new set of data.
- 2014 | **Automated Pipeline for Machine Learning Problems
Mentor: Anirudh Koul, Data Scientist, Microsoft**
*Created a Python command line toolkit using scikit, numpy, pandas and matplotlib libraries to solve machine learning problems automatically. Using imputation and hyper parameteric optimization placed my model among the top 10% of the Titanic kaggle.com challenge (Rank 198/2035 in July 2014). Experimented with large data sets and deployed on Hadoop cluster over AWS. Presented at **Grace Hopper 2015***
- 2014 | **Audio Recognition using Deep Learning**
Recognizes sounds in the vicinity of a smartphone using convolutional neural networks (CNN) in Theano and displays it textually on its screen. Built to help the hearing impaired.
- 2014 | **Twitter Text Normalizer**
Normalizing noisy tweets to their intended form by correcting abbreviations and misspelled words. Best n-gram for a given context was identified based on precomputed clustering of n-grams around their neighborhood. Clustering performed on Hadoop cluster over Amazon Web Services.
- 2014 | **Voice activated Factoid Question Answering System**
Question answering pipeline that accepts fact based questions by voice and replies back by retrieving answer from unstructured Wikipedia text corpus.
- 2014 | **Local Search Indexer**
Voice activated Factoid Question Answering System Built a smart crawler & search indexer for information retrieval using Lucene. Identifies duplicate content, avoids crawling loops and spam sites.
- 2014 | **Who Said it**
Predicts the most likely speaker given an input text, with language models trained using transcripts from politicians during elections. Uses Python's Natural Language Toolkit, libsvm, scikit-learn.
- 2014 | **Kaam Hai. Grace Hopper Conference Hackathon**
Web app connecting low skilled workers with potential employers.