

Siddha Ganju

☎ (412) 789 6469 • ✉ sganju1@cs.cmu.edu • 🔗 <https://www.linkedin.com/in/sidgan>

Education

CARNEGIE MELLON UNIVERSITY (CMU), SCHOOL OF COMPUTER SCIENCE

Master of Computational Data Science

Major: Analytics

Pittsburgh, PA

December 2016 (Expected)

NATIONAL INSTITUTE OF TECHNOLOGY (NITH)

Bachelor of Technology, GPA – 8.08 on 10

Major: Computer Science and Engineering

Hamirpur, India

May 2015

Areas of Interest

Multi-Modal Deep Learning, Machine Learning, Natural Language Processing, Computer Vision, Data Science

Experience

WHAT'S IN A QUESTION: USING VISUAL QUESTIONS AS A FORM OF SUPERVISION

CMU

Research Assistant, Mentors: Olga Russakovsky, Abhinav Gupta

May 2016 - Present

Research focused on an unexplored theme in weak supervision: utilizing visual questions asked about images. We base our work on the key observation that the question itself provides useful information about the image (even without the answer being available). For instance, the question 'what is the breed of the dog?' informs the computer that the animal in the scene is a dog and that there is only one dog present.

EVALUATION OF APACHE SPARK AS AN ANALYTICS FRAMEWORK FOR CERN'S BIG DATA ANALYTICS

CERN

Openlab Intern, Mentors: Valentin Kuznetsov, Tony Wildish, Manuel Martin Marquez, Antonio Romero Marin

June 2015 - August 2015

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. Presented at [Strata+Hadoop World](#). [O'Reilly Data Newsletter I, II](#).

AUTOMATED PIPELINE FOR MACHINE LEARNING PROBLEMS

NITH

Summer Internship, Mentor: Anirudh Koul, Data Scientist, Microsoft

May 2014-August 2014

Created a Python command line toolkit using scikit, numpy, pandas and matplotlib libraries to solve machine learning problems automatically. 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](#).

Publication

Ganju, Siddha *et al.* (2015). Evaluation of Apache Spark as an Analytics framework for CERN's Big Data Analytics.

DOI: [10.5281/zenodo.3186](https://doi.org/10.5281/zenodo.3186)

Skills

Languages

Lua, Python, JAVA, C, R, Octave, C++, L^AT_EX

Libraries

Torch, TensorFlow, Caffe, Theano, scikit-learn, pandas, NLTK, Weka

Frameworks

Spark, Hadoop, Amazon Web Services, .NET, Azure

Select Projects

OPEN ADVANCEMENT OF QUESTION ANSWERING CONSORTIUM

CMU

Capstone Project, Mentors: Eric Nyberg, Matthias Grabmair

Jan 2016-Present

Pursuing active research on QA systems based on an ensemble of Deep Learning and Rule-based systems.

WHAT'S IN THE FUTURE: GENERATING VIDEOS WITH MOTION SENSITIVE ADVERSARIAL NETWORKS

CMU

Course Project, Deep Learning

August-December 2016

Used optical flow to generate future frames using our released FlowGAN architecture. [Website+Github](#).

Achievements

2016 Panelist, [IBM+Apache Spark Maker Community Event](#), San Francisco, USA

2015 [Grace Hopper Conference Scholar](#), Texas, USA

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) *Out of 5000 participants*, Delhi, India

2013 Student Women Representative, Community Volunteers Conference, IET, Sri Lanka