Arjun Datt Sharma

+1 (919) 916-8864 | asharm24@ncsu.edu | www.linkedin.com/in/arjundattsharma | https://github.com/arjundatt

EDUCATION

North Carolina State University
Raleigh, US
MS in Computer Science (Data Science)
Graduation: May 2017
GPA: 3.89/4.0

PEC University of Technology
Chandigarh, India
B.E. in Computer Science
2009-2013
GPA: 8.14/10.0

COURSEWORK

- Algorithms for Data Guided Business Intelligence
- Automated Learning and Data Analysis
- Visual Interface for Mobile Devices
- Foundations of Data Science
- Data Intensive Computing
- Design and Analysis of Algorithms

PROGRAMMING

Proficient: JAVA, Python, Android
Familiar: R, C, C++, JavaScript, HTML5, SQL, RegExp

TECHNOLOGIES

Spark | HBase | Hadoop | Kafka | Linux

AWARDS

- **Best Employee Award** for significant work done at Samsung Research Institute - Noida, India. (*July, 2014*)
- **Best Capstone Project** during undergraduate study from PEC University of Technology, India. (*August, 2013*)

PROJECTS | RESEARCH | PUBLICATIONS

- **Sentiment Analysis** *(Jan 2016)* of Twitter/IMDB data using classical NLP and Doc2Vec techniques for feature vector generation, and Naïve Bayes and Logistic Regression for classification of data. Implemented using Python.
  - **Git link**: https://goo.gl/diYQFt
- **Recommender system** *(Ongoing)*: Development of a music recommender system using implicit feedback of users as training data. Used **Apache Spark's MLlib library** with Python for implementation.
- **One query-multiple databases (OQMD)** *(Sept 2015 – Dec 2015)*: Developed (using JAVA) an interface to mutually access data from heterogeneous databases (like HBase and Oracle), with a single SQL-like query. The application is designed to automate the process of schema unification to map attributes in different databases by recognizing character patterns.
  - **Git link**: https://goo.gl/dFByzF
- **User-Identity Management on Online Social Networks** *(August 2012 - May 2013)*: Constructed a mathematical model to identify and tag suspicious user on online social network (Twitter), by mining meta-data of Twitter users. Used **R** for data mining and visualization.
  - **Publication:** *Friend or Foe: Twitter Users under Magnification*
- **Web based Text Summarizer** [**Independent project**] *(August 2012 - December 2012)*: Developed a web application (using JavaScript) for generating a summary for any given web article using font semantics and other classical methods like *tf-idf* and *POS* tagging.
  - **Publication**: *Too Long-Didn't Read: A Practical Web Based Approach towards Text Summarization*
  - **Link to web app**: http://goo.gl/ICz9K0
- **HeatMap** [**Voluntary work** for Dr. Asmi H. Shah at KIT, Germany]: Developed a web application (using JavaScript) to generate user interactive heat maps and bar plots in order to visualize life science experimental data.
  - **Git link**: https://goo.gl/HxMiO1

WORK EXPERIENCE

Samsung Research Institute, India | Software Developer   *June '13 – Jul '15*
Project **KNOX** – a security framework for android devices for use in enterprises.

- Development of *Bridge* module which is responsible for interaction and sync of data between Knox and personal mode users.
- Enhancement of android *services* framework.
- Development of UI for Knox shortcuts in KNOX (version 2.0) for Samsung Galaxy S5 at **Samsung R&D Center, South Korea (HQ).** *Feb '14–Mar '14*

Karlsruhe Institute of Technology, Germany | **Intern**   *Feb '12 – Jul '12*

- Received **fellowship** from KIT, Germany
- Developed a web application which tags and augments (provides more inline information about) domain specific content in web articles/journals. The project involved application of *tf-idf* techniques for domain specific information retrieval of web articles. The application is currently deployed under ScienceNet - a search engine internal to the KIT scientist community.