

MD IMRUL KAYES

2367 Nautical Way, 111, Winter Park, FL 32792
imrul@mail.usf.edu, (813) 405 6320
<https://imrulkayes.github.io/>

SUMMARY

- 6+ years of combined experience in software development/quality and data science (& engineering): worked for Yahoo, DSG Labs, SoftwarePeople, Delta Life, and Binary Solutions.
- Extensive experience in data science and Engineering from academia and industry: mined large, complex networks (Big Data), such as online social, community Q/A and blogging using data mining techniques; studied user behavior based on psychology and sociology theories; configured and deployed big data infrastructures, built ETL data pipeline, devised and implemented prediction algorithms.
- PhD and MS degrees in Computer Science & Engineering.
- Big Data summer school scholarships and grants from Google, ACM and SoftwarePeople.
- Published academic papers (focusing on Big data analytics, Large-scale systems, Applied machine learning) on top journals (e.g., IEEE transactions) and conferences (e.g., WWW)

EDUCATION

PhD, Computer Science & Engineering	University of South Florida	2015
• Dissertation: Content abuse and privacy concerns in online social networks.		
MS, Computer Science (GPA: 3.85)	University of South Florida	2014
BSc, Computer Sci. & Engg. (GPA: 3.63)	Bangladesh Uni. of Engg. and Tech.	2009

WORK EXPERIENCE

Sonobi Inc., Winter Park	Data & Infrastructure Engineer	March'16 - Present
<ul style="list-style-type: none">• Built predictionViz (Python, MySQL, R, Shiny), an interactive forecast evaluator which analyzes longitudinal prediction results on web publishers? viewable ad spots.• Created and deployed high availability HDFS, Hadoop, Yarn and Spark clusters using Ansible, Salt and Amazon AWS instances.• Contributed to jetstream agency suite product: 1) real-time inventory forecasting using Python, Pandas and Statsmodels 2) devised and implemented algorithms for targeting audience based on users? browsing behavior and spatial distribution of browsing 3) implemented metrics that enable online advertisers to target audiences from high quality publishers.• Developed python ETL processes to consolidate data from Amazon Redshift to MySQL database.• Built ETL data pipeline for forecasting using Amazon S3, HDFS, Spark, Druid and Airflow.		
Yahoo Labs, Barcelona	Research Intern	September'13 - December'13
<ul style="list-style-type: none">• Built a data pipeline to transform raw data of Yahoo Answers for analysis using Hadoop, Pig Latin, Python, and R.• Characterized content abusers in the Yahoo Answers dataset (more than one year recorded activity logs of 1.5 million users), created exploratory graphs, and conducted statistical tests.• Built machine learning models and detected content abusers in Yahoo Answers with 83% accuracy.		
Distributed Systems Lab, Tampa, FL	Research Assistant	May'12 - December'15
<ul style="list-style-type: none">• Designed and Implemented Java and Jena based 'Aegis' privacy framework for social ecosystems.• Analyzed user retention in social blogs and predicted top retained users with 93.62% accuracy using machine learning.• Proposed a network centrality aggregation method to identify influential bloggers and detected influential bloggers.• Analyzed privacy concerns and user behavior in Q/A sites and built machine learning models to predict privacy.• Showed that cross-cultural variations exist in community Q/A sites analyzing user activity logs.		
SoftwarePeople, Copenhagen	Software Engineer	March'10 - July'11
<ul style="list-style-type: none">• Worked as a Scrum team member for the development of Agile-based software and used C#, ASP.NET, MS SQL server to build web-based business applications and service oriented systems.• Researched on software quality and testing; wrote research papers; awarded a travel grant to present the paper.		

Delta Life, Dhaka**Software Engineer****September'09 - December'09**

- Developed “Xfer” using C# and Oracle, which synchronizes data transfer from remote servers to the central server.
- Worked on the development of the official website of Delta Life Insurance Co. Limited.

Binary Solutions, Dhaka**Software Developer****November'08 - August'09**

- Developed a stock market software “EasyShare” using C# and MySQL, which collects data from the Dhaka Stock Exchange server and shows comparative market situation.

TECHNICAL SKILLS

Languages & Frameworks	Python, C, C++, Java, C#, ASP.NET, HTML, JavaScript, CSS
Machine Learning & Data Mining	Time-series , Regression, Classification, Weka, R, Sklearn, Statsmodels
Database	MongoDB, Oracle, MySQL, PostgreSQL
Distributed Data Processing	Hadoop, Pig Latin, Spark, Druid, Airflow
Data Analysis and Visualization	R, Shiny, NetworkX, JUNG, SNAP, Gephi
Deployment and Configuration Management	Ansible, SaltStack
Applications	OmniGraffle, LATEX, Gnuplot, Git

HONORS AND AWARDS

- ACM travel grant to the 26th conference on Hypertext and Social Media, Cyprus, 2015.
- San Diego Supercomputer center scholarship to the Big Data summer school, San Diego, 2014.
- Google travel grant to the 4th SocInfo conference, Switzerland, 2012.
- SoftwarePeople travel grant to the IEEE ICECT conference, India, 2011.

SELECTED PEER-REVIEWED PUBLICATIONS

* Please visit my Google Scholar profile (<https://scholar.google.com/citations?user=ApVhRzEAAAAJ&hl=en>) for a full list of publications.

- [1] **I. Kayes**, N. Kourtellis, D. Quercia, A. Iamnitchi, F. Bonchi, The Social World of Content Abusers in Community Question Answering. *ACM 24th World Wide Web conference (WWW'15), Italy, 2015*. (Acceptance rate: 14.10%)
- [2] **I. Kayes**, J. Chakareski, Retention in Online Blogging: A Case Study of the Blogster Community. *IEEE Transactions on Computational Social Systems (2015)*.
- [3] **I. Kayes**, N. Kourtellis, F. Bonchi, A. Iamnitchi, Privacy Concerns vs. User Behavior in Community Question Answering. *IEEE/ACM Conf. on Advances in Social Networks Analysis and Mining (ASONAM'15), France, 2015*.
- [4] **I. Kayes**, N. Kourtellis, D. Quercia, A. Iamnitchi, F. Bonchi, Cultures in Community Question Answering. *ACM 26th Conference on Hypertext and Social Media (HT'15), Cyprus, 2015*.
- [5] **I. Kayes**, S. Islam, J. Chakareski, The Network of Faults: A Complex Network Approach to Prioritize Test cases for Regression Testing. *Springer Innovations in Systems and Software Engineering Journal (2015)*.
- [6] **I. Kayes**, X. Zuo, D. Wang, J. Chakareski , To Blog or Not to Blog: Characterizing and Predicting Retention in Community Blogs, *ACM 7th Social Computing Conference (SocialCom'14), China, 2014*. (Acceptance rate: 16.6%)
- [7] **I. Kayes**, A. Iamnitchi, Aegis: A Semantic Implementation of Privacy as Contextual Integrity in Social Ecosystems, *IEEE 11th Conference on Privacy, Security and Trust, Spain, 2013*. (Acceptance rate: 28.76%)
- [8] **I. Kayes**, X. Qian, J. Skvoretz, A. Iamnitchi, How Influential are You: Detecting Influential Bloggers in a Blogging Community, *4th International Conference on Social Informatics, Switzerland, 2012*. (Acceptance rate: 25.9%)