

# MD IMRUL KAYES

2367 Nautical Way, 111, Winter Park, FL 32792  
imrulkayes11@gmail.com, (813) 405 6320  
<https://imrulkayes.github.io/>

## SUMMARY

---

- Experienced software engineer with academic and industry research and development experience in distributed systems, big data, machine learning/AI; worked for Sonobi, Yahoo, DSG Labs, SoftwarePeople and Binary Solutions.
- Ph.D. and M.S. degrees in Computer Science & Engineering focusing distributed systems, machine learning, online social networks (privacy, retention and abuse).
- Developed software for diverse domains e.g., Insurance, Banking, Stock market, and online advertising.
- Mined large, complex networks, such as online social, community Q/A and blogging using AI/ML techniques; configured and deployed big data infrastructures, built ETL data pipeline, implemented and deployed AI/ML algorithms.
- Big Data summer school scholarships and grants from Google, ACM, and SoftwarePeople.
- Published academic papers (focusing on big data analytics, large-scale systems, machine learning) on top journals (e.g., IEEE transactions) and conferences (e.g., WWW)

## EDUCATION

---

<b>Ph.D.</b> , Computer Science & Engineering	<b>University of South Florida</b>	<b>2015</b>
• <b>Dissertation:</b> Content abuse and privacy concerns in online social networks.		
<b>M.S.</b> , Computer Science (GPA: 3.85)	<b>University of South Florida</b>	<b>2014</b>
<b>B.Sc.</b> , Computer Sci. & Engg. (GPA: 3.63)	<b>Bangladesh Uni. of Engg. and Tech.</b>	<b>2009</b>

## WORK EXPERIENCE

---

**Sonobi Inc., Winter Park, USA      Software Engineer: Data & Reporting      May'17 - Present**

- Developed systems and machine learning services for audience prediction and ad targeting; each handles billions of daily online auctions (Python, Spark, HDFS, S3, Druid, Airflow, Scikit-learn, StatsModels, Memcached, MySQL).
- Developed and productized frequency capping machine learning algorithms and services for media plans.
- Built pipelines for ML APIs deployment using Jenkins, Docker, and Git; wrote integration tests for APIs.
- Created dashboard and alerting to monitor applications and critical metrics using InfluxDB, Grafana and Slack.
- Maintain, extend, and optimize forecast reporting services and infrastructure.
- Developed ETL data pipelines to shepherd terabytes of daily data via several warehouses and data centers for reporting and forecasting services.
- Introduced statistical sampling in machine learning data pipelines and reduced data processing by up to 90%.
- Performed A/B testing on key features of real-time bidding involving billions of auctions.
- Optimized ML pipelines and APIs which improved service performance by 80%.
- Analyzed large-scale real-time bidding data and devised strategies for revenue, bidding systems, and ad campaigns.

**Sonobi Inc., Winter Park, USA      Software Engineer: Data & Infrastructure      March'16 - April'17**

- Built predictionViz (Python, MySQL, R, Shiny), an interactive forecast evaluator which analyzes longitudinal prediction results on web publishers' viewable ad spots.
- Built 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 3) implemented features that enabled online advertisers targeting 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, Spain      Research Intern      September'13 - December'13**

- 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, USA**      **Research Assistant**      **May'12 - December'15**

- 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, Denmark**      **Software Engineer**      **March'10 - July'11**

- 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.

**Binary Solutions, Dhaka**      **Software Developer**      **November'08 - February'10**

- Developed "Xfer" using C# and Oracle, which synchronizes data transfer from remote servers to the central server.
- 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	Python, Java, Scala, Shell, C, C++, C#, R, HTML, JavaScript, CSS
Machine Learning/AI	Network analysis and mining, Time-series, Regression, Classification, R Scikit-learn, Statsmodels, Deep learning (Keras), NLP
Database and Cache	SQL, Druid, PostgreSQL, MongoDB, Oracle, Memcached, Redis
Distributed Data Processing	HDFS, Hadoop, Spark, Pig Latin, Yarn, Airflow, S3 Amazon Web Services (AWS), Grafana, InfluxDB
Data Analysis and Visualization	Notebook, Pandas, Bokeh, R, Shiny, NetworkX, Matplotlib, Seaborn
Deployment and Configuration Management	Ansible, SaltStack, Jenkins, Docker
Applications	LATEX, 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

---

\* 20+ peer reviewed journal and conference papers. 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, D. Quercia, A. Iamnitchi, F. Bonchi, Cultures in Community Question Answering. *ACM 26th Conference on Hypertext and Social Media (HT'15), Cyprus, 2015*.
- [4] **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)*.
- [5] **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%)

<sup>1</sup>Work authorization: US permanent resident