

Indraneil Bardhan

indraneil.bardhan@tuta.io • 412-961-5794

www.linkedin.com/in/indraneilbardhan • www.github.com/NeilBardhan • neilbardhan.github.io

PROFILE Information Science Graduate with well-developed skills in Big Data Analytics, Data Science and Software Development. Three plus years of experience in Data Science, with special emphasis on data exploration, preparation, transformation, visualization and extraction of actionable insights to deliver timely solutions to challenging problems using statistical methods, machine learning and predictive modeling. Motivated, hard-working individual with excellent oral and written communication skills. Strong team work ethic, presentation skills, and work proficiency.

EDUCATION AND TRAINING

University of Pittsburgh, Pittsburgh, PA

December 2016

- Master of Science in Information Science (Specialization in Big Data Analytics)
- GPA: 3.458

University of Pune, Pune, India

June 2015

- Bachelor of Engineering in Computer Engineering
 - GPA: First Class
-

SKILLS

Proficient in the following programming languages:

- Python
- R
- SQL
- Java
- C/C++
- Scala

Clear understanding of and extensive experience the following frameworks:

- Apache Hadoop
 - Apache Spark
 - D3js
 - Tableau
-

PROFESSIONAL EXPERIENCE

Research Assistant, University of Pittsburgh

October - December 2016

- Designing a web based application to demonstrate how major corporations are perceived on Social Media.
- Optimized back-end processing time by migrating Bash scripts to Python scripts for the web application.

Research Associate, University of Pittsburgh

September - December 2016

- Responsible for data curation from various government sources like NHANES and BRFFS.
- Performed recoding, cleaning, quality checks and statistical analysis operations on the data.
- Created codebooks and data dictionaries for teammates. Developed charts to visualize the information.

Statistical Programming and Analysis Intern, Genentech Inc.

June - August 2016

- Designed and Implemented an experimental Hadoop Ecosystem using Apache Spark for the Statistical Programming and Analysis Department.
 - Migrated legacy clinical trial data from the existing hadoop infrastructure to the experimental hadoop ecosystem.
 - Designed an Apache Solr search template and developed a schema to perform keyword searches on the legacy data files.
 - Analyzed clinical data using conventional methods as well as using SparkR on the experimental Hadoop ecosystem and documented differences in the results obtained.
 - Designed and Applied predictive models to predict adverse events using PySpark and SparkR in Databricks.
-