Shweta Bharti

Ph: (720) 693-6437

10 Night Heron Drive,
Stony Brook, NY – 11790

sbharti@cs.stonybrook.edu
https://www.linkedin.com/in/bhartishweta
https://github.com/shwetabharti

Education

Stony Brook University, New York	**Aug 2017 - May 2019**

- MS in Computer Science (GPA: 3.6/4)

Delhi College of Engineering, University of Delhi, **Delhi**	**Aug 2009 - Jun 2013**

- BE in Electronics & Communication Engineering (Aggregate: 74.24/100)

Relevant Coursework

Big Data Analytics, Data Science Fundamentals, Artificial Intelligence, Data Mining, Algorithms, Database Systems, Programming Languages, Systems Fundamentals.

Work Experience

Regeneron Pharmaceuticals, Big Data Analytics Intern, NY, USA   **3 mos. (May 2018 – Aug 2018)**

- Analyzed public GWAS catalog and compared it with RGC Results Browser associations catalog to find out potential gene mutations for target discovery using **PySpark**.
- Developed ETL pipeline to automate the process of transforming raw data from GWAS catalog and different collaborators into RGC defined schemas, to combine it to the existing data of 90K patients.
- Developed python notebooks (using **Databricks**) to automate the ancestry extraction process of people sequenced for a given gene mutation in public GWAS catalog, using Spacy, to facilitate continent specific target discovery.

Centre for Development of Telematics, Software Developer, Delhi, India   **4 yrs. (Aug 2013 –Aug 2017)**

- Developed a web application using Django, to analyze CDR (Call Detail Records), using **Pyspark**, for more than 2.5 million subscribers of MTNL/BSNL. The analysis was used by the MTNL/BSNL for marketing and promotional purposes.
- Conducted training session about **Hadoop** and **Spark** for the new recruits in the organization. Planned, organized and developed training materials that met specific needs of different teams for new hires.
- Developed and managed different patch versions to upgrade the CMS (Centralized Monitoring System) application and resolve various software issues.
- Performed validation and verification of the CMS application, created their test cases for acceptance testing and demonstrated to the clients.
- Automated some parts of Disaster Recovery process for the CMS application running on Linux OS which included hostname related changes in the DNS and NetApp snapmirror technology for the storage of data.

Centre for Fire, Explosive and Environment Safety, Software Developer Intern   **3 mos. (Jun 2012 – Aug 2012)**

- Implemented k-means clustering algorithm in Python on a dataset of over 5000 images, to classify the skin burns as first, second and third-degree burns.

GrailTest Technologies Pvt. Ltd., Software Development Intern   **3 mos. (Nov 2011-Jan 2012)**

- Developed automated test scripts in python to setup test environment, fetch test data from database and generate the reports as per business requirements & ensure delivery of quality software applications.

Languages and Technologies

C, C++, Python, Java, Scala, Spark, SQL, Javascript, Django, HTML, CSS, Hadoop, MapReduce, D3.js, Tableau, Matplotlib, NLTK, Pandas, NumPy, SciPy, Scikit-Learn, TensorFlow, MATLAB, AWS, Databricks, Linux, Mac OS.

Projects

Analysis of Cyberbullying against Women (Keras, Word2Vec)

Built a sentiment analyzer using Keras Library and Word2Vec, to classify a tweet as violent and non-violent against women and to further classify the severity of the violent tweet on a scale of 10 under the supervision of Prof H. Andrew Schwartz.

New York City Taxi Fare Prediction (Kaggle, Jupyter, Python, Scikit-Learn, Tableau)
Predicted the fare amount for a taxi ride in New York City given the pickup & dropoff longitudinal and latitudinal location and time stamp of the start of the ride. The best prediction model was based on Random Forest approach.

Web Application for Named Entity Recognizer (Django, Python, NLTK)
Developed a simple web application for information extraction using NLTK that identifies the named entities in the text submitted by the user and displays the result.

Women's Safety in Indian Railways (PySpark, Databricks)
Determined the need for female security on railway platforms based on the data provided by Indian Railways with details of departure and arrival timings, origin and destination cities and crime rate in those cities, using Spark.

English Vocabulary Android Application (Android Studio, Java)
Developed a basic android application which provides the mnemonics and relevant images to learn new vocabulary and prepare for GRE vocabulary section, using Android Studio & Java at HackInTheNorth, IIIT Allahabad, India, 2015.

Hand-gesture based mouse control (MATLAB, kNN)
Three gestures were stored corresponding to left click, right click and double click. New gesture was captured through webcam and compared with those three stored gestures using kNN algorithm to find the best match to proceed with the action.

Extra-curricular & Achievements

- *State Topper (Rank-1)* of Delhi NCR Region with 97.4% aggregate marks in tenth grade CBSE Board examination (2007).
- Secured *Rank-4 at the National level* in CBSE Board high school examination (2007).
- Felicitated with *Indira Award* in Mathematics & Sciences for being among top 0.1% students who wrote the CBSE Board Examination (2009).
- Successfully organized three consecutive *Séance* - Annual Day Event at Centre for Development of Telematics as Treasurer to the Staff Recreation Club (2013-2017).