

Saurav Sengupta

ss4yd@virginia.edu | [LinkedIn](#) | ssen7.github.io | +1 (434)-284-3794

EDUCATION

University of Virginia

Masters of Science in Data Science, 3.96/4

July 2018-May 2019

Birla Institute of Technology and Science, Pilani

Bachelors in Electrical and Electronics Engineering, 7.2/10

August 2012-May 2016

PROFESSIONAL EXPERIENCE

Capgemini (Cloud Foundation Services)

Associate Consultant

August 2016 – October 2017

- Responsible for developing REST based APIs using Spring Boot.
- Used cloud platform Pivotal Cloud Foundry® (VMWare) to deploy code to production.
- Worked in an Agile work environment and interacted daily with on-shore clients.
- Worked on making CI/CD pipelines and setting up test environments.
- Wrote Elasticsearch based micro-services.

STMicroelectronics

Trainee

July 2015 - December 2015

- Interned on the team designing an embedded neural network based ASIC.
- Wrote code to interface Random Access Memory with peripherals like I/O and buses.
- Worked in Unix environment and wrote Bash shell and Tcl/Tk scripts for code synthesis and design verification.

TECHNICAL SKILLS

- **Languages:** R, Python, Java 7 (Oracle Certified Associate), C, Verilog
- **Database:** MySQL, MariaDB
- **Packages/Tools:** pandas, numpy, sklearn, caret(R), RShiny, TensorFlow, PyTorch, GitHub, Jenkins, AWS EC2, Apache Spark and EMR (Spark Cluster), Pivotal Cloud Foundry®
- **Operating Systems:** Windows, Linux

PROJECTS

Building CNNs to classify duodenal biopsy images into diseases – Child Health Research Center, UVA

- Used Convolutional Neural Networks (**CNN**) to classify high resolution digitized biopsy images into Celiac Disease, Environmental Enteropathy and Normal tissues.
- Used pretrained **Resnet50** in our analysis, performs well on diverse set of images.
- Achieved close to **98%** biopsy level accuracy. Paper accepted in **IEEE BHI 2019 conference**.
- Backed by the **Bill and Melinda Gates Foundation** and **Aga Khan University**.
- We are also using metabolomics data to find important features of classification.

Music Genre Classification

- Classification of songs into one of 13 genres like rock/pop and country etc. using **ensemble models** for song features and lyrical data.
- Song data features available from [AWS Public Dataset](#). Used genre data from different Kaggle datasets.
- Created **AWS EC2 instance**, mounted snapshot, ran **Jupyter notebooks on EC2** to get data.
- Used Random Forests, SVM, Naïve Bayes for model using song features. Used tf-idf and topic modeling for lyrical data.
- Ensembled both these models. Around 58.06% accuracy, close to the [state of art](#) of 61% but we classify more genres.

Modeling Brain Wave Activity using Muse™ headset

- Used Logistic Regression for real time classification of brain wave signals like alpha waves into left-right, up-down motion. Tested model by moving objects in a game environment.

VOLUNTEER EXPERIENCE

- Deputy Event Manager - Joy of Giving Week 2013 BITS Pilani, Goa Campus for NIRMAAN Organization. Raised around Rs 17,000 for charity.
- Participated in projects to generate employment and self-help services for women from impoverished neighborhoods in Zuarinagar, Goa as part of Nirmaan Organization.