Saurav Sengupta

ss4yd@virginia.edu | GitHub| LinkedIn | Website
+1 (434)-284-3794

EDUCATION

M.S. Data Science – University of Virginia – Charlottesville, VA **July2018-present**
GPA:3.96/4.0
Current Coursework: Linear Modeling and Regression, Data Mining, Deep Learning, Reinforcement Learning, Data Visualization.

B.E. Electrical Engineering – BITS Pilani – India **August 2012-May 2016**

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

PROFESSIONAL EXPERIENCE

Associate Consultant

Cloud Foundation Services – Capgemini **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.

Trainee

STMicroelectronics **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.

PROJECTS

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

- We use Convolutional Neural Networks (CNN) to classify high resolution digitized biopsy images into Celiac Disease, Environmental Enteropathy and Normal tissues.
- We are using existing architectures like Resnet34 and Resnet50 in our analysis, since they are the deep and perform well on a diverse set of images. We use the fabulous fastai library.
- We are also using metabolomics data to find important features of classification.
- https://github.com/UVA-DSI-2019-Capstones/CHRC

Music Genre classification

- Created EC2 AWS instance to get data from the Million Song Dataset.
- Wrote scripts to get data from S3 bucket and run Jupyter notebooks on EC2.
- Used Text mining on song lyrics and song features to predict 13 different genre types. Used feature selection to reduce the number of features and reduce overfitting. 58.6% Accuracy.
- GitHub link.

Modeling brainwave activity

- Used Muse™ headset to collect brain activity data in the form of frequency measurements of alpha, beta, gamma, delta and theta waves to classify actions like moving hands, motor imagination.
- Found correlation between alpha waves and motion.
- Used Logistic Regression for real time modeling and classification of thought and used it to move object in a mobile game environment.

Modeling quality of barbell lifts using accelerometer data

- The goal was to use data from accelerometers on the belt, forearm, arm, and dumbbell of 6 participants, to predict how well they do barbell lifts.
- Tested accuracy of models built using Tree Algorithm, Random Forest, and Model Stacking using the caret package in R. GitHub Pages Link.

CONFERENCE PAPERS/PUBLICATIONS

- ***Deep Learning for Visual Recognition of Environmental Enteropathy and Celiac Disease***
  A. Shrivastava*, K. Kant*, ***S. Sengupta****, S. Kang*, M. N. Khan, S. Moore, S. A. Ali, B. Amadi, P. Kelly, S. Syed, D. Brown (Paper accepted for oral presentation at IEEE-EMBS INTERNATIONAL CONFERENCE ON BIOMEDICAL AND HEALTH INFORMATICS 2019 Chicago, IL) (* equal contribution)
- ***Solving the Stain Dilemma: Computational Image Analyses to Address Differential Tissue Staining Color Bias in Duodenal Biopsies.***
  S. Syed, A. Shrivastava, K. Kant, L. Kang, ***S. Sengupta***, M. Naveed Khan, N. Talat Iqbal, K. Sadiq, C. A. Moskaluk, B. Amadi, P. Kelly, S. Moore, D. Brown. (Abstract accepted for poster presentation at Digestive Disease Week (DDW), May 20th 2019)
- ***Diseased and Healthy Gastrointestinal Tissue Data Mining requires an Engaged Transdisciplinary team.***
  S. Syed, M. N. Khan, A. Shrivastava, K. Kant, L. Kang, ***S. Sengupta***, K. Kowsari, R. Sali, K. Wang, A. Catalano, S. A. Ali, N. Iqbal, K. Sadiq, M. P. Kelly, B. Amadi, C. Moskaluk, J. Papin, S. Moore, D. E. Brown. (Abstract accepted for poster presentation at Association for Clinical and Translation Science: Translational Science 2019 conference, March 6th 2019)

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.