Venkatesh K

Email: vengatesh.vengatesh89@gmail.com
Portfolio: venkatk89.github.io

Education

- **Indian Institute of Technology, Madras** — Chennai, India
  Bachelor of Technology in Engineering Physics; CGPA: 7.18/10 — *Aug 2015 - May 2019*

Experience

- **Centre for Integrative Biology and Systems Medicine, IIT Madras** — Chennai, India
  Junior Research Fellow — *Jan 2021 - Present*
  Advisors: Dr. Himanshu Sinha, Dr. Karthik Raman, Dr. Manikandan Narayanan
  - **GenomeIndia: Cataloguing the genetic variation in Indians**:
    Constructed a pan-genomic graph that acts as a reference for WGS analysis. Integrated multiple bioinformatics tools and developed a computational pipeline to study the whole genome sequences of 10K healthy individuals from the highly heterogeneous Indian populace.
  - **Polygenic Risk Score Analysis**:
    Computed individual-level genetic risk scores for Type 2 Diabetes using SNP effect sizes derived from genome-wide association studies. Employed machine learning techniques to improve the predictive power of the calculated PRSs.
- **Indus Insights and Analytical Services** — Gurgaon, India
  Data Science Consultant — *Sep 2019 - Feb 2020*
  - **Recommender Systems**: Designed cutting-edge collaborative filtering algorithms to recommend destinations for passengers of the biggest airlines in the USA. Leveraged high-performance computing to train and evaluate the deep learning models on over 100 GB of data.
  - **Marketing Models**: Audited the marketing models of a US-based small business lender. Performed detail-oriented examinations on SAS-based customer response and value models to ensure its robustness and performance.
- **Blitzkrieg Retail Private Limited** — Chennai, India
  Data Science Intern — *May 2018 - Jul 2018*
  - **Feature Enhancement**: Developed the product recommendation engine for an online pharmacy app by mining association rules. Created a proprietary image processing application for the company from scratch, using Python and OpenCV.
  - **Data Engineering**: Worked with a team of research and domain experts to cleanse the product database of the app. Built an end-to-end ETL pipeline on Firebase and MSSQL databases to develop the company's dashboard.

Projects

- **DNA Sonification Tool**: Fashioned an auditory display *tool* in R for genetic sequence analysis that incorporated music theory to convert nucleotide sequences into songs. Enhanced the musicality of the tool's output while maintaining its overall analytical capabilities. (October 2020)
- **Genetic Engineering Attribution**: Developed an algorithm that identifies the most likely lab-of-origin for genetically engineered plasmid samples. Utilized machine learning techniques with features extracted from the sequences to outperform the state of the art model based on BLAST. (September 2020)

- **DengAI - Disease Spread Prediction**: Designed an ensemble method to forecast the spread of dengue, using time series analysis and statistical modeling. Trained the models on Collab GPUs with environmental data collected over a decade in two South American cities. (July 2020)

- **Flu Shot Learning - Healthcare Analytics**: Employed deep learning to predict whether people got H1N1 and seasonal flu vaccines. Trained the model using sentiment and behavior data from the National 2009 H1N1 Flu Survey conducted by CDC. (June 2020)

- **Climate Data Analysis - Department of Chemical Engineering, IITM**: Analyzed the data collected by moving and stationary sensors placed across multiple cities in India and validated key insights through hypothesis testing. Created temporal visualizations and geospatial heat maps for radiation and air quality-related parameters. (August 2019)

- **Traffic Management - Sangam ML Hackathon, 2019**: Built statistical models to predict traffic volume using climate and traffic data collected in the US over four years. Finished as one of the top 10 teams among the 450 teams that participated. (July 2019)

- **Fraudulent Transaction Prediction – Department of Computer Science, IITM**: Designed a random forest classifier to predict fraudulent transactions. Conceptualized an online fraud prediction algorithm and won the first position in the Exebit Data Science Challenge, 2018. (April 2018)

- **User Preference Modelling**: Ideated and created a *website* to work with dynamic test data. Created logistic regression models to predict probabilities of the user's phobias using correlating parameters from a survey conducted in the UK. (February 2018)

Technical Skills

- **Languages**: R, Python, Bash, SQL, C, HTML

- **Frameworks**: Advanced Excel, Tidyverse, scikit-learn, Keras, OpenCV, AWS

Positions of Responsibility

- **Student Coordinator and Chief Organizer** Bhoutics 2017
  Led a team of 15 to organize Bhoutics, the physics outreach festival of IIT Madras.

- **Media and Students Relations Coordinator** Shaastra 2017
  Organized Shaastra Spark, a national level quiz competition for school students.