Venkatesh K

Website: venkatk89.github.io Github github.com/venkatk89 Email: vengatesh.vengatesh89@gmail.com

Research Experience

- **Centre for Integrative Biology and Systems Medicine, IIT Madras** — Chennai, India
  Project Scientist — *Jun 2022 - Present*
  Junior Research Fellow — *Jan 2021 - May 2022*
  Advisors: Dr. Himanshu Sinha, Dr. Karthik Raman, Dr. Manikandan Narayanan

  - **Genome Graphs**: Developed scalable computational pipelines by integrating multiple bioinformatics tools to analyze human WGS with genome graphs. Designed novel methods of annotating the genome graph for identifying functionally significant regions. Performed detailed analyses to study the structural complexities in the genome graphs.
  - **GenomeIndia**: Constructed Indian population-specific genome graphs as reference structures to analyze the WGS of 10K healthy individuals from the country's diverse populace. Using the dynamic nature of the genome graphs, captured the common and low-frequency variants in Indians and studied the consequences of genetic variations within and between sub-populations.
  - **Visualization of variants**: Ideated and developed SCI-VCF, a cross-platform application to summarise and compare the genetic variants from VCF files. The online/offline tool can interactively visualize genetic variants, empowering researchers and clinicians to perform exploratory genomic data analysis on VCF files irrespective of their programming expertise.
  - **Polygenic Risk Scores**: Calculated PRS for type-2 diabetes for individuals in an Indian cohort. Multiple GWA Studies were experimented with as base datasets. Employed machine learning techniques to enhance the predictive power of the calculated risk scores.

Education

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

Publications

- **SCI-VCF: A cross-platform application to summarise, compare and design interactive visualisations of the variant call format**: Venkatesh Kamaraj, Himanshu Sinha. bioRxiv 2023 DOI

Presentations

- **GenomeIndia: Cataloguing the genetic variations in Indians**: Poster presentation at the RBCDSAI AI/ML conclave on healthcare, organized by IIT Madras. (August 2023)
- **A comprehensive analysis of human genomes with genome graphs**: Poster presentation at Network Biology Day organized by IMSc, Chennai. (July 2023)
- **Sequence graph representations of yeast and human genomes**: Talk at the GenomeIndia Data Science workshop organized by the GenomeIndia consortium. (April 2021)

Training/ Workshop

- **Data Science-driven solutions to improve maternal and child health**: Reviewed the development and validation of pregnancy dating models for Indian women while emphasizing the ethical and infrastructural challenges in data collection and clinical translation. (February 2023)
- **Microbiomes in Environment, Space, and Human Health**: Explored the core competencies of microbial omics research and explored the advancements in the field. (November 2022)

- **Clinical Genomics to Systems Medicine: Computational Approaches for Transforming Healthcare**: Focused on the novel techniques, methods, and algorithms to dissect omics data at multiple levels, enabling a deeper understanding of complex human diseases. (February 2022)

- **Data Scientist Career track with R – DataCamp**: Gained expertise in data wrangling, visualization, hypothesis testing, modeling, and reporting along with R programming to analyze and interpret complex data. (September 2018)

Work Experience

- **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 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 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)

- **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)

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

Technical Skills

- **Languages/Platforms**: R, Python, SQL, C, HTML, Shell scripting, HPC, AWS, GCP

- **Tools/Frameworks**: Tidyverse, Shiny, Scikit-learn, Keras, OpenCV, Snakemake, Docker, Git