# VENKATESH K

vengatesh.vengatesh89@gmail.com | Website

## Education

**Indian Institute of Technology, Madras**, India (Aug 2015 - May 2019)

Bachelor of Technology, Engineering Physics

CGPA: 7.18/10.00

## Scholastic Achievements

- Secured **All India Rank of 110** in Graduate Aptitude Test in Engineering (Statistics paper), 2019
- Secured **99.6 percentile** score in Joint Entrance Examination, 2015 taken by over 1.5 million students

## Technical Skills

**Languages:** R, Python, SQL, C++, HTML
**Frameworks:** Advanced Excel, Tidyverse, SparklyR, Keras, Tensorflow, Pytorch, OpenCV, GCP, AWS

## Professional Experience

**Data Science Consultant - Indus Insights and Analytical Services**, Gurgaon (Sep 2019 – Feb 2020)

- Designed cutting edge content-based and collaborative filtering algorithms like GMF, Wide&Deep, NeuMF, DeepFM, etc. to recommend domestic and international destinations for the biggest airlines in the USA
- Leveraged cloud machine learning platforms, docker container repositories, and cloud storage services, to train and evaluate the deep learning models on over 100 GB of data
- Audited Marketing models for a US-based small business lender. Performed detail-oriented examinations on SAS-based customer response and value models to ensure its robustness and performance
- Worked closely with the firm's core development team to enhance the in-house Statistical automation tool. Designed and implemented a new module of the tool to train and validate the Light-GBM model

**Data Science Intern - Tabs and Syrups**, Chennai (May 2018 - Jul 2018)

- Developed a product recommendation engine for the online pharmacy app by mining association rules
- Built an end-to-end ETL pipeline on Firebase and MSSQL databases to develop the company's dashboard
- Created a proprietary Image processing application for the company, from scratch, using Python and OpenCV
- Worked with a team of research and domain experts to cleanse the product database of the app

## Projects

**DNA Sonification Tool** (Oct 2020)

- Developed an auditory display tool in R for genetic sequence analysis by designing algorithms that incorporated music theory to convert nucleotide and protein sequences into songs
- Enhanced the musicality of the tool's output while maintaining the overall analytical capabilities.

**Genetic Engineering Attribution** (Sep 2020)

- Developed an algorithm that identifies the most likely lab-of-origin for genetically engineered plasmid samples using machine learning techniques with features extracted from the nucleotide sequences
- Outperformed the existing state of the art model in use that was based on BLAST

**DengAI – Disease Spread Prediction** (July 2020)

- Designed an ensemble method to forecast the spread of dengue, using Time Series Analysis and Statistical modeling, with the aid of environmental data collected over a decade in two South American cities
- Trained the models on Collab GPUs and acquired a top 5% rank globally

**Flu Shot Learning – Healthcare Analytics Challenge** (June 2020)

- Employed deep learning techniques to predict whether people got H1N1 and seasonal flu vaccines, using sentiment and behavior data from the National 2009 H1N1 Flu Survey conducted by CDC
- Implemented a Neural Factorization Machine to attain an AUC of 0.86 and got a top 10% rank

**Climate Data Analysis - Department of Chemical Engineering, IITM** (Aug 2019)

- Analyzed the data collected by moving and stationary sensors placed across multiple cities in India. Performed extensive data cleaning and validated key insights through hypothesis testing
- Created temporal visualizations and geospatial heat maps for radiation and air quality-related parameters

**IITMAA Sangam - ML Hackathon, 2019** (July 2019)

- Developed statistical models to aid in traffic management. Built a deep neural network to predict traffic volume using traffic and climate data collected over four years
- Achieved an accuracy of 93.9% and finished as one of the top 10 teams among the 450 teams that participated

**Data Scientist Career track with R – DataCamp** (May 2018 – Sep 2018)

- Completed a series of 23 courses with case studies and projects to comprehend various fields of data science
- Gained expertise in data wrangling, visualization, hypothesis testing, modeling and reporting along with R programming to analyze and interpret complex data

**Fraudulent Transaction Prediction – Department of Computer Science, IITM** (Apr 2018)

- Designed a random forest classifier to predict fraudulent transactions using R to achieve 81.6% accuracy
- Ideated an Online fraud prediction algorithm & won first position in the Exebit Data Science Challenge, 2018

**User Preference Modelling** (Dec 2017 – Feb 2018)

- Ideated and created a website to work with dynamic test data: *http://fyf.gearhostpreview.com/*
- Built logistic regression models using correlating parameters from a survey conducted in the UK as training data to predict probabilities of the user's phobias

## Positions of Responsibility

**Core, Bhoutics (Physics department Festival), 2017**

- Planned and organized the 2nd edition of Bhoutics, which witnessed a 300+ footfall over two days
- Led a team of three to conduct Physics Bowl, an analytical physics event with over 50 contestants

**Coordinator, Media and Students Relation, Shaastra 2017**

- Organized the Trichy edition of Shaastra Spark, a national level quiz competition for school students
- Coordinated hospitality facilities for over 1000 participants, managing heavy logistical demands