

# Vamshidhar Reddy Gunnala

Aldie, VA-20105 | (571) 523-9365 | [vgunnala@gmu.edu](mailto:vgunnala@gmu.edu) | [linkedin.com/in/vamshidhar-g/](https://www.linkedin.com/in/vamshidhar-g/) | <https://vamshi4067.github.io/>

## SUMMARY

Dedicated analytical professional with excellent problem-solving skills and a Machine learning Enthusiast having over 4+ years of experience in design, development, support testing in the field of Data Analytics in the industrial and service sector. Passionate Data Scientist having a strong background in statistics seeking to solve the boatnecks of the business and increase productivity by implementing the Machine Learning models. Skilled in predictive modeling, data processing, and data mining.

## EDUCATION

**George Mason University, Fairfax,**

**Master of Science in Data Analytics Engineering, GPA: 3.5**

**May 2020**

**Relevant Coursework:** Predictive Analytics, Deep Learning, Operational Research, Big Data to Information

**GITAM, Hyderabad, India**

**Bachelor of Technology in Mechanical Technology, GPA: 8.1/10**

**May 2016**

## TECHNICAL SKILLS

**Languages:** SQL, R, Python, NoSQL, SAS, PL-SQL, T-SQL

**Algorithms:** Linear Regression, Logistic Regression, Polynomial Regression, Lasso Regression, Ridge Regression, Decision Trees, Random Forest, Naive Bayes, Support Vector Machines, K-nearest neighbors, K-means, Hierarchical Clustering, Neural Networks

**Machine Learning Packages :** Scikit-Learn, NumPy, SciPy, TensorFlow, Pandas, Matplotlib, dplyr, Caret, tidyverse, ggplot2.

**Deep Learning:** Artificial Neural Networks, Convolutional Neural Networks, Multi-Layer perceptron's, Recurrent Neural Networks, LSTM, GRU, SoftMax Classifier, Back Propagation, Chain Rule, Choosing Activation Functions, dropout, Optimization Algorithms, Vanishing and Exploding Gradient, Striding, Padding, Optimized weight Initializations, Batch Normalizations, Max Pooling.

**Statistical packages & Business Intelligence tools:** R, Tableau, SAS, Excel (Pivot Tables, Power Pivot, VLOOKUP), SSIS, SSRS

**Database Systems:** Microsoft SQL Server, Oracle SQL Developer, MySQL, Redshift, DynamoDB

**Tools:** Tableau, Alteryx, Excel, Microsoft Power BI, Orange, Weka, Jupiter, Spyder, RStudio, PyCharm, SSIS, JIRA.

**Big Data Technologies:** Hadoop, Map Reduce, Spark, Hive, Pig

## PROFESSIONAL EXPERIENCE

**Uber, Data Analyst**

**Jun 2016 – Apr 2018**

- Performed **Extraction, Transformation and Loading (ETL)** of the financial data using Microsoft SSIS and IBM DataStage.
- Developed automated solutions to address operational weaknesses identified through monitoring of weekly trends. Design and analyze experiments to measure efficacy of various product/ process measures on rider.
- Developed SQL stored **procedures, views, indexes and complex SQL queries** for extracting, manipulating and loading the customer and transaction data.
- Created visually impactful reports and dashboards for real time insights on key performance metrics using Tableau, SSRS, Excel and provided actionable insights to retain customers; **increased product sales by 7%**
- Improved the query latency by ~ 25%** by optimizing existing stored procedures in **Oracle database**; created data flows & identified the inter-dependencies among stored procedures to remove redundancies
- Built, maintained, and automated detailed models for city-level, regional, and global reporting with quality assurance as top priority.
- Extensively used **JIRA** as defect tracking system to track issues and to configure various workflows, customizations.
- Worked closely with analysts across multiple cross functional teams such as Product Categories, Com- ops, Quality, and Safety to ensure seamless integration of processes and reporting standards.

**George Mason University, Data Analyst (Graduate Teaching Assistant)**

**Sep 2019 – May 2020**

- Conducted quality assurance checks on data, reports, and presentations and created **BI reports** for alumni data to help University Life office find target alumnus for research funds.
- Performed hypothesis testing to analyze and validate trends among different groups.
- Performed exploratory data analysis and provided impactful insights using tableau like scatter plots, dot plots, box and whisker charts, geographic maps and density charts.
- Directed pre-college study and developed course work. Provided Academic Projects support for students in Data Warehousing in visualization using **tableau** and creating tables using **Microsoft SQL server**.

## ACADEMIC PROJECTS

---

### Realtime detection of anomalies in Steel Manufacturing using Deep Learning-based Computer Vision (Sponsored Project) Jan 2020 - Present

This project is sponsored by **ACCURE** and the main aim of this project is to create a computer vision model that is fast enough to determine visual defects in real-time by analyzing a live stream of images from a camera. Our project is an attempt to leverage and possibly improvise **YOLO** machine learning algorithms, extract valuable information from a stream of pictures using advanced data science techniques and utilize deep learning-based computer vision to detect anomalies, check parts for many different defects (e.g., Inclusion, scratches, pitted surfaces, or deformations caused by faults in production) and specifications (mainly dimensional abnormalities) in real-time on finished products to minimize the time & cost spent for dimensional measurement and surface inspection while maintaining the quality standards by automating the inspection process in manufacturing.

### Malaria Cell detection using Convolution Neural Networks: Aug 2019 – Nov 2019

In this **Deep learning** project, the objective is to detect the cells effected from malaria, so that the infected person can be informed at the initial stages of the infection before it turns endemic. By looking at the images in the dataset a human can easily say which cells are affected by malaria and which are not affected, but it would be a tiresome job for a human to check all the endemic cases. So, in order to automate this process, we applied various architectures that gives us the best modelling technique in deep learning to correctly classify the new or unseen image into parasitized and uninfected categories. We tried using a few of the pre-trained models like **ResNet-18** and **VGG-16** which has their parameters pre-trained.

### Predictive Modelling on Young People Survey: Jan 2019 – May 2019

The objective of the project is interpreting data and find out the differences in interests of the male and female based on their demographics, music interests and health habits. To predict about the willingness to put money into healthy food based on the hobbies and interests, health habits and spending habits. In addition to this, dataset is used to find out the habituate of people based on gender in prevalence to phobias.

- Performed data preprocessing including outlier treatment, missing value treatment, variable transformation and imputations in **Python**.
- Performed Exploratory data analysis (**EDA**) to observe trends and patterns in the data.
- Used the **SVM, KNN, Ridge Regression, Random Forest predictive and Lasso Regression** models.

### Anime Recommendation: Aug 2019 – Nov 2019

- Created a recommendation system using **k-nearest neighbor** and performed **semantic analysis** to recommend the shows based on the names and it is more useful to recommend the shows by mapping with the partition as partial names and similarity within the names.
- Performed partial semantic analysis for our data where we tried to recommend the shows based on the similarity in the semantics of the names and based on same type, genre.
- Applied different imputation techniques (**MICE**) to handle NULL values of each column based upon the correlation of the attributes.
- Calculated the distance between the categorical variables using **Binary Euclidean Distance, coefficient of similarity** and **Hamming distance**. Used **clustering technique** for recommending the shows.

### Statistical Analysis of US employment and the poverty rate: Aug 2018 – Nov 2018

- Our project goal was to analyze the employment and the poverty rate in all the states of United States of America. The problem we are solving is to evaluate the role of the various sectors that affect the employment.
- We considered the parameters like manufacturing, transportation, construction, government and service providing sectors. We have analyzed the dataset and applied all the regression techniques to know which the stronger variables in the employment are and for visualization we have used Orange and tableau.
- **Python** and **R** was used for extracting the US employment data for each state.
- Linear regression, Multiple Regression, Zero-R and k-Nearest Neighborhood algorithms was implemented to know which factors are highly responsible for the employment.

## HONORS AND ACTIVITIES:

---

- Our Capstone project "Realtime Detection of anomalies in Steel Manufacturing" had been published in the George Mason University's Annual Report and Awards.
- Active member of GMU Cricket Club.
- Designed and fabricated an eco-friendly go-kart and represented my University at National level competition.
- Active member at society of Data Analytics Engineers at Volgenau School of Engineering.