VAMSHIDHAR REDDY.GUNNALA

(571)-523-9365 | vgunnala@gmu.edu | Aldie, VA-20105

LinkedIn : https://www.linkedin.com/in/vamshidhar-reddy-gunnala/

GitHub: https://github.com/vamshi4067 |**Portfolio** : https://vamshi4067.github.io/vamshidhar.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.
- Ability to work effectively in cross-functional team environments and experience of providing training to business users. highly analytical and process-oriented data analyst with in-depth knowledge of data warehousing concepts, ETL, data manipulation, data modeling, reporting and visualization.
- Liaising with Business users, Project management team and other stakeholders regarding requirement gathering.
- Interpret problems and provide solutions to business problems using data analysis, data mining, optimization tools, machine learning techniques and statistics by logical and physical data models putting to practice.

EDUCATION

George Mason University, Fairfax, VA
Master of Science, Data Analytics Engineering – GPA: 3.51/4 **May 2020**

GITAM University, Hyderabad, Telangana, India
Bachelor of Technology, Mechanical Engineering - GPA: 8.1/10 **May 2016**

TECHNICAL SKILLS

- **Programming:** Python, R, SAS, SQL, PL-SQL, T-SQL,
- **Database Technologies**: MySQL, Oracle.
- **Big Data Technologies:** Hadoop, Map Reduce, Spark.
- **Tools**: Tableau, Alteryx, Excel, Microsoft Power BI, Orange, Weka, Jupiter, Spyder, RStudio, PyCharm, SSIS, , JIRA.
- **Operating Systems**: Linux, Windows and Mac.
- **Packages and Cloud**: NumPy, Pandas, Scikit-learn, TensorFlow, Matplotlib, Seaborn, NLTK, Spacy, Beautiful Soup, AWS Services (Good Knowledge on services)
- **Software**: Microsoft Office Suite, MS Visio, MS Access, Frontline Analytic Solver.
- **Machine Learning Algorithms**: Logistic Regression, Linear Regression, Support Vector Machines, Decision Trees, K-Nearest Neighbors, Random Forests, Ensemble methods, Naïve Bayes, K-Means Clustering, Hierarchical Clustering.
- **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.
- **Related Skills**: Categorical Data Analysis, Non- Parametric Analysis, Multivariate Analysis, Forecasting, Optimization, statistics, Predictive modeling, Hypothesis Testing like ANOVA, CHI-SQ Test, A/B Testing.

PROFESSIONAL EXPERIENCE

Data Analyst (GTA): George Mason University **Sep 2019 – Present**

- Gathered requirements and Conducted surveys on the students experience in the Volgenau school of engineering at Mason.
- 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 and visualized data in Tableau using 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.
- Responsible for guiding 56 Undergraduate students in Advanced Database management systems which includes Triggers, Stored Procedures, Indexes, Performance tuning and Security.
- Scheduled and prompted meetings, office hours for students to clarify their doubts.

Data Analyst – Uber **June 2016 -April 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 and Excel (Pivot tables, Pivot charts).
- Coordinated with other members of the Community Operations team to measure the impact and efficiency of new products and business processes by various monitoring tools through APM like AWS CloudWatch, AppDynamics, Application Insights in Visual Studio.
- Built, maintained, and automated detailed models for city-level, regional, and global reporting with quality assurance as top priority.
- Pre-trip experience: App errors, reliability (surge and ETAs), ETA accuracy, pricing anomalies.
- Troubleshooting Payment: Payment gateway errors/ bugs, sifting through multiple payment profiles U4B, Zaakpay, Braintree and Build tracking dashboards and reports.
- 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.

Graduate Engineering Trainee - Parker Hannifin **Nov 2015-Mar 2016**

- Worked with quality assurance team to check the optimum tolerances and clearances of the hose pipe.
- Ensuring the manufacturing of Polyethylene Instrument Grade Tubing - E/EB Series with desired client requirements.

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.
- Tools: R and Tableau for visualization & Weka.

Formula SAE Project- Automotive Design:

- Performed 3-D modeling using Creo and FEA using ANSYS workbench, on Aluminum uprights, to reduce the weight by 50%.
- Designed chassis and crash analysis/worthiness using LS-DYNA.
- Simulated the suspension geometry using Optimum Kinematics and applied Tire Data into the design to model vehicle dynamics using MATLAB and Microsoft Excel.