VAMSHIDHAR REDDY GUNNALA

DATA ENTHUSIAST

CONTACT

vgunnala@gmu.edu
(248) 516-9320
23659 Glenmallie CT, Aldie, VA
https://www.linkedin.com/in/vamshidhar-g/
https://github.com/vamshi4067
https://vamshi4067.github.io/

PROFILE

Data Scientist with expertise in Machine learning and Deep learning and building data models to solve business problems. Aiming to use my knowledge of Data Analytics, Statistics and Machine Learning to transform your data and drive business growth.

EDUCATION

George Mason University
[Fairfax, VA] 2018 - 2020
Master of Science – Data Analytics Engineering GPA – 3.5/4

GITAM University
[INDIA] | 2012 - 2016
Bachelor of Engineering – Mechanical Engineering GPA – 8.1/10

MANAGEMENT SKILLS

- Agile Development
- Critical thinking & problem solving
- Communication & Decision-making
- Git, Excel, PowerPoint

TECHNICAL SKILLS

- **Programming Languages:** R, Python, SQL, T-SQL, PL-SQL
- **Machine Learning:** Regression Models, Time series Analysis, Classification, Clustering, PCA, Random Forest, Naïve Bayes, NLP.

EXPERIENCE

Jan 2020– Present
Machine Learning Engineer – Intern | Accure AI. Inc

- Implemented Deep **learning** model using **YOLO** algorithm to predict Realtime detection of anomalies in Steel Manufacturing and performed various hyperparameter tuning to increase the performance of the model
- Developing a customised Deep Learning based **OCR** model to capture the right information from a document that needs to be processed and approved.

Sep 2019 – May 2020
Data Analyst (Graduate Teaching Assistant) - George Mason University,

- 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. Created tables using **Microsoft SQL server.** Performed exploratory data analysis and provided impactful insights using tableau like scatter plots, box and whisker charts, geographic maps and density charts.
- Collaborated with various other teams and departments to share and discuss issues through MS Teams and WebEx meetings.

June 2016 – April 2018
Data Analyst | UBER

- 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 (Pivot Tables, V-lookup and macros)** 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
- Extensively used **JIRA** as defect tracking system to track issues and to configure various workflows, customizations.
- Strong proofreading skills and overall consistent attention to details provided and enhanced modules such as managed document review, tax and audit processes to reduce recurring patterns.

VAMSHIDHAR REDDY GUNNALA

DATA ENTHUSIAST

- **Statistics:** Hypothesis testing, Confidence Interval, Distributions, ANOVA, A/B testing, CHI-SQ Test

- **Deep Learning:** ANN, Convolutional Neural Networks, Recurrent Neural Networks, LSTM, GRU.

- **BI Tools and Data Visualization:** Alteryx, Tableau, Power BI, SAS, Excel (Pivot Tables, Power Pivot, VLOOKUP, Frontline Analytical Solver, Macros), SSIS, SSRS, matplotlib, seaborn, plotly, ggplot, ggplot2.

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

CERTIFICATES

- ETL extraction using Apache Spark – Databricks
- AWS Foundations: Getting Started with the AWS Cloud Essentials
- Tableau Analyst
- AMCAT Certified Data Processing Specialist
- AWS Foundations: The Elements of Data Science
- NLP with Python for Machine Learning Essential Training

ACTIVITIES

- Active member of GMU Cricket Club.
- Designed and fabricated an **eco-friendly go-kart** and represented my University at National level competition (**SAE Baja**).
- Active member at society of Data Analytics Engineers at **Volgenau School of Engineering.**

ACADEMIC PROJECTS

Malaria Cell detection | Deep Learning, CNN, Resnet-18, VGG-16, Python, Tableau

- Objective of this project is to detect the cells infected by Malaria using 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.
- Classification problem which has only two different classes to predict parasitized and uninfected.
- Built different Deep Learning models with different hyperparameters and performed Transfer Learning to use the pre- trained models like **Resnet-18** and **VGG-16** to compare the performance of the various models.
- Measured and evaluated model performance and the model classified the cells with an **accuracy of 95%, precision of 97%, recall of 94% and with F1 score of 95.**

Predictive modelling on Young People survey | SVM, KNN, Ridge Regression, Random Forest and Lasso Regression

- The goal of this 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.
- 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| KNN, Semantic Analysis, MICE, Python, Power BI, Tableau

- 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.

Car Price Prediction | XGBoost, Random Forest, Flask, Python, Heroku

- Performed Exploratory data analysis (**EDA**) to observe trends and patterns in the data.
- Built a model to predict car price using **XGBoos**t library and **Random forest** algorithm.
- Tried using **PyCaret** library to automate the pipeline of ML model.
- Model predicted the price of a car with an **R-square** of **0.89** and deployed the model using **Flask** framework on **Heroku** which is a platform as service.