Venkata Guru Siva Sai Nagarapu

Location: Baltimore, MD |Email: nagarapu37@gmail.com | Phone- (941)-216-8112
LinkedIn: https://linkedin.com/in/Venkata-guru-siva-sai-nagarapu| GitHub: - https://github.com/Sivasai998
Website: **https://siva-portfolio-website.herokuapp.com/**

EDUCATION

University of Maryland, Baltimore County, MD, USA   **Jan 2020-Dec2021**
Master of Science, in **Data Science**   **CGPA: 3.89/4**
Key specialized courses: Algorithms and Data Structures**,** Intro to Data Science, Database Management systems, Intro to Machine Learning, Platforms for Big Data, Intro to Natural Language processing, Artificial Intelligence.

TECHNICAL SKILLS

General skills: Data Visualization, Machine Learning, Statistical Analytics, Agile Methodologies, Data Preparation, Quality Management, Business Intelligence, Pattern recognition, Data Mining.
Libraries: NumPy, Pandas, Scikit-Learn, Matplotlib, Seaborn, SciPy, NLTK, TensorFlow, Keras, OpenCV, Django, Flask.
Programming Languages: Python Spark, HTML, CSS, SQL, SAS, R, Java, C, C++.
Databases: MySQL, MS SQL Server, Teradata, Oracle, MongoDB, PostgreSQL.
Visualization/ Big Data tools: MS-Power BI, MS Excel (Analysis ToolPak), Plotly, Hadoop, Google Data Studio, Tableau, ETL.
Other skills: Git, Jupiter, Anaconda, AWS (EC2, S3), JIRA, Shell Scripting, selenium, unit testing, Microsoft Azure.

WORK EXPERIENCE   *1 year*

Index- Analytics LLC, Maryland, USA   June 2021- Present
Data Analyst Intern
- Developed and automated a **r**isk-scoring **model** for a federal healthcare agency using PySpark to predict risk score by analyzing a patient's demographic variables, medical diagnostic history, Hierarchical Condition Categories (HCC), and insurance claims data using a multivariate linear regression model that helps stakeholders strategize health insurance pricing and medical coverages.
- Employed the Gradient Descent algorithm to **minimize the loss function**, accelerate convergence, and optimize parameters for better performance
- Reduced processing time by **75%** (20 minutes in SAS to under % minutes in **Pyspark**) and saved approximately **$50,000** in labor and computational costs

InterCurve LLC, Hyderabad, India   May 2019- Dec 2019
Junior Python Developer
- Developed Backend Components and used **Django**, **MYSQL** database and different python libraries for development of application.
- Handled programming tasks for and maintained **5 internal websites** with a high success rate (97%) in product update deployment.
- Automated and optimized collecting data using **SQL**, reaching over a **35%** system's response time boost.
- Performed **Unit Testing** and used **Jira** for bug-tracking and used **GIT** as version control.

python Developer Intern, InterCurve LLC, Hyderabad, India   June 2017 – August 2017
- Worked on building modern single page web applications and utilized **Flask** Framework to implement and develop the reusable code in the application there by improving the reusability of code by **40%.**
- Worked on high-impact projects, e.g., expense tracker, delivering solutions with lower than **15% code churn**.
- Used Python programming to implement **algorithms**, data processing and various automation tasks.

ACADEMIC PROJECTS

Netflix Movies and TV Show Analysis in Different Countries of the world – [Plotly, Python-Pandas, Matplotlib]
- Used python 3.6 to analyze the Netflix Dataset found in Kaggle and performed cleaning, data characterization, data visualization on movies and Tv shows released in different countries across the world.

Sentiment analysis of Election Prediction using NLP- [Python-Pandas, NLTK, Text-blob, Topic-modelling]
- Collected the data from **twitter** and News articles and performed data Cleaning, data visualization like word cloud on election data and used **text blob** to find the polarity and calculated the sentiment. And used **NLTK** library performed different techniques like stemming, Lemmatization, TF-IDF transform and used to **Naïve-bayes algorithm** to predict the percentage of tweet. And performed the **topic modelling** on News articles and calculated the sentiment of every month. This project helps to predict the final winner of election prediction

English Premiere league Analysis using Big-Data Tools- [Python, MongoDB, Pyspark, Tableau]
- Used different datasets from Kaggle and stored in MongoDB. And using **Pymongo** connecting the **Jupyter** and **MongodB** database performed Data preparation, data characterization and with the help of Naive bayes algorithm I have trained the model and predicted the final winner of the season and using Tableau performed the visualizations.