Venkata Guru Siva Sai Nagarapu

nagarapu37@gmail.com | (941)-216-8112 |Arbutus, MD,21227 |
linkedin.com/in/Venkata-guru-siva-sai-nagarapu|https://github.com/Sivasai998
website: https://siva-portfolio-website.herokuapp.com/

OBJECTIVE

A Data Scientist enthusiast with a career in Data Science, Machine Learning, Big Data and Information technology with experience as a python developer and Passionate about capturing hidden patterns and values through a blend of business acumen & data Science techniques. Proficient in data processing, data mining, data Visualization and algorithms as well as scripting languages, including Python and SQL.

EDUCATION

University of Maryland, Baltimore County, MD, **USA** **Jan 2020-current**

Master of Science, in Data Science **CGPA: 3.89**

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

Operating Systems: Android, iOS, Windows 7/8/10, Mac OS X, Linux, Ubuntu
General skills- Data Visualization, Machine Learning, Statistical Analytics, Agile Methodologies, Data Preparation, Quality Management, Business Intelligence, Pattern recognition, Data Mining.

Programming Language: Python libraries- (NumPy, Pandas, Scikit-Learn, Matplotlib, Seaborn, SciPy, NLTK, TensorFlow, Keras, OpenCV), Spark, OpenCV, Django, Flask, HTML, SQL, SAS, R.
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

Data Analyst Intern, Index- Analytics LLC, Maryland, USA June 2021- Present

- Developed and automated a risk-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

Junior Python Developer, InterCurve(LLC), Hyderabad, India May 2019- Dec 2019

- Developed Backend Components and used Django, MYSQL database and different python libraries for development of application.
- Coordinate, communicate and provide technical support to other functional groups relating to web applications.
- Followed and improved established processes for software development life cycle with an agile approach to delivery software
- Performed Unit Testing and used Jira for bug-tracking and used GIT as version control.
- Worked with different Datasets and performed data cleaning and visualization.

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 the model view control architecture.
- Wrote python routines to log into the websites and fetch data for selected options.
- Used Python programming to implement algorithms, data processing and various automation tasks.

KEY 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 Jupiter 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.

CERTIFICATIONS

- PCEP - Certified Entry-Level Python Programmer, PYTHON INSTITUTE Certification Code-KxLw.sfJb.8KQX
- Python for Data Science and AI, COURSERA Certification Code- B8XHHT6LQWQZ
- Python Data Structures, UDEMY
- Interactive Python Dashboards with Plotly and Dash, UDEMY
- Fundamentals of deep learning for Computer Vision, NVIDIA