VINAY KIRAN RAJU









in Vinay Kiran Raju | 🔀 raj.vinay2408@gmail.com | 🕻 +1 (201)-492-8306 | 🗑 vinaymeesaraganda.github.io

TECHNICAL SKILLS

- Big Data Technologies: Hadoop, MapReduce, Apache Spark, Apache Kafka, HDFS, Sqoop, Ni-Fi, Hive, Cloudera, HBase.
- Programming Languages: Python (Pandas, Beautiful Soup, NumPy, Matplotlib, Seaborn, scikit Learn), Scala, SQL.
- **Cloud Services :** Amazon Web Services, Microsoft Azure.
- Databases: MySQL, PostgreSQL SQL Server, Oracle, HBase.
- Visualization Tools: Tableau, MS EXCEL.
- Machine Learning Algorithms: Random Forest, Fb Prophet, Naive Bayes, Logistic Regression, XG Boost, Decision-Tree.
- **Version Control:** Git, GitHub.

EDUCATION

Pace University, Seidenberg School of Computer Science and Information Systems

New York, USA

Master of Science in Data Science | Concentration: Data Visualization & Machine Learning | GPA: 3.71/4

May 2023

Gandhi Institute of Technology and Management

Andhra Pradesh, India

Bachelor of Technology in Electronics and Communication Engineering | GPA: 8.38/10

Apr 2021

WORK EXPERIENCE

Company: Natsoft Corporation Role: Data Engineer (Internship)

Feb 2023 - Apr 2023

New Jersey, USA

- Performed ETL pipelines to extract data from RDBMS and load the data into HDFS and Hive using SQOOP and perform aggregations and transformations using Spark.
- Employed AWS services like EMR and Amazon S3 to handle large data sets for development, processing and storage.
- Generated data visualizations using Tableau, communicating actionable insights and analytics based on Medicaid data.
- Implemented Git version control to manage the source code and integrating **Git** with **Jenkins** to support build automation.

Company: Natsoft Corporation

Sep 2020 - Dec 2021

Role: Data Engineer

Hyderabad, India

- Established EC2 instances and EMR clusters for development, testing and deployment as per the requirements.
- Developed Extract, Transform, and Load (ETL) operations, including Import and Export of data into HDFS and Hive using **Sqoop** and managed data within the environment.
- Processed web URL data using Scala and converted to data frames for further transformations and engaged POCs using Apache Spark SQL.
- Loaded and processed semi-structured data such as XML, JSON, Avro and Parquet and optimized Spark SQL queries to enhance data access.
- Engaged in the development of Code & peer review of assigned tasks and Bug fixing

ACADEMIC PROJECTS

British Airways Analysis - virtual Experience

Jun 2023

- Scraped over 1000 customer reviews from Skytrax website using Python library Beautiful Soup and performed sentiment analysis, revealing trends: 48.9% positive and 43.2% negative and 7.9% neutral sentiments.
- Prepared customer booking dataset for predictive modeling by identifying key variables like purchase lead time, route, flight hour, and length of stay.
- Developed Random Forest model in Python predicting customer bookings with 85% accuracy, evaluated model and feature importance, identified improvements by adding promotional offers, payment type.

Loan Default Prediction Using Machine Learning Models

Feb 2023

- Conducted analysis on loan data from 10,000 applicants for a Nairobi-based digital lending company, employing Random Forest and XG Boost algorithms to build predictive models.
- Applied data cleaning, feature engineering, and evaluation techniques using python libraries to identify patterns and select most effective model, achieving a remarkable accuracy of 94% with the Random Forest Classifier.

Time Series Forecasting Using FB Prophet

May 2022

- Analyzed Kaggle data on sunspot counts and implemented Time Series Forecasting using FB Prophet in Python.
- Cleaned and prepared the data, conducted exploratory data analysis, and predicted future sunspot counts with a mean absolute percentage error of 4.40.

CERTIFICATIONS

Agile Methodology Virtual Experience Program - Cognizant

Jun 2023

Data Science Virtual Experience Program - British Airways

Jun 2023

Azure Data Fundamentals - Microsoft

Apr 2023

Tableau Desktop Specialist - Tableau

Mar 2023

Google Data Analytics -Coursera

Nov 2022