

SAI KRISHNA VIVEK Y

Email: yskvivek9@gmail.com
Phone: 9701508517

EDUCATION

GEORGIA INSTITUTE OF TECHNOLOGY
MS in Data Analytics
2021(Expected) | CGPA – 3.5/4.0

GOKARAJU RANGARAJU INSTITUTE OF ENGINEERING AND TECHNOLOGY
B.TECH - CSE
2018 | 86%

JUBILEE HILLS PUBLIC SCHOOL
10th [CBSE]
2012 | CGPA – 10/10

CERTIFICATIONS

TECHNOLOGY ENTREPRENEURSHIP PROGRAM | INDIAN SCHOOL OF BUSINESS, Hyderabad.
ORACLE CERTIFIED ASSOCIATE SQL | ORACLE

TECHNICAL SKILLS

BIG DATA TECHNOLOGIES
Spark 3 • Hadoop • Hive • Impala
CLOUD SERVICES
AWS Redshift • EMR • S3 • Athena
PROGRAMMING LANGUAGES
Python • R • Java • C++
PYTHON AND R PACKAGES
numpy • pandas • scikit • scipy
keras • CRAN • mgcv • tseries
DBMS
SQL – MARIADB, POSTGRES SQL
VISUALIZATION
Tableau, D3

COURSEWORK

- Data Mining with Python
- Introduction to Machine Learning
- Computational Analysis
- Visual Analytics
- Time Series Analysis

PROFILE

- Around 3 years of experience in Data Engineering and Machine Learning using Python Spark and AWS technologies with strong knowledge in finance and healthcare.
- Currently pursuing Masters in Data Analytics at Georgia Institute of Technology.
- Excellent communicational, organizational, analytical, problem-solving and presentation skills.

EXPERIENCE – 3yrs and 2months

DATA ENGINEER

PHILIPS INDIA LIMITED – Royal Philips Healthcare (Domain: Healthcare)

June 2021 – Present | Bangalore

Technologies: Python, Apache Airflow, AWS Redshift, S3, EMR, SQL, Shell Scripting.

Involved in designing, building, testing and maintaining data pipeline using python and AWS cloud services.

- Involved in setting up of Airflow on docker container host and monitoring tools such as flower and Prometheus
- Building configurable, parametrized, scalable, robust ETL pipelines using SQL.
- Migrating data from various sources such as salesforce, google 360, adobe analytics etc. to Redshift.

DATA ENGINEER

DEVELOPMENT BANK OF SINGAPORE – DBS (Domain: Banking, Finance)

July 2018 – May 2021 | Hyderabad

Technologies: Spark 3, Python Spark, AWS, Hadoop, Hive, Impala, Airflow, MariaDB SQL, Shell Scripting, Jupyter Notebook, Drools, Tableau.

- Developed a Strategic Cost Allocation product using Pyspark, Hadoop, Impala and SQL.
- Performed extensive Exploratory Data Analysis on close to 90 billion data using SQL in hive.
- Provided support in ops using Jenkins CI/CD pipeline and scheduling of jobs using Apache Airflow.

PROJECTS IN MASTERS

1. Density estimation using multidimensional Gaussian Kernel to study whether the two brain regions are likely to be independent of each other.
2. Expectation Maximization using Gaussian Mixture Model on handwritten digits of MNIST dataset in python.
3. Comparative study of performance of Football Teams using Spectral Clustering(graph) in python.