

BIRAJDAR WAMAN

wamanbirajdar@gmail.com | +91 9881045931

EDUCATION

PG-DIPLOMA

C-DAC NOIDA

Post Graduate Diploma In Big data Analytics And ML | 2022

B.E. COMPUTER SCIENCE ENGINEERING

UNIVERSITY OF PUNE | 2021

CGPA: 7.16/10

HSC IN COMPUTER SCIENCE

STATE BOARD | 2015

Grade: 69.23%

SKILLS

PROGRAMMING

Python and iPython Notebook

Core JAVA

SQL

LINUX

HTML, and CSS

BIG DATA FRAMEWORKS

Hadoop

Apache Spark

Apache Hive

Kafka

DATABASES

MySQL

Cassandra

HBase

MongoDB

AWS AND SCHEDULERS

AWS S3, EMR, GLUE, LAMBDA, CloudWatch, AIRFLOW.

CI CD PIPELINES

GIT And DOCKER

HANDS ON

Statistics, Machine Learning, Power BI

LINKS

Github:// [wamanbirajdar_me](#)

LinkedIn:// [wamanbirajdar](#)

Medium:// [Tech_Blog](#)

PROJECT

ETL DATA PIPELINE | MYSQL, CONFLUENT KAFKA, PYSPARK, AWS S3, AWS REDSHIFT

- **Mysql-to-Kafka-to-AWS S3**
 - Producer Code Read all tables from db having Raw Data and create topics in Kafka, do some transformation.
 - In AWS S3 created bucket read tables inside S3 bucket we get folder respect to table written by Consumer Code.
- **AWS S3-to-Redshift**
 - PySpark Code read all folder from S3 dumps clean data into Redshift DB.
 - Incremental load, created one table for checking offset of records.

EMPLOYEE ATTRITION PREDICTION | PYTHON, ML, STREAMLIT

- **Exploratory Data Analysis and Model Building**
 - In EDA investigations on 7 lakh data so as to discover patterns, to spot anomalies, to test a hypothesis and to check assumptions with the help of summary statistics and graphical representations.
 - After Data Cleaning we build diff models Like logistic Regression, Decision tree, Random Forest, XGboost and Gradient Boosting Classifier.
- **Model Choosing and User Interface**
 - XGboost model give highest AUC so we used XGboost model for employee attrition datasets.
 - Built User Interface for XGBoost model. Deploy ML Model On Streamlit for ease of access to End User.

DATA PIPELINE | CASSANDRA, APACHE KAFKA, PYSPARK

- PySpark Producer fetch data from cassandra db and created topic in Kafka.
- PySpark Consumer Script continue running and write data into local system in parquet file format.

DATA PIPELINE | API-TO-SPARK-TO-HFDS

- Have open API including Date parameters Fetching data and dump into HDFS.
- On top of that perform some spark sql queries.

ACHIEVEMENTS | CERTIFICATION

- Python Certified By Google.
- ML Certified By Perfect E-Learning.
- Completed 3 Months Data Associate Bootcamp conducted By Symbiosis Skills and professional University.
- Write Tech blog on LinkedIn and on medium.
- 1 Gold 1 SILVER 3 BRONZE Medals By Hacker Rank.