Big Data Analytics

A Project Submitted

By

Batch 2

Under the Guidance of

Kunal

BMS INSTITUTE OF TECHNOLOGY AND MANAGEMENT

YELAHANKA BANGALORE

Phase 5 Documentation

Project: Big Data Analytics

Phase 4 Submission by: Batch 2
Year: 4th Year, Computer Science and Engineering (CSE)

1.Introduction to Deployment

Deployment bridges the gap between developing a machine learning model and making it usable in a real-world application. In this project, after data preprocessing and model training, we partially implemented deployment steps to integrate the system with a cloud database and prepared for further extensions like API or web app integration.
In this project, the deployment included:
Connecting to IBM Db2 on Cloud for storing rainfall data.
Inserting cleaned data into the database for persistent storage.
Validating data upload through query-based retrieval.
Model creation and evaluation, ready for serialization and future API deployment.

2. Database Deployment using IBM Db2

a. Connection Establishment
The project connects to the IBM Db2 Cloud Database using the ibm_db Python library. Connection credentials were securely passed into the ibm_db.connect() function.

```
conn=
ibm_db.connect("DATABASE=bludb;HOSTNAME=...;PORT=...;UID
=...;PWD=...;SECURITY=SSL", '', '')
```

This code successfully connects to the cloud database hosted on IBM Cloud.

b. Data Insertion

The cleaned rainfall dataset (df) was uploaded into a table (RAINFALL2) in IBM Db2 using SQL INSERT INTO queries inside a loop.

```
for i in range(len(df)):
    insert_query = f"INSERT INTO RAINFALL2 (STATE, DISTRICT,
TEMPERATURE, HUMIDITY, PRESSURE, WINDSPEED, RAINFALL)
VALUES ('{df.loc[i, "STATE"]}', '{df.loc[i, "DISTRICT"]}', {df.loc[i,
"TEMPERATURE"]}, {df.loc[i, "HUMIDITY"]}, {df.loc[i,
"PRESSURE"]}, {df.loc[i, "WINDSPEED"]}, {df.loc[i,
"RAINFALL"]})"
    ibm_db.exec_immediate(conn, insert_query)
```

This ensures that preprocessed data is stored in a structured and queryable format in the cloud.

c. Data Retrieval for Verification

To verify data upload, a SELECT * FROM RAINFALL2 query was executed and records were fetched using ibm_db.fetch_assoc().

python

CopyEdit

```
select_query = "SELECT * FROM RAINFALL2"
stmt = ibm_db.exec_immediate(conn, select_query)
result = ibm_db.fetch_assoc(stmt)
while result:
    print(result)
    result = ibm_db.fetch_assoc(stmt)
```

Successful retrieval confirms that data was correctly inserted into IBM Db2.

3. Model Deployment Strategy

Though the model is not yet deployed via an API or interface, the code structure makes it deployment-ready.

a. Model Training and Evaluation

Three models were trained and evaluated:

Random Forest Classifier

Logistic Regression

LSTM

Model accuracy was measured using accuracy_score, classification_report, and confusion_matrix.

4. Cloud and Production-Ready Considerations

While full deployment to cloud or web platforms was not done, the code sets up the following possibilities:

IBM Cloud Integration

Cloud-based storage of rainfall data in IBM Db2 provides real-time access for future inference or applications.