Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

Department of Computer Science and Engineering (Data Science)

Subject: Applied Data Science (DJ19DSL703)

Experiment - 9

(Data Science Project Architecture)

Bhuvi Ghosh
60009210191

Aim: To implement Data Science Project Architecture.

Theory:

DevOps (Development Operations):

DevOps is a set of practices, principles, and cultural philosophies that aim to enhance collaboration and communication between software development (Dev) and IT operations (Ops) teams. The goal is to automate the processes of software development, testing, and infrastructure management to deliver high-quality software more quickly and reliably.

- Code Building: Managing the process of compiling and building code into executable files.
- Continuous Integration (CI): Integrating code changes into a shared repository frequently and automatically.
- Continuous Deployment (CD): Automating the deployment of code changes to production environments.
- Infrastructure as Code (IaC): Managing and provisioning infrastructure through code to ensure consistency.
- Configuration Management: Maintaining and updating system configurations across different environments.
- Monitoring and Logging: Continuous monitoring of application and system performance with logging and alerting.
- Collaboration: Encouraging collaboration between development and operations teams to enhance communication.

MLOps (Machine Learning Operations):

- Data Management: Ensuring efficient data acquisition, cleaning, and storage for machine learning models.
- Model Training: Automating the training of machine learning models using diverse datasets.
- Model Deployment: Deploying models into production environments and managing their lifecycle.
- Monitoring and Logging: Continuous monitoring of model performance, logging, and alerting for any anomalies.
- Version Control: Managing versions of both code and machine learning models to track changes.
- Collaboration: Facilitating collaboration between data scientists, engineers, and other stakeholders in the ML workflow.
- Automation: Automating repetitive tasks to streamline the ML pipeline and reduce manual errors.

Difference between MLOps and DevOps:

- **Focus:**

MLOps: Primarily focuses on the machine learning lifecycle, including model development, training, and deployment.
DevOps: Primarily focuses on the overall software development lifecycle, from code writing to deployment and operations.

- **Nature of Artifacts:**

MLOps: Involves artifacts like machine learning models, datasets, and experimentation logs.
DevOps: Involves artifacts like source code, binaries, and configuration files.

- **Workflow:**

MLOps: Involves specialized steps for data pre-processing, feature engineering, and model evaluation.
DevOps: Involves steps like code compilation, testing, and deployment.
Tools:
MLOps: Uses tools specific to machine learning, such as TensorFlow, PyTorch, and specialized model deployment tools.
DevOps: Uses a broader set of tools for version control, CI/CD, infrastructure management, like Git, Jenkins, Docker, and Kubernetes.

Collaboration:
Both MLOps and DevOps emphasize collaboration between different teams involved in the software development and deployment processes. Collaboration ensures better communication, faster feedback loops, and more effective problem-solving.

Scalability:
MLOps: Focuses on the scalability of machine learning workflows, ensuring that models can handle larger datasets and diverse environments.
DevOps: Focuses on the scalability of software applications, infrastructure, and deployment processes to handle increased loads and user demands.

Reusability:
MLOps: Encourages the reuse of machine learning components, models, and workflows to save time and resources.
DevOps: Encourages the reuse of code, configurations, and deployment scripts to maintain consistency across different environments and applications.

Lab Assignment:
Using any dataset relevant to data science project implement in data preprocessing, feature engineering, and model training on.
The chosen machine learning model will be serialized and deployed through a Flask or FastAPI REST API endpoint, containerized using Docker for consistency. The assignment emphasizes MLOps practices by integrating the project into a CI/CD pipeline (e.g., GitHub Actions).

Furthermore, students will implement basic monitoring and logging functionalities to track the model's performance in a production-like environment. Clear and comprehensive documentation covering development environment setup, model training, deployment, and monitoring is an integral part of the evaluation. This assignment offers a hands-on, practical introduction to MLOps concepts while addressing the real-world problem of predictive maintenance**.**

app.py

```
Capp. route(" / predict" , methods = ['GET', 'POST']

def predict():
    if request "GET":
        return render_template("predict.html")

df = pd.read_csv('spam.csv', encoding = 'latin-1')
df.drop(['Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4'], axis = 1, inplace = True)
df['label'] = df['type'].map({"ham": 0, spam :1})
X = df['text']
y = df['label']


from sktearn.feature-extraction.text import CountVectorizer
from sklearn. model_selection import GridSearchCV

cv = CountVectorizer()
X = cv.fit_transform(X)

from sklearn. model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size = 0.2, random_state = 42
)

from naive_bayes import MultinomialNB

clf = multinomialNB(alpha = 0.1, force_alpha = True)

clf.fit(X_train, y_train)
ctf.test(X_test, y_test)

message = request.form['message']
data = [message]
vect = cv.transforn(data). toarray()
my_prediction = ctf. predict(vect)
return render_template('predict.html', prediction = my_prediction)
```

req.py

```python
import requests

url = "http://localhost:5000/api"

test_data = {
    "buying_low": 0.0,
    "buying_med": 1.0,
    "buying_vhigh": 0.0,
    "maint_low": 0.0,
    "maint_med": 0.0,
    "maint_vhigh": 1.0,
    "doors_3": 0.0,
    "doors_4": 1.0,
    "doors_5more": 0.0,
    "persons_4": 1.0,
    "persons_more": 0.0,
    "lug_boot_med": 0.0,
    "lug_boot_small": 0.0,
    "safety_low": 1.0,
    "safety_med": 0.0,
}

# 'unacc'

r = requests.post(url, json=test_data)
print("Predicted class: ",r.json())

```