News Classification MLOps Pipeline
User Manual

1 Introduction

This document provides a comprehensive guide to setting up and using the News Classification MLOps Pipeline. The pipeline automates:

- News data collection and classification
- Model training, fine-tuning, and serving
- Web application hosting
- Real-time monitoring and visualization

2 System Architecture Overview

The system consists of two major parts:

- **Host Side Services**: Model training, serving, monitoring (Prometheus, Grafana)
- **Docker Side Services**: Web application (FastAPI + HTML templates)

3 Components

3.1 Model Pipeline (Host Side)

- **Stage 1 - Data Extraction**: Extracts data for model training.
- **Stage 2 - Hyperparameter Tuning**: Selects the best parameters to optimize model performance.
- **Stage 3 - Model Fine-tuning and Serving**: Best model is fine-tuned and served on port `5002`.
- **model/run.py**: Starts the MLflow server on port `8080` and initiates DVC pipeline runs.

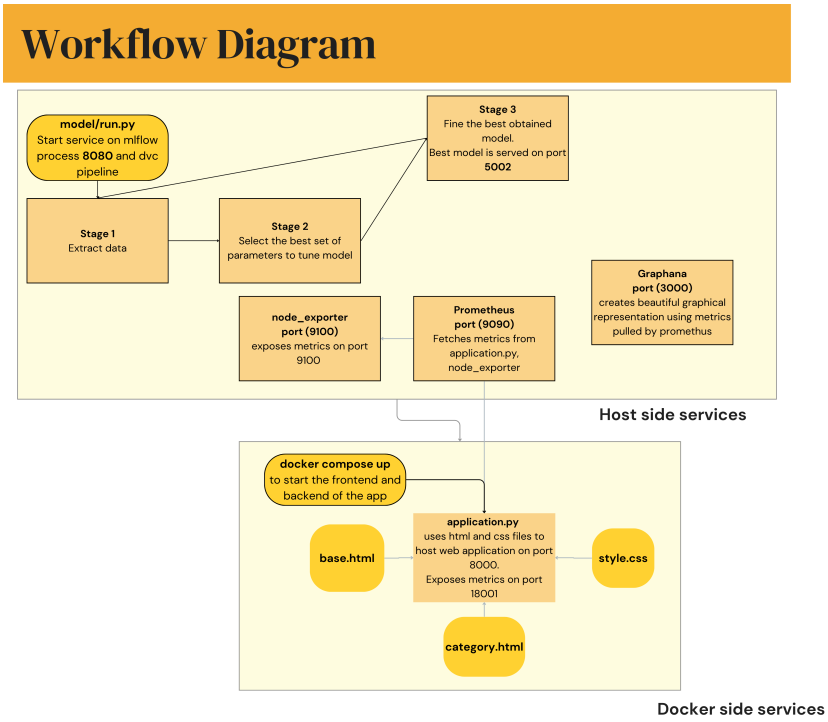

Figure 1: System Workflow Diagram

3.2 Monitoring Stack

- **node_exporter (port 9100)**: Exposes system metrics.
- **Prometheus (port 9090)**: Collects metrics from `application.py` and `node_exporter`.
- **Grafana (port 3000)**: Visualizes collected metrics with customizable dashboards.

3.3 Web Application (Docker Side)

- **application.py**: Serves the frontend using HTML (Jinja2 templates) and CSS.
- Templates used:
  - `base.html`
  - `category.html`
  - `style.css`

- Hosted on port `8000` for user access.
- Metrics for Prometheus exposed on port `18001`.

4 Installation

4.1 Prerequisites

- Docker and Docker Compose
- Python 3.9 or higher
- (Optional) NVIDIA Docker for GPU acceleration

4.2 Setup Instructions

1. Clone the repository.
2. Run the datapipline:

```
    python3 model/run.py

```

3. Start the web application:

```
    docker-compose up -d

```

5 Configuration

5.1 Environment Variables for the dockerised app

Variable	Description
DB_HOST	PostgreSQL database host address
DB_NAME	Database name
DB_USER	Database user
DB_PASSWORD	Database password
MODEL_SERVER_URL	URL of the ML model inference server
PYTHONUNBUFFERED	Set to 1 for real-time logging

6 Usage Guide

6.1 Web Interface

- Access homepage: `http://localhost:8000`
- View news by category: `http://localhost:8000/category/<category>`

6.2 API Endpoints

Endpoint	Function
GET /	Homepage listing all news
GET /category/{category}	News articles filtered by category
POST /submit-feedback	Endpoint to submit corrections/feedback for retraining

7 Monitoring Setup

7.1 Prometheus Configuration Example

```
scrape_configs:
  - job_name: "news_app"
    scrape_interval: 1s
    static_configs:
      - targets: ["localhost:18001"]
```

7.2 Metrics Tracked

- `api_call_counter`: Number of API calls
- `num_tags_feed`: Distribution of news categories

7.3 Starting Prometheus Service

- `Start Prometheus server`: Start on port 9090

- `Start node`$_{e}xporter$: $To extract system metrics$`Open Graphana` : $Open port 3000 to access grapahan UI$.

8 Model Training and Update Cycle

8.1 Training Workflow

1. Data collected by `extract_data.py`.
2. Best parameters selected using `register_best_model.py`.
3. Best model fine-tuned using `fine_tune_best_model.py`.
4. Model registered and served via MLflow.

8.2 Serving

- ML model is served on port `5002`.
- Accepts JSON input (news text) and returns category prediction.

9 Maintenance Guide

9.1 Model Updates

1. Update model parameters in `params.yaml`.
2. Re-run the DVC pipeline:

```
dvc repro
```

3. Alternatively, manually start:

```
python3 model/run.py
```

4. Restart the model serving container if needed.

9.2 Scaling Tips

- Increase Docker replicas for high availability.
- Use GPU acceleration for model training.
- For production scaling, migrate to Kubernetes.

10 Feedback Integration

- User corrections are collected from the web interface.
- Stored under `/data/news_feed.csv`.
- Automatically incorporated during the next training cycle.

11 Application User Interface

11.1 Home page

The app is served on port 8000. In the homepage 2 the navigation menu allows to switch between different categories:

- Home
- Business
- Politics
- Sport
- Tech
- Entertainment

For each feed on the homepage, the category is shown in Figure 2.

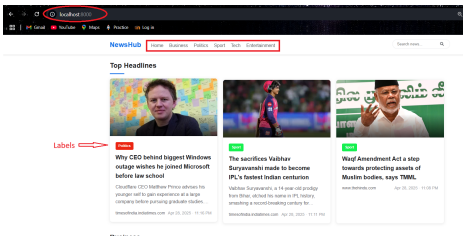

Figure 2: Home page of app

11.2 Category page

All the other categories are served at the address `http://localhost:8000/category/category_name`. For each category, users can send feedback to each news article, which is used to fine-tune the ML model used for prediction. Figure

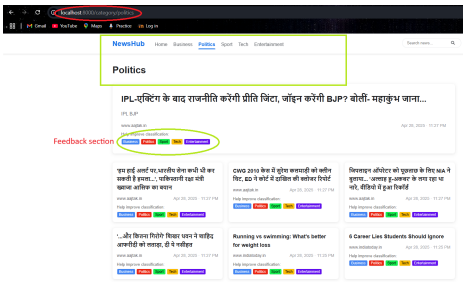

Figure 3: Caption

3 shows the same.