

Drug A Eligibility Prediction System

This project builds a machine learning-based system to identify patients diagnosed with Disease X who are eligible for Drug A but are unlikely to receive treatment. It includes a data pipeline, ML model, REST API using FastAPI, and Docker containerization.

Project Structure

```
.
├── app
│   └── api
│       └── main.py           # FastAPI server
├── model
│   ├── model.pkl            # Trained ML model
│   └── preprocess.py        # Feature engineering logic
├── data
│   └── model_table.csv       # Reference table for expected dataset structure
├── requirements.txt          # Python dependencies
├── Dockerfile                # Docker container setup
├── README.md                 # Project documentation
└── train_model.py            # ML model training pipeline
```

Setup Instructions

1. Clone the Repository

```
git clone https://github.com/yourusername/drug-a-eligibility.git
cd drug-a-eligibility
```

2. Create Virtual Environment

```
python3 -m venv venv
source venv/bin/activate
```

3. Install Dependencies

```
pip install -r requirements.txt
```

4. Train and Save the Model

```
python train_model.py
```

5. Run FastAPI Server

```
uvicorn app.api.main:app --host 0.0.0.0 --port 8000
```

API Usage

Endpoint

```
POST /predict
```

Request Body (JSON)

```
{
  "age": 68,
  "gender": "F",
  "high_risk": 1,
  "days_since_symptom": 3,
  "multiple_meds": true,
  "frequent_visits": false
}
```

Response

```
{
  "likelihood_of_treatment": 0.32
}
```

Run with Docker

1. Build Docker Image

```
docker build -t drug-a-api .
```

2. Run Docker Container

```
docker run -p 8000:8000 drug-a-api
```

Version Control Strategy

- **Git branches:**

- `main` : stable codebase
- `feature/<name>` : feature development
- `model/<version>` : trained models with tracked metadata

- **Git tags:**

- `v1.0.0-model` : model versions
- `v1.0.0-api` : API release

- **Conflict Resolution:**

- Use feature branches
- Pull Requests with code reviews
- Rebase before merging for clean history

Features Engineered

- `high_risk` : based on age ≥ 65 or comorbidities
- `days_since_symptom` : treatment only valid in first 5 days
- `multiple_meds` : multiple contraindicated medications
- `frequent_visits` : indicator of patient's physician contact

Evaluation Metrics

- Model: Random Forest Classifier
- Metrics: Accuracy, Precision, Recall, F1 Score

Assumptions

- Patients with contraindicated medications are less likely to be treated.
 - High-risk patients should be prioritized for alerting.
 - The likelihood score can be used to trigger EMR alerts.
-

Contact

For questions or contributions, please contact: [\[your.email@example.com\]](mailto:your.email@example.com)

Note: Please include the `model_table.csv` file when submitting or publishing.