

# XGBoost 糖尿病预测 API 文档

## 基本信息

- 基础URL: `http://1.94.9.72:5000`
- 所有请求和响应均使用 JSON 格式（除了文件上传）
- 所有响应都包含适当的 HTTP 状态码

## API 端点

### 1. 健康检查

检查API服务的运行状态。

- **URL:** `/health`
- **方法:** GET
- **响应示例:**

```
{
  "status": "healthy",
  "model_loaded": true,
  "model_file_exists": true
}
```

### 2. 单个预测

对单个样本进行糖尿病风险预测。

- **URL:** `/predict`
- **方法:** POST
- **Content-Type:** application/json
- **请求体格式:**

```
{
  "age": 45,                // 年龄（数值）
  "gender": "Male",         // 性别（"Male"/"Female"/"Other"）
  "bmi": 28.5,              // 体重指数（数值）
  "HbA1c_level": 6.5,       // 糖化血红蛋白水平（数值）
  "blood_glucose_level": 140, // 血糖水平（数值）
  "smoking_history": "former", // 吸烟史
                             ("current"/"never"/"former"/"ever"/"No Info")
  "hypertension": 1,        // 是否有高血压（0/1）
  "heart_disease": 0        // 是否有心脏病（0/1）
}
```

- **响应示例:**

```
{
  "prediction": 1,
  "prediction_text": "糖尿病",
  "probability": {
    "non_diabetes": 0.2,
    "diabetes": 0.8
  },
  "confidence": 0.8,
  "input_data": {
    // 输入的原始数据
  }
}
```

### 3. JSON批量预测

对多个样本进行批量预测。

- **URL:** /predict\_batch
- **方法:** POST
- **Content-Type:** application/json
- **请求体格式:**

```
{
  "data": [
    {
      "age": 45,
      "gender": "Male",
      "bmi": 28.5,
      "HbA1c_level": 6.5,
      "blood_glucose_level": 140,
      "smoking_history": "former",
      "hypertension": 1,
      "heart_disease": 0
    },
    // ... 更多样本
  ]
}
```

- **响应示例:**

```
{
  "total_samples": 2,
  "successful_predictions": 2,
  "results": [
    {
      "index": 0,
      "prediction": 1,
      "prediction_text": "糖尿病",
      "probability": {
        "non_diabetes": 0.2,
        "diabetes": 0.8
      },
      "confidence": 0.8
    },
    // ... 更多结果
  ]
}
```

```
]
}
```

## 4. CSV文件批量预测

通过上传CSV文件进行批量预测。

- **URL:** `/predict_csv`
- **方法:** POST
- **Content-Type:** multipart/form-data
- **参数:**
  - `file`: CSV文件 (必需)
- **CSV文件要求:**
  - 必须包含以下列:
    - age
    - gender
    - bmi
    - HbA1c\_level
    - blood\_glucose\_level
    - smoking\_history
    - hypertension
    - heart\_disease
  - 第一行必须是列名
  - 数据格式必须符合单个预测的要求
- **响应示例:**

```
{
  "total_samples": 100,
  "successful_predictions": 98,
  "failed_predictions": 2,
  "results": [
    {
      "row_index": 0,
      "prediction": 1,
      "prediction_text": "糖尿病",
      "probability": {
        "non_diabetes": 0.2,
        "diabetes": 0.8
      },
      "confidence": 0.8
    },
    // ... 更多结果
  ],
  "errors": [
    {
      "row_index": 98,
      "error": "数据格式错误"
    },
    // ... 更多错误信息
  ]
}
```

# 错误处理

所有接口在发生错误时都会返回适当的HTTP状态码和错误信息：

- 400: 请求参数错误
- 404: 接口不存在
- 500: 服务器内部错误

错误响应格式：

```
{
  "error": "错误描述信息"
}
```

# 数据格式说明

## 字段说明

### 1. age

- 类型：数值
- 范围：0-100
- 说明：患者年龄

### 2. gender

- 类型：字符串
- 可选值：
  - "Male": 男性
  - "Female": 女性
  - "Other": 其他

### 3. bmi

- 类型：数值
- 范围：10-60
- 说明：体重指数

### 4. HbA1c\_level

- 类型：数值
- 说明：糖化血红蛋白水平

### 5. blood\_glucose\_level

- 类型：数值
- 说明：血糖水平

### 6. smoking\_history

- 类型：字符串
- 可选值：
  - "current": 当前吸烟
  - "never": 从不吸烟
  - "former": 以前吸烟
  - "ever": 曾经吸烟
  - "No Info": 无信息

### 7. hypertension

- 类型：整数
- 可选值：

- 0: 无高血压
- 1: 有高血压

#### 8. heart\_disease

- 类型：整数
- 可选值：
  - 0: 无心脏病
  - 1: 有心脏病

## 使用示例

### Python示例

```
import requests
import pandas as pd

# 单个预测
def predict_single(data):
    response = requests.post(
        "http://1.94.9.72:5000/predict",
        json=data
    )
    return response.json()

# CSV批量预测
def predict_csv(csv_file):
    with open(csv_file, 'rb') as f:
        files = {'file': (csv_file, f, 'text/csv')}
        response = requests.post(
            "http://1.94.9.72:5000/predict_csv",
            files=files
        )
        return response.json()

# JSON批量预测
def predict_batch(data_list):
    response = requests.post(
        "http://1.94.9.72:5000/predict_batch",
        json={"data": data_list}
    )
    return response.json()
```

### cURL示例

#### 1. 健康检查：

```
curl http://1.94.9.72:5000/health
```

#### 2. 单个预测：

```
curl -X POST \
  http://1.94.9.72:5000/predict \
  -H "Content-Type: application/json" \
  -d '{
    "age": 45,
    "gender": "Male",
    "bmi": 28.5,
    "HbA1c_level": 6.5,
    "blood_glucose_level": 140,
    "smoking_history": "former",
    "hypertension": 1,
    "heart_disease": 0
  }'
```

### 3. CSV文件批量预测:

```
curl -X POST \
  http://1.94.9.72:5000/predict_csv \
  -F "file=@data.csv"
```

## 注意事项

1. 所有数值型字段必须提供有效的数值，不能为空
2. 字符串字段必须使用规定的可选值
3. CSV文件必须包含所有必需字段，且列名必须完全匹配
4. 建议在进行批量预测时，先使用少量数据测试
5. 如遇到错误，请检查请求数据格式是否正确