

糖尿病预测平台

项目报告

Team 10

2253323 陶子芾

2252835 胡家豪

2253640 黄俊哲

2253320 张茁含睿

目录

- 1.项目概述 4
 - 1.1 项目背景与意义..... 4
 - 1.2 目标用户 4
 - 1.3 项目目标与功能定位 5
- 2.系统架构设计 5
 - 2.1 架构总览 5
 - 2.2 前端展示..... 6
- 3.数据集介绍..... 7
 - 3.1 数据集基本信息..... 7
 - 3.2 字段结构与含义说明 7
 - 3.3 特征工程与预处理策略..... 8
 - 3.4 标签分布与类别不平衡问题..... 9
 - 3.5 特征选择与建模基础 9
 - 3.6 数据集价值与适配性 10
- 4.模型介绍 10
 - 4.1MindSpore 深度学习模型构建与训练..... 10
 - 4.1.1 模型开发背景..... 10
 - 4.1.2 模型结构设计（DiabetesNet） 11
 - 4.1.4 数据处理与加载 11
 - 4.1.5 训练流程与监控 13
 - 4.1.6 模型评估与报告生成..... 13
 - 4.1.7 技术优势与亮点总结 14
 - 4.2XGBoost 模型构建与训练..... 14
 - 4.2.1 数据加载与清洗 14
 - 4.2.2 特征选择与标准化 15

4.2.3 类别不平衡处理 (SMOTE)	15
4.2.4 模型构建与超参数搜索	16
4.2.5 可视化分析	16
4.2.6 模型保存	17
4.2.7 技术总结与亮点	17
4.3 模型性能评判标准	18
4.3.1 准确率 (Accuracy)	18
4.3.2 精确率 (Precision)	19
4.3.3 召回率 (Recall)	19
4.3.4 F1-score (调和均值)	20
4.3.5 指标选择策略总结	20
5.系统部署方案	21
5.1 模型推理服务部署 (云端)	21
5.1.1 云平台环境	21
5.1.2 服务划分	21
5.1.3 网络配置	22
5.2 前端运行环境 (本地)	22
5.3 部署优势与后续可拓展性	22
6.项目创新点总结	23
6.1 多模型融合支持, 提高模型泛化与稳定性	23
6.2 精准控制数据质量, 增强模型可靠性	23
6.3 前后端解耦 + 云端模型服务, 实现高扩展部署架构	23
6.4 指标体系科学, 关注医学风险特性	24
6.5 易用性与可视化并重, 提升用户交互体验	24
6.6 平台具备可移植性与可重用性	24

1.项目概述

1.1 项目背景与意义

糖尿病是一种代谢性慢性疾病，主要表现为血糖水平异常升高，若未能及时发现并干预，将可能引发多种严重并发症（如心脑血管疾病、视网膜病变、肾病等），严重影响患者生活质量。根据世界卫生组织（WHO）统计，全球已有超过4亿人患有糖尿病，其中相当比例患者在患病多年后才被确诊，错过了最佳治疗时机。

我国正处于糖尿病高发阶段，居民健康意识不断增强，公众对疾病的“早预警、早干预”需求日益迫切。然而，传统的糖尿病诊断依赖实验室化验与专业医生判断，效率低、周期长、且对医疗资源依赖强，不适用于快速、大范围的筛查与评估。

在此背景下，人工智能特别是机器学习与深度学习技术，为糖尿病的智能预测提供了新路径。它们可以通过对个人健康数据的学习，建立预测模型，实现无创、快速的患病风险预测，有助于推动疾病筛查从“诊断驱动”向“预测驱动”转变。

本项目基于这一背景，设计并实现一个**基于机器学习与深度学习算法的糖尿病预测平台**，集数据输入、模型预测、交互展示、预测记录管理于一体，旨在实现低门槛、高精度、易扩展的智能预测系统，为慢病管理和健康干预提供信息化支持。

1.2 目标用户

本平台面向多类用户，具有广泛适用性：

- **普通公众与糖尿病高风险人群**

可通过平台录入基础健康信息（如体重、年龄、血压、血糖等）进行自我检测，获取个体化患病风险评分和提示建议，从而提前干预、预防发病。

- **基层医疗机构与健康管理人员**

可借助平台实现患者糖尿病筛查与健康状态跟踪，辅助初步判断，为后续检查提供参考依据，提升基层服务效率。

- **科研人员与数据分析师**

可基于平台提供的接口或导出数据开展模型优化、特征研究与人群分

析，推进糖尿病相关人工智能研究。

- **教育培训机构**

可作为 AI+医疗交叉领域的教学工具，支持学生和研究人员进行人工智能建模与医疗数据分析实验。

1.3 项目目标与功能定位

本项目旨在开发一个以 Vue 前端为基础，集成 XGBoost 与 MindSpore 两种智能预测模型的 Web 平台系统，实现糖尿病风险预测服务的数字化与智能化。具体目标包括：

- **平台化设计：**提供一个用户友好、交互直观的网页系统，支持用户通过图形界面输入健康参数，发起预测请求，并直观查看预测结果。
- **双模型支持：**分别集成 XGBoost（经典机器学习模型）与 MindSpore 实现的深度学习模型，支持用户自主选择预测模型，比较不同算法结果。
- **高可视化分析：**支持预测结果以概率值、等级分类、图形图表等形式展示，如雷达图、条形图等，以增强理解性和可操作性。
- **历史记录管理：**用户可以查看过往预测记录，对自身健康变化进行追踪和比较。
- **模块化架构：**采用前后端分离、模型服务独立部署的架构，为未来扩展到其他疾病预测（如高血压、心脏病等）提供良好基础。

2.系统架构设计

2.1 架构总览

本项目采用**前后端分离 + 模型服务独立部署**的架构设计，以提升系统的可维护性、扩展性和部署灵活性。系统整体可划分为四大子模块：

1. **前端展示层（Vue）：**负责用户交互、数据输入与预测结果展示；

2. **后端服务层**：负责接收前端请求、数据验证、调用模型预测服务、记录历史数据等；
3. **模型服务层 (XGBoost & MindSpore)**：独立部署的机器学习与深度学习推理接口，封装模型加载与预测逻辑；
4. **数据存储层**：负责用户信息、输入记录、预测结果的存储与管理。

2.2 前端展示

糖尿病预测系统
Diabetes Prediction System

系统概览

数据输入

预测结果

模型分析

数据可视化

健康教育

历史记录

首页 / 数据输入

模型运行正常 | 新建预测 | 帮助

患者数据输入

请输入患者的基本健康指标进行糖尿病风险评估

* 年龄

45

+

* 性别

男性

女性

其他

身高 (cm)

+

体重 (kg)

+

计算BMI

填入表单

* BMI

28.5

+

* 糖化血红蛋白 (%)

6.5

+

* 血糖水平 (mg/dL)

140

+

* 既往史

former

* 是否患有高血压

是

否

* 是否患有心脏病

是

否

提交预测

清空

填入示例数据

BMI计算器

身高

cm

+

体重

kg

+

参考范围

血糖浓度:

正常: 70-99 mg/dL
糖尿病前期: 100-125 mg/dL
糖尿病: ≥126 mg/dL

血压(舒张压):

正常: <90 mmHg
高血压前期: 90-99 mmHg
高血压: ≥100 mmHg

BMI分类:

偏瘦: <18.5
正常: 18.5-24.9
超重: 25-29.9
肥胖: ≥30

糖尿病预测系统
Diabetes Prediction System

系统概览

数据输入

预测结果

模型分析

数据可视化

健康教育

历史记录

首页 / 预测结果

模型运行正常 | 新建预测 | 帮助

预测结果

基于机器学习模型的糖尿病风险评估结果

健康状态

未检测到糖尿病风险

风险低率

88.6%

风险较高, 建议及时咨询医生

风险等级评估

低风险

0-20%

中风险

20-70%

高风险

70-100%

操作建议

+ 新建预测

+ 保存结果

+ 导出报告

+ 健康指导

预测时间: 2025/6/17 22:25:12

3.数据集介绍

在本项目中，所使用的数据来源于 Kaggle 平台公开发布的结构化医疗数据集 **Diabetes Prediction Dataset**，该数据集广泛用于糖尿病预测与健康风险评估研究，包含大量真实或模拟生成的个体特征信息及其对应的糖尿病患病状态标签。

3.1 数据集基本信息

- 数据集名称:** Diabetes Prediction Dataset
- 数据格式:** CSV（逗号分隔表格）
- 样本总数:** 原始样本约 100,000 条
- 特征数量:** 共包含 8 个输入特征列 + 1 个目标标签列
- 任务类型:** 二分类（0 = 未患糖尿病，1 = 已患糖尿病）
- 应用目标:** 基于个体体征、疾病史与生活习惯信息，预测是否存在糖尿病患病风险。

3.2 字段结构与含义说明

下表总结了原始数据集中的所有字段：

字段名	类型	描述	示例值
gender	类别型	性别，包含 Male、Female、Other	"Male"
age	数值型	年龄（单位：岁）	46
hypertension	二值型	是否患有高血压（0=否，1=是）	1
heart_disease	二值型	是否患有心脏病（0=否，1=是）	0
smoking_history	类别型	吸烟历史，包括 never、current、former 等	"never"

字段名	类型	描述	示例值
bmi	数值型	身体质量指数（体重 / 身高 ² ，单位 kg/m ² ）	28.3
HbA1c_level	数值型	糖化血红蛋白浓度（%）	6.4
blood_glucose_level	数值型	静脉空腹血糖（mg/dL）	145
diabetes	标签	是否患糖尿病（0 = 否，1 = 是）	1

这些特征综合考虑了生理健康状态（如 BMI、HbA1c、血糖）、慢病史（高血压、心脏病）、行为习惯（吸烟）、人口属性（性别、年龄）等，是临床判断糖尿病风险的关键因素。

3.3 特征工程与预处理策略

为了保证数据质量，增强模型对不同变量的学习能力，本项目对原始数据集进行了如下预处理：

（1）缺失值处理

- 所有包含缺失值的记录被剔除；
- 处理后剩余约 **93,000 条有效样本**。

（2）异常值裁剪

- 将 age 限制在 [0, 100] 岁，过滤极端老龄值；
- 将 bmi 限制在 [10, 60] 区间，避免极端体型对模型产生误导。

（3）类别变量编码

- 对 gender, smoking_history, age_group 等离散变量使用 LabelEncoder 编码为整数；
- 构造衍生变量 age_group（年龄段划分），将 age 分为四类：young（≤30）、middle_aged（31-45）、senior（46-60）、elderly（>60）。

（4）数值特征标准化

- 使用 StandardScaler 对数值型特征（如 bmi, HbA1c_level,

blood_glucose_level) 进行 Z-score 标准化处理;

- 使不同量纲的变量具有相似分布, 加快模型收敛速度并提升性能稳定性。

3.4 标签分布与类别不平衡问题

在原始数据中, 糖尿病患者 (label = 1) 约占全部样本的 8%~12%, 存在显著的类别不平衡问题:

类别	含义	样本数量	占比
----	----	------	----

0	非糖尿病	~84,000	~90%
---	------	---------	------

1	糖尿病患者	~9,000	~10%
---	-------	--------	------

这种不平衡容易导致模型过度偏向多数类 (非糖尿病), 忽视关键的少数类 (真实患者), 从而在医学场景中带来较大的漏诊风险。因此, 在训练阶段使用 **SMOTE** (合成少数类过采样技术) 生成等量的少数类样本, 使模型更均衡地学习两类分布。

3.5 特征选择与建模基础

根据医学研究与特征相关性分析, 最终选定以下特征作为模型训练的输入变量:

- age, bmi, HbA1c_level, blood_glucose_level
- hypertension, heart_disease
- gender_encoded, smoking_history_encoded, age_group_encoded

这些特征具有以下优势:

- 与糖尿病高度相关, 具备生理基础;
- 易于收集和量化, 便于线上部署与问卷录入;
- 同时包含静态特征 (如性别、慢病史) 与动态特征 (如血糖、HbA1c), 利于综合建模。

3.6 数据集价值与适配性

该数据集作为平台的训练与评估数据基础，具备以下优势：

- **特征维度合理**：不多不少，便于快速建模与可视化；
- **变量代表性强**：贴近糖尿病早期风险评估场景；
- **兼容性高**：适用于 XGBoost、深度学习、逻辑回归等多种建模方式；
- **可拓展性好**：可在后续集成更多慢病或个性化变量；
- **高可解释性**：各特征对模型输出的影响可被追踪与分析。

4.模型介绍

4.1MindSpore 深度学习模型构建与训练

4.1.1 模型开发背景

为提升糖尿病预测的非线性建模能力，本平台基于华为自研的深度学习框架 **MindSpore** 开发了一套多层感知机（MLP）结构的神经网络模型，支持可调激活函数、Dropout 防过拟合、早停机制、日志与报告自动记录，具有良好的训练稳定性与上线部署能力。

4.1.2 模型结构设计 (DiabetesNet)

DiabetesNet 是一个多层前馈神经网络 (Feedforward Neural Network)，主要特点如下：

层级	类型	参数说明
输入层	Dense	输入特征维度自动推断 (如 8维)
隐藏层1	Dense + 激活函数 + Dropout	神经元个数默认为 64
隐藏层2	同上	默认神经元数为 32
隐藏层3	同上	默认神经元数为 16
输出层	Dense + Sigmoid	输出单个值 (预测为糖尿病的概率)

4.1.3 模型参数配置

模型相关参数通过命令行传入并统一管理，支持灵活定制训练过程。

参数	含义	默认值
--hidden_sizes	隐藏层维度列表	[64, 32, 16]
--dropout_rate	Dropout概率	0.2
--activation	激活函数类型	'relu'
--epochs	训练轮数	100
--batch_size	每批样本数	32
--learning_rate	初始学习率	0.001
--early_stopping_patience	验证指标无提升的最大等待轮数	10

所有参数通过 argparse 管理，便于命令行运行和自动化脚本调用。

4.1.4 数据处理与加载

由 data/data_processor.py 提供接口 prepare_data_for_training 完成数据预处理与加载：

- 数据格式：CSV 文件，包含若干糖尿病相关特征（如 BMI、血糖、年龄

等)与目标标签;

- 数据处理流程包括:
 - 缺失值处理
 - 特征归一化(如 Min-Max 归一)
 - 划分数据集为训练集 / 验证集 / 测试集
 - 使用 MindSpore 数据集接口构建 GeneratorDataset 对象

由于在糖尿病预测任务中,数据集中往往存在**类别不平衡**的问题:

- 非糖尿病(负类)样本数量远多于糖尿病(正类)样本。
- 如果直接用不平衡数据训练模型,模型可能只学会预测大多数类别(非糖尿病),从而导致漏诊。

在未处理不平衡之前,模型运行分数如下:

```
验证集指标:  
accuracy: 0.9690  
precision: 0.9737  
recall: 0.6529  
f1: 0.7817  
  
测试集指标:  
accuracy: 0.9682  
precision: 0.9667  
recall: 0.6482  
f1: 0.7761
```

在处理不平衡后,对应结果为:

```
📁 保存最终模型...  
📁 保存模型到: ./checkpoints/diabetes_model.ckpt  
✅ 模型保存完成!  
  
📁 最终模型评估...  
验证集指标:  
accuracy: 0.8985  
precision: 0.8997  
recall: 0.8971  
f1: 0.8984  
  
测试集指标:  
accuracy: 0.9007  
precision: 0.8932  
recall: 0.9103  
f1: 0.9017  
📄 训练报告已保存: ./logs/training_report_20250617_211740.txt
```

可以看到得到了显著提升。

4.1.5 训练流程与监控

训练流程由 `train_model()` 函数驱动：

- **环境配置：**通过 `setup_environment()` 设置设备类型（CPU/GPU/Ascend）、随机种子、目录创建；
- **模型实例化：**调用 `DiabetesNet(config)` 构造模型实例；
- **训练器创建：**调用 `create_trainer()`，内部封装优化器（Adam）、损失函数、监控回调；
- **回调机制：**
 - `LossMonitor`：实时打印 Loss 值；
 - `TimeMonitor`：每轮耗时记录；
 - `ValidationCallback`：定期评估验证集指标，并执行早停判断；
- **训练控制：**
 - 用户可随时中断（支持 `KeyboardInterrupt`）；
 - 每轮训练完成后保存日志与验证评估结果；
 - 最终保存最优模型权重与结构；

4.1.6 模型评估与报告生成

训练完成后将调用 `evaluate_model()` 分别在验证集与测试集上评估指标，包括但不限于：

- Accuracy（准确率）
- Precision（精确率）
- Recall（召回率）
- F1-score
- AUC（曲线下面积）

最终调用 `save_training_report()` 生成详细训练报告文本文件，内容包括：

- 训练时间
- 所用设备
- 模型结构与参数

- 验证集、测试集指标汇总

保存路径如：./logs/training_report_20250616_212345.txt

4.1.7 技术优势与亮点总结

1. **结构灵活**：支持任意层数与隐藏层尺寸组合
2. **可移植性强**：可运行在 CPU、GPU 与 Ascend 芯片上
3. **可复现性高**：设定随机种子，输出一致
4. **完整监控**：训练日志、验证指标、模型结构统一记录
5. **自动报告生成**：方便性能记录与项目审阅
6. **支持早停机制**：有效避免过拟合

4.2 XGBoost 模型构建与训练

为提升模型的泛化能力与决策透明性，项目引入了 **XGBoost (eXtreme Gradient Boosting)**，作为传统机器学习模型对比方案。XGBoost 具有结构清晰、效率高、可解释性强等优点，尤其适用于结构化医疗数据的分类任务，在本项目中用于预测个体是否存在糖尿病患病风险。

4.2.1 数据加载与清洗

模型输入数据来自 diabetes_prediction_dataset.csv，该数据集为结构化表格数据，包含以下关键字段：

- **基本特征**：年龄 (age)、BMI (bmi)、糖化血红蛋白水平 (HbA1c_level)、血糖水平 (blood_glucose_level)；
- **疾病特征**：是否患有高血压 (hypertension)、心脏病 (heart_disease)；
- **行为特征**：吸烟史 (smoking_history)、性别 (gender)；
- **目标变量**：是否患有糖尿病 (diabetes)，为二分类标签。

处理流程包括

步骤	内容
空值处理	删除所有缺失值行
极端值处理	对年龄进行上限裁剪（最大值为100），BMI裁剪范围10~60
年龄分组	年龄被划分为四个阶段：young, middle_aged, senior, elderly
类别编码	使用 LabelEncoder 将 gender、smoking_history 和 age_group 编码为整数

4.2.2 特征选择与标准化

选取 9 个特征作为模型输入：

1. 'age'
2. 'hypertension'
3. 'heart_disease'
4. 'bmi',
5. 'HbA1c_level'
6. 'blood_glucose_level',
7. 'gender_encoded'
8. 'smoking_history_encoded'
9. 'age_group_encoded'

使用 StandardScaler 对所有数值型特征进行归一化处理，使其具有零均值和单位方差，确保模型训练收敛更快、效果更稳定。

4.2.3 类别不平衡处理（SMOTE）

由于原始数据中阳性（糖尿病）样本占比较低，为避免模型对多数类过拟合，引入 SMOTE（Synthetic Minority Oversampling Technique）方法对少数类进行过采样。

在未引入时，对应的

结果示例

原始分布：{0: 16000, 1: 4000}

SMOTE 后分布：{0: 16000, 1: 16000}

有效提升了模型在正类样本上的识别能力。

4.2.4 模型构建与超参数搜索

使用 XGBClassifier 构建基础模型，并通过 RandomizedSearchCV 进行超参数优化，搜索空间如下：

参数名称	范围说明
n_estimators	[50, 200], 树的数量
max_depth	[3, 7], 树的最大深度
learning_rate	0.05 - 0.2, 学习率
subsample	0.7 - 1.0, 样本采样比例
colsample_bytree	0.7 - 1.0, 特征采样比例

交叉验证轮数为 3，评分标准使用 F1-score，最终输出最佳模型及参数组合。

输出示例：

```
最佳参数: {
    'n_estimators': 120,
    'max_depth': 5,
    'learning_rate': 0.11,
    'subsample': 0.81,
    'colsample_bytree': 0.94
}
```

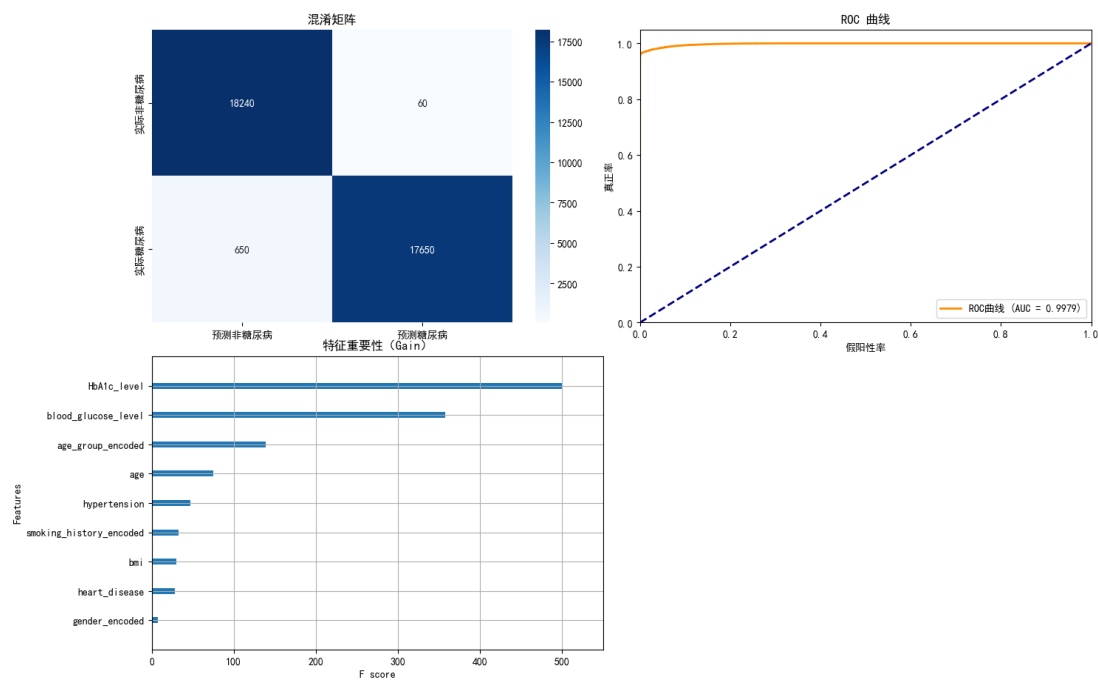
最佳交叉验证 F1 分数: 0.8541

4.2.5 可视化分析

训练完成后进行结果可视化：

图表类型	说明
混淆矩阵	显示真实与预测标签的分布
ROC曲线	展示模型在不同阈值下的性能表现
特征重要性图	使用 Gain 作为重要性衡量指标，解释模型决策依据

可视化结果如下：



4.2.6 模型保存

模型训练完成后，使用 `joblib` 和 `xgboost` 原生方法保存模型：

```
joblib.dump(best_xgb_model, "xgboost_model.pkl")
```

```
best_xgb_model.save_model("xgboost_model.json")
```

便于后续平台部署或模型推理 API 接入。

4.2.7 技术总结与亮点

1. 高效建模：XGBoost 对结构化数据效果优异，计算效率高；
2. 自动调参：结合 `RandomizedSearchCV`，实现高效超参数优化；
3. 可解释性强：通过特征重要性图可清晰展现模型推理依据；

4. 跨平台易部署：模型支持 .pkl 和 .json 格式保存，适配前端服务部署。

4.3 模型性能评判标准

为了全面评估糖尿病预测平台中所集成的两种模型（XGBoost 与 MindSpore）在实际应用中的性能表现，项目采用了四项核心评估指标，包括**准确率**

（Accuracy）、**精确率**（Precision）、**召回率**（Recall）和**F1-score**。这些指标共同构成了模型能力评估的基础框架，涵盖整体正确性、正例识别能力、误报与漏报的平衡关系等多个方面，尤其适用于医疗领域中高度敏感的分类问题。

4.3.1 准确率（Accuracy）

概念定义

准确率表示模型预测正确的样本数量在总预测样本中的比例，是最常用的整体性能衡量指标之一。

计算公式

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

其中：

- TP：真正例（实际为糖尿病，预测也为糖尿病）
- TN：真反例（实际为非糖尿病，预测也为非糖尿病）
- FP：假正例（实际为非糖尿病，预测为糖尿病）
- FN：假负例（实际为糖尿病，预测为非糖尿病）

医学场景解释

准确率高，意味着模型整体判断正确率较高。然而，在糖尿病预测这类类别高度不平衡的任务中，准确率可能掩盖模型对关键阳性（糖尿病患者）的识别不足。

应用价值

适合作为一个总体性能的参考指标，但不应作为唯一依据，尤其在医疗筛查任务中，模型对“阳性样本”的识别效果（召回率）更加关键。

4.3.2 精确率 (Precision)

概念定义

精确率衡量的是模型预测为糖尿病的所有样本中，实际患有糖尿病的比例，反映“预测为阳性”的可信度。

计算公式

$$\text{Precision} = \frac{TP}{TP + FP}$$

医学场景解释

若一个模型精确率高，说明其误报率低，即不会轻易误判健康人群为糖尿病患者，从而减少资源浪费与患者焦虑。

应用价值

适用于控制误诊风险，尤其在医疗服务资源紧张时，建议选择高精确率模型以优先处理更可能为真实阳性的样本。

4.3.3 召回率 (Recall)

概念定义

召回率衡量的是所有真实糖尿病患者中，被模型成功识别出来的比例，反映模型的“检出能力”。

计算公式

$$\text{Recall} = \frac{TP}{TP + FN}$$

医学场景解释

高召回率意味着模型几乎不会漏掉糖尿病患者，对于疾病早筛与预防具有重要意义。

应用价值

在糖尿病预测场景中，召回率尤为重要。漏诊一个真实患者的代价远高于误判一个健康人。因此，对于高敏感任务，召回率是核心评估指标之一。

4.3.4 F1-score（调和均值）

概念定义

F1-score 是精确率与召回率的调和平均值，综合考量了模型的误报与漏报性能，特别适用于类别不平衡问题。

计算公式

$$F1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

医学场景解释

F1-score 高表示模型在不牺牲召回率的前提下，也具备良好的精确率，适合医疗类问题中对“假阳性”和“假阴性”都敏感的情况。

应用价值

本项目中，F1-score 被设为模型调优与交叉验证的主评估指标。它在平衡识别能力与误报控制之间提供了一种客观量化方式，是最终模型比较和选择的主要依据。

4.3.5 指标选择策略总结

评估指标	衡量维度	医学意义	项目中用途
Accuracy	总体判断正确性	粗略评估整体效果	辅助参考
Precision	预测为糖尿病的准确性	降低误诊率	控制误报
Recall	对真实糖尿病的检出能力	降低漏诊率	主要关注指标
F1-score	精确率与召回率的平衡	诊断效能综合评估	模型选择核心指标

本平台统一采用上述四项指标作为评估标准，对比 MindSpore 与 XGBoost 在不同数据切片与阈值下的性能表现，确保模型不仅具有优秀的算法指标，同时具备可解释性、实用性与医学适配性。

5.系统部署方案

为了实现糖尿病预测平台的高效、稳定运行，本项目采用了“前后端分离 + 模型远程推理”的部署架构。前端界面部署于本地终端（如开发电脑或校园网内服务器），两个核心预测模型（XGBoost 和 MindSpore）分别部署于 **华为云弹性云服务器（ECS）** 上，作为独立的模型推理服务提供预测结果。

该部署结构兼顾了本地开发的灵活性与云端计算的稳定性和可扩展性，为后续系统扩展与多用户接入提供了良好基础。

5.1 模型推理服务部署（云端）

5.1.1 云平台环境

- 云服务商：华为云（Huawei Cloud）
- 服务类型：弹性云服务器（ECS）
- 系统环境：Ubuntu 20.04
- Python 环境：3.9 + pip + virtualenv
- 部署工具：Gunicorn + Flask/FastAPI + Nginx（可选）

5.1.2 服务划分

- **XGBoost 模型服务**
 - 模型加载：xgboost_model.pkl / xgboost_model.json
 - 接口路径：/api/predict/xgboost
 - 返回内容：预测概率 + 是否糖尿病
- **MindSpore 模型服务**
 - 模型加载：.ckpt 文件 + 配套网络结构
 - 接口路径：/api/predict/mindspore
 - 返回内容：预测概率 + 是否糖尿病

5.1.3 网络配置

- 所有服务通过公网 IP 暴露端口，或配置安全组以供前端访问
- 服务端口统一规范（如 8500 / 8600），保证系统通信一致性
- 若需支持 HTTPS，可通过华为云 SSL 服务配置证书并由 Nginx 提供加密通道

5.2 前端运行环境（本地）

前端采用 **Vue.js 3 + Vite** 构建的单页应用（SPA），主要部署在本地机器上（可为开发电脑或校内服务器）。

部署方式

- 项目目录打包命令：npm run build
- 本地测试命令：npm run dev
- 可通过 localhost:5173 或局域网 IP 访问前端系统界面

功能说明

- 表单填写：支持录入 BMI、血糖、年龄等健康参数
- 模型选择：切换 XGBoost 或 MindSpore 推理
- 接口调用：向远程云端服务器发起预测请求，接收结果
- 数据展示：以概率 + 标签方式呈现预测结果，并结合图表辅助展示

5.3 部署优势与后续可拓展性

优势	说明
计算解耦	推理计算集中在云端，减轻本地前端负担
易于维护	模型服务独立，便于后续优化和更新
灵活切换	支持模型选择，便于用户体验对比或AB测试
可扩展性	可横向扩展更多模型服务，如支持心脏病预测模块等

6.项目创新点总结

本项目围绕“糖尿病早期风险预测”这一关键公共健康问题，融合了人工智能建模、现代 Web 技术与云计算资源，构建了一套高可用、高精度、易交互的糖尿病预测平台。相较于传统医疗数据应用或单一模型预测工具，本平台在以下六个方面展现了显著的创新性与实用价值：

6.1 多模型融合支持，提高模型泛化与稳定性

平台同时集成了两类性能优异的预测模型：

- **XGBoost**：适合结构化数据，具备优秀的泛化能力与训练效率；
- **MindSpore 深度学习模型**：通过多层神经网络增强非线性关系建模能力，适应复杂特征空间。

用户可自由切换模型进行对比预测，结合实际应用场景选择最合适模型。该设计打破了单一算法依赖的局限性，提升了平台的**灵活性与稳健性**。

6.2 精准控制数据质量，增强模型可靠性

在数据预处理阶段，本项目采取了系统化的数据工程策略：

- 异常值剔除与归一化处理（如 bmi, age, HbA1c_level）；
- 类别变量合理编码，并构造 age_group 等**衍生变量**；
- 应用 SMOTE 技术处理类别不平衡，增强模型对“糖尿病患者”这一少数类的识别能力；
- 使用特征标准化实现对异构数据的统一建模。

通过这一系列细致处理，保障了模型输入的**稳定性、一致性与可解释性**。

6.3 前后端解耦 + 云端模型服务，实现高扩展部署架构

平台采用“**本地前端 + 云端后端模型服务**”的部署结构：

- 所有模型部署于 **华为云弹性服务器**，实现独立推理服务化；

- 前端为本地 Vue SPA 应用，低门槛运行、快速响应；
- 各模型封装为 RESTful API，方便集成与后续多模型并行部署。

该架构支持后期平滑扩展其他疾病预测模块（如高血压、肾病），具备**横向可扩展性**和**异构模型集成能力**。

6.4 指标体系科学，关注医学风险特性

在模型评估中，平台综合使用：

- **准确率**：反映整体正确率；
- **精确率**：控制误诊风险；
- **召回率**：防止漏诊；
- **F1-score**：综合评价指标；

其中，**F1-score 被设为模型主评判标准**，特别适用于类别不平衡的医疗预测任务。平台通过分类报告、混淆矩阵等多维度反馈机制，使模型性能评估更加全面严谨，符合实际临床需求。

6.5 易用性与可视化并重，提升用户交互体验

- 设计简洁清晰的输入表单，用户仅需填写基本健康参数即可获取预测结果；
- 支持预测历史回顾，后续可拓展“健康趋势追踪”功能；
- 适配移动与 PC 端，便于医护人员与普通用户使用。

这种“**可理解 + 可操作**”的界面设计，兼顾医学应用的严肃性与大众使用的便捷性。

6.6 平台具备可移植性与可重用性

- 所有模块采用模块化设计，便于快速替换模型、调整结构；
- 预测服务兼容 REST API 标准，可对接其他前端（如移动端、微信小程序）；
- 支持数据接口导入/导出，方便对接电子健康档案系统或第三方应用；
- 模型保存格式支持 .pkl, .json, .ckpt 等，方便迁移至其他 AI 框架部署。

平台整体具有良好的**通用性、移植性与教育科研价值**，可作为面向慢病筛查的基础模块化工具原型。