

# 2025 数据分析与挖掘 - 期末项目

## 1. 任务目标:

根据提供的数据，建立糖尿病预测模型（二分类）并评估其性能。模型占总分的 80%，其余内容占 20%。

## 2. 数据集

Diabetes prediction dataset: <https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset>.

## 3. 评价指标

请使用 **F1-Score** 评测模型性能。

**[注意]** 鼓励自主开发模型。尽你所能优化开发的模型，模型性能将在以下两种数据中进行最终测试：

（1）公开数据：数据分布及变量与 Diabetes prediction dataset 严格一致。

（2）真实数据：变量与 Diabetes prediction dataset 一致。

以上两种测试数据将会在后续发布。

## 4. 其他

### （1）强制性要求：

- 使用华为云技术（如 MindSpore）。
- 在华为云服务器上部署模型

**注：**如果你无法使用华为技术实现你的想法，请声明。

### （2）加分项（可选）：

- 将预测扩展到其他疾病（如高血压、脂肪肝等）
- 开发互动界面/工具（如 Web App）

### （3）提交指南：

组长必须提交一个名为 *Group\_ID\_Final.zip* 的 zip 文件，其中包含：

- 详细介绍项目的报告文件（PDF 格式，简洁而全面）
- 幻灯片文件（PPT，用于课堂演示）
- 带有 README 文件的项目代码
- 分工文件：使用 txt 文件陈述组员的工作内容
- 补充文件，但应避免大型模型文件

**提交截止日期：** 2025/06/17 23:59

**演示日期：** 2025/06/18

### 重要说明：

- 严禁抄袭和人工智能生成
- 报告可以用中文或英文撰写
- 文档无需过长，简洁明了是关键
- 请将文件提交到 [Canvas](#)
- 请将逾期提交的文件通过电子邮件发送至 1461715017@qq.com

# 2025 Data Analysis and Mining - Final Project

## 1. Task Objective:

Develop a diabetes prediction model (binary classification) and evaluate its performance based on the data provided. The model is worth 80% of the total score and the remaining components are worth 20%.

## 2. Datasets

Diabetes prediction dataset: <https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset>.

## 3. Evaluation Metric

Please use **F1-Score** to evaluate the model performance.

[Note] Self-developed models are encouraged. Optimize the developed model to the best of your ability.

The model performance will be finally tested on the following two types of data:

- (1) Public data: data distribution and variables are strictly consistent with Diabetes prediction dataset.
- (2) Real-world data: the variables are consistent with the Diabetes prediction dataset.

The above two types of test data will be released later.

## 4. Additional Instructions

### (1) Mandatory Requirements:

- Use Huawei Cloud technologies (e.g., MindSpore).
- Deploy the model on Huawei Cloud servers

Note: If you are unable to implement your idea using Huawei's technology, please make a statement.

### (2) Bonus Tasks (Optional):

- Extend predictions to other diseases (e.g., hypertension, fatty liver disease)
- Develop an interactive interface/tool (e.g., Web App)

### (3) Submission Guidelines:

The group leader must submit a *zip* file named *Group\_ID\_Final.zip* containing:

- A report file (PDF, concise but comprehensive) detailing the project
- A slide file (PPT, for in-class presentation)
- Project code with a README file
- Division of labor document: use a txt file to state what work the member has done
- Supplementary files (e.g., training loss plots), but avoid large model files

**Submission Deadline:** 2025/06/17 23:59

**Presentation Date:** 2025/06/18

## Important Notes:

- Plagiarism and AI generation are strictly prohibited.
- The report can be written in Chinese or English.
- The document doesn't need to be overly lengthy—clarity and conciseness are key.
- Please submit your file to [Canvas](#).
- Please email late submissions to 1461715017@qq.com.