2025 Data Analysis and Mining - Final Project

1. Task Objective:

Based on the provided health examination data (including local datasets and publicly available online data), build a diabetes prediction model and evaluate its performance. Finally, we will release several new data points to test the constructed model. The model accounts for 80% of the total score, with the remaining 20% allocated to the PPT presentation and report.

2. Datasets

(1) Local Dataset: Available in the course files on Canvas: final_project_dataset.zip.

(2) Online Datasets: You may collect data through web scraping or use public datasets (e.g., UCI Diabetes Dataset, Kaggle datasets, etc.).

Feel free to leverage any additional data sources you can find to optimize your model.

3. Evaluation Metric

You must use the **F1-Score** to evaluate your model:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}.$$

The final model will be tested on new data points released later.

4. Additional Instructions

(1) Mandatory Requirements:

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

(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
- 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.