ourdiu.com | Shared by Rafsan Bin Habib

Daffodil International University

Department of Computing and Information Systems

Lab Final: Summer-2026

Program: B.Sc. in CIS Course Code: BI334L

Course Title: Data Analysis and Business Modeling Lab

Task: Maternal Health Risk Assessment Total Marks: 40

Submission Type: Individual

Submission Format: Sharing Jupyter Notebook in Kaggle (Handle: MAK) and pdf file on AI Professor

Exam Scenario

You are given the Maternal Health Risk Factors Dataset [1] collected from pregnant patients at General Hospital, Kurigram, Bangladesh. Build a reproducible data-analysis workflow that explores maternal health indicators and develops a model for predicting the binary Risk Level (Low/High).

Dataset Variables

Age, Systolic BP, Diastolic BP, Blood Sugar (BS), Body Temperature, BMI, Previous Complications, Pre-existing Diabetes, Gestational Diabetes, Mental Health, Heart Rate, and Risk Level.

Instructions

- Use the provided maternal-health dataset only; do not substitute another dataset.
- Write all code in Python. The notebook must run from top to bottom without errors.
- For every analysis, show the relevant output and provide a brief interpretation in your own words.
- Use a fixed random_state where applicable so results are reproducible.
- Do not copy numerical findings from the accompanying dataset article; compute results from the supplied data.
- Clearly justify preprocessing and modeling choices. Avoid unnecessary preprocessing.

Marking Rubric

Component	Marks	Assessment focus
Data understanding & quality	6	Inspection, cleaning decisions, observations
Exploratory data analysis	8	Relevant visualizations and interpretation
Statistical analysis	8	Correct tests, hypotheses, computation, interpretation
Preprocessing & modeling	12	Sound pipeline, three classifiers, metrics, comparison
Interpretation & conclusion	6	Risk-factor evidence, critical discussion, conclusion
Total	**40**	

Task 1 — Data Understanding and Quality Assessment [6 marks]

- Load the dataset and report its shape, column names, data types, and target variable. (2 marks)
- Check missing values and duplicate rows. Report what you find and explain any action taken.
- Produce appropriate descriptive statistics and state two initial observations. (2 marks)

ourdiu.com | Shared by Rafsan Bin Habib

Task 2 — Exploratory Data Analysis [8 marks]

- Visualize the distribution of Risk Level and comment on class balance. (2 marks)
- Create suitable plots to examine BMI and Blood Sugar across Risk Level. Interpret the patterns. (2 marks)
- Create a scatter plot for Age versus BMI, distinguishing observations by Risk Level. Interpret it. (2 marks)
- Create a correlation heatmap for appropriate numeric variables and identify two notable relationships.

Task 3 — Statistical Analysis [8 marks]

- Construct a contingency table for Pre-existing Diabetes versus Risk Level and perform a Chi-Square test of independence. State H0/H1 and interpret the p-value at $\alpha = 0.05$. (4 marks)
- Compare BMI between patients with and without Pre-existing Diabetes using an appropriate two-sample test. State H0/H1, report the test statistic and p-value, and interpret the result. (4 marks)

Task 4 — Preprocessing and Predictive Modeling [12 marks]

- Prepare X and y. Apply only necessary encoding/scaling and justify the steps. Use a stratified train/test split. (3 marks)
- Train any three suitable classifiers. At least one must be a linear model and at least one a tree-based model. (4 marks)
- Evaluate all models using Accuracy, Precision, Recall, F1-score, and a Confusion Matrix. Present a comparison table. (3 marks)
- Select the best model and justify your choice, giving particular attention to correctly identifying high-risk cases. (2 marks)

Task 5 — Interpretation and Viva Readiness [6 marks]

- Identify the three most influential or informative risk factors using an appropriate evidence-based method (e.g., coefficients, feature importance, or permutation importance). (3 marks)
- Write a concise conclusion covering the main findings, one limitation, and one practical implication for maternal-health risk assessment. (3 marks)

Academic Integrity and Submission Notes

- This is an individual examination. Identical or substantially copied notebooks may be treated as plagiarism.
- You must be able to explain your code, statistical tests, preprocessing choices, model results, and conclusions during the viva.
- Results shown without executable code or meaningful interpretation may receive reduced marks.

Academic Integrity and Submission Notes

- This is an individual examination. Identical or substantially copied notebooks may be treated as plagiarism.
- You must be able to explain your code, statistical tests, preprocessing choices, model results, and conclusions during the viva.
- Results shown without executable code or meaningful interpretation may receive reduced marks.

References:

[1] Mojumdar, M. U., Sarker, D., Assaduzzaman, M., Shifa, H. A., Sajeeb, M. A. H., Islam, O., Bari, M. S., Alam, M. J., & Chakraborty, N. R. (2025). Maternal health risk factors dataset: Clinical parameters and insights from rural Bangladesh. Data in Brief, 59, 111363. https://doi.org/10.1016/j.dib.2025.111363