**A REPORT On**

**Maternal Health Risk**

Submitted to

**Prof. Debasmita Pradhan**

**SCA, KIIT\_DU**

In Fulfillment of the Requirement for the Completion of

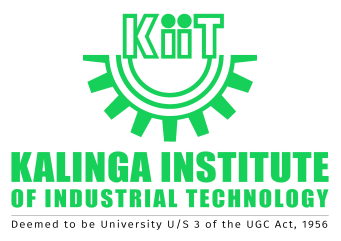
**Big Data Analysis (BDA) PROJECT**

Submitted by

**Aniket Kumar Gupta 2370043**

**Ashutosh Pradhan 2370092**

**Ayush Kumar Mishra 2370104**

****

SCHOOL OF COMPUTER APPLICATION

**KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY**

BHUBANESWAR, ODISHA -751024

7th, November 2024

**Table Of Contents:**

1. **Project Overview.**
2. **Problem Statement.**
3. **Methodologies.**
4. **Data Pre-processing.**
5. **Feature Extraction.**
6. **Model Training and Evaluation.**
7. **Libraries and Prerequisites.**
8. **Result Analysis.**
9. **Conclusion.**

Maternal Health Risk:

Maternal health refers to the health of women during pregnancy, childbirth and the post-natal period. Each stage should be a positive experience, ensuring women and their babies reach their full potential for health and well-being. Although important progress has been made in the last two decades, about 295 000 women died during and following pregnancy and childbirth in 2017. This number is unacceptably high. The most common direct causes of maternal injury and death are excessive blood loss, infection, high blood pressure, unsafe abortion, and obstructed labour, as well as indirect causes such as anemia, malaria, and heart disease. Most maternal deaths are preventable with timely management by a skilled health professional working in a supportive environment. Ending preventable maternal death must remain at the top of the global agenda. At the same time, simply surviving pregnancy and childbirth can never be the marker of successful maternal health care. It is critical to expand efforts reducing maternal injury and disability to promote health and well-being. Every pregnancy and birth is unique. Addressing inequalities that affect health outcomes, especially sexual and reproductive health and rights and gender, is fundamental to ensuring all women have access to respectful and high-quality maternity care.

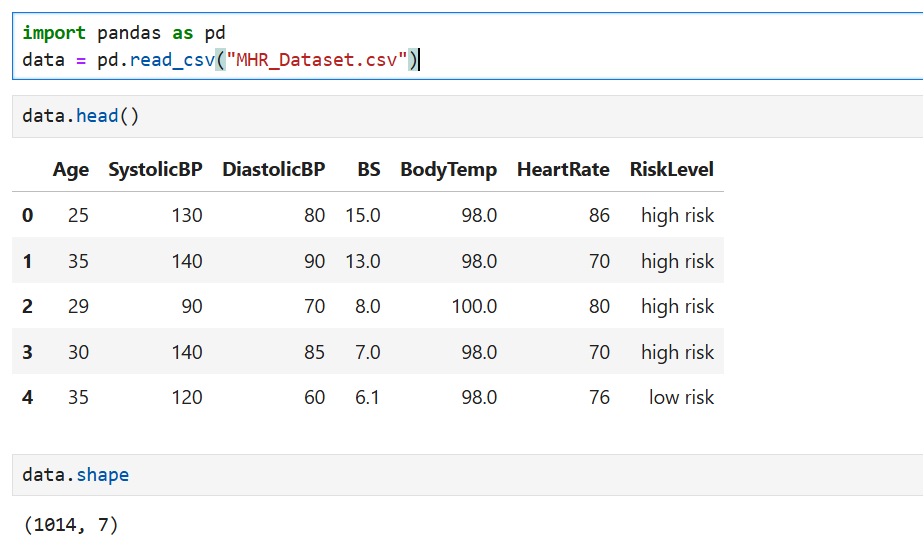
The goal of the project is to apply what has been learned during the MLOps Zoomcamp course to build a MLOps pipeline for woman health risk prediction during pregnancy.

Dataset:

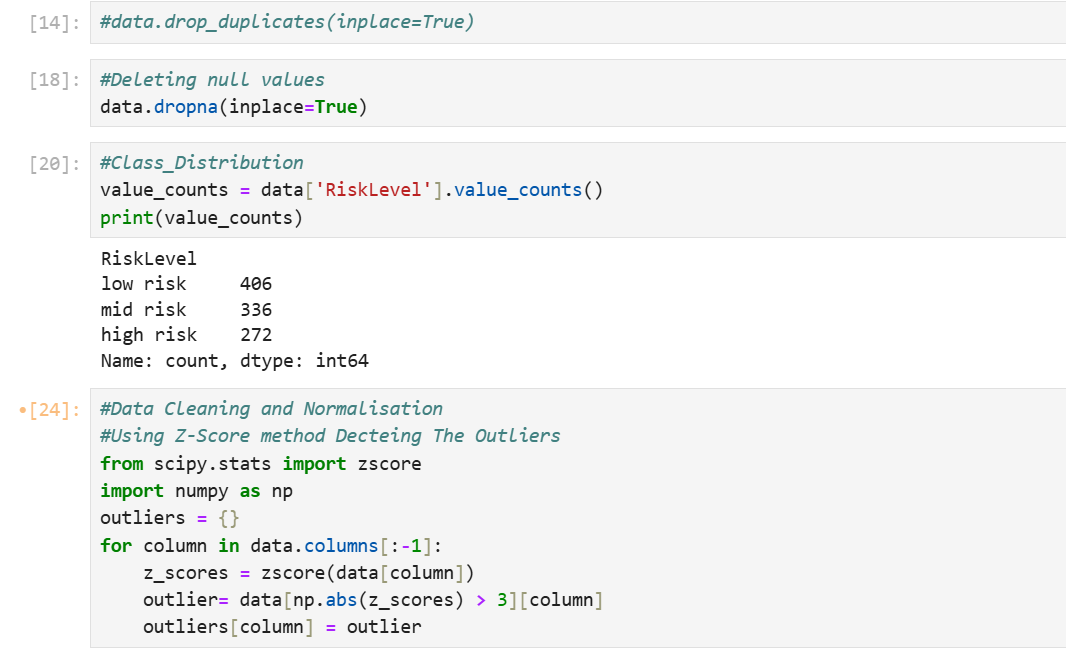
The dataset used to feed the MLOps pipeline has been downloaded from Kaggle and contains data collected from several hospitals, community clinics and maternal health cares through an IoT-based risk monitoring system. The dataset is updated daily and is characterized by the following features:

|  |  |
| --- | --- |
| **Feature** | **Description** |
| Age | Age when a woman is pregnant. |
| SystolicBP | Upper value of blood pressure. |
| DiastolicBP | Lower value of blood pressure. |
| BS | Blood glucose levels in terms of molar concentration. |
| HeartRate | A normal resting heart rate. |
| BodyTemp | Average human body temperature. |
| Risk Level | Predicted risk intensity level during pregnancy considering the previous attributes. |

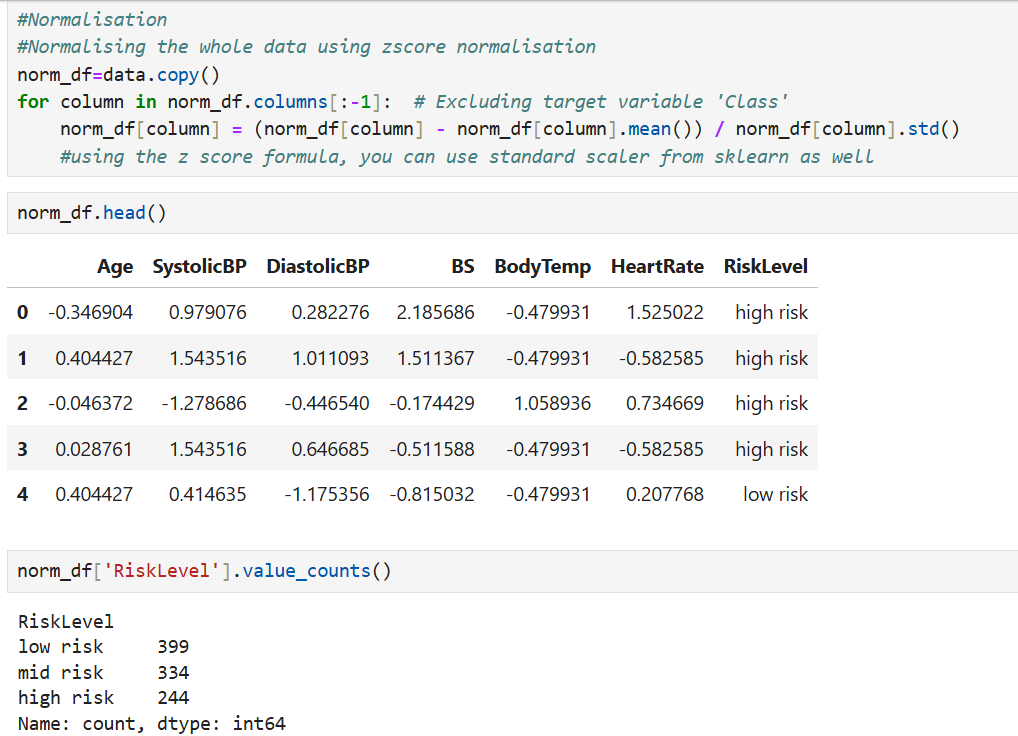
Loading the dataset as a pandas dataframe object:

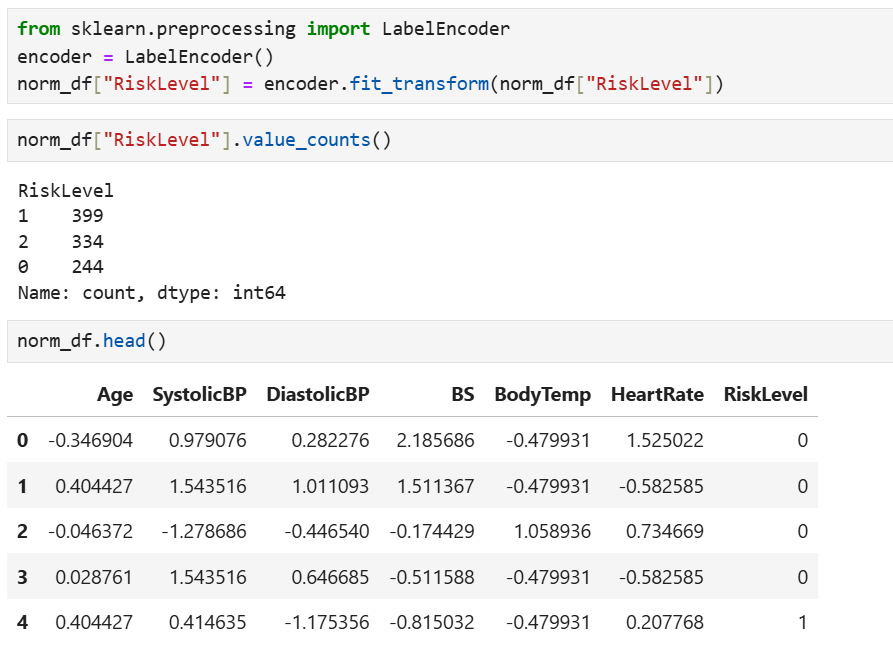


Outlier Analysis and Removal:

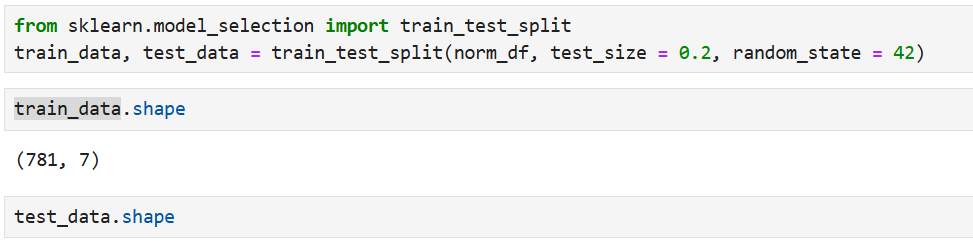


Normalization using Z score method:

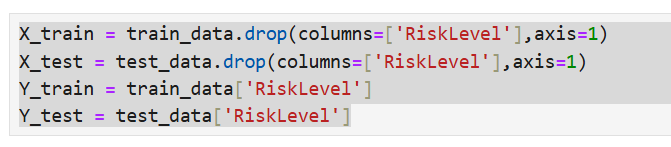




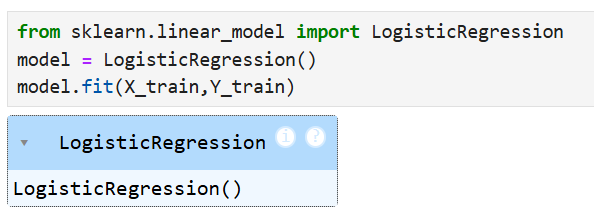
Spilting the data set into train and test data:



Feature matrix and target vector creation:



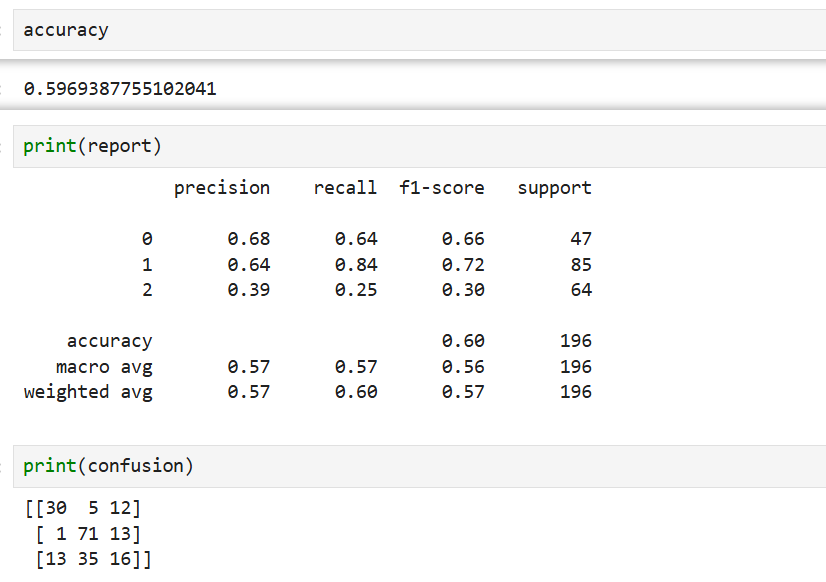
Selecting the model and training the test data set into that model:



Predicting the result:



Model Evalution:



Result Analysis:

* Data Quality: The notebook's success largely depends on data quality, including feature distribution, correlation strength with the target, and presence of outliers.
* Model Performance: The random forest model often performs well in classification tasks with tabular data; it may be the best choice here if it achieves high accuracy and recall. Performance metrics across models should be reviewed to confirm this.
* Feature Importance: Understanding which features contribute most to predictions is critical in maternal health; the random forest model may provide this information via feature importance scores.

Conclusion and Recommendations:

Based on the notebook structure and typical machine learning workflows, here’s a general conclusion:

1. Best Performing Model: The model with the highest accuracy, recall, and F1-score (often a random forest in health prediction models) should be selected for maternal health risk prediction.
2. Feature Insights: The most influential health measurements identified by the model should be noted as they provide actionable insights for healthcare professionals.
3. Model Application: If model metrics are strong (high precision and recall), it could be deployed in clinical settings to identify high-risk maternal cases early, improving patient outcomes.