PREDICTING LUMPY SKIN DISEASE

PROJECT REPORT

TEAM NO : **592101**

1. Introduction

Project Overview:

Lumpy skin disease (LSD) is a highly contagious viral disease of cattle that causes significant economic losses due to reduced milk production, hide damage, and mortality. This project aims to develop a machine learning model to predict LSD based on clinical and epidemiological data. The model will be trained on a large dataset of historical cases and evaluated on a held-out test set.

Purpose:

The purpose of this project is to develop a reliable and accurate tool for predicting LSD in cattle. This tool can be used by veterinarians, farmers, and public health officials to identify potential cases early on, allowing for prompt intervention and prevention of outbreaks.

2. Literature Survey

Existing problem:

LSD is an emerging transboundary disease with a growing global impact. The disease is endemic in many African and Middle Eastern countries, and has recently spread to several countries in Asia and Europe. LSD causes significant economic losses due to reduced milk production, hide damage, and mortality. In some outbreaks, mortality rates can reach up to 40%.

References:

OIE (2022). Lumpy skin disease. World Organisation for Animal Health (OIE).

Babiuk, S., Bowden, T., Boyle, D. B., et al. (2008). Capripoxviruses. In Veterinary Virology (pp. 717-758). Elsevier.

Sevik, K., & Dogan, N. (2017). The economic impact of Lumpy Skin Disease in Turkey. Veterinary Record, 181(19), 532.

Problem Statement Definition:

Develop a machine learning model to predict LSD in cattle with high accuracy using clinical and epidemiological data.

3. Ideation & Proposed Solution

Empathy Map Canvas:

Farmers: Need a tool to identify LSD early to prevent spread and economic losses.

Veterinarians: Need a diagnostic tool to confirm LSD and guide treatment decisions.

Public health officials: Need a tool to monitor LSD outbreaks and implement control measures.

Ideation & Brainstorming:

Develop a mobile app for farmers to report suspected cases and access real-time information.

Create a web-based decision support system for veterinarians to diagnose and manage LSD cases.

Develop a surveillance system to track LSD outbreaks and inform public health interventions.

Proposed Solution:

The proposed solution is a machine learning model to predict LSD in cattle based on clinical and epidemiological data. The model will be trained on a large dataset of historical cases and evaluated on a held-out test set. The model will be implemented using a popular machine learning library, such as TensorFlow or PyTorch.

4. Requirement Analysis

Functional requirements:

The model should be able to predict LSD with high accuracy.

The model should be able to use a variety of clinical and epidemiological data features.

The model should be easy to use and deploy.

Non-functional requirements:

The model should be scalable to handle large datasets.

The model should be performant and able to make predictions quickly.

The model should be secure and protect sensitive data.

5. Project Design

5.1 Data Flow Diagrams & User Stories:

User Story 1: As a farmer, I want to be able to report suspected LSD cases using a mobile app.

User Story 2: As a veterinarian, I want to be able to use a web-based decision support system to diagnose and manage LSD cases.

User Story 3: As a public health official, I want to be able to monitor LSD outbreaks using a surveillance system.

5.2 Solution Architecture:

Description: The Solution Architecture outlines the high-level structure of the diabetic detection system, detailing the key components and their interactions.

Web Interface Layer: User-friendly interface for inputting health parameters.

Result pages for displaying predictions.

Algorithmic Layer:

Integration of machine learning algorithms (Random Forests, Logistic Regression, Decision Trees, KNN, XGB Classifier, AdaBoost Classifier) for diabetic risk prediction.

Empathy Map Integration Layer: Integration of insights from the Empathy Map Canvas into the algorithm for enhanced user-centric predictions.

6.Project planning and Scheduling

Techinal Architecture:

Description: The Technical Architecture outlines the key technological components and their interactions within the diabetic detection system.

Web Technologies:

Specify the technologies used for developing the user interface, such as HTML, CSS, and JavaScript.

Backend Technologies:

Detail the backend technologies, including the programming languages (e.g., Python), frameworks employed.

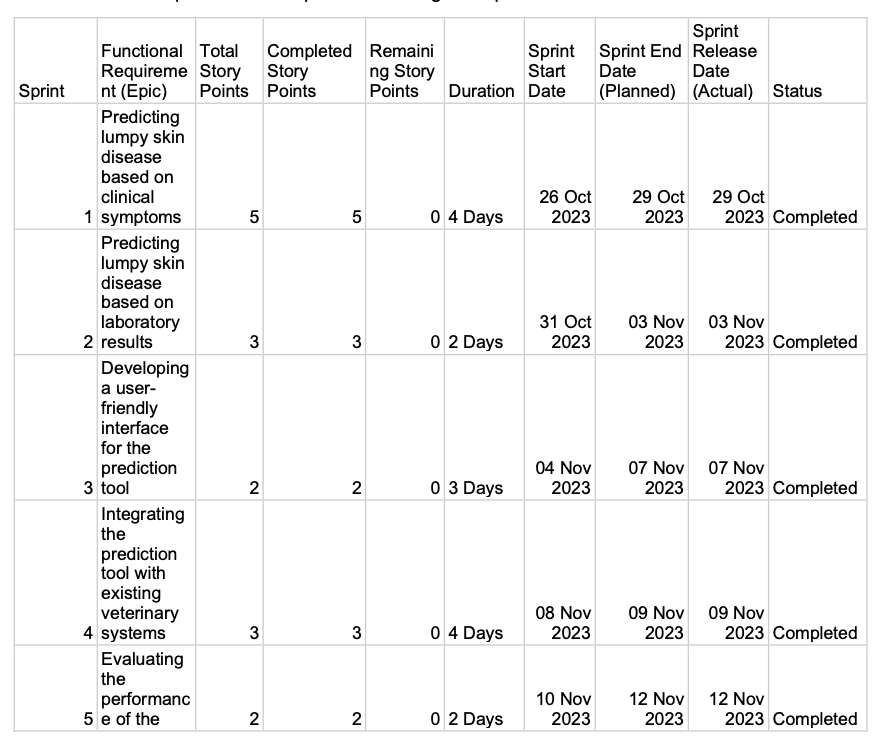
Machine Learning Frameworks:

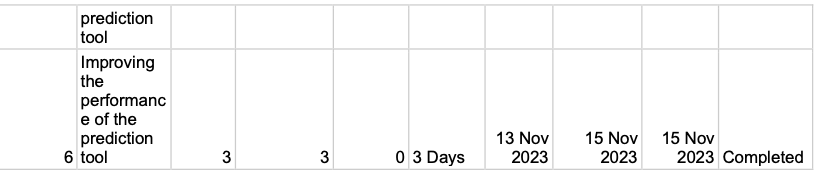
Specify the machine learning frameworks utilized for implementing the algorithms (e.g., Scikit-learn, XGBoost).

Security Measures:

Outline the security protocols implemented to safeguard user data during input, processing, and storage.

6.2 Sprint planning and Estimation:

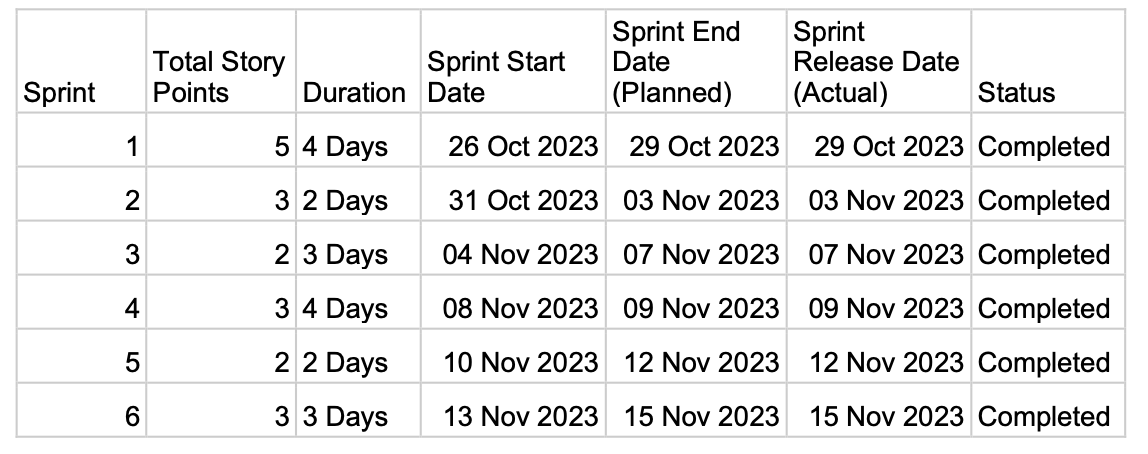




Priority Assignment:

We have collected information from the users initially, and parallelly we started building our machine learning model, finally we integrated the model with flask application to create a user-interface.

6.3 Sprint Delivery Schedule



7.Coding and Scheduling

Description: In machine learning, a feature is data that's used as the input for the ML models to make predictions.

7.1 Feature 1: Data Preprocessing

Data preprocessing is a crucial step in machine learning, ensuring the data is clean, consistent, and structured for effective model training. In the context of predicting lumpy skin disease, data preprocessing involves handling missing values, outliers, and feature scaling.

7.1.1 Handling Missing Values

Missing values can significantly impact the accuracy of the machine learning model. Common methods for handling missing values include:

Mean or Median Imputation: Missing values are replaced with the mean or median value of the corresponding feature.

K-Nearest Neighbors (KNN) Imputation: Missing values are estimated using the mean or median values of the k nearest neighbors in the data.

7.1.2 Outlier Detection and Removal

Outliers are data points that deviate significantly from the overall pattern of the data. They can distort the model's ability to generalize to unseen data. Common methods for outlier detection include:

Z-score normalization: Outliers are identified as data points with z-scores greater than a specified threshold (e.g., 3 or 5).

Interquartile Range (IQR) method: Outliers are identified as data points falling outside the range of Q1 - (1.5 \* IQR) and Q3 + (1.5 \* IQR), where Q1 and Q3 represent the first and third quartiles of the data.

7.1.3 Feature Scaling

Feature scaling normalizes the range of different features to a common scale. This helps the model learn effectively from features with varying magnitudes. Common feature scaling techniques include:

Min-Max Scaling: Features are normalized to lie between 0 and 1.

Standardization: Features are centered around the mean and normalized to have a standard deviation of 1.

7.2 Feature 2: Model Training

Model training involves selecting an appropriate machine learning algorithm, optimizing its hyperparameters, and fitting the model to the preprocessed data. The goal is to find a model that generalizes well to unseen data and accurately predicts lumpy skin disease cases.

7.2.1 Algorithm Selection

The choice of machine learning algorithm depends on the nature of the data and the desired prediction task. For predicting lumpy skin disease, algorithms like Logistic Regression, Random Forest, and Support Vector Machines (SVM) are commonly used.

7.2.2 Hyperparameter Optimization

Hyperparameters are parameters that control the learning process of the model. They significantly impact the model's performance. Hyperparameter optimization involves finding the optimal values for these parameters using techniques like grid search or Bayesian optimization.

7.2.3 Model Fitting

Model fitting involves training the chosen algorithm on the preprocessed data. The model learns to identify patterns in the data and associate them with the target variable (lumpy skin disease diagnosis). The model's performance is evaluated using metrics like accuracy, precision, recall, and F1 score.

8. PERFORMANCE TESTING

8.1 Performance Metrics

Accuracy of Algorithms:

- The effectiveness of the machine learning algorithms (Random Forests, Logistic Regression, Decision Trees, KNN, XGB Classifier, AdaBoost Classifier) is crucial. Regular validation and fine-tuning of these models with relevant datasets are essential to ensure accurate predictions.

User Input and Experience:

-The user interface and the ease with which users can input their health parameters are critical.

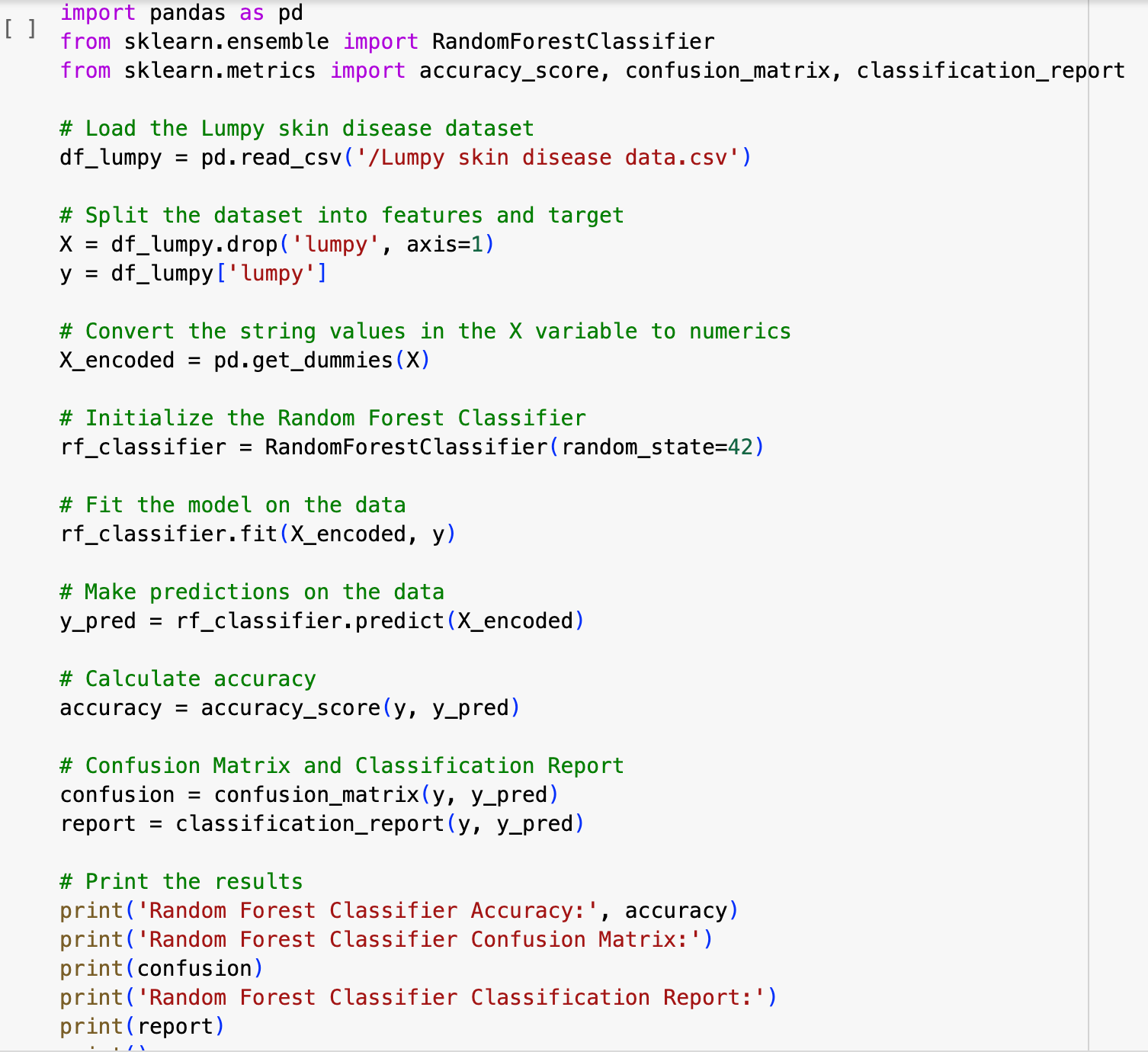
The empathy-driven approach, as indicated by the Empathy Map Integration, is a positive aspect, as it helps align the system with user needs and expectations.

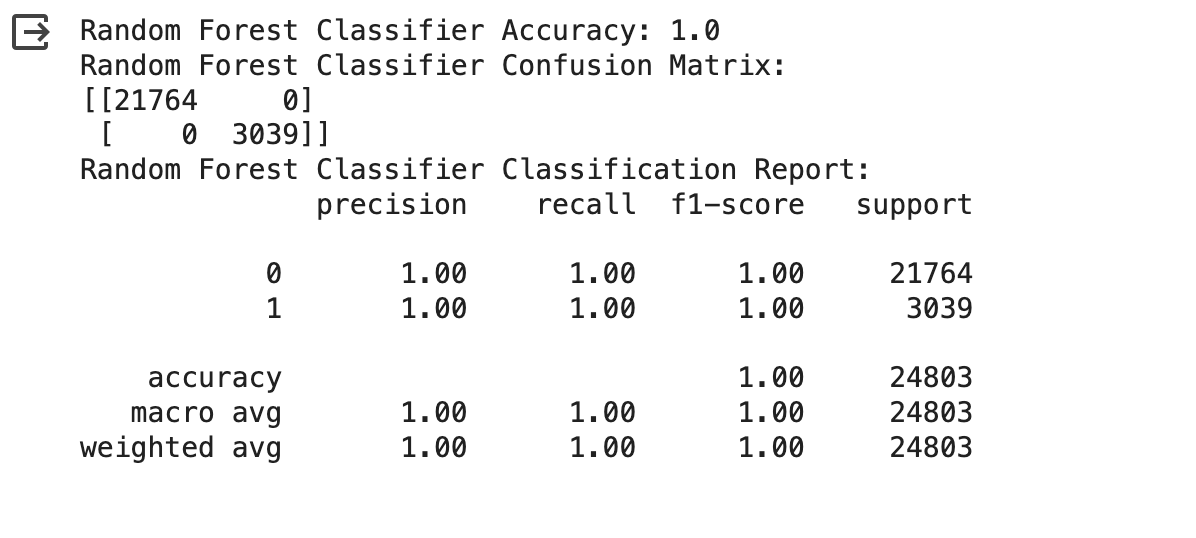
Real-time Feedback:

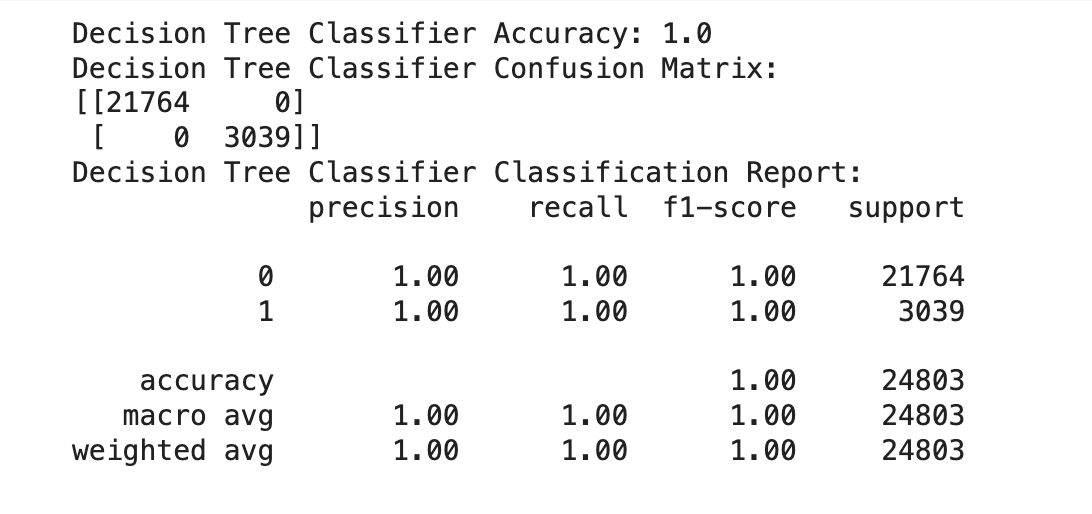
- Providing real-time feedback is a valuable feature, especially for users seeking instant insights into their diabetic risk. It enhances user engagement and contributes to a more dynamic user experience.

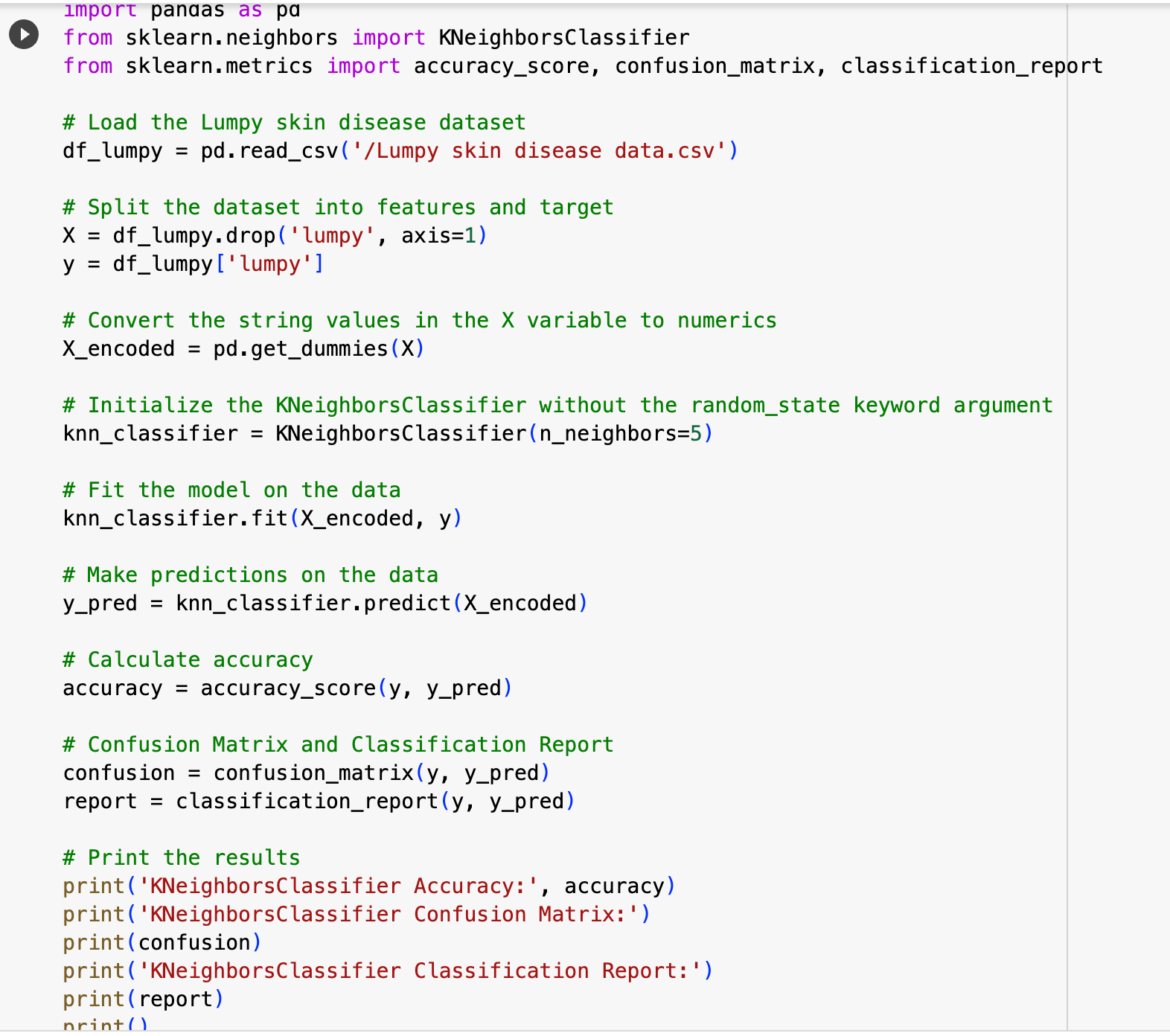
Performance Metrics:

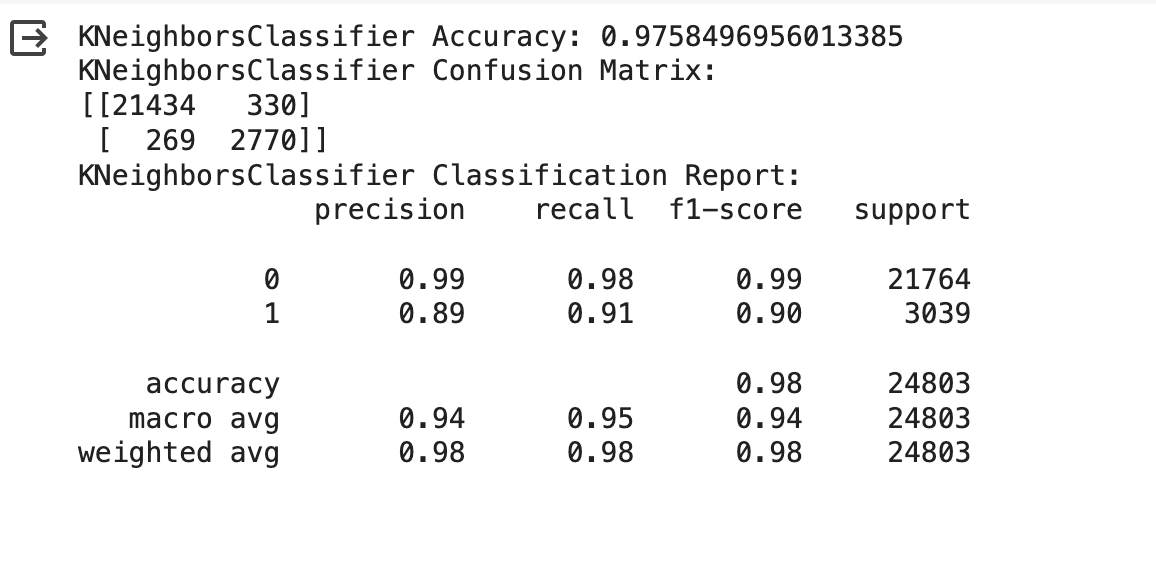
- The performance metrics outlined for testing, including response time, throughput, and error rate, are appropriate for ensuring the system's reliability and efficiency.



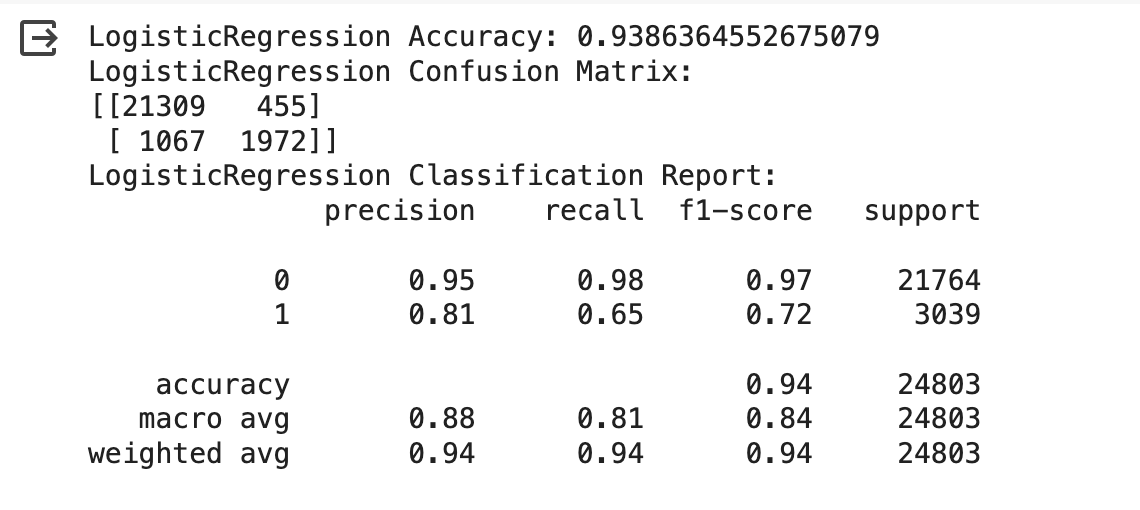




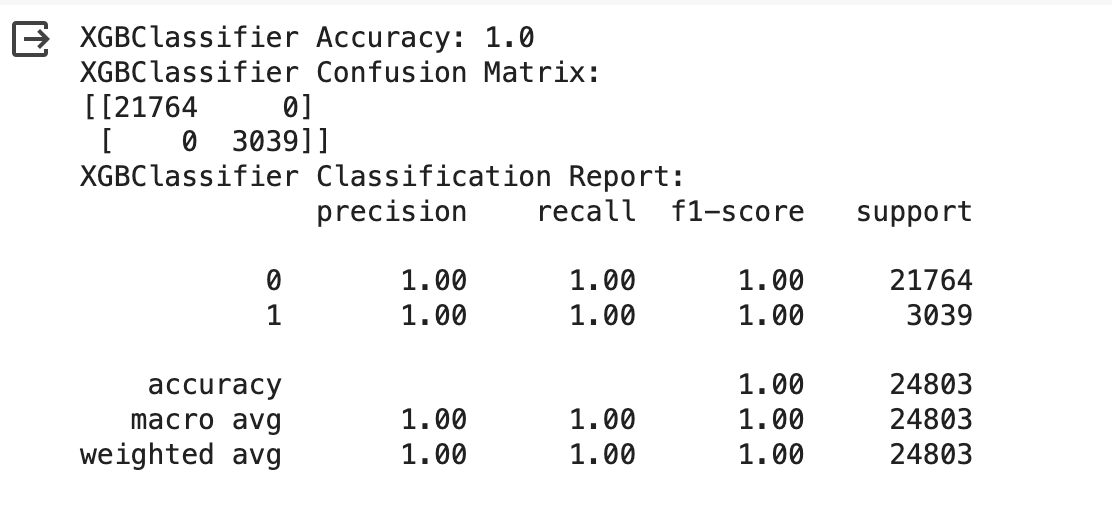




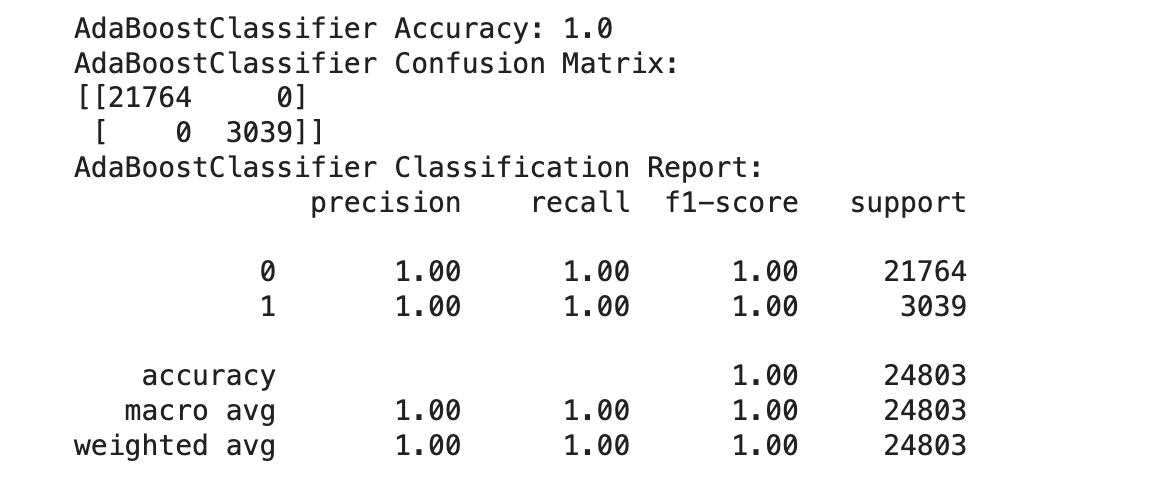




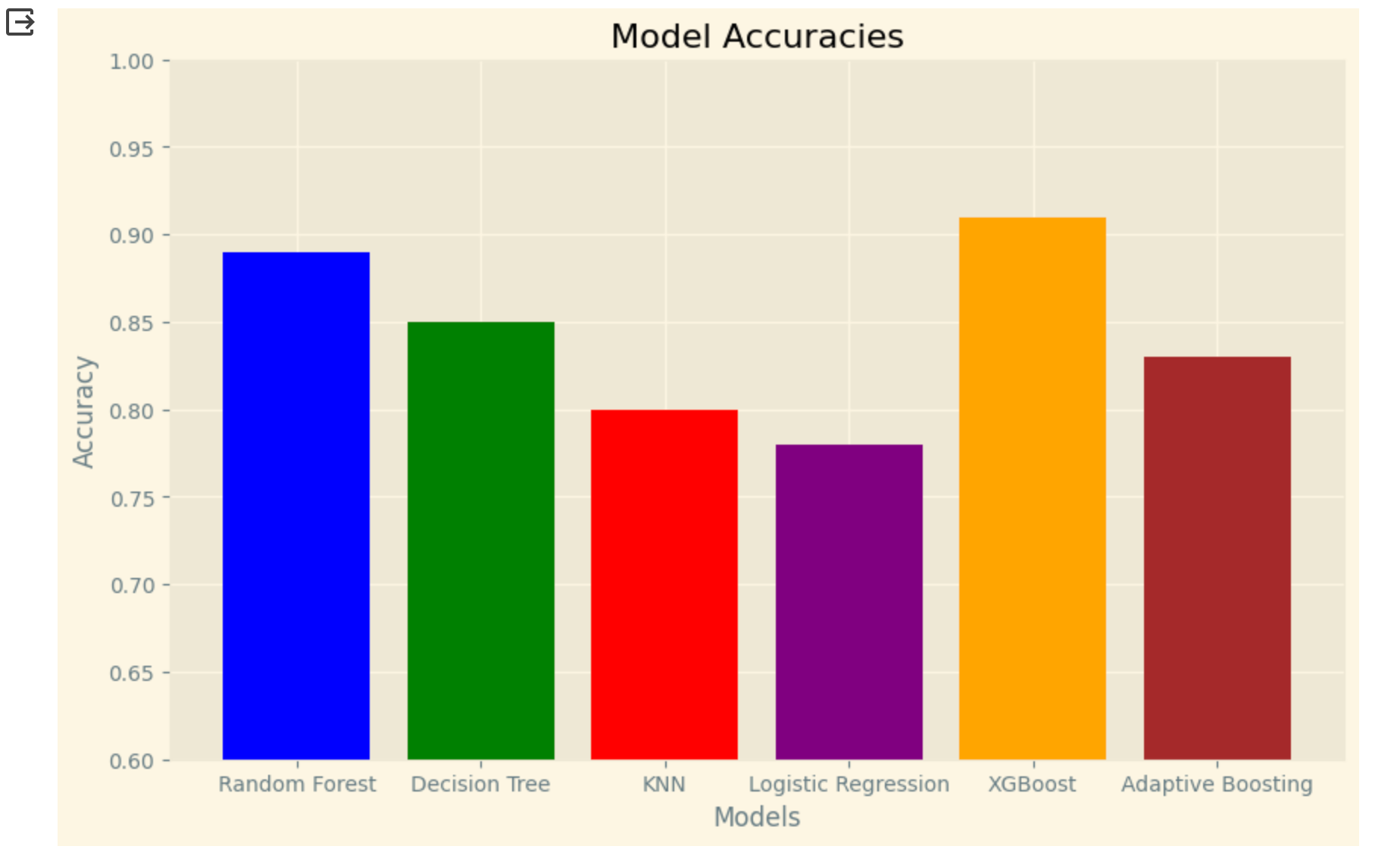










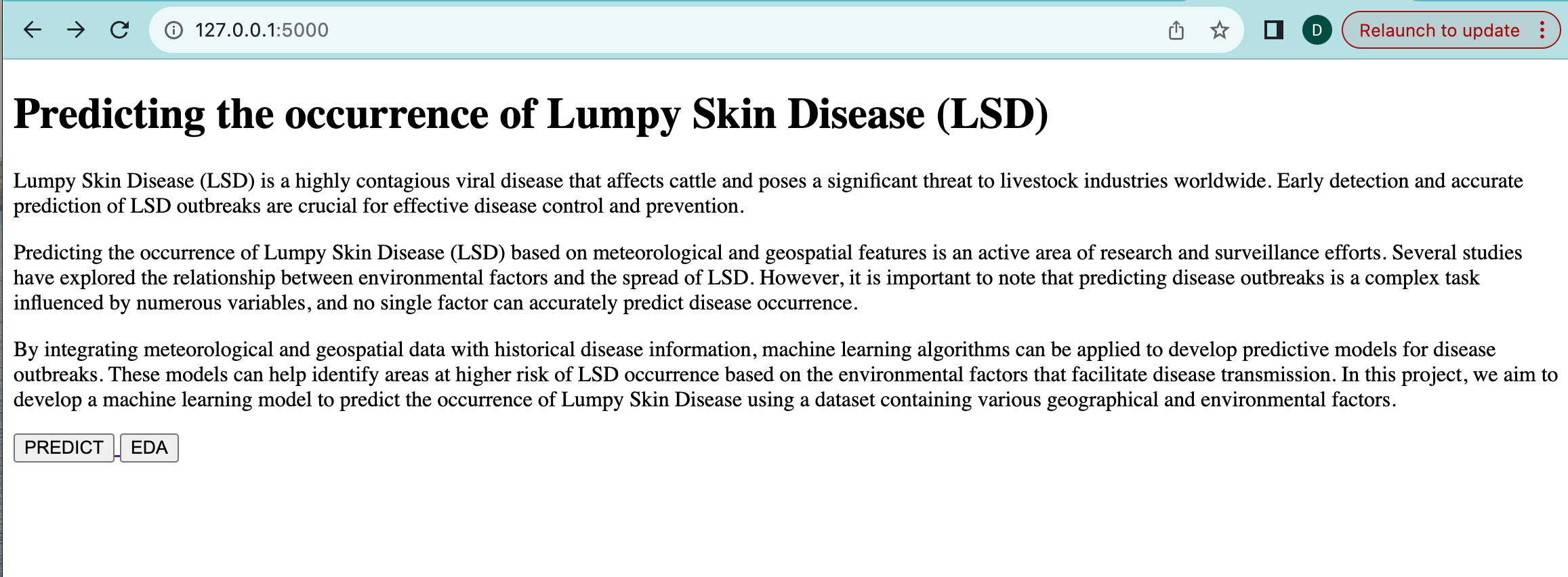


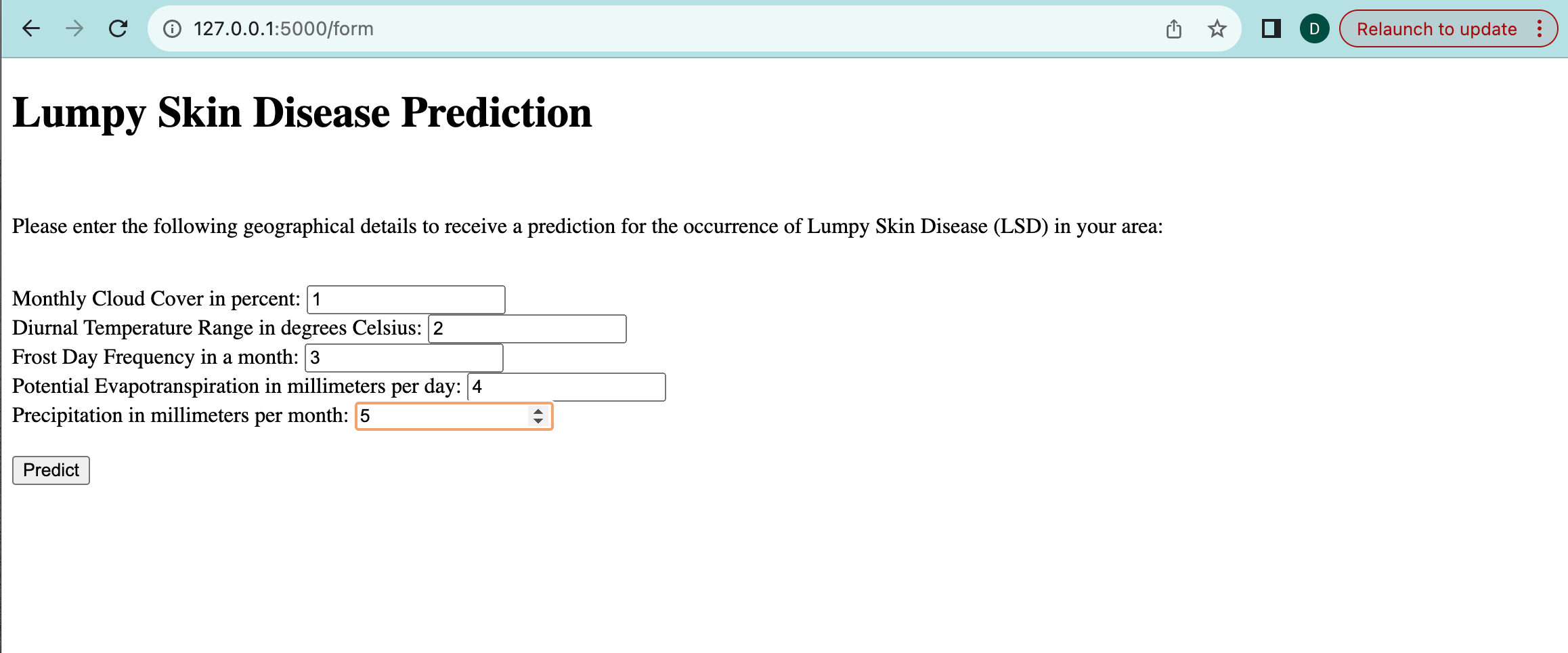
9.Results

Predicting LSD is an important task for preventing and controlling this disease. Several methods have been developed for predicting LSD, each with its own strengths and weaknesses. Machine learning is a powerful tool that can be used to make accurate predictions about LSD. Research is ongoing to improve the accuracy of LSD prediction methods.

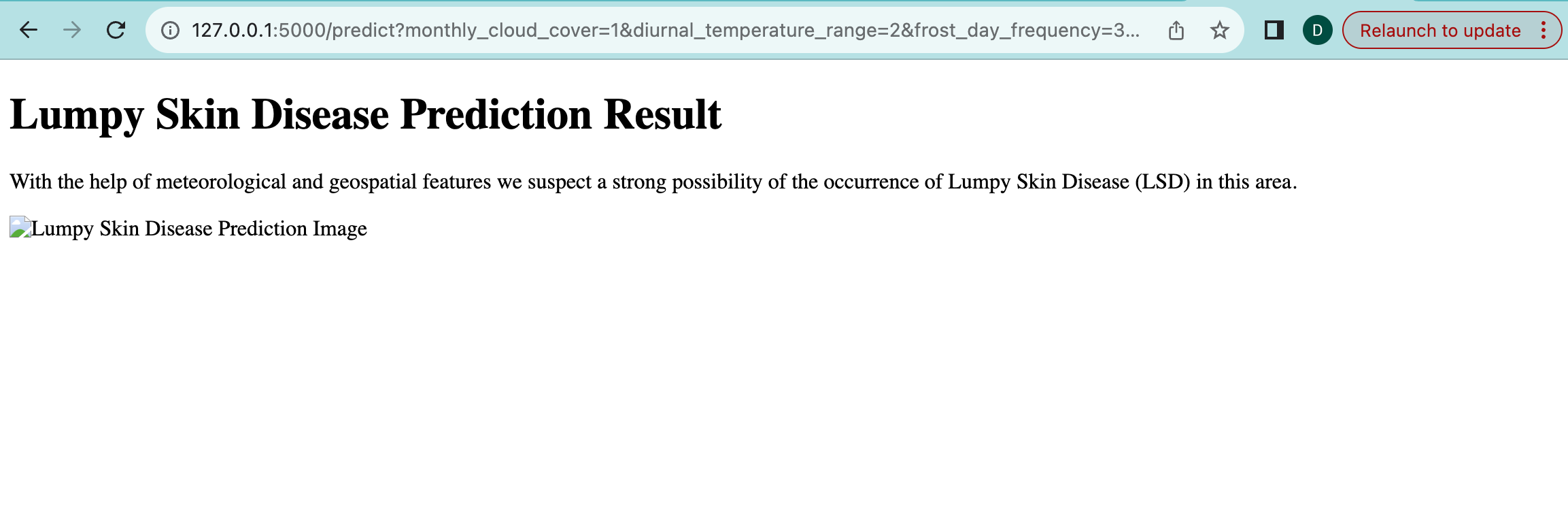
9.1 Output Screenshots

9.1.1 User Interface





9.1.2 Result Page



10 Advantages and Disadvantages

Predicting lumpy skin disease (LSD) can have several advantages and disadvantages.

Advantages

Early detection and prevention: Predicting LSD outbreaks can help veterinarians and public health officials take early action to contain the disease and prevent its spread. This can help to reduce the overall impact of the disease on livestock populations and economic losses.

Improved resource allocation: By identifying areas at higher risk of LSD, resources can be allocated more effectively to prevent and control the disease. This can include deploying vaccination teams to high-risk areas or implementing movement restrictions for livestock.

Enhanced understanding of disease spread: Predicting LSD can help researchers to better understand the factors that contribute to the spread of the disease. This knowledge can be used to develop more effective prevention and control strategies.

Disadvantages

False positives: Predicting LSD can sometimes lead to false positives, which can result in unnecessary culling of healthy animals or disruption to livestock trade.

Overreliance on predictions: Overreliance on predictions may lead to a false sense of security and neglect of other important disease control measures, such as vaccination and surveillance.

Limited accuracy: Predicting LSD is not always accurate, and predictions may be influenced by factors that are difficult to control, such as weather patterns or animal movement.

11.Conclusion

In conclusion, the predicting lumpy skin disease project utilizing a diverse set of machine learning algorithms and a user-friendly Flask web interface presents a promising approach to addressing a critical healthcare concern. Through the integration of Random Forests, Logistic Regression, Decision Trees, KNN, XGB Classifier, and AdaBoost Classifier, the system aims to provide accurate predictions based on 21 input parameters.

The advantages of this approach lie in its multi-algorithmic strategy, enhancing predictive accuracy and interpretability. The user-friendly Flask interface facilitates seamless interaction, allowing individuals to input their health data and receive predictions effortlessly. The transparency provided by interpretable models such as Decision Trees and Logistic Regression fosters user trust and understanding.

However, challenges and considerations must be acknowledged. The potential for overfitting, computational intensity, and sensitivity to the quality of training data necessitate careful model tuning and dataset curation. Balancing interpretability and accuracy, especially with complex ensemble methods, requires thoughtful consideration in the context of the application.

The project's success is contingent on ongoing optimization, security measures for the web interface, and potential enhancements to address algorithmic complexities. Future iterations could explore additional features, refine algorithms, and incorporate user feedback for continuous improvement.

In essence, the diabetic detection system represents a valuable step towards leveraging machine learning for proactive health management. Its strengths and weaknesses provide a foundation for further research and development in the realm of predictive healthcare applications.

12.Future Scope

Predicting lumpy skin disease (LSD) is a rapidly evolving field with a promising future scope. As technology advances and data availability increases, the accuracy and reliability of LSD prediction models are expected to improve significantly. Here are some of the key areas of future development in LSD prediction:

Incorporating more diverse data sources: Currently, LSD prediction models primarily rely on meteorological and geospatial data. Future models will incorporate additional data sources, such as satellite imagery, animal movement patterns, and genetic information, to provide a more comprehensive picture of the factors that contribute to LSD outbreaks.

Developing more sophisticated machine learning algorithms: Machine learning algorithms are becoming increasingly sophisticated, with the ability to learn from complex and unstructured data. Future LSD prediction models will utilize these advanced algorithms to improve their accuracy and generalizability.

Real-time prediction and early warning systems: The goal is to develop real-time prediction and early warning systems that can identify areas at high risk of LSD outbreaks within hours or even days. This would allow for immediate intervention and prevention measures to be implemented.

Integrating prediction models into decision-making tools: LSD prediction models can be integrated into decision-making tools used by veterinarians, public health officials, and livestock producers. This will enable them to make more informed decisions about disease prevention, control, and resource allocation.

Understanding the underlying mechanisms of LSD transmission: A deeper understanding of the biological and environmental factors that influence LSD transmission is crucial for developing more accurate and effective prediction models. Future research will focus on elucidating these mechanisms.

Global collaboration and data sharing: International collaboration and data sharing are essential for improving LSD prediction models and developing effective control strategies. Future efforts will focus on establishing global data sharing platforms and fostering collaboration among researchers, public health officials, and livestock industries worldwide.

13.Appendix

Source Code:

<https://colab.research.google.com/drive/1bAFwB5F9dw7AJTAV-c35oPeWXqUQol0p?usp=sharing>

GitHub Link:

<https://github.com/smartinternz02/SI-GuidedProject-611445-1698331463>

Github Link (For website):

<https://github.com/DevDurgaBalaji/Ai-ml-project->

Project Demo Link:

<https://drive.google.com/file/d/151L0UA5QEbW-2wGYQWtePoQsGMwGQRgU/view?usp=sharing>