DOCUMENTATION - PD150

INTRODUCTION:

What if you could know that you are about to have a heart attack or a deadly disease before hand? Not only would it save your life but also allow you to take **preventive and corrective** measures. In this document, we explore the scope of detecting and preventing fatal and deadly diseases and providing immediate solutions.
Artificial Intelligence and machine learning provides the perfect getaway for achieving such wonders. Using these techniques and algorithms, we have created **machine learning models capable of predicting and detecting** the following diseases :

- Diabetes
- Heart Diseases
- Liver Diseases
- Breast Cancer
- Sepsis Detection
- Skin Cancer
- Skin Lesions
- Pneumonia
- Retina damage detection

ALGORITHMS USED:

Multiple machine learning algorithms have been employed based on the nature of datasets used and implementation factors.

1. **LightGBM**: This is a gradient boosting framework that uses tree based learning algorithms. It is designed to be **distributed and efficient** with the following advantages:
    - Faster training speed and higher efficiency.

- Lower memory usage.
- Better accuracy.
- Support for parallel and GPU learning.
- Capable of handling large-scale data.

2. **MobileNet**:
   MobileNet (Keras) is a general architecture and can be used for multiple use cases. Depending on the use case, it can use **different input layer size and different width factors**. This allows different width models to reduce the number of multiply-adds and thereby reduce inference cost on mobile devices. The number of parameters and number of multiply-adds can be modified by using the `alpha` parameter, which increases/decreases the number of filters in each layer.

3. **DenseNet**:
   This is a convolutional network architecture algorithm.It introduces **direct connections between any two layers** with the same feature-map size. DenseNets scale naturally to hundreds of layers, while exhibiting no optimization difficulties.

4. **Random Forest**:This is a **meta estimator** that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and **control over-fitting**.This is an **ensemble learning method for classification**, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the predicted class.

DEPLOYMENT:

- **Flask** is the primary framework which has been used to deploy the machine learning models on platforms such as **OpenStack** and Heroku.
- In the iOS application, **CoreML** frameworks have been used to implement this on mobile platforms. This is also the foundational machine learning framework from Apple that builds on top of Accelerate, BNNS, and Metal Performance Shaders.

DATASETS USED:

The nature and linearity of the datasets differ vastly due to which different algorithms best suited for each dataset have been used.
Following are the datasets used and their source.

S.No	Disease	Organisation	Origin
1	Diabetes	UCI	Indian
2	Liver	UCI	Indian
3	Kidney	UCI	Indian
4	Pneumonia	Mendeley	Global
5	Skin Lesion	Harvard	Global
6	Heart Disease	UCI	Global
7	Breast Cancer	UCI	Global
8	Skin cancer	ISIC	Global
9	Sepsis	Physionet	Global
10	Retina oct	Mendeley	Global

RESULTS:

The solution designed has the capability to generate results in two formats based on user preferences. The first is the basic format which predicts the probability of having a given disease and the fitness quotient. The second is a much more integrated and detailed result which consists of a report having dedicated columns and probability of all the diseases covered along with the fitness quotient of the user.The first format has been designed for average internet users while the second format of result is useful for commercial applications such as hospitals and pharmacies, however there is no limitation, the user has added feature to choose this format.

CONCLUSION:

Predicting and detecting a disease can save millions of lives and predicting 11 different and deadly diseases has a multifold result.The diseases covered by the application kill around 10 million people every year and also called silent killers meaning the symptoms show too late when the damage is already done. This app eliminates this factor and once detected, early warning measures and precautions may be taken to safeguard lives.