Revolutionizing Liver Care : Predicting Liver Cirrhosis using Advanced Machine Learning Techniques
Revolutionizing Liver Care : Predicting Liver Cirrhosis using Advanced Machine Learning Techniques
Liver cirrhosis is a chronic and progressive liver disease characterized by the irreversible scarring of liver tissue, which can lea
Project Flow:
User interacts with the UI to enter the input.
Entered input is analysed by the model which is integrated.
Once model analyses the input the prediction is showcased on the UI
To accomplish this, we have to complete all the activities listed below,
Define Problem / Problem Understanding
Specify the business problem
? Business requirements
? Literature Survey
? Social or Business Impact.
? Data Collection & Preparation
? Collect the dataset
? Data Preparation
Exploratory Data Analysis
Descriptive statistical
? Visual Analysis
Model Building
Training the model in multiple algorithms
? Testing the model
Performance Testing & Hyperparameter Tuning
Testing model with multiple evaluation metrics
? Comparing model accuracy before & after applying hyperparameter tuning
Model Deployment
Save the best model
? Integrate with Web Framework
Project Demonstration & Documentation
Record explanation Video for project end to end solution
? Project Documentation-Step by step project development procedure
Prior Knowledge:
You must have prior knowledge of following topics to complete this project.
ML Concepts
Supervised learning:
Unsupervised learning:
Decision tree:
Random forest:
KNN:
Xgboost:
Evaluation metrics:
Flask Basics :
Project Structure:
Create the Project folder which contains files as shown below
We are building a flask application which needs HTML pages stored in the templates folder and a python script app.py for scri
Static folder and assets folder contain the CSS and JavaScript files along with images.
rf_acc_68.pkl and normalizer.pkl are our saved models. Further we will use these models for flask integration.
Training folder contains a model training file that are jupyter notebooks.
Here are some Diabetes Results for an Liver Cirrhosis predictor using machine learning:
Introduction
Considerations for deploying the predictive model into a real-world clinical setting.
Integration of the model with existing healthcare systems for seamless data exchange and decision support.
Limitations and Future Directions
Discussion of Overview of liver cirrhosis and its impact on public health.
The importance of early detection and prediction for effective treatment.
Introduction to machine learning and its potential in healthcare.
Dataset Acquisition and Preprocessing
Selection of a suitable dataset containing liver-related features and cirrhosis labels.
Data preprocessing steps, including cleaning, handling missing values, and feature selection.
Splitting the dataset into training and testing sets.
Exploratory Data Analysis
Statistical analysis of the dataset to gain insights into the distribution and correlation of features.
Visualization techniques to understand the patterns and trends in the data.
Identification of potential risk factors and predictors for liver cirrhosis.

Feature Engineering
Transformation and creation of new features to enhance the predictive power of the model.
Techniques such as scaling, normalization, and feature extraction.
Handling imbalanced data issues if applicable.
Machine Learning Models
Introduction to various advanced machine learning algorithms suitable for liver cirrhosis prediction.
Implementation of algorithms such as logistic regression, decision trees, random forests, support vector machines (SVM), and
Model training, hyperparameter tuning, and evaluation metrics selection.
Model Evaluation
Evaluation of the trained models using appropriate metrics (e.g., accuracy, precision, recall, F1-score, ROC curves).
Cross-validation techniques to ensure model robustness and generalizability.
Comparison of different models and selection of the best performing model.
Interpretability and Explainability
Techniques for interpreting machine learning models and understanding the underlying factors contributing to predictions.
Feature importance analysis to identify the most influential features for liver cirrhosis prediction.
Deployment and Integration
limitations and challenges associated with liver cirrhosis prediction using machine learning.
Potential solutions and future research directions to address these limitations.
Conclusion
Summary of the study's findings and the potential impact of using advanced machine learning techniques for liver cirrhosis pre
Emphasis on the importance of early detection and personalized interventions for improved liver care.
By following this documentation, healthcare professionals, researchers, and data scientists can gain insights into predicting liv
Activity 3: Literature Survey (Student Will Write)
A literature survey for a liver cirrhosis Prediction project would involve researching and reviewing existing studies, articles, and
Milestone 2: Data Collection & Preparation
ML depends heavily on data. It is the most crucial aspect that makes algorithm training possible. So, this section allows you to
Activity 1: Collect the dataset.
Activity 1: Collect the dataset.
There are many popular open sources for collecting the data. Eg: kaggle.com, UCI repository, etc.
In this project we have used .csv data. This data is downloaded from kaggle.com. Please refer to the link given below to down
Link:
As the dataset is downloaded. Let us read and understand the data properly with the help of some visualisation techniques an
Note: There are a number of techniques for understanding the data. But here we have used some of it. In an additional way, y
Activity 1.1: Importing the libraries
Import the necessary libraries as shown in the image.
Activity 1.2: Read the Dataset
Our dataset format might be in .csv, excel files, .txt, .json, etc. We can read the dataset with the help of pandas.
In pandas we have a function called read_csv() to read the dataset. As a parameter we have to give the directory of the csv fil
Activity 2: Data Preparation
Activity 2: Data Preparation
As we have understood how the data is, let's pre-process the collected data.
The download data set is not suitable for training the machine learning model as it might have so much randomness so we ne
Handling missing values
Handling categorical data
Handling Outliers
Note: These are the general steps of pre-processing the data before using it for machine learning. Depending on the condition
Activity 2.1: Handling missing values
Let’s find the shape of our dataset first. To find the shape of our data, the df.shape method is used. To find the data type, df.in
For checking the null values, df.isnull() function is used. To sum those null values we use .sum() function. From the below ima
Activity 2.2: Handling Categorical Values
As we can see our dataset has categorical data we must convert the categorical data to integer encoding or binary encoding.
To convert the categorical features into numerical features we use encoding techniques. There are several techniques but in c
Activity 2.3: Handling Outliers in Data
With the help of boxplot, outliers are visualized. And here we are going to find upper bound and lower bound of numerical feat
From the below diagram, we could visualize that Discount_offered, Prior_purchases features have outliers. Boxplot from matp
To find upper bound we have to multiply IQR (Interquartile range) with 1.5 and add it with 3rd quantile. To find lower bound ins
To handle the outliers transformation technique is used. Here L1 transformation is used.
Data splitting
The data was split into train and test variables as shown below using the train_test_split() method of scikitlearn module with a
Normalization
The data will be normalized using L1 regularisation that will be applied on x_train and x_test variables separately.
Milestone 3: Exploratory Data Analysis
Data analysis
Activity 1: Descriptive statistical

Activity 1: Descriptive statistical

Descriptive analysis is to study the basic features of data with the statistical process. Here pandas has a worthy function calle

Activity 2: Visual analysis

Visual analysis is the process of using visual representations, such as charts, plots, and graphs, to explore and understand da

Activity 2.1: Univariate analysis

In simple words, univariate analysis is understanding the data with single feature. Here we have displayed two different graphs

Seaborn package provides a wonderful function histplot. With the help of histplot, we can find the distribution of the feature. To

From the plot we came to know

In our dataset we have some categorical features. With the countplot function, we are going to count the unique category in th

Activity 2.2: Bivariate analysis

To find the relation between two features we use bivariate analysis. Here we are visualizing of data

Outliers were found for 2 features as visualized above using box plots. To be specific using IQR (inter quartile range) it was ob

Activity 2.3: Multivariate analysis

In simple words, multivariate analysis is to find the relation between multiple features. Here we have used heatmap from seab

• From the below image, we came to a conclusion that the product discount is the feature that most highly correlates to if a pro

Splitting data into train and test

Now let’s split the Dataset into train and test sets. First split the dataset into x and y and then split the data set.

Here x and y variables are created. On x variable, data is passed with dropping the target variable. And on y target variable is

Milestone 4: Model Building

Model building

Activity 1: Training the model in multiple algorithms

Activity 1: Training the model in multiple algorithms

Now our data is cleaned and it’s time to build the model. We can train our data on different algorithms. For this project we are

Activity 1.1: Writing function to train the models

A function named models_eval_mm is created and train, test data are passed as parameters. In the function, logistic regressio

Activity 1.2: Calling the function

The function is called by passing the train, test variables. The models are returned and stored in variables as shown below. Cl

Milestone 5: Performance Testing & Hyperparameter Tuning

Performance Testing