RevolutionizingLiverCare:PredictingLiverCirrhosisUsing Advanced Machine Learning Techniques

1. Introduction

1.1. ProjectOverview

This project aims to build a machine learning model to predict liver cirrhosis based on various patientfeatures.Byanalysingthesefeatures,themodelwillclassifypatientsintoriskcategories, aiding in early diagnosis and treatment.

1.2. Purpose

This project aims to develop a machine learning-based predictive model for early detectionof liver cirrhosis using clinical and biochemical data. By applying advanced algorithms and performance tuning, the model identifies patients at risk with high accuracy. The bestperforming model is deployed using Flask, enabling integration into real-time diagnostic systems.Thisapproachsupportsfaster,data-drivendecisionsinliverhealthcareandimproves patient outcomes through early intervention.

2. IdeationPhase

2.1. ProblemStatement

Thegoalistoclassifypatients'risklevelsforlivercirrhosisbasedontheirmedicaldata.Accurate prediction will support better management and early intervention for liver health.

2.2. ProjectProposal(ProposedSolution)

Thesolutioninvolvesdevelopingseveralmachinelearningmodelstopredictlivercirrhosis.We will select and optimize the best model based on performance metrics to achieve the highest accuracy.

2.3. InitialProjectPlanning

Initialplanningincludedsettinguptheprojectenvironment,definingobjectives,andoutliningthe workflow for data collection, preprocessing, model development, and evaluation.

3. DataCollectionandPreprocessingPhase

3.1. DataCollectionPlanandRawDataSourcesIdentified

Thedatasetforthisprojectwassourcedfrom Kaggle,containingpatientdatarelevanttoliver cirrhosis prediction(Dataset link: https://www.kaggle.com/datasets/bhavanipriya222/livercirrhosis-prediction).

3.2. DataQualityReport

- **DataShape:**Thedatasetinitiallycomprised[numberofrows,numberofcolumns]rows and columns.
- **MissingValues:**Handledbydroppingrowswithmissingvalues.

3.3. DataExplorationandPreprocessing

- **UnivariateAnalysis:**Histogramswereplottedfornumericalfeatures.
- **BivariateAnalysis:**Scatterplotsandpairplotsexploredrelationshipsbetweenfeatures.
- **OutlierHandling:**OutlierswеredetectedandmanagedusingtheIQRmethod.

4. ModelDevelopmentPhase

4.1. FeatureSelectionReport

Featuresrelevanttolivercirrhosisprediction wereselected,anddatascalingwasappliedto standardize the input.

4.2. ModelSelectionReport

- **ModelsTested:**NaiveBayes,RandomForest,LogisticRegression,RidgeClassifier,Support Vector Classifier, KNN, XG Boost.
- **EvaluationMetrics:**Accuracy,ConfusionMatrix,ClassificationReport.

4.3. InitialModelTrainingCode,ModelValidationandEvaluationReport

- **Code:**Modeltrainingandevaluationstepswereimplementedforeachalgorithm.
- **Validation:**Modelswerevalidatedusingatestset,withperformancemetricsrecorded.The KNN model achieved the highest accuracy of 86.32%.

5. ModelOptimizationandTuningPhase

5.1. HyperparameterTuningDocumentation

- **KNN:**Optimizedbytuningthenumberofneighborsanddistancemetrics.
- **XGBoost:**Hyperparameterstunedforlearningrate,maxdepth,andn_estimators.

5.2. PerformanceMetricsComparisonReport

Model	Accuracy	Precision	Recall	F1-Score
NaiveBayes	35.79%	0.00	0.00	0.00
RandomForest	35.79	0.00	0.00	0.00
LogisticRegressionCV	81.58%	91.80	79.43%	86.49
RidgeClassifier	84.21%	93.44	83.82	88.37
SupportVectorClassifier	35.79%	0.00	0.00	0.00
LogisticRegression	79.47%	91.80	79.43	85.58
KNN	86.32%	94.26	85.82	89.84
XGBoost	35.79%	3.28	50.00	6.15

5.3. FinalModelSelectionJustification

The K-Nearest Neighbors (KNN) model was selected as the final model due to its superior accuracyof86.32%.KNNexcelledinhandlingcomplexdatarelationshipsanddemonstratedthe best performance in terms of precision, recall, and F1 score. This makes it a robust choice for predicting liver cirrhosis, aligning well with the project's goals.

6. Results

6.1. OutputScreenshots

Thesourcecodeandoutputscreenshotsareavailableintheaccompanyingfiles.

7. Advantages&Disadvantages

- **Advantages:**Highaccuracy,effectiveathandlinglocaldatavariations,robust performance.
- **Disadvantages:**Canbecomputationallyintensive,requirescarefultuning.

8. Conclusion

The project successfully developed a machine learning model to predict liver cirrhosis with high accuracy.TheKNNmodel,afterhyperparametertuning,providedthebestresultsandwaschosenfor its robustness.

9. FutureScope

- Furtherdatacollectiontoincludemorefeaturesandincreasedatasetsize.
- Explorationofadditionalfeaturesandengineeringtechniques.

- Experimentationwithdeeplearningmodelstopotentiallyoutperformtraditionalmodels.
- Integrationwithareal-timepredictionsystemforpracticaldeployment.

10. Appendix

10.1. SourceCode

CodeFile:model.ipynb

10.2. GitHub&ProjectDemoLink

GitHubRepository:

https://github.com/CHENNAMSETTYGURUTEJA/Liver_Cirrhosis_Prediction

Demo Link:
https://drive.google.com/file/d/1NghxRleC98qL7087w2Tn9uYrcFqPNC5m/view?usp=drivesdk