Using Machine Learning for Disease Diagnosis

Eren Kızılırmak 210704025

Abstract

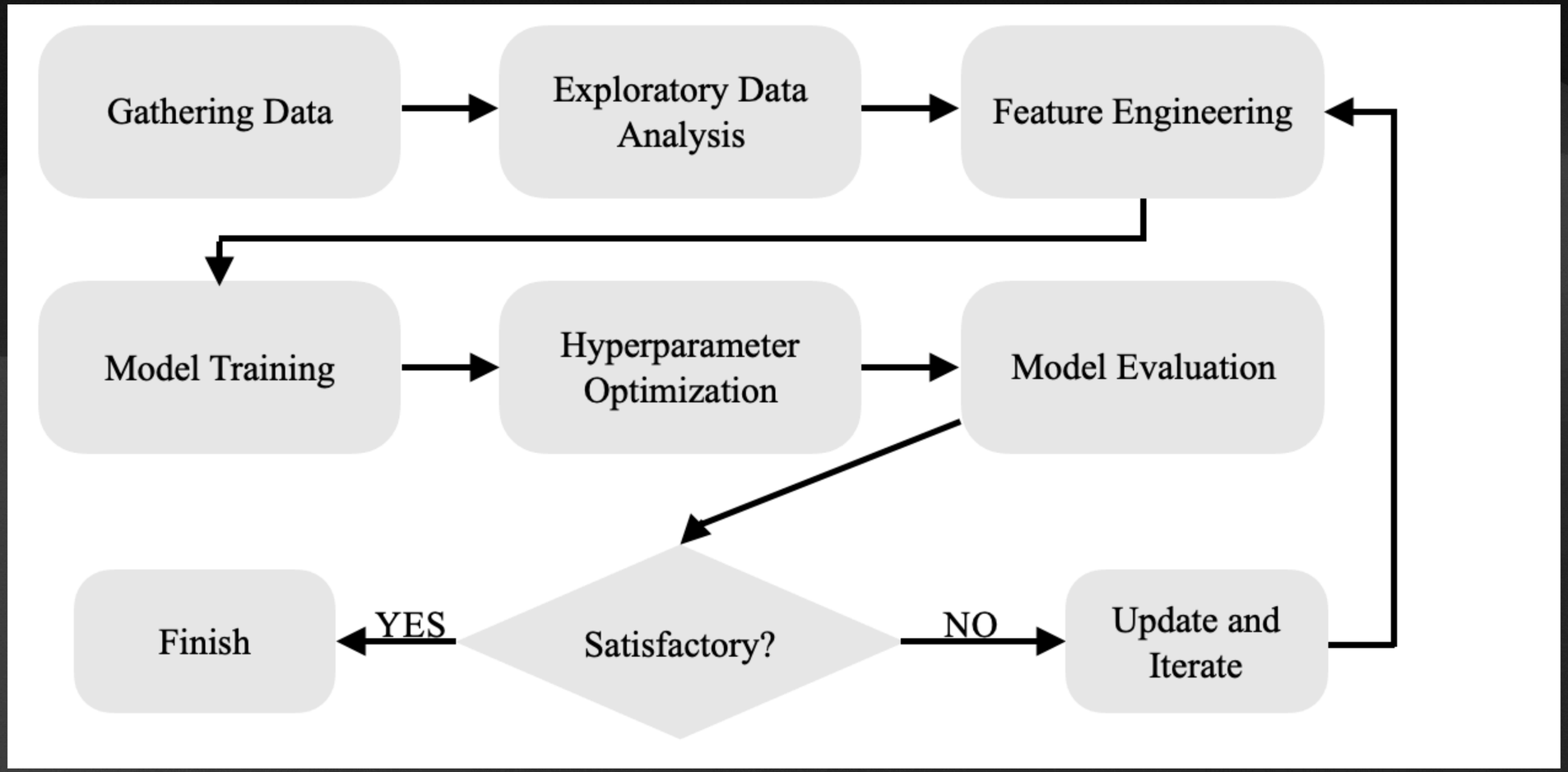
- Early and accurate diagnosis is important for effective treatment of diseases.
- Machine learning (ML) has potential to improve speed and accuracy of diagnoses.
- Evaluated 7 ML models for diagnosing heart disease using patient data.
- Performance metrics: ROC-AUC and cross-validation score.
- Conclusion: SVC and XGB classifiers best performers for this dataset and can be used for heart disease diagnosis.

Literature Review

- previously implementation of fuzzy logic system.
- Methods used by other researchers (fuzzy logic, SVM 83%, Naive Bayes 84%).
- XGB and PCA not utilized in previous works.

Methodology

Diagram

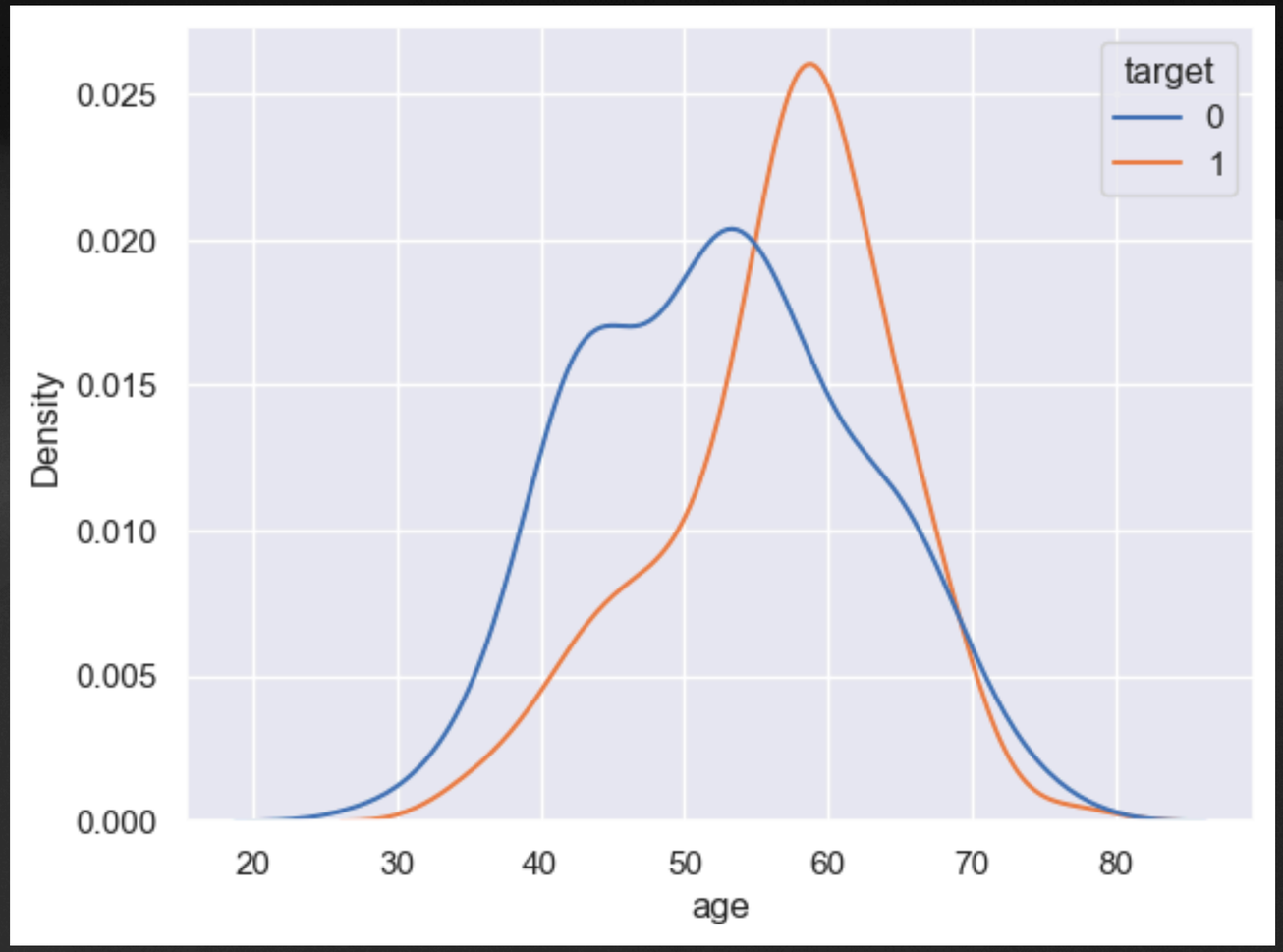


Exploratory Data Analysis

to undertand data

- Observation: Older people prone to heart attacks.

- Observation :around age 45 there is curve and can be used for binning



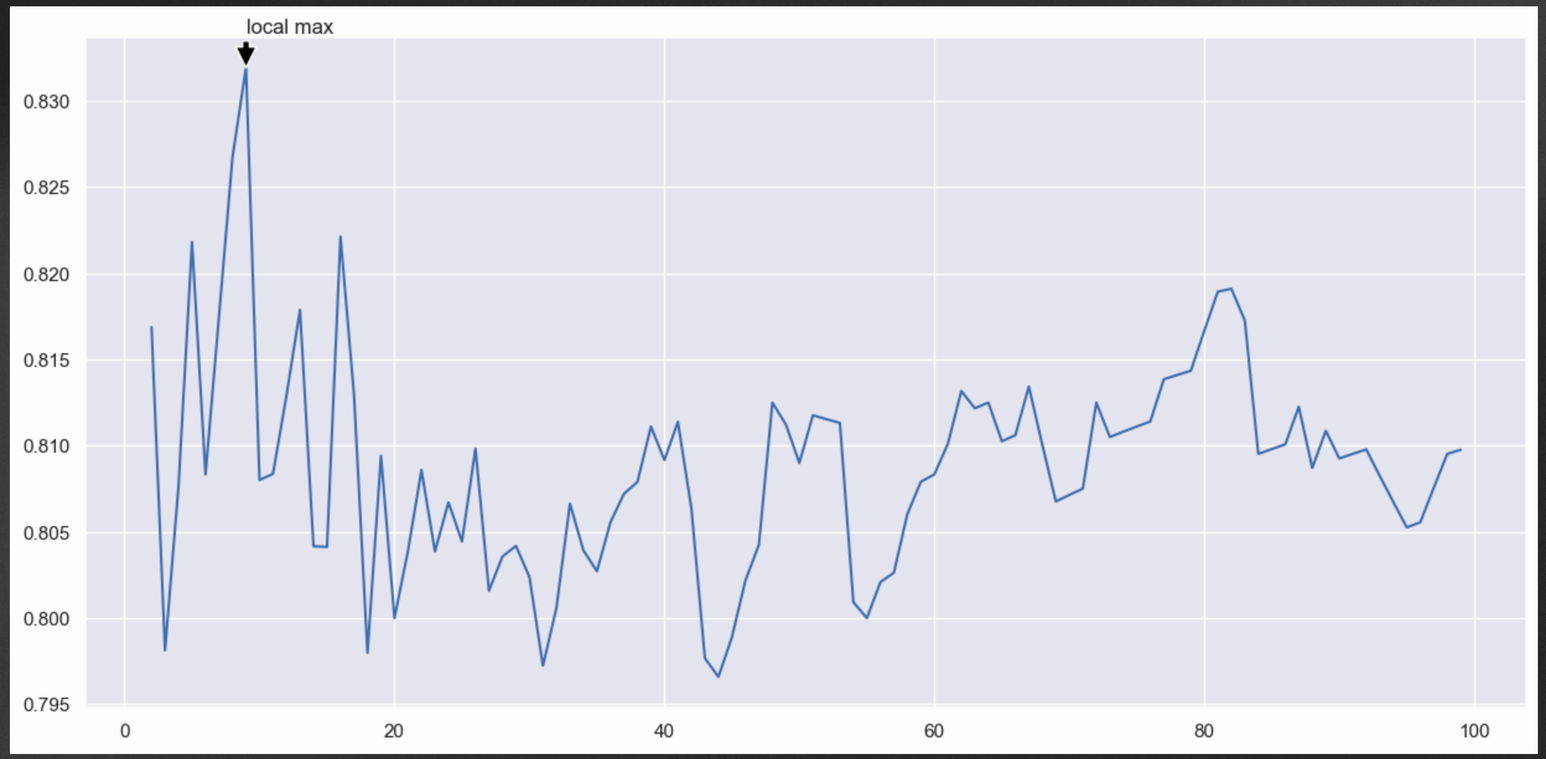
Feature Engineering

- Binning (age, trestbps, chol). but only binned age values were helpful.
- PCA for dimensionality (5 features). pca reduced correlation. and new five features are combined with the original dataset.
- Encoding and scaling.

Model Training

SVC fold split graph

- Models: SVM, ANN, DTC, RFC, XGB, KNN, ADA-BOOST
- finding best Kfold(cross validation) split for each model

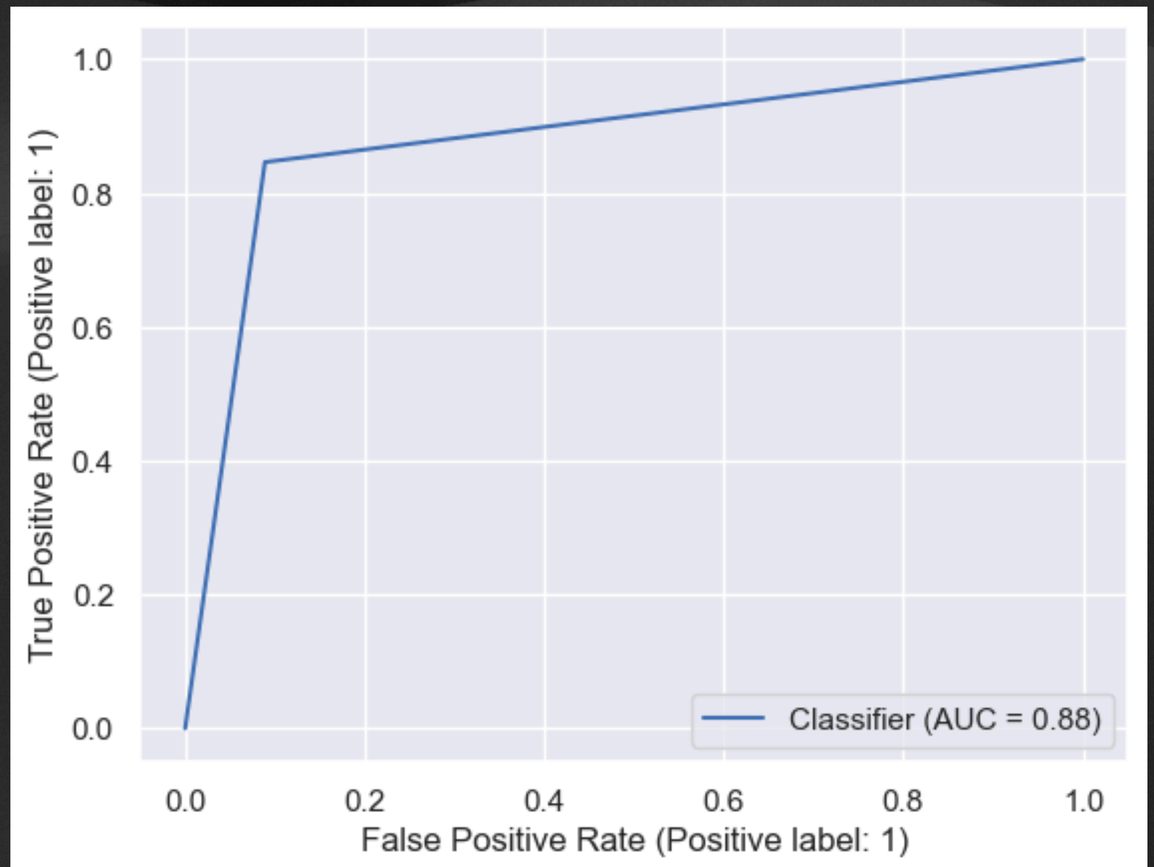


Results and Conclusion

Evaluation metrics: ROC-AUC, cross-validation score.

MODEL	CROSS VALIDATION	ROC-AUC
SVM	0.833	0.840
ANN	0.816	-
DTC	0.769	0.746
RFC	0.833	0.829
XGB	0.825	0.878
KNN	0.805	0.798
ADA-BOOST	0.819	0.820

- Conclusion: XGB with PCA enhanced data effective for CHD diagnosis.



Future Work

- Weakness: Risk of overfitting with PCA. because dimensionality got higher
- To overcome this, Future research might be Explore alternative dimensionality reduction other than PCA.
-

Thanks for Listening