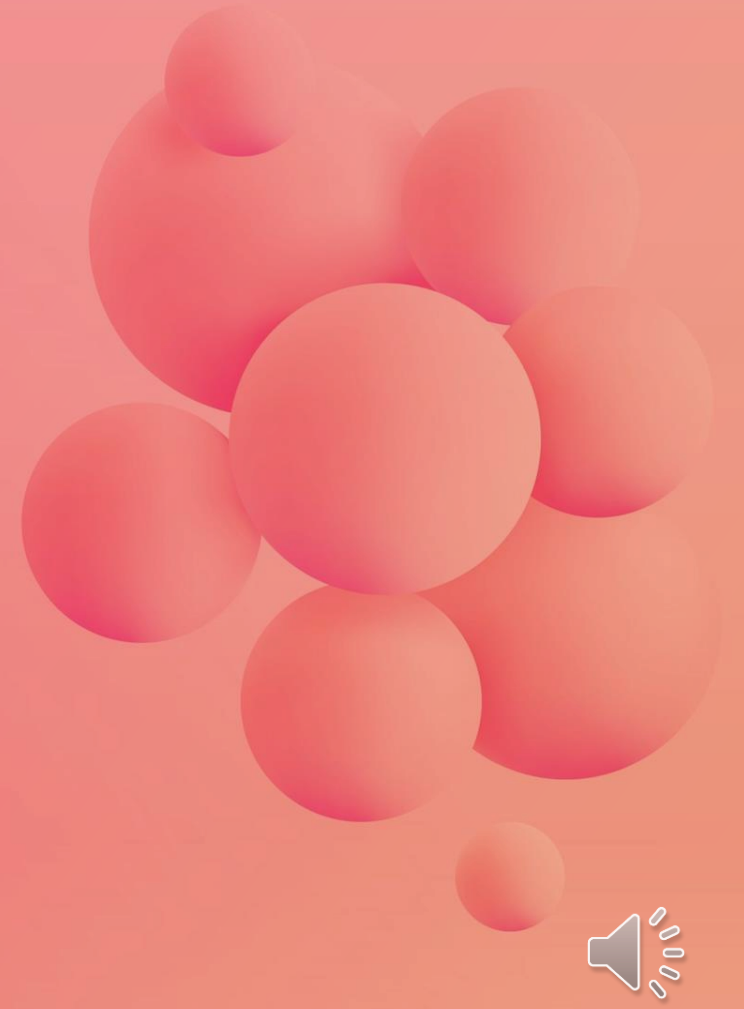


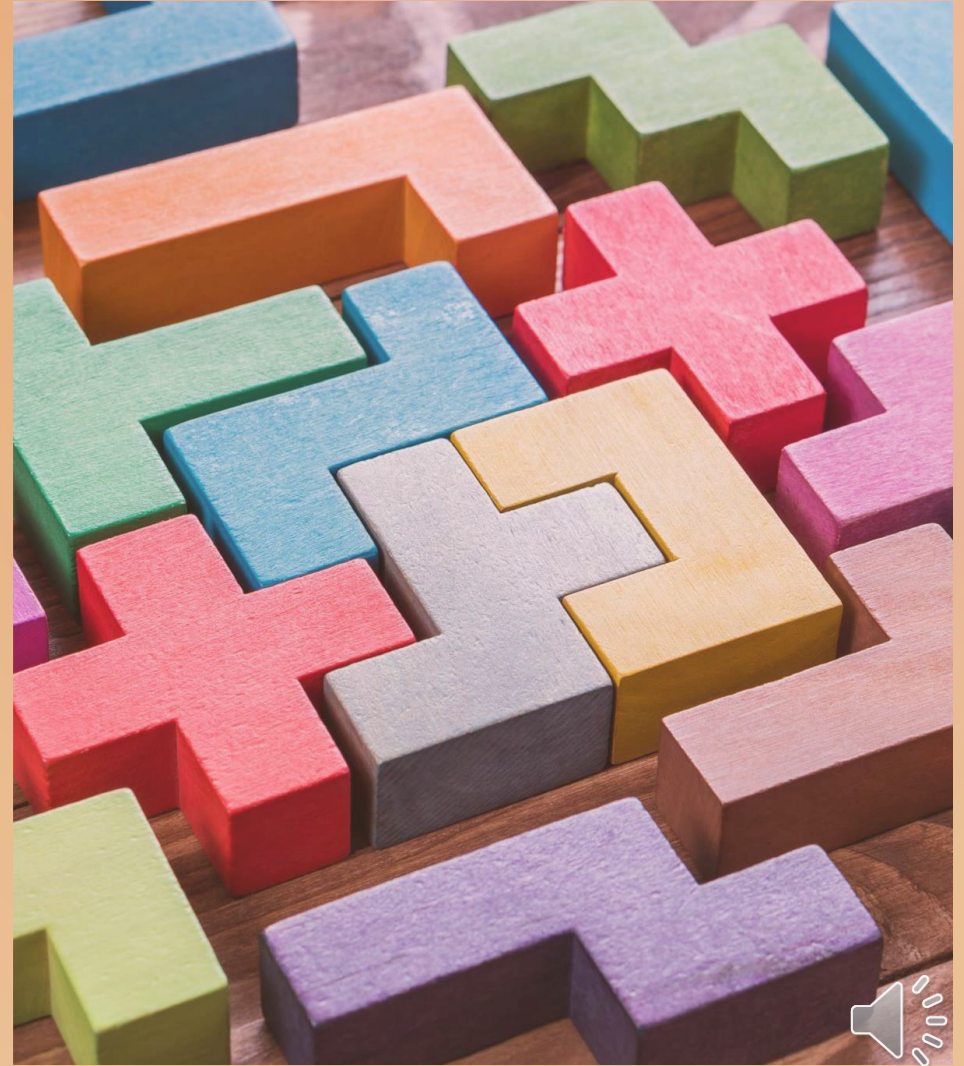
Disease prediction task

By/Ali Sameh



Introduction

- Applying ensemble learning techniques for enhanced predictions.
- Leveraging hyperparameter tuning for optimized model performance.
- Incorporating cross-validation to ensure model generalizability.

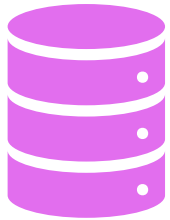


Data description

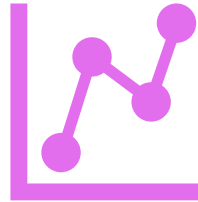
- Describing data distribution characteristics in training and test sets
- Examining feature importance and correlation for model relevance.



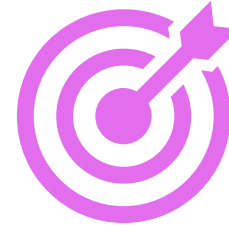
Preprocessing



Applying label encoding for categorical data transformation



Utilizing feature scaling to standardize variable ranges

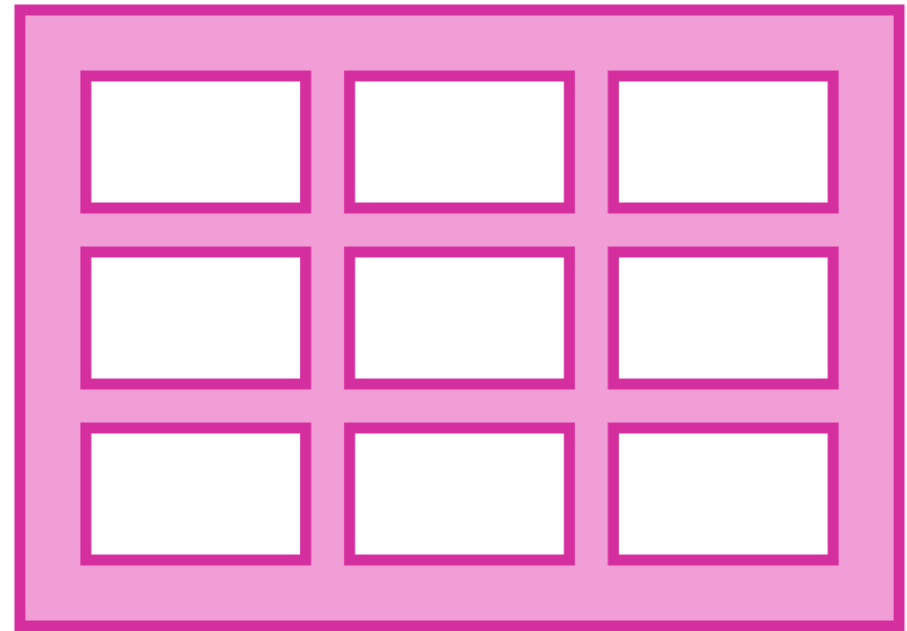


Ensuring data preprocessing consistency for model efficiency



Train-validation split

- Implementing the 80-20 rule for data partitioning.
- Leveraging 80% data for model training and 20% for validation.



Random forest classifier

1

**CONFIGURING
HYPERPARAMETERS
FOR OPTIMAL
MODEL
PERFORMANCE**

2

**ANALYZING BIAS-
VARIANCE
TRADEOFF TO
IMPROVE MODEL
RELIABILITY**

3

**COMPARING
DIFFERENT
PERFORMANCE
METRICS FOR
COMPREHENSIVE
EVALUATION**



Bagging classifier



Analyzing model validation accuracy for performance assessment.



Evaluating test accuracy to assess model generalizability and reliability.



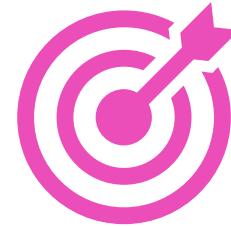
Ensemble method



Comparing ensemble methods for optimal predictive performance

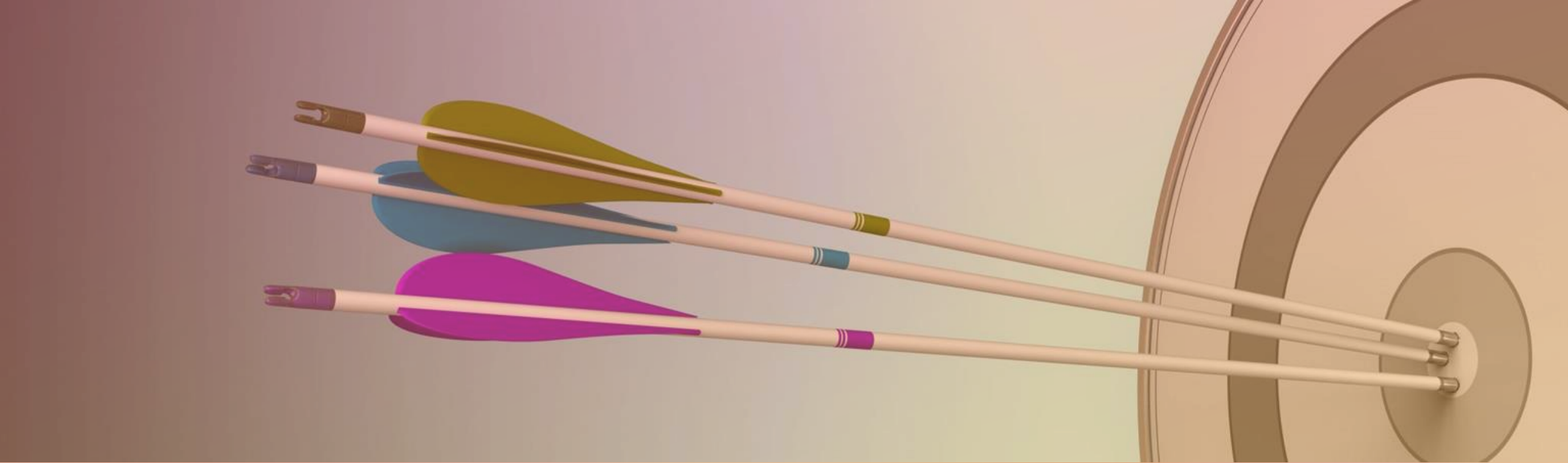


Assessing the impact of voting strategies on model accuracy



Evaluating validation and test accuracy for model reliability assessment





Feature selection

- Evaluating feature importance through Recursive Feature Elimination (RFE)
- Selecting optimal feature subset for improved model performance.
- Assessing the impact of feature selection on model generalizability.



Hyper parameter tuning

- Identifying best parameters through GridSearchCV.
- Evaluating model's performance with best parameters.



SMOTE-Handling imbalance



**APPLYING SMOTE TECHNIQUE FOR
IMBALANCED DATA HANDLING**



**ENHANCING MODEL'S
GENERALIZABILITY WITH
AUGMENTED DATA SAMPLES**



**IMPROVING MINORITY CLASS
PREDICTION ACCURACY THROUGH
OVERSAMPLING**



Data visualization

Utilize box plots to identify outliers

Leveraging histograms to visualize data distribution

Utilizing correlation heatmap for feature relationship analysis



Thank you

