Assignment 4

Deep Salunkhe. 21102 A0014

Dif In the context of a financial institution, you are tasked with predicting loan details risk. Build a classification and organisation tree (CART) model using historical loan Explain has you would split the data, choose the feartures and evaluate the model's performance using the confusion matrix and roc conve. Discuss any ethical consideration orelated to using such a model.

Objective: Develop CART model for loan default prediction and evaluate it using various performance matrics while considering ethical implication

=> Building a CART Model for loan Default prediction

Data splitting Wom Pan Al Millian

The first step is to divide the historical loan data in 2 sets

Toring set > This is used to build the cast model, it-Should be separesentation of over all population

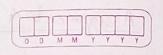
Testing set > This dataset is used to evaluate the models

performan on unseen data

The common approach is to split the dat in

Feature selection

Feature. selection is courial for building an altertie.
CART man, Relevant feature con significanty



improve modul performance and interpretability, some key features for loan detailst prediction include

- Demographic information
- -> Financial information
- -> be havious information.
- -> Economic indicator

Feature selection can be done using technique like

- -> Cornelation anlaysis
- -> information gain
- Recurre Featux Elimination

Buildmy the CART model

- a cost mosel can be constructed. The CART' algorithm recursily partition the data into subsiderated on fearther, values.
- The spliting contenion is typically the Gini impunity.

 or entropy The process continus until a stopping
 contenion is mot such as reaching a maximum
 depth or minimum number of observation in
 a not

Model Evaluation

poediction on the testing set. It helps to cake lake acressey, parcision, recall and FI scon



-> ROC : The receiver operating characteristic (ROC)
cuove plots the true positive rate against.
the false positive seite at differen classification
thousholds. The Azea under curve CAVC) provides
an overall measure of model pertermance
Ethical consideration
-> Fairnes: The model should not discorminate bard
on protected attribute such as race, gender
or age.
-> Transpareny: The model's decision-making proces
should be understandable and explainable
-> Parray: Sensitive personal Information. Should be.
nant atro
-> Social impact: The model's outcomes should not
exactbak existing social inequalities