Assignment: 3



Oil You are a data Scientist at a real estate company your task is to predict house prices based on various features such as square footage, number of bedrooms and location. Using linear regression, build a model to make these prediction. Disuss the steps you would address them. Objective: Apply linear regression to a real-world problem and discuss data preparation and.

validation techniques

Data Preparation

- 1. Data Collection: Gather a comprehensive data set footage, number of bedrooms, and location
 - information
- 2. Data cleaning .:
 - Handle missing values: Impute missing values using. techniques like mean, median mad.
 - outlier detection: Identify and handle autlier using statistical method or visualization
 - Feature Engineering: Create new feature it recessary
- 3. Data exploration:
 - Descriptive statistics: Calculate rummany statistics
 for numerical feature to
 understand, distribution.



Visualization: Create histogram, scatter plots, and correlation matrices to explore relationships.

3. Feature Selection

- 1. Feature Importance: Use techniques like. romelation analysis, frature. Importance.

 or statistical. tests to identify the most relevantfratures
- 2. Dimensionally reduction: if dealing with a large number of features, consider techniques like Poincipal component analysis to reduce dimensionality

Model Building.

linear regression: Create a linear regression, model using selected fractur.

Model Fraining: Fit the model to the baining date not



Model Evaluation

Splitting Data: Divide the dataset into training and testing set to evaluate model performance on unseen data.

Model Evaluation Metrics: Use metrics like Meansmont.

Error (MSE), Root mean

squared. Error (RMSE), Meen. absolute Error (MAE),
and R-squard.

Cooss-validation: Employ cross-validation to get.

a more robust estimate of
model performana.

Potential lissues and solutions

- Non-linearty: If the relationship between fourter.

and the target variable is nonlinear, consider transformation, usry non-linear module

- Overfithing: If the model performs well on
the touining data but possy on the
testing data, consider orgalorization, frature relection of
incoming the Lateralt site.



	under fitting: If the model performs poorly on both training and testing data, consider adding more fracture, increasing model complexity or improving data quality
	both training and testine data,
	consider adding mon fracture, increasing model
bar	complexity or immorry data avaling
	Talmar strators of the mother
	able nastru no aspenioliza
VIII Ind	rost salit esistem sell : misser auticule of labora
	reso took, (1211) kora
.(34	My over I styled a sight (CEMB) were brown to
	Annances has
day	at notto liber-esses polymit involentiles - 2200 1 1888
	Senite turing the party of the senior
	no consider the
	andtelor has zon of districted
SALL	en ex aldibus book at his stand - note - 1
4.4.	
Medical	Candillenson versity and one is to assist as but to a series
7	Use granding lellow with the williams -
14-	no telescy tel state president and
6	alterative and a control of the lines and more and the stand
_	