Car Dheko - Used Car Price Prediction

A project on converting “RAW ,UNFORMATED DATA” to a “STRUCTURED DATA” and performing

“A STREAMLIT APPLICATION “for a Model that’s built on that Data.

**Step 1: Data Cleaning and Preprocessing**

Import the excel files

Since the data is in unstructured comma separated type like json type convert that to a normal rows and column type using Use ast.literal\_eval to convert the string to a dictionary.

Eg: data\_dict = ast.literal\_eval(data\_str)

Convert the unstructured data into structured data. Handling missing values, standardizing data formats, encoding categorical variables, and normalizing numerical features.

**Step2: Exploratory Data Analysis**

We can’t have like 27 columns for a model,the model will suffer from over fitting ,under fitting etc.so with the domain knowledge have only 7-10 columns ,also in an application its hard for a user to fill 25 columns so make it simple.

From the selected columns check for outliers ,class imbalance etc and treat them.You can use Inter Quartile range,mean meadian,visualization,etc.

**Step:3 Machine Learning Model Development**

Build 5 different Regression model that you have in mind and cross check with their MAE , MSE ,r2 score etc. and find a better model.

After choosing a model note the R2 score ,MAE ,MSE and again drop some rows ,columns and check if accuracy increases or not. Note: model should not be also overfitting.

Do the process until the R2 score is high.

**Step 4: Model Deployment**

Pickle the model ,as we are going to use the same in another python file for “Stream lit” application.

**Step 5: Streamlit Application Development**

Develop a user friendly streamlit application that gets the input of selected column and use the pickled model by inserting the values got and predict the price.