Script for ChatGPT and YouChat

Part 1:

I want to give you a dataset, and the link on the Google drive <https://drive.google.com/file/d/1NGLzYvjJXuULWB2t1DvcXATOHFil40Ej/view?usp=sharing> I have give you the permission to read it. The topic of the data is about E-Commerce Shipping data. Please read the data ‘Train.csv’ and I want you to analyze this data by using python.

Given the dataset uploaded named ‘Train.csv’ that contains the variables, you are to predict ’Reached.on.Time\_Y.N’. Since you already have the data, can you list the steps I have to follow to develop an end-to-end project, and pass me the final accuracy result of the best model?

--If the output is not showing and the code snippet is incomplete, you can follow by “what is the output of your result”, but usually the result is still not completed.

--The simple end-to-end project code is not working on python, error occurred as the data preparation steps is not enough. For example, the categorical features was not transformed to numerical values.

Part 2:

Please write a Python code to load and perform Exploratory data analysis (EDA) on the ‘Train.csv’ dataset I uploaded. From the EDA that you performed, can you suggest any observation of the data?

Write a python code to perform feature engineering to the ‘Train.csv’ data I uploaded, only add feature engineering part, and then drop the columns that is not necessary to predict ‘Reached.on.Time\_Y.N’.

After steps of feature engineering you provided, write a python code to clean and preprocess the dataset “Train.csv”, and check the outliers and balance the data if the target variable is imbalanced.

Also, the PCA is required to reduce the data “Train.csv” dimensionality.

Part 3:

Next, write a follow-up python code for model selection on the cleaned data ‘Train.csv’. Try decision trees, random forests, logistic regression, or support vector machines (SVM). Note: only include the model selection part. Please split the dataset into training and testing data before using models.

Choose RandomForestClassifier, then create a Python method to evaluate the model. To make sure the model is not overfitting to the training data, use cross-validation, accuracy, precision, recall, and F1 score as metrics. Include hyperparameter adjustment as well in the code above, and save the model with the best performance.

Please save the model.

Create a Gradio web application for “Train.csv” data classifiers by writing the Python code. We are not using columns ['ID’]

just include the gradio app part.

modify the code to show classification instead of class probabilities.

update the Gradio code with component from gradio.components

Deploying the Web App on Spaces

how to deploy gradio app to huggingface spaces.

You have to keep remind Chatgpt to follow the data I uploaded, otherwise it will generate a new data and columns.

Cannot remember exact the column name, always confused with other dataset

And drop the wrong columns, when check again, the system chose to keep the variable

Mistakes from ChatGPT: there is no negative correlation between customer\_care\_calls and prior\_purchase. It also didn’t catch the stronger correlation between the variables?

|  |  |  |
| --- | --- | --- |
|  | Chat GPT | YOU CHAT |
| User Experience |  | Fail to copy & paste the last prompt |
|  |  |  |