🚗💥 Vehicle Claims Prediction PreProject 🚀

Hello Class!

Welcome to the optional Vehicle Claims Prediction Project, where you get the chance to apply your machine learning skills in predicting vehicle claims! 🚗💥 This project is designed to challenge your problem-solving abilities and sharpen your data science toolkit. The students with the best solutions will be asked to present their project in class and explain how they beat the challange. Ready to dive in? Let's go!

🔍 What's the Goal?

Your mission is to predict the missing labels for a dataset of vehicle claims. You'll be working with real-world data which we messed up, so careful data preprocessing, smart feature engineering will be crucial for success.

The evaluation metric will be this f1_score with average='macro':
**from sklearn.metrics import accuracy_score, classification_report, f1_score **
**score = f1_score(y_true, y_pred, average='macro') **

Note: Reaching a Score of 65 % is possible with basic code

Update: You can test your prediction against the evaluation data on this website:
https://f24redimlchallange-hcqihppmsjbuekyubre4du.streamlit.app/

🎞 Files Overview

You'll be working with three key files:

– prediction_unlabeled.csv: This file has 30% of the dataset, but no labels. Your task? Predict those missing labels using the magic of machine learning! 🔮

– train_labeled.csv: Here's where the training happens! This file contains 70% of the dataset with labels, giving you the data needed to build and fine-tune your model.

– example_prediction.csv: This is your guide! It shows how your final submission should be structured, including the ID column from prediction_unlabeled.csv and a placeholder label column set to 0. Your final predictions need to follow this structure, but with your predicted labels instead!

🚀 How to Upload Your Data & Follow the Structure

1. Prepare Your Predictions: After running your model on the prediction_unlabeled.csv data, save your output in the same format as example_prediction.csv.

2. Follow the Structure: Make sure your final file includes:

   o ID Column: Identical to prediction_unlabeled.csv.

   o Label Column: Replace the placeholder 0s with your actual predicted labels.

3. Upload Your File: Once your prediction file is ready and follows the correct format, upload it with you name in filename to this link for evaluation. The format matters—mismatched files

may not be tested, so double-check!
https://drive.google.com/drive/folders/1tI-C1iHqgdrx4-kw0mdNBlvmvDiOtrWz?usp=sharing

🗓 Important Dates

–   Deadline for Model Submission: Get your prediction models in by 11.11.2024! So mark your calendars and plan ahead. 🗓

🛠 Teamwork

You want to work in a group? Feel free to share ideas and help but everybody should write there own code.

🛠 Resources

You can choose any ML methode you want.
Have questions? Stuck on a challenge? We're here to support you! Don't hesitate to reach out for guidance along the way.

Best of luck—get ready to dive into the data, explore possibilities, and create your best model yet! 💪🎉