Dear Professor,

We are writing to submit our dataset proposals for the upcoming Machine Learning 2 course projects. We have selected two datasets from Kaggle that are well-suited for applying the advanced techniques covered in our course.

Below are the details for each proposed project:

1. Regression Project: Used Cars Dataset

- **Dataset Source:** Kaggle (Used Cars Dataset) - Used Cars Dataset
- **Description:** This extensive dataset contains information on used vehicles, including features such as manufacturer, model, year of production, mileage, fuel type, and transmission.
- **Objective:** The goal is to predict the **price** of a used car based on its attributes.
- **Proposed Methods:** We plan to start with **Regression Trees** to establish a baseline. Subsequently, we will focus on ensemble methods, specifically **Random Forest (Bagging)** and **Boosting** algorithms, to improve predictive performance. Time permitting, we also intend to explore **Neural Networks** for this regression task.

2. Classification Project: Rain in Australia

- **Dataset Source:** Kaggle (Rain in Australia) - Rain in Australia
- **Description:** This dataset contains daily weather observations from various Australian locations. Features include temperature, rainfall, sunshine, wind gusts, humidity, and pressure.
- **Objective:** The goal is to build a binary classification model to predict the **RainTomorrow** target variable.
- **Proposed Methods:** We will use **Classification Trees** as an initial model. Our primary analysis will involve more complex models such as **Random Forests** and **Boosting**. We also plan to try to implement a **Neural Network** approach to compare its performance against the tree-based ensembles.

Please let us know if these proposals meet the requirements.

Best regards,

Wojciech Hrycenko, 473173