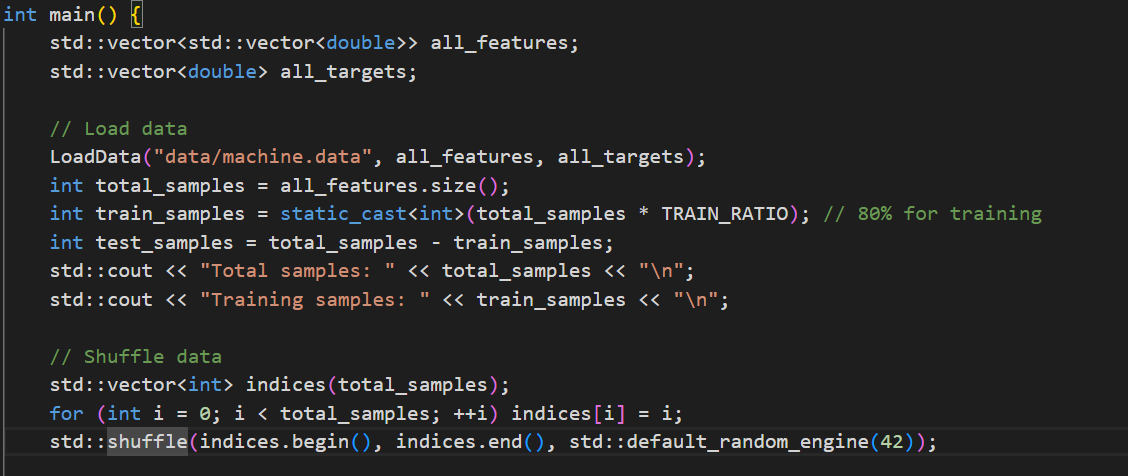
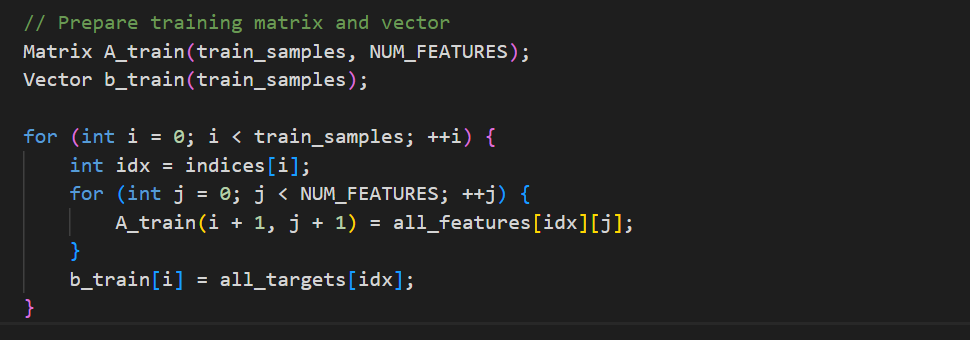
**Tiny Project**

This C++ program implements a linear regression model using the normal equations method to predict a target variable from multiple features. The program begins by reading data from a CSV file ("data/machine.data") containing various attributes of machines, with a focus on 6 specific predictive features and a target variable named PRP.

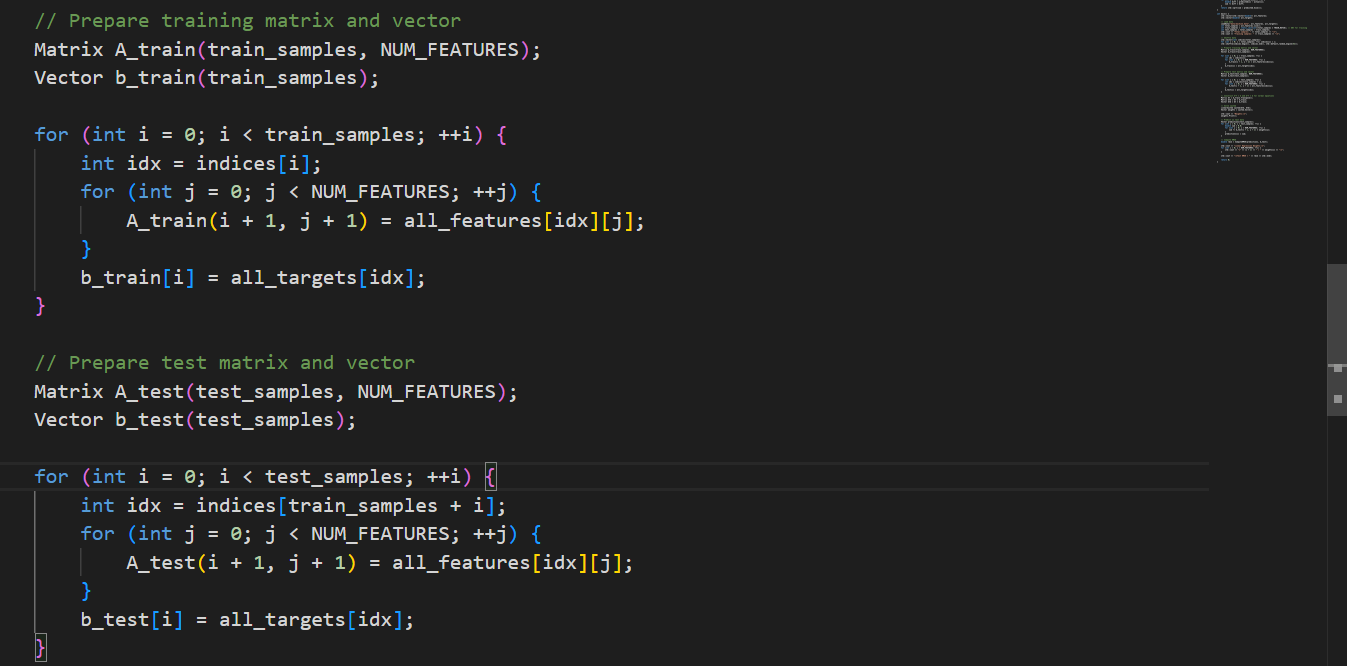
**Data Loading and Preprocessing:** The LoadData function reads the dataset line by line, skipping the first two categorical columns (vendor and model) and extracting the subsequent six numerical features alongside the PRP target variable. These are stored in features and targets vectors, respectively. The data is then randomly shuffled to ensure unbiased training and testing splits.



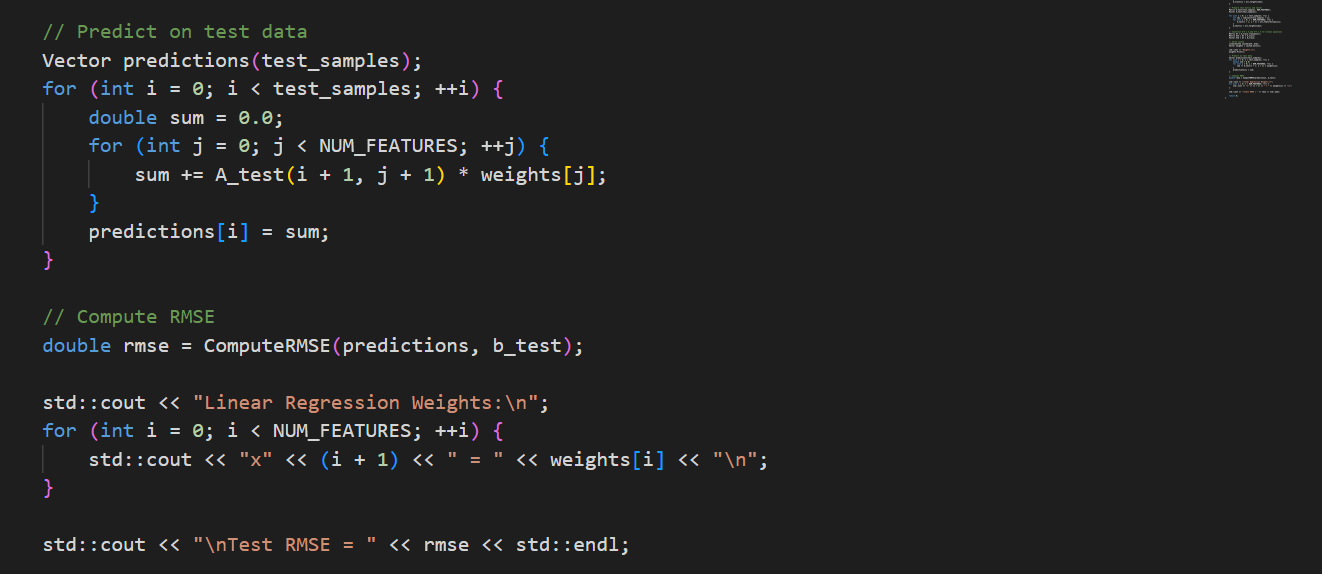
**Dataset Splitting:** The data is split into training and testing sets using an 80-20 ratio. The training data populates matrix A\_train and vector b\_train, while test data fills A\_test and b\_test. Note that indexing into the matrix uses 1-based indexing (i+1, j+1), consistent with my custom Matrix class design.



**Model Training Using Normal Equations:** The model computes the weights (coefficients) by solving the normal equations:  
  
 Here, At is the transpose of the training matrix, and matrix multiplication operators are used to compute AtA and Atb. The linear system is solved via a LinearSystem solver, yielding the weight vector.

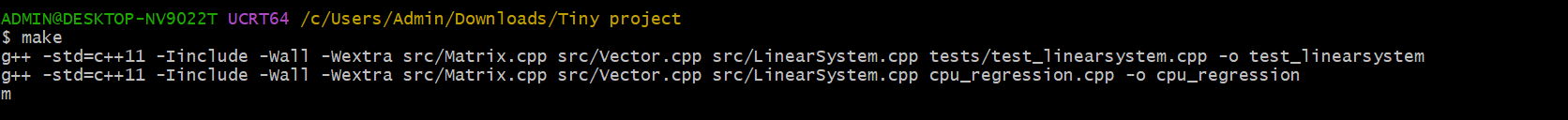


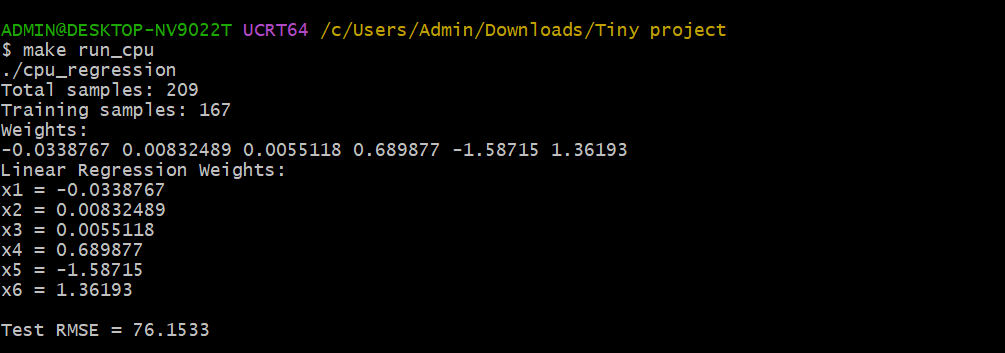
**Prediction and Evaluation:** Using the computed weights, predictions are generated for the test set by performing a dot product between test features and weights. The Root Mean Squared Error (RMSE) metric quantifies the prediction accuracy by comparing predicted and actual values.



**Makefile execution**

This Makefile compiles and runs a C++ project with Matrix, Vector, and LinearSystem classes. It builds two executables: test\_linearsystem for testing and cpu\_regression for running a regression model. It supports cleanup with make clean, and platform-specific delete commands for Windows and Unix. Targets run\_test and run\_cpu execute the respective binaries.



****

**Summary:** This program is a complete pipeline from data loading, preprocessing, model training using linear algebra techniques, to performance evaluation and prediction for the dataset.