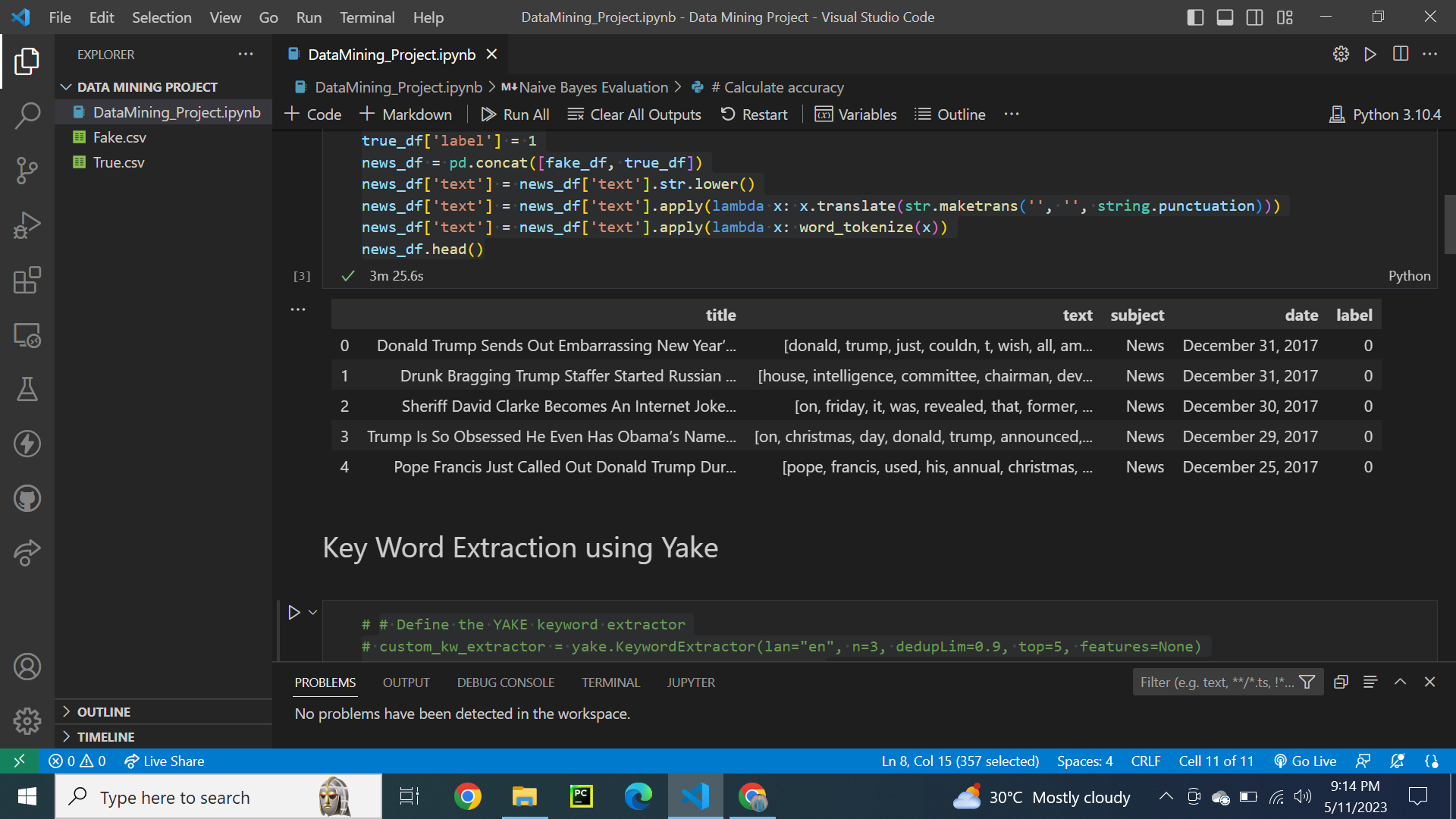
**Project: Text Classification using CAR  
Project Report  
Group Members: 19i-1986\_19i-0485**[Talha Rashid](mailto:i191986@nu.edu.pk)

**Waleed Mukhktar**

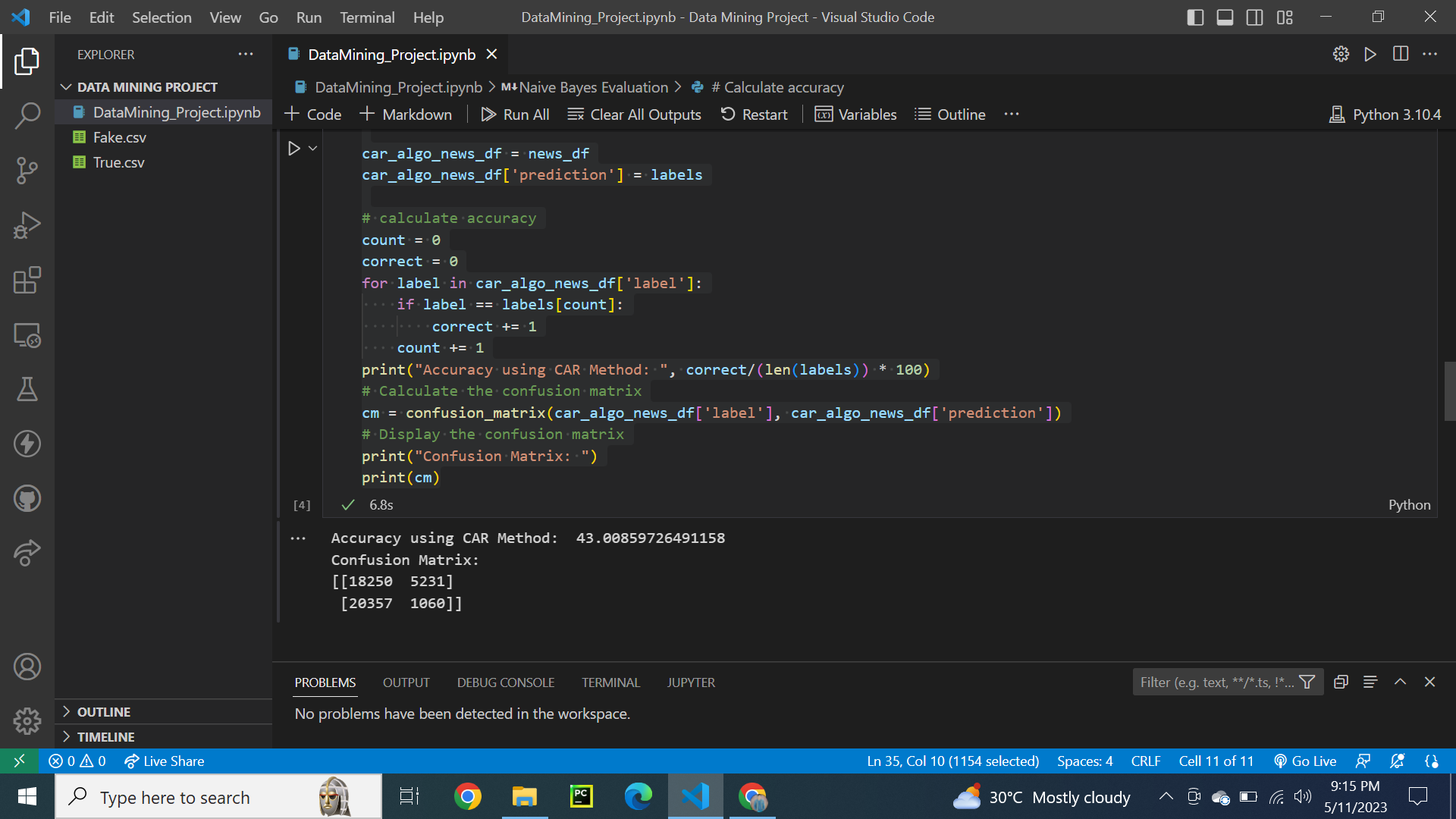
The given code performs text classification on two datasets containing news articles, one with real news and the other with fake news. The code first preprocesses the text data by converting it to lowercase, removing punctuation, and tokenizing the text. It then extracts the top 5 keywords from each article using the YAKE keyword extractor and adds them to the dataframe. The code then defines the keywords for each class, 'true' and 'fake', and assigns a label to each article based on the occurrence of these keywords. It calculates the accuracy of this classification method and displays the confusion matrix.

**Pre Processing Results:**



The code then splits the data into training and testing sets, vectorizes the text data using CountVectorizer, and trains a Multinomial Naive Bayes model on the training set. It predicts the labels for the test set and calculates the accuracy, precision, recall, and F1-score of the model. It also displays the confusion matrix.

**Accuracy using Car Algorithm:**



Overall, the code performs text classification on news articles using two different methods: a custom approach based on the occurrence of specific keywords and a machine learning approach using the Multinomial Naive Bayes model. It evaluates the performance of both methods by calculating various metrics such as accuracy, precision, recall, and F1-score.

**Naive Bayes Evaluation:**

