**CSE523 Machine Learning**

**Prof. Mehul Raval**

**Product Classification using their Ingredient**

**Week 1 Report**

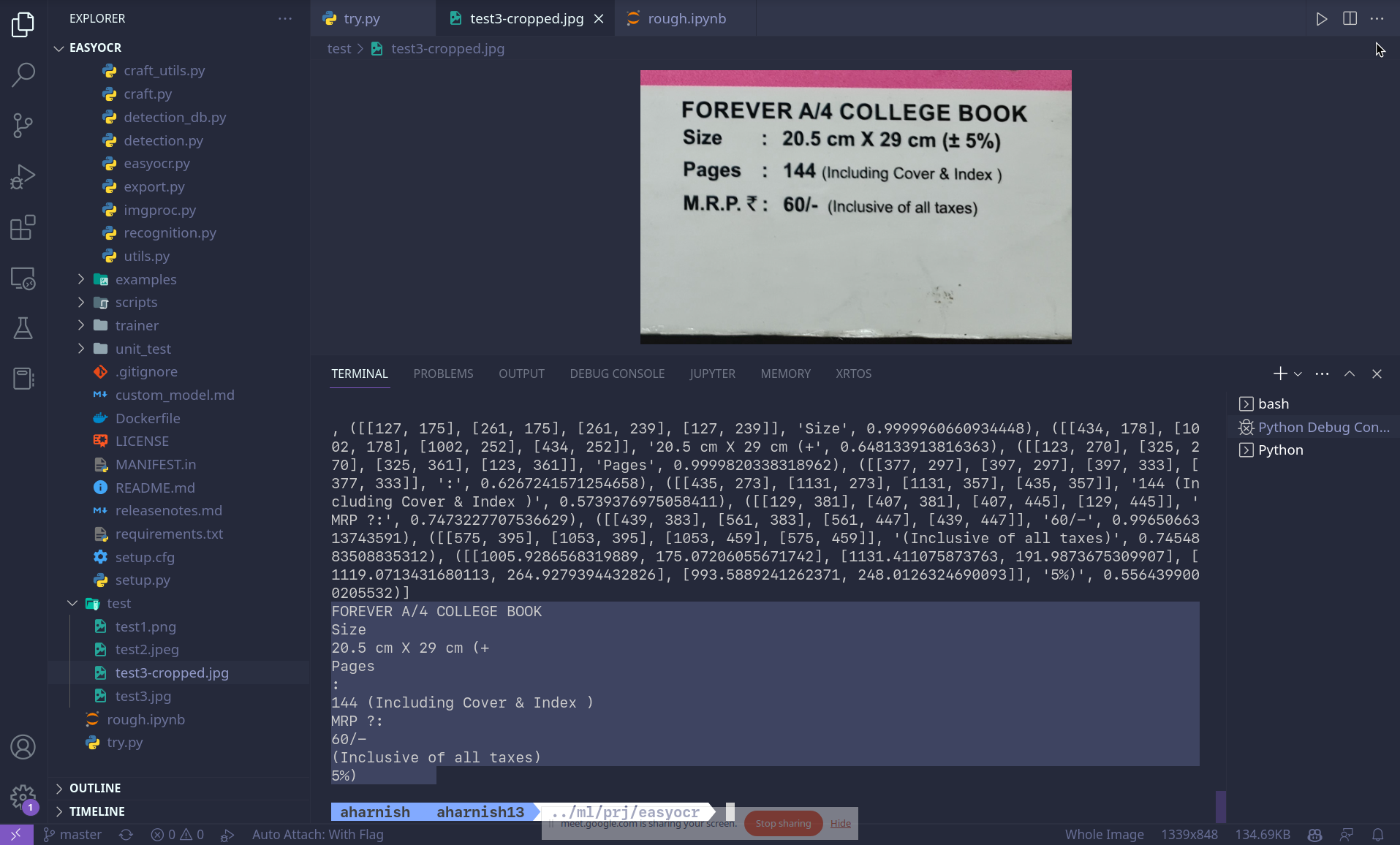
| **Name** | **Enrolment Number** |
| --- | --- |
| Aharnish Pithva | AU2040022 |
| Jevin Jivani | AU2040051 |
| Astha Bhalodiya | AU2040067 |
| Yug Patel | AU2040181 |

**1) Tasks Performed in the week.**

We searched libraries capable of extracting text data from images, including the position and size of each text element. As a result of our investigation, we discovered an open-source library named easyOCR for python. The library can be used for image text extraction, providing the output of an array containing the words and their respective positions.

**2) Outcomes of the tasks performed.**

We have received the following results, but with some spelling inaccuracies and unrecognised text.



**3) Tasks to be performed in the upcoming week.**

We are exploring the possibility of grouping strings together (clustering) based on their coordinates to search classified text. It should be noted that the classification process may be performed using a bag-of-words approach, which places little emphasis on the order or composition of sentences.