

Agroknow Data Team Assignment

At Agroknow, we are constantly looking to increase our data coverage by always including new sources of information.

One such source is the Japanese food safety authority that announces its border rejections on its webpage:

https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/shokuhin/yunyu_kanshi/index_00017.html under the **Recent Cases of Violation of the Food Sanitation Act** block.

These announcements are in the form of xls files and are published on a monthly basis.

The goal of this assignment is split into 3 subtasks:

- The announced xls should be collected and ideally the collection process should be easily automated,
- The collected xls should be transformed to an internal JSON schema, there are no limitations to the schema,
- A dataset should be created from the transformed JSON files and a classifier should be trained with the goal of identifying the reason behind the border rejection (CONTENTS OF VIOLATION column).

There are no limitations in the submission of the assignment. Although the end to end implementation is encouraged, one can choose to implement the first and the second subtask and present an approach for the third.

The suggested format of the submission is that of a github repository, shared with [mihalispap](#). However a zip file is also acceptable.

For any questions during the assignment you can contact: mihalispapak@agroknow.com