Industry Attachment/ Industry Project

**SG United Programme (SGUS): Upskill in Business Analytics**

**Specialist Diploma in Business Analytics**

**AY20/21, SGUS Aug Intake1**

Final Report

| **Project Title: Text classification for primary school model** | |
| --- | --- |
| **Company (for IA only): Apventure** | |
| **Student Name: Ng Shuling** | **Adm No: 2070694a** |
| **Supervisor Name: Ethan** | |
| **Liaison Officer Name (for IA only): Lim Kok Yau** | |

Abstract

[Enter your abstract here.]

To build a text classification model based primary science school topics

INTRODUCTION

Background

-opportunity for students/parents to experience the teaching through video recordings from our tutors answering the questions.

-to create a lead funnel for future users to be converted to paying users.

-get as much keywords from test papers that will help increase searchability

to our platform for relevant keywords. e.g test papers, math subjects, algebra questions,

Objectives

TO create a free resources website for students to learn and revise their schoolwork.

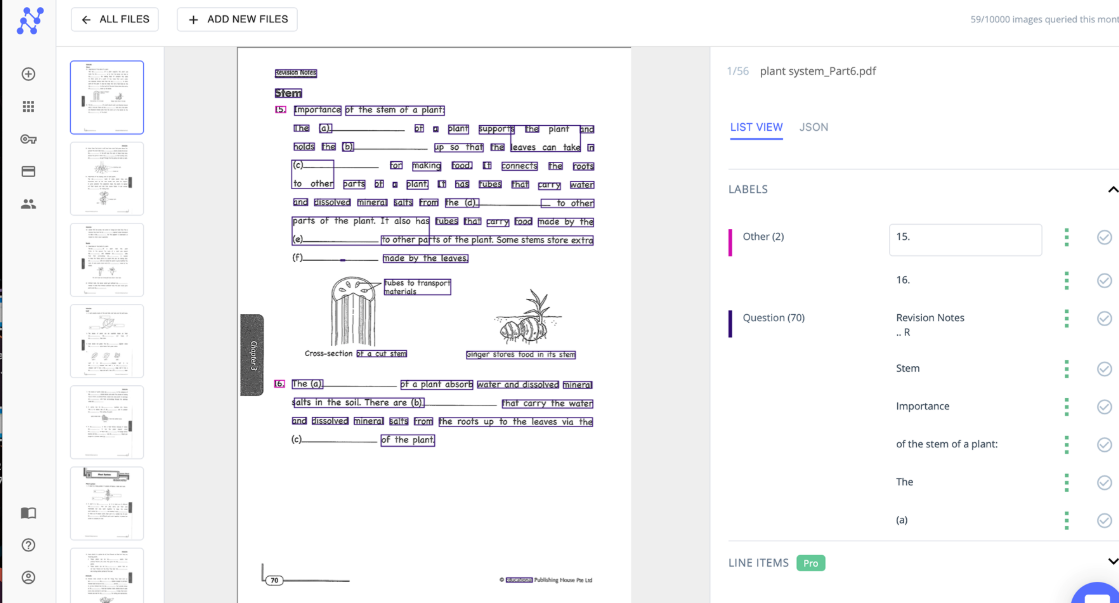
PROJECT DESCRIPTION

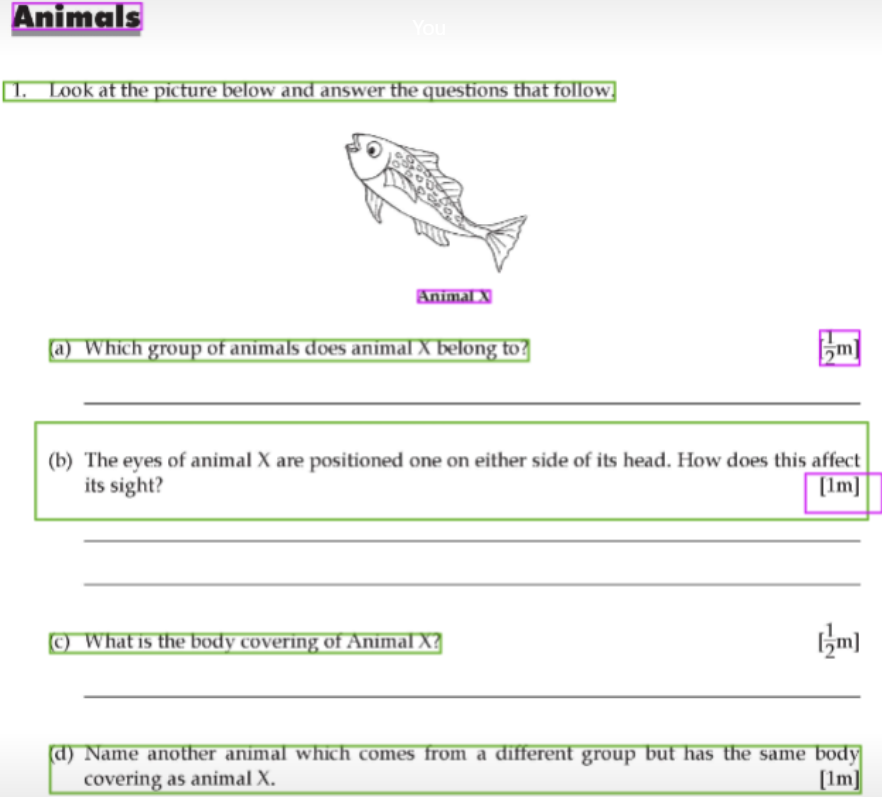
Using text mining,AI, supervised and unsupervised machine learning to create a text classification model for primary school subjects. Data extracted from past year papers are then fed into the model and placed under the correct classification.

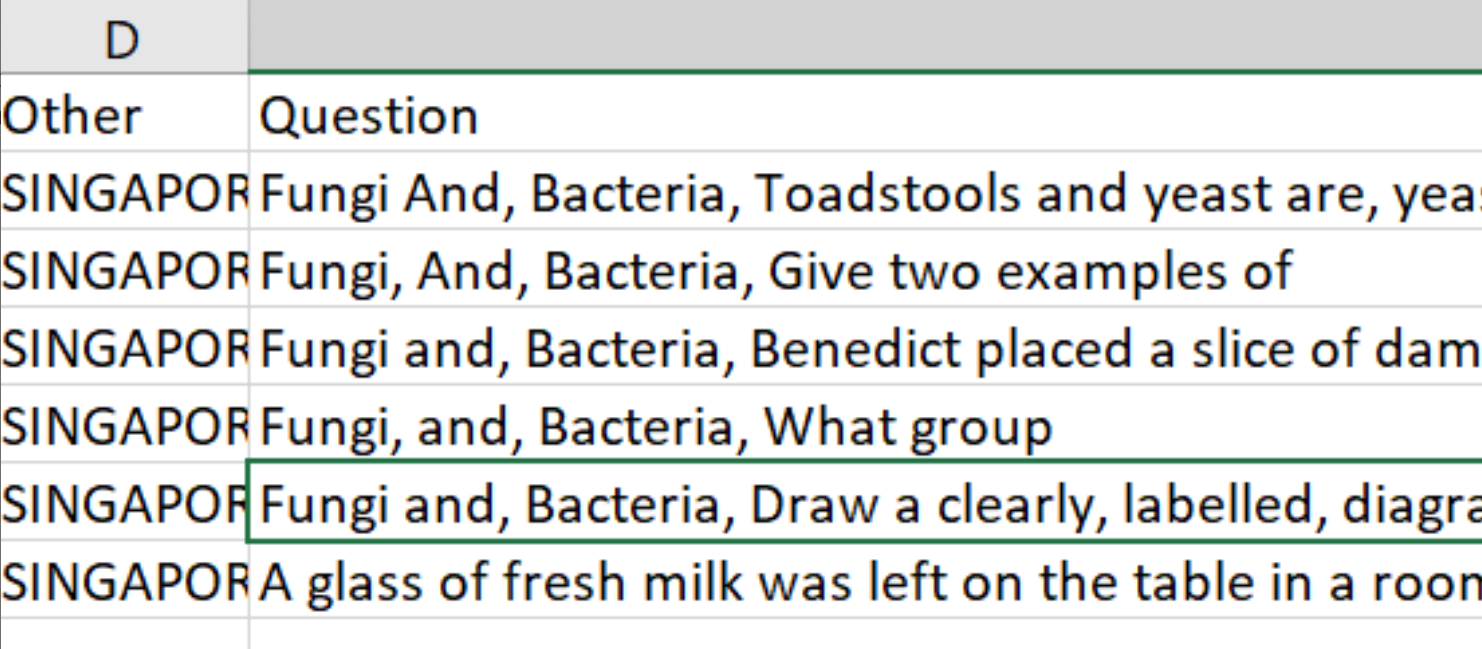
PROJECT DEVELOPMENT

[Describe the methodology, process, etc, to achieve the project objectives/deliverables. Include relevant screenshots. ]

Nanonet OCR was used to extract text data from past year papers and education papers. the data was then separated into training data and testing data. Data is first split into separate pdfs. Then a training model is created in nanonet to train the model which data we required. 2 Labels Question and others were created.

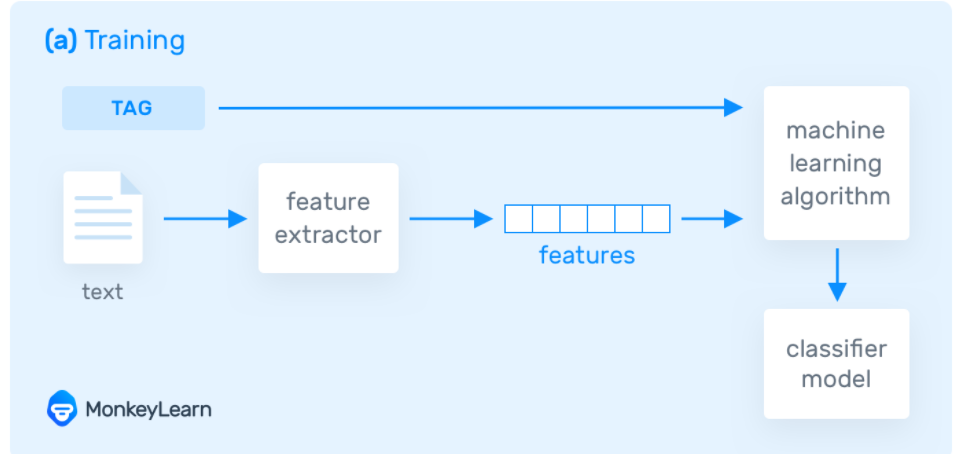


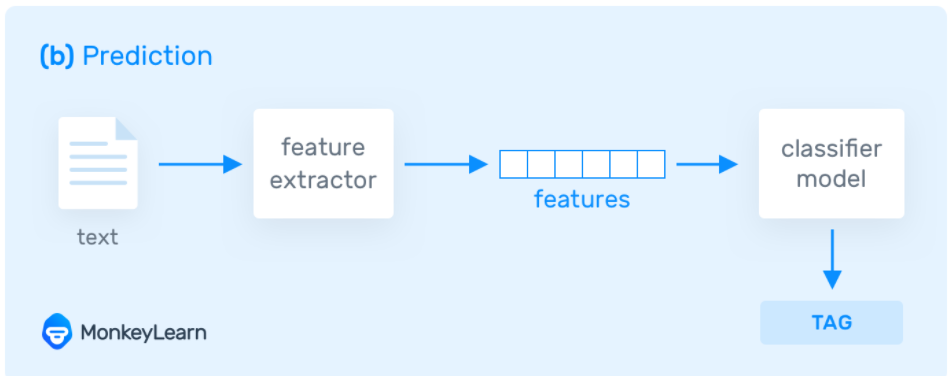


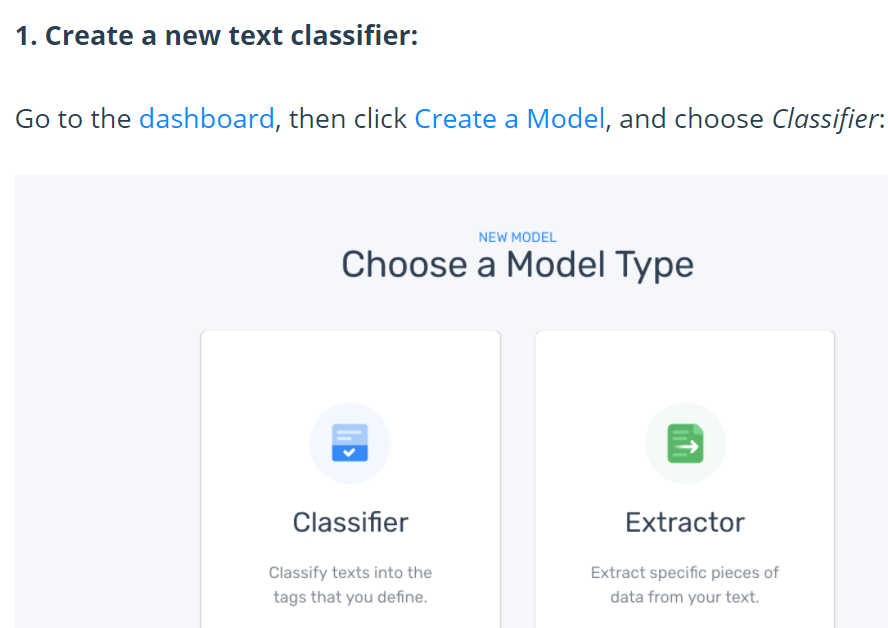


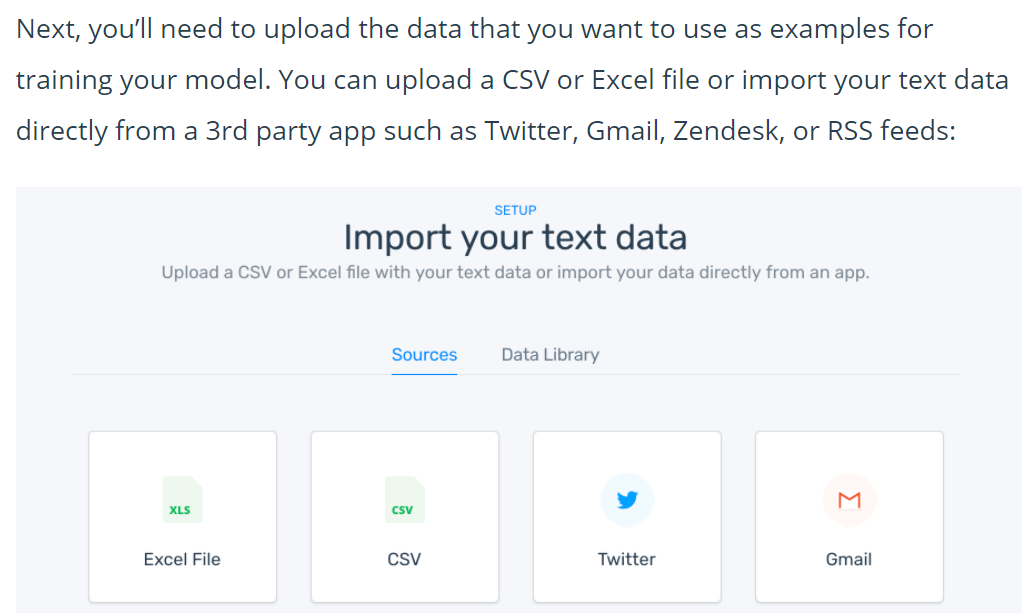
The new data is fed into the model and we extract the data into a csv file.

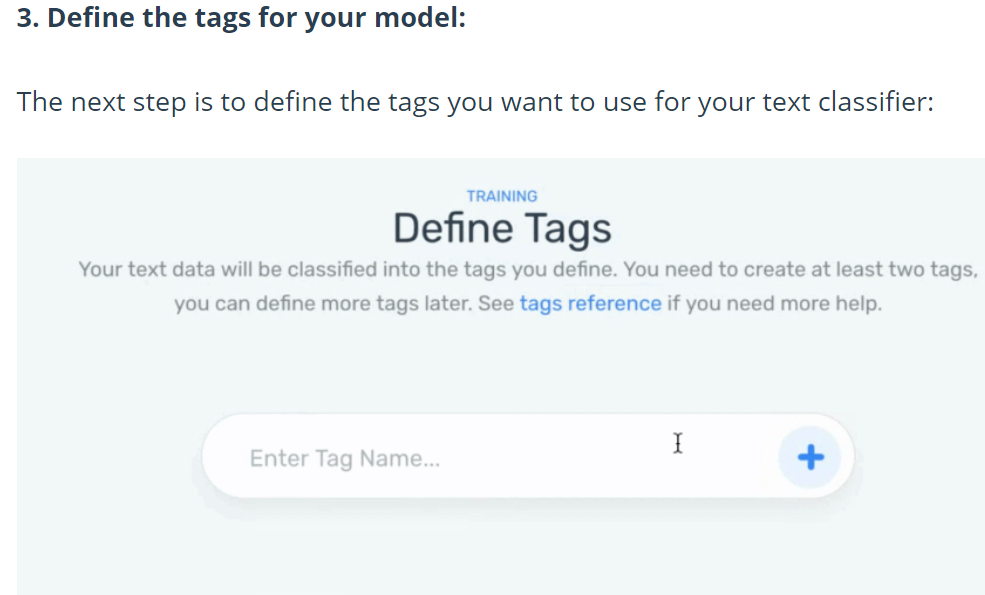
We used several applications such as Orange, Monkey Learn and KNIME to build the text classification model.

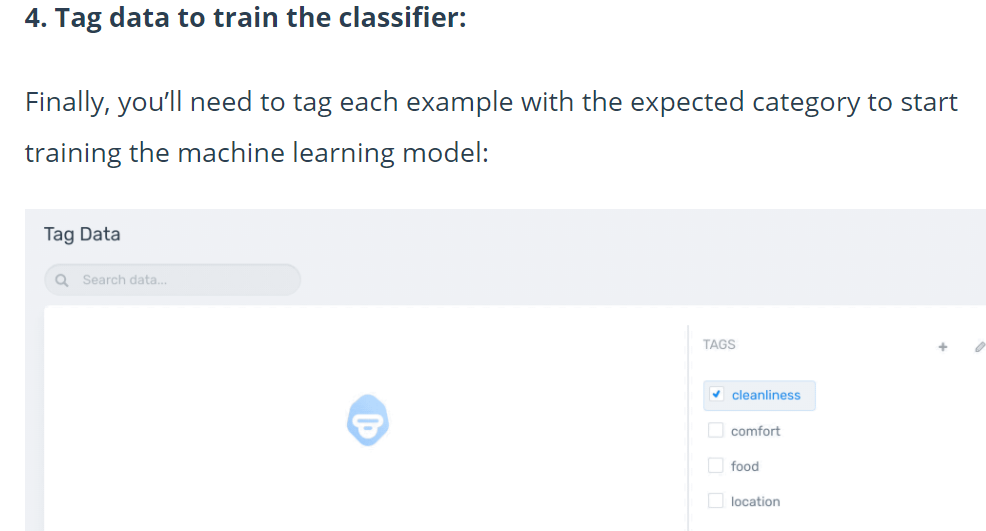


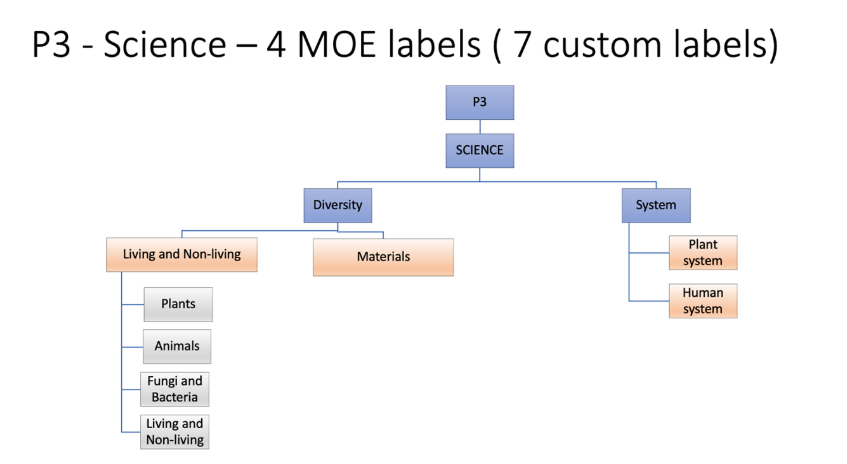


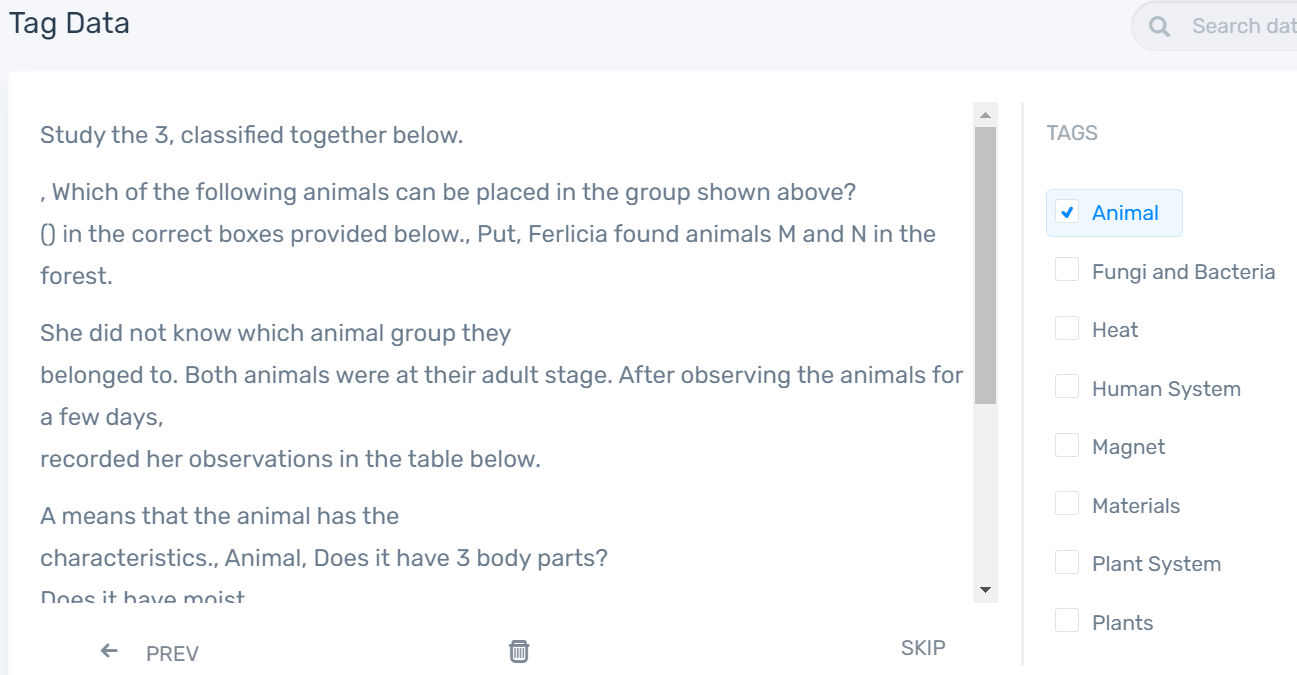








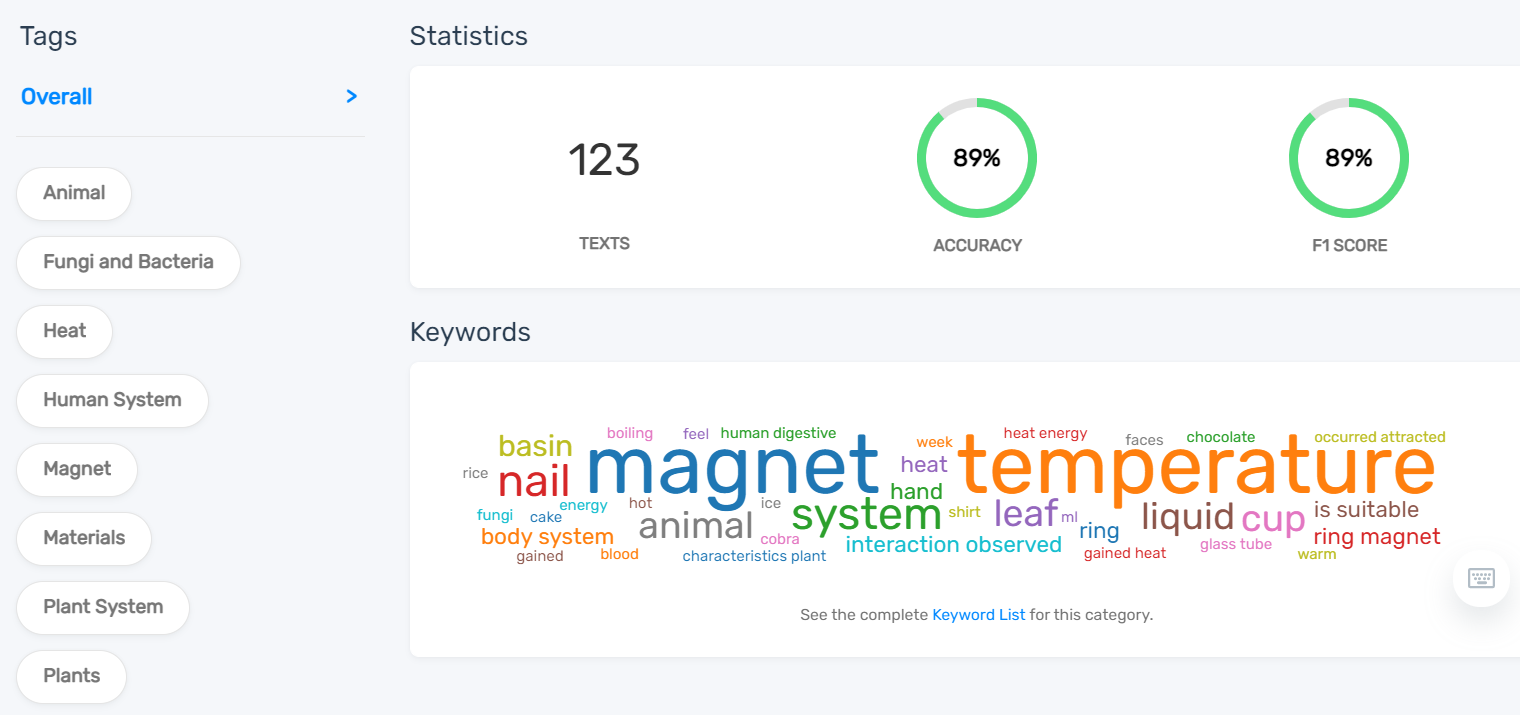


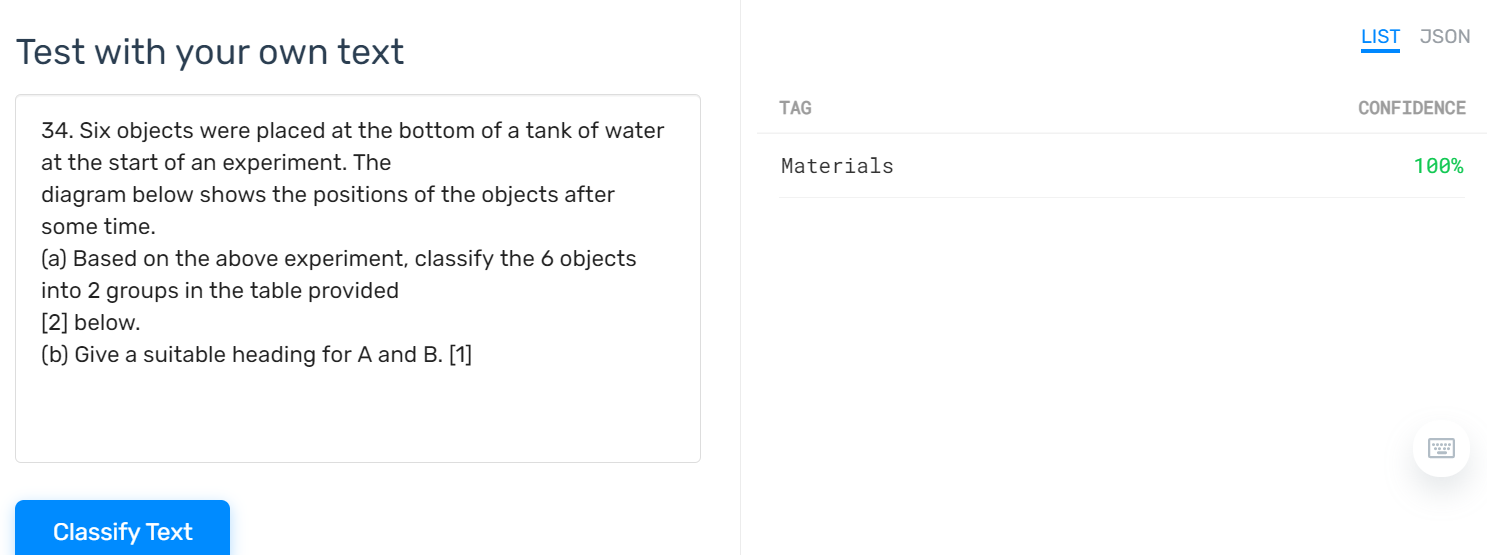


This is the model i build based on data extracted from past year papers

RESULTS

[Include description, analysis, and relevant screenshots]





I keyed in a sample to test my model and the text is classified as Materials which is correct. To achieve a higher accuracy for my model, I would need to train more data.

| Question | Classification | Confidence |
| --- | --- | --- |
| Nigel rinsed his mouth thoroughly. Next, he put a large chunk of plain white bread into his mouth and chewed it slowly and thoroughly. He was unable to detect any taste until a few seconds later., plain white bread, Why did Nigel rinse his mouth?, How did the plain bread taste after a few seconds?, Why did the bread taste like this?, Which part of Nigel's body enabled him to, the bread?, Nigel finally swallowed the chunk of bread. Outline the path taken by the bread through the digestive system. | Fungi and Bacteria | 1 |
| Tom tried to put a metal cube into a glass beaker and it fitted into the glass beaker nicely, as shown in the diagram below., cube, glass beaker, Tom then took out the metal cube and heated it over a fire for 10 minutes, and immediately tried to put the metal cube into the glass beaker., Would he be able to put it into the glass beaker? Why?, What can he do to the glass beaker to fit in the, cube | Magnet | 0.615 |
| Study the two animals shown below., Animal X, Animal Y, How are Animals X and Y similar in the way they reproduce?, How are Animals X and Y different in their body coverings? | Animal | 1 |
| Fill in the blanks with the correct answers., Mammals, to their young., They produce, for their young to feed on. They breathe, through their, The diagrams below show two animals, P and Q., P, Animal Q, Complete the table below using the correct words., P, Animal Group, Fish, Outer covering, Hair | Animal | 1 |
| Plant A is a fern that grows on the trunk of a tree., trunk of tree, plant A (fern), Fill in each blank with a suitable word., Plant A has a, stem which grows around, the tree trunk for support., It climbs up the tree trunk and spread, its leaves so that its leaves can get enough, to make food., The diagram below shows plant B., B, pole, state the difference in the way plants A, reproduce. | Plant System | 0.883 |

The table above is a batch sample for my model to test whether random data is sorted correctly and the data is labelled correctly.

CONCLUSION

[Provide a conclusion/future work/ recommendations here.]

Using applications like Monkeylearn enables non-tech users to build their text classifiers with Machine Learning models without coding. Organising and sorting unstructured data is extremely time consuming. However with Machine Learning models, these data can be sorted within a few minutes and it is consistent with its accuracy.

Acknowledgements

I thank and acknowledge Apventure for the learning experience.

References

https://monkeylearn.com/text-classification/

APPENDIX (if any)