

Datamininghw1

January 31, 2022

Name:

Mohith Krishna Behata

Email:

mbm2b@mst.edu

Course:

CS 5402

Assignment:

Programming Assignment 1

Date:

2022-01-27

GitHublink:

<https://git-classes.mst.edu/mbm2b/dataminingprog1>

```
[1]: # Imported for data management (dataframes)  
import pandas as pd  
  
# Imported to create word clouds  
# This package needs part of anaconda and needs to be installed  
# conda install -c conda-forge wordcloud  
from wordcloud import WordCloud  
  
# Imported to allow for the display of word clouds  
import matplotlib.pyplot as plt
```

0.1 Concept Description:

The goal of this task is to identify the differences and similarities between the words used to describe the definitions of data mining and artificial intelligence. We will examine the differences in the language used to discuss data mining and the language used to discuss artificial intelligence. For finding the similar language and different language used between data mining and artificial intelligence we would generate a image with help of matplotlib and wordcloud.

0.2 Data Collection:

Data has been collected using the following procedure:

First, open Google incognito mode because the results should not be based on the previous search criteria of the machine or account.

Google the terms “What is Data Mining” and “What is Artificial Intelligence” separately..

Examine the sites that provided the returned results and choose the ones that produced results for both search phrases.

Try to search the two phrases from similar sites in order to understand how different each one of them expresses the detail about the phrase

Browse the corresponding results pages for each search phrase from the sites and try to find a block of text that began with something like “Data Mining is...” or “Artificial Intelligence is...”.

Record text blocks are used as examples, along with the page’s link, the search phrase, the date the results were retrieved, and the name of the site from which the results were retrieved.

The results for each search phrase are saved in one excel file per each search phrase.total two excel files one for data mining and one for artificial intelligence

0.3 Example Description:

Category

Examples are recorded with appropriate labels indicating the search phrase. The Category is a nominal data type.

PullDate

The date the search result was obtained. The pull date is a date.

Source

Recorded the name of the sites that returned the text block about the search phrases. The source is a character string.

Link

The link that was returned as a result. The link is a character string.

Text

The text block that is was selected from the page that was returned as a result of the search. The text is a character string.

0.4 Data Import and Wrangling:

The results of each search phrase has been read from the respective Microsoft Excel Spreadsheets (xlsx) into separate dataframes. Careful attention is paid to make sure the data is read in as character strings.

```
[2]: #With help of pandas the code below helps in reading the excel file that has
      ↳been imported into jupyter files
```

```
#This code will store excel file (datamining) into data_mining
```

```
data_mining = pd.read_excel("DataMining.xlsx",dtype=str)
```

```
#This code will store excel file (Artificialintelligence) into
↳data_Intelligence
```

```
data_intelligence=pd.read_excel("Artificialintelligence.xlsx",dtype=str)
```

```
[3]: #show data mining dataset
data_mining
```

```
[3]:
```

	Catogery	Pull Date	Source \
0	Data Mining	2022-01-27 00:00:00	IBM
1	Data Mining	2022-01-27 00:00:00	ORACLE
2	Data Mining	2022-01-27 00:00:00	STUDY
3	Data Mining	2022-01-27 00:00:00	DATAVERSITY
4	Data Mining	2022-01-27 00:00:00	IBERDROLA
5	Data Mining	2022-01-27 00:00:00	SIMPLILEARN
6	Data Mining	2022-01-27 00:00:00	INDEED
7	Data Mining	2022-01-27 00:00:00	GURU99
8	Data Mining	2022-01-27 00:00:00	SAP
9	Data Mining	2022-01-27 00:00:00	COGNIZANT
10	Data Mining	2022-01-27 00:00:00	JAVATPOINT
11	Data Mining	2022-01-27 00:00:00	ANALYTICSVIDHYA
12	Data Mining	2022-01-27 00:00:00	TECHOPEDIA
13	Data Mining	2022-01-27 00:00:00	TALEND
14	Data Mining	2022-01-27 00:00:00	BRITANNICA

```
Link \
```

0	https://www.ibm.com/cloud/learn/data-mining
1	https://docs.oracle.com/database/121/DMCON/GUI...
2	https://study.com/academy/lesson/what-is-data-...
3	https://www.dataversity.net/what-is-data-mining
4	https://www.iberdrola.com/innovation/data-mini...
5	https://www.simplilearn.com/what-is-data-minin...
6	https://www.indeed.com/career-advice/career-de...
7	https://www.guru99.com/data-mining-tutorial.html
8	https://www.sap.com/westbalkans/insights/what-...
9	https://www.cognizant.com/us/en/glossary/data-...
10	https://www.javatpoint.com/data-mining
11	https://www.analyticsvidhya.com/blog/2021/05/i...
12	https://www.techopedia.com/definition/1181/dat...
13	https://www.talend.com/resources/what-is-data-...

14 <https://www.britannica.com/technology/data-mining>

Text

```
0 Data mining, also known as knowledge discovery...
1 Data mining is a technique that discovers prev...
2 Companies amass a great deal of data in the co...
3 Data Mining is an older (and now allied) subse...
4 Data mining is an automatic or semi-automatic ...
5 Typically, when someone talks about "mining," ...
6 Data mining is the process of examining data s...
7 Data Mining is a process of finding potentiall...
8 Data mining is the process of extracting usefu...
9 Data mining is the process of examining large ...
10 The process of extracting information to ident...
11 Data Mining", that mines the data. In simple ...
12 Data mining is the process of analyzing hidden...
13 Data mining isn't a new invention that came wi...
14 data mining, also called knowledge discovery i...
```

```
[4]: #show artificial intelligence dataset
data_intelligence
```

```
[4]:
```

	Catogery	Pull date	Name \
0	Artificial Intelligence	2022-01-27 00:00:00	IBM
1	Artificial Intelligence	2022-01-27 00:00:00	ORACLE
2	Artificial Intelligence	2022-01-27 00:00:00	STUDY
3	Artificial Intelligence	2022-01-27 00:00:00	DATAVERSITY
4	Artificial Intelligence	2022-01-27 00:00:00	IBERDROLA
5	Artificial Intelligence	2022-01-27 00:00:00	SIMPLILEARN
6	Artificial Intelligence	2022-01-27 00:00:00	INDEED
7	Artificial Intelligence	2022-01-27 00:00:00	GURU99
8	Artificial Intelligence	2022-01-27 00:00:00	SAP
9	Artificial Intelligence	2022-01-27 00:00:00	COGNIZANT
10	Artificial Intelligence	2022-01-27 00:00:00	JAVATPOINT
11	Artificial Intelligence	2022-01-27 00:00:00	ANALYTICSVIDHYA
12	Artificial Intelligence	2022-01-27 00:00:00	TECHOPEDIA
13	Artificial Intelligence	2022-01-27 00:00:00	TALEND
14	Artificial Intelligence	2022-01-27 00:00:00	BRITANNICA

Link \

```
0 https://www.ibm.com/design/ai/basics/ai/
1 https://www.oracle.com/artificial-intelligence...
2 https://study.com/academy/lesson/what-is-artif...
3 https://www.dataversity.net/artificial-intelli...
4 https://www.iberdrola.com/innovation/what-is-a...
5 https://www.simplilearn.com/tutorials/artifici...
6 https://www.indeed.com/career-advice/career-de...
```

```

7  https://www.guru99.com/artificial-intelligence...
8  https://www.sap.com/insights/what-is-artificia...
9  https://www.cognizant.com/us/en/glossary/artif...
10 https://www.javatpoint.com/artificial-intellig...
11 https://www.analyticsvidhya.com/blog/2021/09/i...
12 https://www.techopedia.com/definition/190/arti...
13 https://www.talend.com/resources/what-is-artif...
14 https://www.britannica.com/technology/artifici...

```

Text

```

0  Any system capable of simulating human intelli...
1  In the simplest terms, AI which stands for art...
2  Artificial intelligence (AI), in its simplest ...
3  "Artificial intelligence (AI) makes it possibl...
4  Artificial intelligence has long since abandon...
5  Artificial Intelligence is a method of making ...
6  Artificial intelligence is valuable for increa...
7  AI (Artificial Intelligence) is the ability of...
8  An early definition of artificial intelligence...
9  AI is the area of computer science that focuse...
10 The Artificial Intelligence tutorial provides ...
11 We have come a long way in the field of Machin...
12 Artificial intelligence (AI), also known as ma...
13 Artificial Intelligence (AI) is waking up and ...
14 artificial intelligence (AI), the ability of a...

```

```

[5]: # Convert all Text attribute values to text strings.
str_mining = ''.join(list(data_mining['Text']))

# Replace any Na values with a blank text string.
str_intelligence = ''.join(list(data_intelligence['Text']))

# We need to use .fillna('') if there are empty spaces in the dataset

# for the the current datasets no need to fill na in any cell because data is
→available for all attributes

```

0.5 Exploratory Data Analysis:

Word Clouds are generated for each search result.

```

[7]: #The code below will generate word cloud
wordcloud_mining = WordCloud().generate(str_mining)
wordcloud_intelligence = WordCloud().generate(str_intelligence)

```

```

[8]: #display word cloud of phrase "data mining" with matplotlib
plt.figure(figsize=(15,8))

```

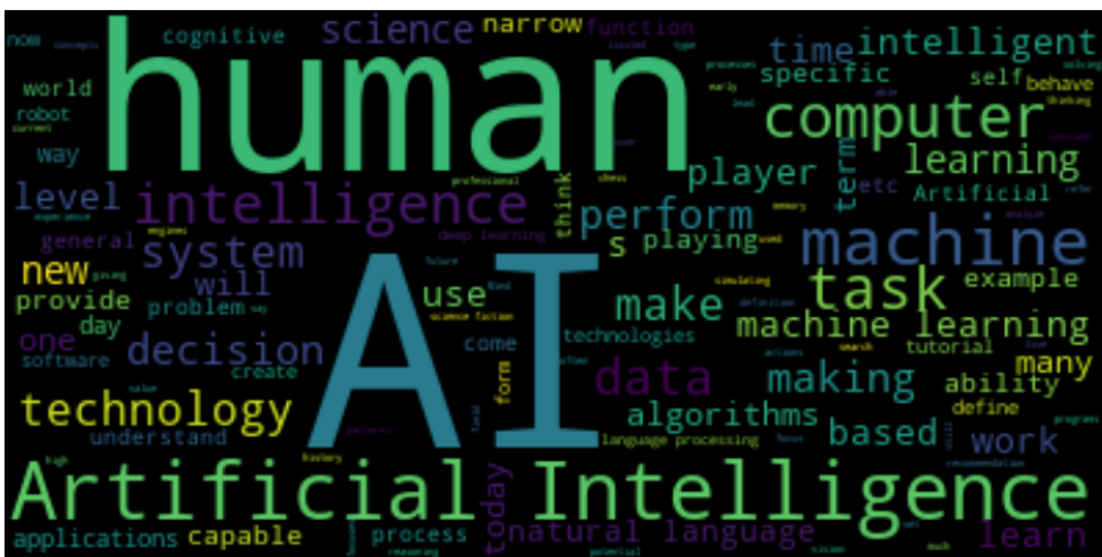
```
plt.axis('off')
plt.imshow(wordcloud_mining, interpolation='bilinear')
```

```
[8]: <matplotlib.image.AxesImage at 0x19c300b2190>
```



```
[9]: #display word cloud of phrase "artificial intelligence" with matplotlib
plt.figure(figsize=(15,8))
plt.axis('off')
plt.imshow(wordcloud_intelligence, interpolation='bilinear')
```

```
[9]: <matplotlib.image.AxesImage at 0x19c30d36d90>
```



0.6 Mining or Analytics:

Not applicable.

0.7 Evaluation:

Not applicable.

0.8 Results:

The language used for Data Mining is different compared to that of artificial intelligence.

The Most concentrated words used for defining data mining by various sites in the internet are data, data mining, pattern, process, information, set, business, technique, large, find etc.

The Most concentrated words used for artificial intelligence by various sites in the internet are Artificial intelligence, human, computer, system, human, algorithm, technology, decision, time, new etc.

Some words that appeared to be common in data mining and artificial intelligence are machine, data, intelligence, machine learning, algorithm, science etc.

0.9 References:

Anaconda Cloud. Wordcloud. Retrieved (2022, January 27) from <https://anaconda.org/conda-forge/wordcloud>

Muller, Andreas. (2021). A little word cloud generator in python. Retrieved (2022, January 27) from https://github.com/amueller/word_cloud

Python pool. (2021). How to change matplotlib size?. Retrieved (2022, January 27) from <https://www.pythonpool.com/matplotlib-figsize> Python guide. (2021). Python concatenate list. Retrieved (2022, January 27) from <https://pythonguides.com/python-concatenate-list/>