

# Programming Assignment 1 - Data Mining

January 25, 2020

## 1 Programming Assignment 1 - Data Mining

Author: Skylar Trendley Course: CS5402 Date: January 25, 2020

```
In [31]: #Imported for data management (dataframes)
import pandas as pd

#Imported to create word clouds
from wordcloud import WordCloud

#Imported to allow for the display of word clouds
import matplotlib.pyplot as plt

#read in the .csv file
df_mining=pd.read_csv("data_mining.csv",dtype=str)

#turn any instances of floats into strings
df_mining = df_mining.applymap(str)

#join the columns of text to make a "word glob"
str_mining=''.join(list(df_mining['Text']))

#generate the word cloud counting common occurrences
wordcloud_mining=WordCloud().generate(str_mining)

#display the word cloud
plt.figure(figsize=(10,8))
plt.axis('off')
plt.imshow(wordcloud_mining,interpolation='bilinear')
```

```
Out[31]: <matplotlib.image.AxesImage at 0x7f744eec08d0>
```





### 1.0.5 Exploratory Data Analysis:

N/A

### 1.0.6 Mining or Analytics:

Wordcloud and matplotlib allowed for the data to be transformed into a visual format. In the visualization, the largest words are the words that appear the most, and as they appear smaller in size, their occurrences in the follow a linear correlation.

### 1.0.7 Evaluation:

N/A

### 1.0.8 Results:

The language used for data mining and data science appear to be very similar. Common occurrences in both include words such as “information, extracting/extraction, useful, analysis, and statistics”. An interesting trend in the data science category is that it often mentioned data mining in its definition, whereas data mining hardly mentions data science. This could be due to the fact that data mining is a field of data science. Data science was often described to be multidisciplinary and an umbrella term for subcategories in its definition, whereas data mining was often more specific in what it actually meant.

### 1.0.9 References:

Websites used to collect Data: <https://www.datasciencecentral.com/profiles/blogs/difference-of-data-science-machine-learning-and-data-mining> <https://data-flair.training/blogs/data-mining-and-data-science/> <https://www.datasciencegraduateprograms.com/data-mining/> <https://www.houseofbots.com/news-detail/11973-1-clarifying-differences-between-data-analysis-data-mining-data-science-machine-learning-and-big-data> <https://www.newtechdojo.com/what-is-data-mining/> <https://www.edureka.co/blog/what-is-data-science/> <https://www.coursehero.com/file/p5p65iv/What-is-Data-Science-Data-Science-is-a-blend-of-various-tools-algorithms-and/> <https://medium.com/tech-in-200-words/what-is-data-mining-understanding-in-200-words-cd6ddf4fcf11> <https://www.talend.com/resources/what-is-data-mining/> <https://www.educba.com/advantages-of-data-mining/> [https://pinformatics.tamhsc.edu/phpm631/ppt/lec1\\_ds.pdf](https://pinformatics.tamhsc.edu/phpm631/ppt/lec1_ds.pdf) <https://365careers.net/data-science-explained-predictive-modeling-data-mining-data-warehousing-querying/> <https://becominghuman.ai/8-key-differences-between-data-science-and-data-mining-674f09599df2> <https://www.microstrategy.com/us/resources/introductory-guides/data-mining-explained>

Imports used to shape and analyze data: <https://pandas.pydata.org/> <https://matplotlib.org/> [https://github.com/amueller/word\\_cloud](https://github.com/amueller/word_cloud)

In [ ]: