## Exercise 1 example ideas

Lecture 1 – slide 19

Use of Unsupervised Learning

* Outlier detection – to detect outlier in noisy datasets
* Detect fraud, find anomalies in the credit card usage patterns of customers
* Detect abnormal behaviour (malicious or unintentional) in systems/networks

Below examples taken from [ <https://datafloq.com/read/7-innovative-uses-of-clustering-algorithms/6224> ]

1. Fake news
2. **Spam filter**
3. Marketing and sales
4. **Classify network traffic**
5. **Document analysis**
6. Fantasy football and sports
7. **Hospital locations within a city**

More examples which might be better: <https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/>

Even more examples [ <https://blogs.oracle.com/datascience/introduction-to-k-means-clustering#:~:text=sample%20dataset%20here.-,Introduction%20to%20K%2Dmeans%20Clustering,represented%20by%20the%20variable%20K>. ]

* Behavioural segmentation:
  + Segment by purchase history
  + Segment by activities on application, website, or platform
  + Define personas based on interests
  + Create profiles based on activity monitoring
* Inventory categorization:
  + Group inventory by sales activity
  + Group inventory by manufacturing metrics
* Sorting sensor measurements:
  + Detect activity types in motion sensors
  + Group images
  + Separate audio
  + Identify groups in health monitoring
* Detecting bots or anomalies:
  + Separate valid activity groups from bots
  + Group valid activity to clean up outlier detection

More examples [ <https://dzone.com/articles/10-interesting-use-cases-for-the-k-means-algorithm> ]

## Spam email to potentially add

<https://ijcert.org/ems/ijcert_papers/V3I1210.pdf>

Email can be labelled as a spam email only depending on these properties:

• Uninvited Emails: E-mails that are received from contacts that are not known to the user.

• Bulk Mailing: The kind of emails that are sent in mass or bulk to multiple account holders at the same time.

• Unknown Mails: In this type of mails in the identity and the details of the sender are not revealed

or demonstrated.

## Main Kmeans fraud sources

<http://ijcsit.com/docs/Volume%206/vol6issue02/ijcsit20150602177.pdf>

<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.680.1195&rep=rep1&type=pdf>

<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.323.6870&rep=rep1&type=pdf#:~:text=Classification%20and%20Outlier%20Detection,monitored%20in%20the%20same%20network>.

<https://www.researchgate.net/publication/272863425_Fraud_Detection_in_Credit_Card_by_Clustering_Approach>