



# 7 Innovative Uses of Clustering Algorithms in the Real World



Whittaker / 6 min read.

, 2019



floq.to/Z4H2E



In order to optimize the website and to continuously improve



into the problem. Before diving into the innovative uses of clustering algorithms, I will first share an overview of the two algorithms.


## What is unsupervised learning?

Before we get started, let me first introduce the concept of unsupervised learning. Unsupervised learning is where you train a machine learning algorithm, but you don't give it the answer to the problem.

### 1) K-means clustering algorithm

The K-Means clustering algorithm is an iterative **process** where you are trying to minimize the distance of the **data** point from the average data point in the cluster.

#### **hierarchical clustering**

 hical clustering algorithms seek to **create** a hierarchy of clustered data points. The algorithm aims to minimize the number of clusters by merging those closest to one another using a distance measurement such as **Euclidean distance** for numeric clusters or **Hamming distance** for text.



## Here are 7 examples of clustering algorithms in action.

### 1. Identifying Fake News

Fake **news** is not a new phenomenon, but it is one that is becoming prolific.

What the problem is: Fake news is being created and spread at a rapid rate due to technology innovations such as social media. The issue gained attention recently during the 2016 US presidential campaign. During this campaign, the term Fake News was referenced an unprecedented number of times.

How clustering works: In a **paper recently published** by two **computer science** students at the University of California, Riverside, they are using clustering algorithms to identify fake news based on the **content**.

The goal of the algorithm is to identify the content of the fake news articles. The authors explain that in order to optimize the website and to continuously improve



You know the junk folder in your **email** inbox? It is the place where emails that have been identified as spam by the algorithm.

Many machine learning courses, such as Andrew Ng's famed Coursera course, use the spam filter as an example of unsupervised learning and clustering.

**What the problem is:** Spam emails are at best an annoying part of modern day marketing techniques, and at worst, an example of **people** phishing for your personal data. To avoid getting these emails in your main inbox, **email companies** use algorithms. The purpose of these algorithms is to flag an email as spam correctly or not.

**How clustering works:** **K-Means clustering techniques** have proven to be an effective way of identifying spam. The way that it works is by looking at the different sections of the email (header, sender, and content). The data is then grouped together.



groups can then be classified to identify which are spam. Including clustering in the classification



ss improves the accuracy of the filter to 97%. This is excellent news for **people** who want to be sure

are not missing out on your favorite newsletters and offers.



## Marketing and Sales

Personalization and targeting in marketing is big business.

This is achieved by looking at specific characteristics of a **person** and sharing campaigns with them that have been successful with other similar people.

**What the problem is:** If you are a business trying to get the best return on your marketing investment, it is crucial that you target people in the right way. If you get it wrong, you risk not making any sales, or worse, damaging your **Customer trust**.

**How clustering works:** Clustering algorithms are able to group together people with similar traits and likelihood to purchase. Once you have the groups, you can run tests on each group with different marketing copy that will help you better target your messaging to them in the future.

---

Interested in what the future will bring?  
Download our 2023 Technology

In order to optimize the website and to continuously improve



LITIGANT

☒ Subscribe to our newsletter.

[DOWNLOAD NOW](#)

## 4. Classifying network traffic

Imagine you want to understand the different types of traffic coming to your website. You are particularly **f**ocused in understanding which traffic is spam or coming from bots.

**the problem is:** As more and more **services** begin to use APIs on your application, or as your website **in**creases, it is important you know where the traffic is coming from. For example, you want to be able to block **in**valid traffic and double down on areas driving growth. However, it is hard to know which is which when it **in**volves **in**teractions to classifying the traffic.

**How clustering works:** K-means clustering is used to group together characteristics of the traffic sources. When the clusters are created, you can then classify the traffic types. The process is faster and more accurate than the previous Autoclass method. By having precise information on traffic sources, you are able to grow your site and plan capacity effectively.

## 5. Identifying fraudulent or criminal activity

In this scenario, we are going to focus on fraudulent taxi driver behavior. However, the technique has been used in multiple scenarios.

**What is the problem:** You need to look into fraudulent driving activity. The challenge is how do you identify what is true and which is false?

**How clustering works:** By **analysing the GPS logs, the algorithm** is able to group similar behaviors. Based on the characteristics of the groups you are then able to classify them into those that are real and which are fraudulent.

In order to optimize the website and to continuously improve



documents and classify it.

**How clustering works:** Hierarchical clustering has been used to solve this problem. The algorithm is able to look at the text and group it into different themes. Using this technique, you can cluster and organize similar documents quickly using the characteristics identified in the paragraph.

## 7. Fantasy Football and Sports

Ok so up until this point we have looked into different business problems and how clustering algorithms have been applied to solve them.

But now for the critical issues – fantasy football!

**What is the problem:** Who should you have in your team? Which players are going to perform best for your team and allow you to beat the competition? The challenge at the start of the season is that there is very little data available to help you identify the winning players.

**How clustering works:** When there is little performance data available to train your model on, you have an edge for unsupervised learning. In this type of machine learning problem, you can find similar players based on some of their characteristics. This has been done using K-Means clustering. Ultimately this means you can get a better team more quickly at the start of the year, giving you an advantage.

## How will you use clustering algorithms?

So there you have it, those were 7 innovative uses of clustering algorithms. As you can see, while the technique remains reasonably constant, you can apply it to many different scenarios.

Looking at the characteristics of different groups of data can help you make better predictions of behavior. In this scenario, the real value of the algorithms is to help you create the best possible groups of data.

Once you have a solid foundation of grouped data to work with, the opportunities become infinite.

Categories: [Artificial Intelligence](#)

In order to optimize the website and to continuously improve

