What is Association Rule Learning?

Supriyo Ain · Follow

5 min read · Jul 6, 2023

Before we dive deep into ***Association Rule Learning***, let me explain what ***Unsupervised Learning*** is. In this type of learning, the training data is unlabeled, i.e. the system tries to learn the information without a teacher. ***Association Rule Learning*** is an example of ***Unsupervised Learning***, it is used in Market Basket Analysis, Intrusion Detection, Web Usage Mining etc. This algorithm's goal is to dig into large amounts of data and discover interesting relations between attributes.

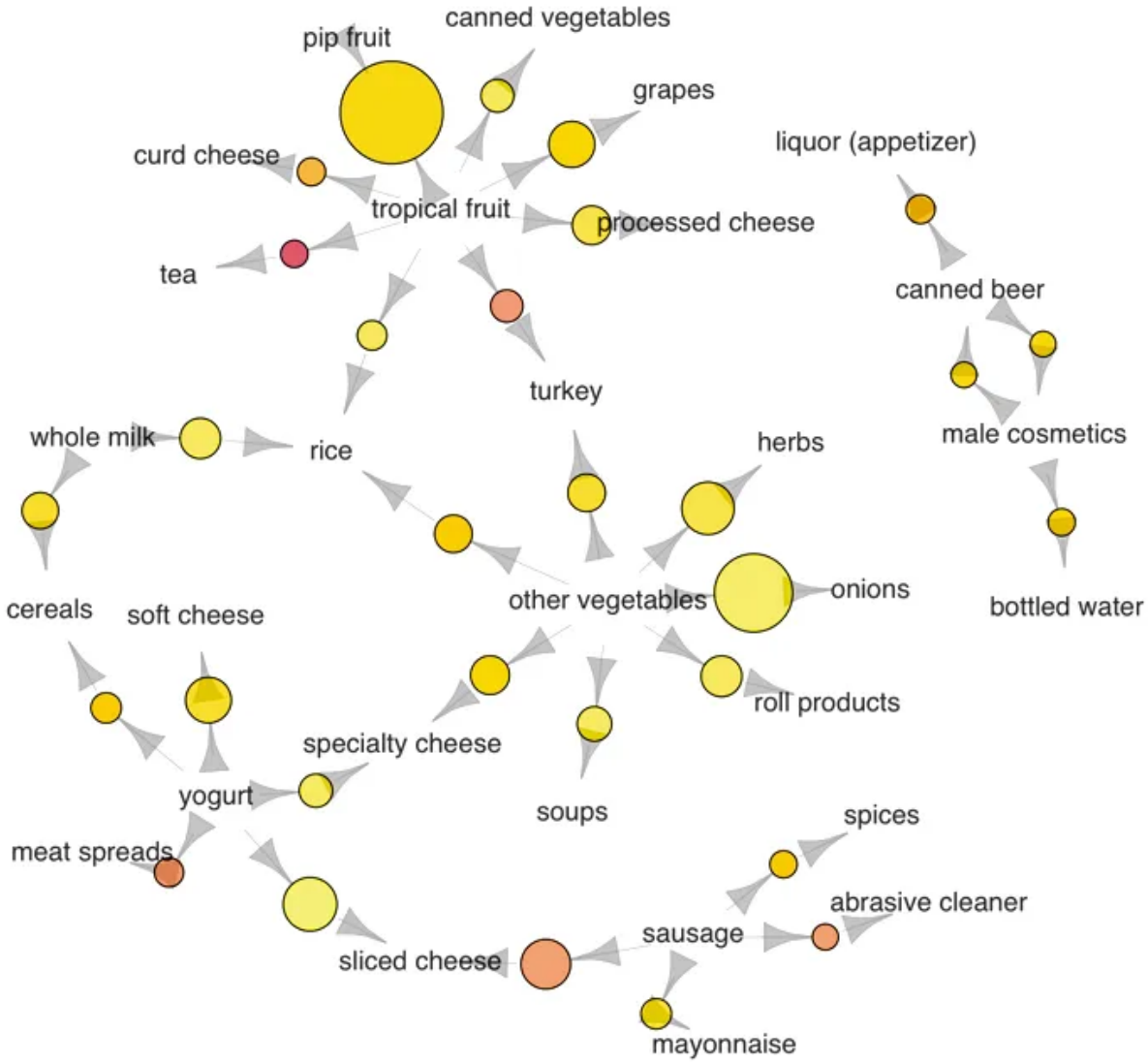
Why do we use Association Rule Learning?

Well, the answer lies in the task this algorithm performs. Say, you are the owner of a Supermarket and you want people to buy your products easily. What you can do is, you can run this algorithm on your sales log and find interesting relations between the items. For example, you find out that people who purchase milk and bread, also tend to purchase butter. Thus you may want to do the following things to improve the quality of your mart:

- You can place milk, bread and butter on the same shelf, so that buyers of one item would be prompted to buy another item.
- You can put milk, bread and butter on discount to increase your item sales.
- You can also target the buyers of milk or bread with the advertisement of butter.
- Or you can also combine bread and butter into a whole new product i.e. buttery bread with slightly milky flavor and then put it on sale.

Not only in increasing sales, association rules can also be used in other fields, for example, in medical diagnosis, understanding which symptoms tend to co-morbid can help to improve patient care and medicine prescription.

How does Association Rule Learning Works?



This algorithm counts the frequency of complimentary occurrences, or associations, across a very large dataset with over thousands of attributes. The goal is to find associations that take place together far more often than you would find in a random sampling of possibilities. So, to measure the associations between thousands of data items, there are several metrics. These metrics are given below:

- **Support —** This says how popular an itemset is, i.e. it is used to find the frequency of a certain itemset appearing in the dataset.

$$Support(A) = Frequency(A)$$

- **Confidence —** This says how likely item B is purchased when item A is purchased, expressed as (A -> B).

$$Confidence(A \rightarrow B) = \frac{Support(A \rightarrow B)}{Support(A)}$$

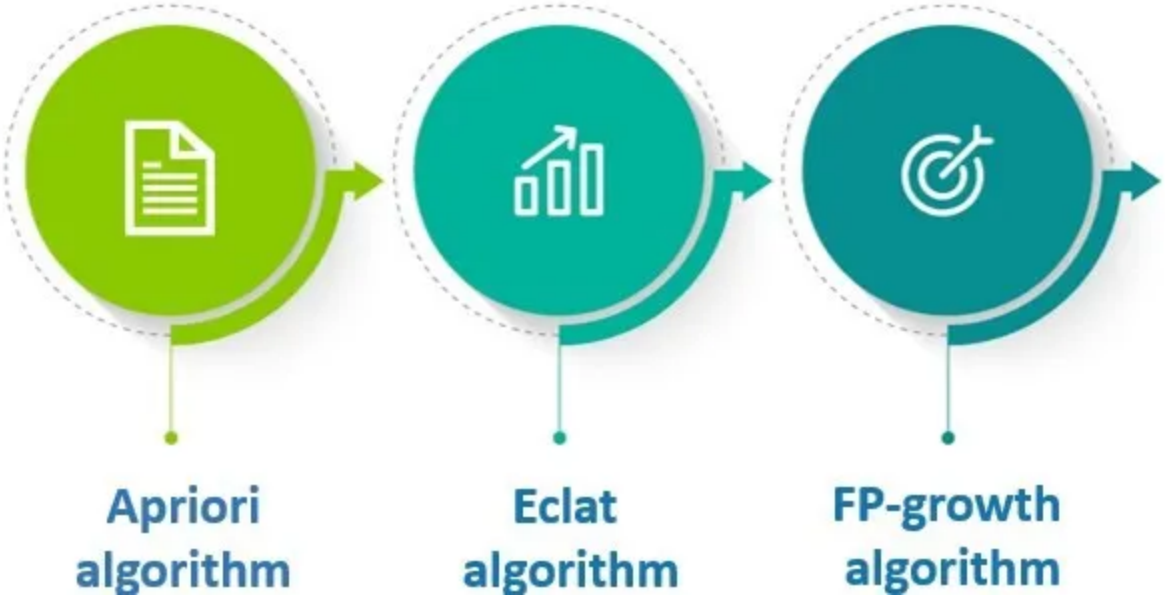
- **Lift —** This says how likely an item A is purchased while controlling how popular item B is.

$$Lift(A \rightarrow B) = \frac{Confidence(A \rightarrow B)}{Support(B)}$$

Lift has three possible values —

- **Lift = 1 —**The probability of occurrence of A and B is independent of each other.
- **Lift > 1 —**It determines the degree to which A and B are dependent on each other.
- **Lift < 1 —** It tells us that A is a substitute for B, which means A has a negative effect on item B.

Different types of Association Rule Learning



Association Rule Learning can be divided into three algorithms —

- **Apriori—** This algorithm uses frequent datasets to generate association rules. We apply an iterative approach or level-wise search where k-frequent itemsets are used to find k+1 itemsets. This algorithm uses a ***Breadth-First Search*** algorithm and ***Hash-Tree*** to calculate the itemset efficiently.

```
!pip install apyori
from apyori import apriori

transactions = []

#convert the dataset into a list of arrays and the data must be in strings

for i in range(len(df)):
    transactions.append([str(df.values[i, j]) for j in range(len(df[0]))])

rules = apriori(transactions = transactions,
                min_support = 0.003,
                min_confidence = 0.2,
                min_lift = 3,
                min_length = 2,
                max_length = 2)
```

```
results = list(rules)
```

- **Eclat —** Eclat algorithm stands for ***Equivalence Class Transformation.*** While the Apriori algorithm works in a horizontal sense imitating the ***Breadth-First Search*** of a graph, the ECLAT algorithm works in a vertical manner just like the ***Depth-First Search*** of a graph.. It performs faster execution than Apriori Algorithm.

```
transactions = []

#convert the dataset into a list of arrays and the data must be in strings

for i in range(len(df)):
    transactions.append([str(df.values[i, j]) for j in range(len(df[0]))])

rules = apriori(transactions = transactions,
                min_support = 0.003,
                min_length = 2,
                max_length = 0)

results = list(rules)
```

- **F-P Growth—** The F-P Growth algorithm stands for ***Frequent Pattern,*** and it is the improved version of the Apriori Algorithm. The FP-Growth Algorithm is an alternative way to find frequent item sets without using candidate generations, thus improving performance. It uses a ***Divide-and-Conquer*** strategy and the core of this method is the usage of a special data structure named ***Frequent-Pattern Tree (FP-tree)***, which retains the item set association information. The purpose of this frequent tree is to extract the most frequent patterns.

```
!pip install pyfpgrowth
import pyfpgrowth

transactions = []

#convert the dataset into a list of arrays and the data must be in strings

for i in range(len(df)):
    transactions.append([str(df.values[i, j]) for j in range(len(df[0]))])

patterns = pyfpgrowth.find_frequent_patterns(transactions, 2)
rules = pyfpgrowth.generate_association_rules(patterns, 0.7)
```

Applications of Association Rule Learning

- **Basket Data Analysis —** It is one of the most popular examples of Association Rule Learning. Big retailers use this algorithm to determine association between items or what customers are looking for.
- **Web Usage Mining —** Finding the hidden correlations is a powerful predictive tool to discover brand new security threats and network performance issues.
- **Medical Diagnosis —** With the help of association rules, it will be alot easier to cure patients, as it helps in identifying the probability of illness for a particular disease.

and many more...

Thanks for reading!

Machine Learning | Association Rule Learning | Unsupervised Learning

Written by Supriyo Ain

Follow

8 Followers

Hi, I am Supriyo Ain. I am a 3rd year Computer Science Engineering student. My interests lie in the field of AI, ML and DL, especially in NLP and LLMs.

More from Supriyo Ain

Supriyo Ain

Unsupervised Learning—Clustering Algorithms

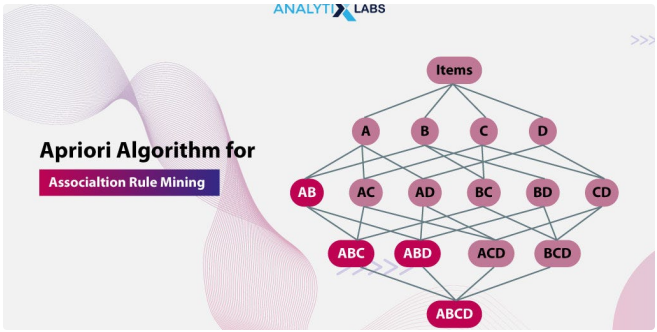
You have probably heard the quote—"Cluster together like stars." Cluster means a group of similar things or people positioned or...

6 min read · Jul 13, 2023

29

See all from Supriyo Ain

Recommended from Medium



AnalytixLabs

Apriori Algorithm In Data Mining: Implementation, Examples, and...

The accumulation of data has increased rapidly in recent times. The development in...

18 min read · Jan 24, 2024

1

Everton Gomede, PhD

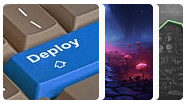
CARMA Algorithm: Bridging the Gap Between Association Rule...

Introduction

5 min read · Oct 6, 2023

5 1

Lists



Predictive Modeling w/ Python

20 stories · 931 saves

Practical Guides to Machine Learning

10 stories · 1094 saves

Natural Language Processing

1224 stories · 704 saves

The New Chatbots: ChatGPT, Bard, and Beyond

12 stories · 314 saves

BetterLife

Eclat Algorithm in Machine Learning

The Eclat algorithm is a frequent itemset mining technique used in machine learning...

3 min read · Sep 12, 2023

51

IUYasik

Market Basket Analysis & Apriori Algorithm using Zhang's Metric

Have you ever chuckled at the age-old adage that claims fathers who purchase diapers als...

9 min read · Oct 4, 2023

66

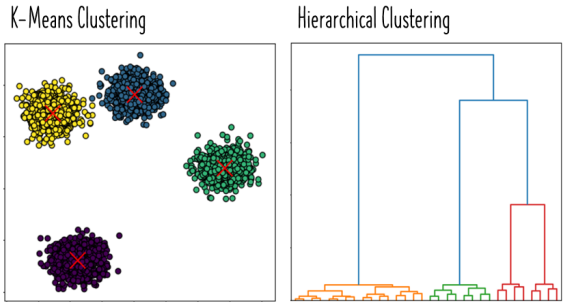
Idil Ismiguzel in Towards Data Science

A Guide to Association Rule Mining

Create insights from frequent patterns using market basket analysis with Python

· 10 min read · Apr 5, 2023

111 4



Mbali Kalirane

Hierarchical Clustering vs K-Means Clustering: How do the Clustering...

K-Means and hierarchical clustering are both techniques for clustering datapoints...

6 min read · Oct 16, 2023

106

See more recommendations