



Targeted Advertising

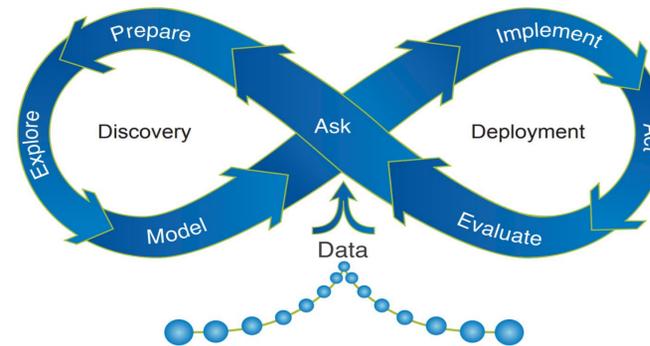
- Companies are always looking for more efficient methods of advertising
- Mediums of advertising in the 20th century such as newspapers, radios, and televisions were able to target interest groups, but not individuals. This is due to the lack of information accessible to companies about their consumers.
- The emergence of the internet and social media has filled that lack of knowledge and has opened the door for companies to finally target individuals through vast amounts of big data

Advertising on Facebook

Facebook is a social media platform for friends, loved ones, and acquaintances to keep in touch with one another and post photos and updates about their lives. With over one billion users, Facebook provides a large audience for companies to advertise to. Facebook, and social media in general, creates a unique opportunity for companies to showcase their products and services to people who would likely be interested in them. They are able to do this (advertise to a specific group of people) thanks to social media data mining. Corporations are able to access terabytes of user data at a time from Facebook in order to make conclusions about what demographics and types of users would be best for their advertisements to be shown. Big corporations often hire analytics companies to dredge through all the data from Facebook and make the conclusions for them. Smaller businesses also have an option to manually select demographics and groups they want their advertisements to appear to and let Facebook do that for them. The process for these small businesses is as simple as checking a few boxes on a website and providing an advertisement.



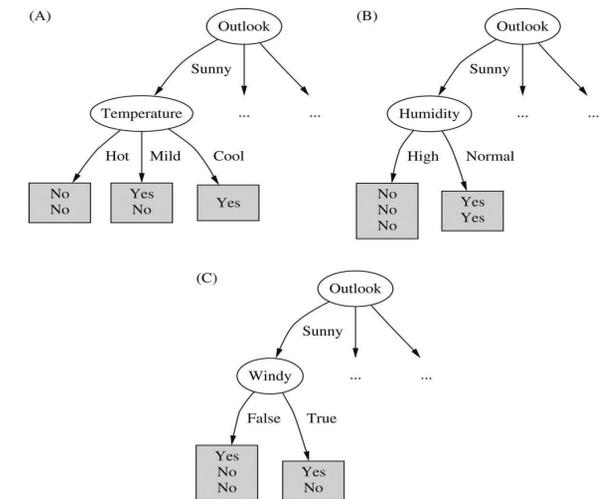
Cycle of Data Mining



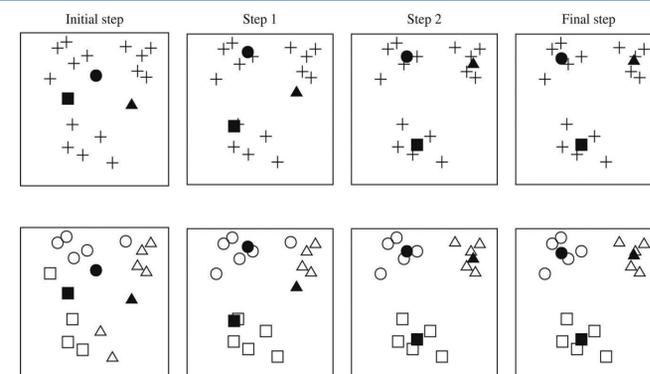
The first stage, usually done by data scientists, is to find different data sources, geared around answering a specific question, and figuring out a way to compile different forms of raw data from different sources into data that can be used as input for data mining. The second stage is to explore the data by using visualization tools to improve and refine the initial hypothesis and to transform the data to better answer our questions. The third stage is to model the data by creating and applying various analytical algorithms that will find patterns and draw predictions based on said data. More specifically, the three common types of data modeling are descriptive, predictive, and prescriptive. Thousands of models are created and by using machine learning algorithms, the most valid and effective models are found. The last stage is to deploy the best models by making strategic decisions, made by humans, and operational decisions, made by machines, which will answer the initial question. Finally, with continuous monitoring and measuring of the models, the success of the models' outcomes are evaluated and the four stages are altered slightly to progressively increase the effectiveness of the specific data mining application in question.

Decision Tree Algorithm

Decision trees, a predictive modeling technique, are tree-shaped diagrams where each branch represents a probable occurrence. Decision trees work by first placing an attribute at the root node and creating a branch for each possible value. This splits up instance sets into instance subsets. This process is recursively repeated for each branch, using only the instances that reach that part of the branch, until the recursive process terminates and the tree is finalized. The smaller the tree-diagram is, the better the decision tree is, and therefore the trees that reach recursive termination sooner are sought after. To find this tree, the attribute that produces the purest daughter nodes are selected. Purity is measured in bits and represents the expected amount of information needed to specify whether an instance should be classified as yes or no. The more pure a node is, the less bits of information are needed. The figure on the right provides a great example of selecting the best attribute.



K-Means Clustering Algorithm



Clustering, a descriptive modeling technique that groups similar records together, is a common technique because it is simple and effective. The classic clustering approach is called k-means. First, you define how many clusters should be searched for, this is the parameter k. Then, k amount cluster centers are selected at random points. Based on the ordinary Euclidean distance metric, all instances are then assigned to their closest cluster. Next, the centroids, or mean, of all the instances in each cluster are calculated. These centroids now serve as the new centers for their particular cluster. Lastly, the full process is repeated with the new cluster centers. This repetition continues until all points are consistently and repeatedly assigned to the same cluster, this means that the algorithm has converged and the cluster centroids have stabilized and are finalized. For situations in which the cluster count is unknown beforehand, different k sizes are tested to see which is most optimal. The figure on the left side visualizes the 4-step k-means algorithm.

Facebook's revenue and net income from 2007 to 2017 (in million U.S. dollars)

