**Predictive Analytics: Interweaving of Statistics, Computer Science and Business**

[G R Chandrashekhar](https://expersight.com/author/chandrashekhar/) May 14, 2020 [No Comments](https://expersight.com/predictive-analytics-interweaving-of-statistics-computer-science-and-business/#respond) 0 likes



Predictive analytics, the art and science of ferreting out actionable insights from data, has been in vogue for some now. Predictive analytics combines multiple techniques along its process of data exploration, data wrangling, model development, and validation. Open source and proprietary tools are both available for the various tasks involved in predictive analytics. Predictive analytics provides insights from past data that could be used for future initiatives.

What are the components of predictive analytics?

Statistical techniques form an important and significant part of predictive analytics, especially in data exploration. A new or unknown dataset should ideally be explored first. In the process of exploration, we may find missing elements in a dataset or some likely anomalous entries. We may also find different types of data — ratio scale, categorical, images etc., present in a raw form with different orders of magnitude. Thus data-wrangling becomes an essential part of the exploration phase of predictive analytics. This could possibly include imputation of missing values, normalization of ratio scale data, encoding categorical data amongst other aspects. Data visualization techniques are very handy to visualize a dataset after data wrangling. The data visualization phase not only helps understand the broad trends but also helps us develop apriori hypotheses.

The second important aspect of predictive analytics is to build a model or develop an algorithm after the data exploration phase. In this phase both statistical techniques and computer science based machine learning, deep learning algorithms are used either individually or sometimes in combination. In many cases, predictive analytics is used to do one of exploring, classify or predict. In classification approaches, achieving linear separation is ideal but not always straightforward. Techniques such as decision trees, random forest, logistic regression, support vector machine etc. are normally used depending on the type of data and the problems being inquired into. The figure below indicates the broad classification of the types of approaches under machine learning.

Data exploration as already broached before is normally considered unsupervised, as we do not have a target class or target variable in mind while exploring data. The model development phase involves one of prediction or classification and is classified under supervised learning. In prediction approaches, the objective is normally to predict a target variable, regression techniques such as linear, polynomial, and logistics regression amongst others are used for the purpose.

The ratio scale and categorical data can be explored and modeled using the above-mentioned techniques, however, images would need different techniques to analyze. Deep Learning approaches such as Convolutional Neural Network (CNN) and Capsule Nets are used for developing insights from image data.

Diagram

Description automatically generated

Applications of predictive analytics

So far we have seen the use of statistics and the use of computer science based algorithms in predictive analytics. However without including the business side predictive analytics would not be complete. What are some applications of predictive analytics that have become common now? Let us consider some of them here:

Supply chain

Traditional supply chain models seem to focus on distribution efficiencies and largely employ spreadsheet based dashboards for monitoring vendors and inventory amongst others. However, the interrelatedness of various aspects of supply chain is quite complex to be monitored by spreadsheets. Ideally organizations should be able to run ‘What If Analysis’, have a data based approach for managing trade-offs amongst customer service, inventory, and supply chain costs, and lastly be able to model profitability. As we can see all of these point towards an organization becoming competitive. The machine learning techniques mentioned before would allow for such analyses to be done.

Customer segmentation

What would interest a marketer? Very likely the following — knowing her customer’s purchases, types of products purchased, value and volume of purchases, the geographical distribution of these purchases, and which customers to target for her future campaigns. All of these form part of the customer segmentation process. Machine learning techniques are very appropriate for such a purpose.

Anomaly detection

 What would a credit card company be interested in? Probably know its customers as mentioned before, in addition, it would also like to know about the likely misuse of its card. A credit card misuse probably occurs a few times in a million transactions. How does one pick these misuse or anomalies and how do one alert a bank and its customers. Machine learning techniques could be put to productive use to pick such anomalies.

Predictive analytics is unlikely to solve all the problems a company or an organization faces. However, if an organization bases its decisions on data and insights from data it only likely to become more competitive in its activities.