

Introduction to Data Mining

8 V's of Big Data

1. Volume
2. Value
3. Veracity
4. Visualization
5. Variety
6. Velocity
7. Viscosity
8. Virality

Knowledge Discovery in Databases (KDD) is “the nontrivial process of identifying valid, potential useful, and ultimate understandable pattern of data”. Dr's Mark relates the recommendation system as an example of KDD. Furthermore, he highlighted the knowledge discovery and decision making under uncertainty, with illustration of financial market, IoT and epidemiology, etc. In the classical science, we treated the prediction as a white-box models, which indicates the relationship between inputs and outputs. It can be consider as deterministic model. However, in fact it is hard to formulate the relationship between inputs and outputs. One remedy for this issue is make use of statistical model, which we assume based on prior belief. Moreover, the black-box model is more sophisticated.