

クレジット:

UTokyo Online Education データマイニング入門 2018 森 純一郎

ライセンス:

利用者は、本講義資料を、教育的な目的に限ってページ単位で利用することができます。特に記載のない限り、本講義資料はページ単位でクリエイティブ・コモンズ 表示-非営利-改変禁止 ライセンスの下に提供されています。

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

本講義資料内には、東京大学が第三者より許諾を得て利用している画像等や、各種ライセンスによって提供されている画像等が含まれています。個々の画像等を本講義資料から切り離して利用することはできません。個々の画像等の利用については、それぞれの権利者の定めるところに従ってください。



Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems.

It is an essential process where intelligent methods are applied to extract data patterns.

It is an interdisciplinary subfield of computer science.

The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.

Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.

Data mining is the analysis step of the knowledge discovery in databases process, or KDD.