



Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Assignment Project Exam Help

Data Mining
Data vs Information
What is Data Mining?

Machine Learning
Definitions
DM vs ML

Tasks and
Problems
ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources

Books
Tools

Introduction to Data Mining and Machine Learning

<https://eduassistpro.github.io>

Sarah Erfani and Karin Verspoor, CIS

Add WeChat [edu_assist_pro](#)

THE UNIVERSITY OF
MELBOURNE



Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and
Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro

[http://www.innovation.gov.au/Science/PMSEIC/Documents/
DataForScience.pdf](http://www.innovation.gov.au/Science/PMSEIC/Documents/DataForScience.pdf)

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Assignment Project Exam Help

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and
Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Assignment Project Exam Help

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and

Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Assignment Project Exam Help

Information interpreted with respect to a user's context to extend human

<https://eduassistpro.github.io>

Increasing insight into data, based on a user
given context.

Add WeChat edu_assist_pro

Resources

Books

Tools

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining
Data vs Information
What is Data Mining?

Machine Learning
Definitions
DM vs ML

Tasks and
Problems
ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources
Books
Tools

Tackling the challenge of knowledge management and discovery at a massive scale

Assignment Project Exam Help

<https://eduassistpro.github.io>

business and enterprise applications

applications

Exponential growth of data with the rise of the multitudes of automatic data gen

Add WeChat `edu_assist_pro`

This trend is expected to continue in the foreseeable future.

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining
Data vs Information
What is Data Mining?

Machine Learning
Definitions
DM vs ML
Tasks and
Problems
ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources
Books
Tools

Importance of Problem

- Current computational methods cannot handle magnitude and dimensionality of the data
- Decision makers and Scientists need techniques to help form

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro

Tools are required to integrate, distill, and make sense of data.

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Assignment Project Exam Help

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and

Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro

Assignment Project Exam Help

- implicit,
-
-

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro

- Needed: programs that detect patterns
- Strong patterns → good predictions
- Problem 1: most patterns are not interesting
- Problem 2: patterns may be inexact (or spurious)
- Problem 3: data may be garbled or missing

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Assignment Project Exam Help

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and

Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

<https://eduassistpro.github.io>

“being explicitly programmed”

Add WeChat edu_assist_pro

Arthur Sandel (1959),



<https://eduassistpro.github.io>

- “A computer program is said to learn from experience E if it changes its behavior in respect to some class of tasks T and performance metric P, as measured by its anticipated future performance, as a result of the experience E.”

Add WeChat edu_assist_pro

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining
Data vs Information
What is Data Mining?

Machine Learning
Definitions
DM vs ML
Tasks and
Problems
ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources
Books
Tools

Assignment Project Exam Help

Algorithms for acquiring structural descriptions from examples

<https://eduassistpro.github.io>
derived

(may be even more important)

Add WeChat `edu_assist_pro`

Methods originate from artificial intelligence
databases

Assignment Project Exam Help

Dictionary definitions of “learning”:

- To get knowledge of by study, experience, or being taught



<https://eduassistpro.github.io>

→ Difficult to measure; Trivial for comput

Add WeChat edu_assist_pro

- Operational definition:
- Things learn when they change their behavior so that they perform better in the future.
 - Does learning imply intention?

Assignment Project Exam Help

Supervised learning

- Teach the computer how to do something (*by example*), then let it

<https://eduassistpro.github.io>

Unsupervised learning

Add WeChat edu_assist_pro

- Let the computer *learn how to do so*
- Determine structure and patterns in data
- Unlabeled data: Don't give the computer "the answer"

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining
Data vs Information
What is Data Mining?

Machine Learning
Definitions
DM vs ML

Tasks and
Problems
ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources
Books
Tools

The distinctions between Data Mining and
Machine Learning are not cut-and-dried.

Data mining is primarily about discovering
something hidden in your data, that you did

<https://eduassistpro.github.io>

new data, as accurately as possible.

Techniques used to learn from data.

Data mining applications typically use a lot
of machine learning techniques. For
example a pattern in a data set that is
useful for generalisation might represent
new knowledge.

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining
Data vs Information
What is Data Mining?

Machine Learning
Definitions
DM vs ML

Tasks and
Problems

ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources

Books
Tools

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro



Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining
Data vs Information
What is Data Mining?

Machine Learning
Definitions
DM vs ML

Tasks and
Problems

ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources

Books
Tools

Supervised learning

- Classification
predicting a discrete class

-

<https://eduassistpro.github.io>

- Association
detecting associations between fe
- Information organisation; Clustering
grouping similar instances into clust
- Reinforcement learning
- Recommender systems
- Anomaly/outlier detection

Add WeChat edu_assist_pro

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining
Data vs Information
What is Data Mining?

Machine Learning
Definitions
DM vs ML

Tasks and
Problems
ML applications
Types of ML

Examples

ML concepts
core definitions
attributes

Resources

Books

Tools

Can we predict housing prices?

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro

A friend has a house which is 750 square feet – how much do you expect to get?

(draw a straight line vs. fit a curve)

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining
Data vs Information
What is Data Mining?

Machine Learning
Definitions
DM vs ML

Tasks and
Problems
ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources
Books
Tools

- Given gene expression data for individuals, cluster based on expression profiles
- Group newspaper articles into cohesive groups
- Credit card fraud
-

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and
Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

Assignment Project Exam Help



<https://eduassistpro.github.io>

also known as feature



Concepts: things that we aim to learn

Add WeChat `edu_assist_pro`

Example: Supervised Learning (Classification)

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Assignment Project Exam Help

Data Mining
Data vs Information
What is Data Mining?

Machine Learning

Definitions
DM vs ML

Tasks and
Problems

ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources

Books

Tools

Outlook	Temperature	Humidity	Windy	Play
sunny	hot	high	FALSE	no
				no
				yes
				yes
rainy	cool	n		no
⋮	⋮	⋮		⋮

Add WeChat edu_assist_pro

Given information about current weather conditions and the forecast, can we determine whether we will go out to play?

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Assignment Project Exam Help

Outlook	Temperature	Humidity	Windy	Play
---------	-------------	----------	-------	------

no	1
no	2
yes	3

rainy	cool
rainy	cool

<https://eduassistpro.github.io>
Add: WeChat edu_assist_pro

Data Mining
Data vs Information
What is Data Mining?

Machine Learning

Definitions
DM vs ML

Tasks and
Problems

ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources

Books
Tools

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining
Data vs Information
What is Data Mining?

Machine Learning
Definitions
DM vs ML

Tasks and
Problems
ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources
Books
Tools

Assignment Project Exam Help

Outlook	Temperature	Humidity	Windy	Play
rainy	cool	high	false	no
rainy	cool	high	true	yes

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Assignment Project Exam Help

Outlook Temperature Humidity Windy Play

no no no no no

no no no no yes

yes no no yes yes

yes yes yes yes yes

<https://eduassistpro.github.io>

rainy cool

rainy cool

Add WeChat edu_assist_pro

Data Mining
Data vs Information
What is Data Mining?

Machine Learning
Definitions
DM vs ML

Tasks and
Problems

ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources

Books

Tools

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining

Data vs Information

What is Data Mining?

Machine Learning

Definitions

DM vs ML

Tasks and

Problems

ML applications

Types of ML

Examples

ML concepts

core definitions

attributes

Resources

Books

Tools

Assignment Project Exam Help

<https://eduassistpro.github.io>

- ordinal
- continuous

Add WeChat edu_assist_pro

Assignment Project Exam Help

- Values are distinct symbols (e.g. sunny, overcast, rainy})

<https://eduassistpro.github.io>

- Special case: dichotomy (“boolean” attribute)
- No relation is implied among nominal values (distance measure), and only equality

Add WeChat edu_assist_pro

Assignment Project Exam Help

- An explicit order is imposed on the values (e.g. hot,mild,cool}

<https://eduassistpro.github.io>

- Example rule: temperature < hot play = yes
- Distinction between nominal and ordinal
(e.g. oullook)

Add WeChat edu_assist_pro

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining

Data vs Information
What is Data Mining?

Machine Learning

Definitions
DM vs ML

Tasks and
Problems

ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources

Books
Tools

Assignment Project Exam Help

▪ <https://eduassistpro.github.io>

Distance between an obj

▪ All mathematical operations are allowed

Add WeChat **edu_assist_pro**

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining
Data vs Information
What is Data Mining?

Machine Learning
Definitions
DM vs ML

Tasks and
Problems
ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources
Books
Tools

Assignment Project Exam Help

How might you approach data mining the Weather dataset?

<https://eduassistpro.github.io>

- Attributes
 - Are there regularities among the attr
 - Are there different ways you could m
(e.g. different combinations? diff

Add WeChat edu_assist_pro

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining
Data vs Information
What is Data Mining?

Machine Learning
Definitions
DM vs ML

Tasks and
Problems
ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources
Books
Tools

Assignment Project Exam Help

Introduction to Data Mining

<https://eduassistpro.github.io/>

Data Mining: Practical Machine Learning

Ian Witten, Eibe Frank, Mark Hall

<http://www.cs.waikato.ac.nz/ml>

Add WeChat edu_assist_pro

Introduction to
Data Mining and
Machine Learning

COMP90049
Knowledge
Technologies

Data Mining
Data vs Information
What is Data Mining?

Machine Learning
Definitions
DM vs ML

Tasks and
Problems
ML applications
Types of ML
Examples

ML concepts
core definitions
attributes

Resources
Books
Tools

Assignment Project Exam Help

WEKA Toolkit

<https://eduassistpro.github.io>

Data sets

UC Irvine Machine Learning Data Repository

<http://archive.ics.uci.edu/ml/>

Add WeChat edu_assist_pro