Data Mining

Thiyanga S. Talagala

2025-10-11

Table of contents

Preface

1 Introduction to Data Mining

1.1 What is Data Mining?

- Process of discovering interesting patterns of knowledge from huge amounts of data.

1.2 What do we mean by interesting patterns?

- Interesting patterns: Valid, Novel, Useful, Understandable

Example

- Retailers collect data about customer purchases at the checkout counters
- Customer purchasing patterns: Identify which items are frequently sold together?
- Products that are likely to be purchased together.

Why it is useful?

- Can make a purchase suggestion to their customers
- Gives an idea that how we can arrange items in a store to as a strategy for boosting sales.

1.3 Characteristics of Big Data: 5 V's of Big Data

1. Volume: size
2. Velocity: how quickly data is generated?
3. Variety: diversity
4. Veracity: quality of data
5. Value: how useful?

1.4 What motivates the development of data mining field?

- Scalability
- High dimensionality
- Heterogeneous and complex data
- Data ownership and distribution

1.5 Data Mining Tasks

1. Predictive tasks: Predict the value of a particular attribute based on the values of other attributes
2. Descriptive tasks: Find human-interpretable patterns that describe data

1.6 Data Quality

1. Range: How narrow or wide of the scope of these data?
2. Relevancy: Is the data relevant to the problem?
3. Recency: How recent the data is generated?
4. Robustness: Signal to noise ratio
5. Reliability: How accurate?

1.7 Applications

1. Web mining: recommendation systems
2. Screening images: Early warning of ecological disasters
3. Marketing and sales
4. Diagnosis
5. Load forecasting
6. Decision involving judgement

Many more...

2 Summary

In summary, this book has no content whatsoever.

```
1 + 1
```

```
[1] 2
```