

Data Mining: Entscheidungsbäume

Baum für den Wetter-Datensatz

In der Vorlesung wurde der Algorithmus ID3 vorgestellt und ein Baum für den Wetter-Datensatz erstellt.

- Bestätige die Werte, die als Informationsgewinn für die Attribute *Outlook*, *Temperature* und *Humidity* als Baumwurzel berechnet wurden, indem Du den Informationsgewinn selbst von Hand ausrechnest.
- Baue den Baum zu Ende auf, bis sich unter allen Blättern nur Instanzen einer einzigen Kategorie befinden.

Ein linguistisches Datensatz

In vielen Sprachen sind die Artikel ambig. Zum Beispiel wird im Deutschen *der* vor maskulinen Substantiven im Nominativ Singular und vor weiblichen Substantiven im Dativ Singular verwendet, und *den* vor maskulinen Substantiven im Akkusativ Singular und vor Substantiven allen Genera im Dativ Plural.

Auf Grundlage der folgenden Daten soll der richtige Artikel vor einem Substantiv in einer unbekanntenen Sprache bestimmt werden. Die Attribute beziehen sich jeweils auf das Substantiv. Konstruiere einen Entscheidungsbaum auf Basis des Datensatzes.

Genus	Numerus	Kasus	Klasse
Maskulin	Singular	Dativ	lo
Maskulin	Plural	Akkusativ	la
Maskulin	Singular	Akkusativ	lo
Feminin	Singular	Nominativ	le
Feminin	Singular	Dativ	le
Feminin	Plural	Nominativ	le