Fundamentals of Data Science

Semester B 20-21

Tutorial 1

1   a. This is just a simple arithmetic operation and not a data mining task.
    b. This is just a simple sorting task and cannot be regarded as data mining.
    c. This is a data mining task known as predictive modeling.

2. Each article can be represented as a data record, where each attribute of a record corresponds to a unique word that appears in the table. The value of each attribute is the number of times the corresponding word occurs in the document.

3. a. All home owners can successfully repay their loans, while additional information for non-home owners is required to determine whether they will default or not.

    b. All married non-home owners can successfully repay their loans, while additional information for the single non home owners is required to determine whether they will default or not.

    c. Making use of the attribute Annual Income, we can determine that, for each of the single non-home owner, if his/her annual income is greater than 80K, they can successfully repay the loan. On the other hand, if the annual income is equal to or less than 80K, they cannot successfully repay the loan.