**CPS685 Pattern Recognition and Data Mining**

**Homework week 1**

**2A: Data Mining as misnomer**

Data Mining as the name might suggest in literal term would be that we are mining for the data, that the processes involved would ultimately lead us to mine data, as we refer to other kind of mining in general. As the instructor gave example, if we refer to gold mining, our ultimate goal is to mine the earth to find gold and hence the name ‘gold mining’ or any natural resource mining for the fact. This analogy isn’t followed in Data Mining, as we are actually looking for knowledge in data that we have, Data is just earth as we would compare it above gold mining example. Data is being processed to find knowledge which can then be used to make Data-Driven decisions in business operations. Data Mining is more of Knowledge Mining or as termed KDD(Knowledge Discovery in Databases)

**2B: Coincidence, Correlation and Causality**

**Own Answers**

**Coincidence**: Coincidence in data mining refer to relation or pattern in data which could be coincidently mined but doesn’t hold any significance for the analysis.

**Correlation**: Correlation shows us the relation or pattern between the data mined that one attribute of the data can be dependent on the other attribute. It is useful for the analysis as it shows the pattern between two set of attributes which might hold some underlying relationship.

**Causality**: Causality shows us there is a cause-and-effect relation between one attribute of data on the other. It shows a direct relationship between the attributes and can be taken as fact rather observing it as simple association as one attribute directly influences the other attribute in the relationship established.

**Chatgpt answers**

**Coincidence**: In data mining, a **coincidence** typically refers to a situation where two or more events or patterns appear to be related or occur together by chance, rather than due to any actual causal relationship or underlying connection. This can happen when analyzing large datasets, where random associations or patterns can emerge simply due to the volume of data, rather than indicating any meaningful relationship.

**Correlation**: In data mining, **correlation** refers to a statistical measure that describes the extent to which two or more variables move in relation to each other. It helps identify and quantify the strength and direction of a linear relationship between variables in a dataset. Correlation is used to discover patterns, trends, and associations between variables, which can provide insights for making predictions and identifying underlying relationships.

**Causality**: In data mining, **causality** refers to the relationship between two or more variables where one variable (the cause) directly influences or brings about a change in another variable (the effect). Establishing causality means determining that changes in one variable are responsible for changes in another, rather than simply observing that the variables are correlated or associated.

**Updated answers**

**Coincidence**: Coincidence in data mining refer to relation or pattern in data which could be coincidently mined and doesn’t have any actual relationship or underlying connection occurring together by chance. This can happen while analysing large datasets where random patterns might emerge which might not hold any significance in for analysis.

**Correlation**: Correlation refers to the statistical measure describing the extend at which one attribute of the data can be dependent on the other attribute. It helps identify and quantify the strength and direction of a linear relationship between variables in a dataset. Correlation is a great measure which can provide insights to make predictions and pattern between two set of attributes which might hold some underlying relationship.

**Causality**: Causality shows us there is a cause-and-effect relation between one attribute of data on the other. Establishing causality means determining that changes in one variable are responsible for changes in another and can be taken as fact rather observing it as simple association as one attribute directly influences the other attribute in the relationship established.