

Computer Science M5 T2 Final Exam Short

Data Analysis

Data analysis is a process of turning information into a valuable resource by using the results obtained from the analysis. When we analyze something, we always have 3 important questions:

1. What happened?
2. What will happen?
3. What should I/we do?

The answers can be obtained by analyzing 3 types of data:

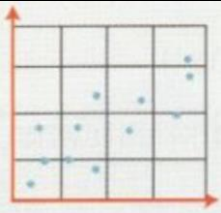
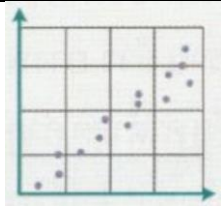
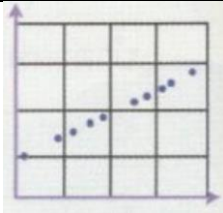
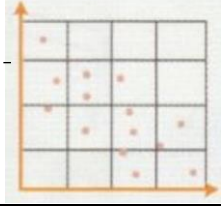
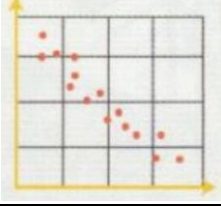
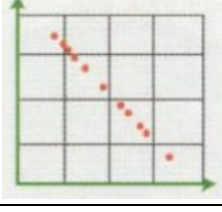
- **Descriptive analytics**
- **Predictive analytics**
- **Prescriptive analytics**

Descriptive analytics – is a fundamental analysis, it gives an overview of data and the relationship between them. Explains what happened in the past so it will help us make choices in the future.

Predictive analytics - is an analysis that helps predict or predict what is likely to happen in the future. By using historical data, we can possibly predict what can happen.

Prescriptive analytics - is an analysis to build on the predictions that may happen. We simulate the possible options of a simulation and predict the outcome of each situation. This way we can suggest which situation would be the best and its possible results.

Dataset relationship

Direction/Degree	Weak	Moderate	Strong
Positive			
Negative			

Numerical prediction

Numerical data prediction is the use of historical data to analyze trends in order to predict the future. The predictions can only be made by using historical data. That data has to be analyzed so that we can know the relationship between data sets and that we can create a prediction model that gives numerical results.

We can predict the data set of interest (y) when the other data set (x) is known to be related. We have to draw a straight line through as many points in the distribution diagram as possible. We call this line a **"Trendline"**. Data is never perfect so we have to use a **"best fit line"**. The **best fit trendline** is a line that should have the same number of points above and below it.

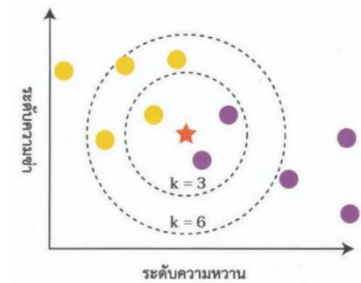
We predict data in 3 steps:

1. Create a distribution diagram
2. Draw the trendline
3. Predict the value from the trendline

Category prediction

Category prediction is the prediction of information that is not numerical from another set of related data. The main prediction concept relies on using historical data that has been already categorized. To predict a new set of unknown data we can use many different techniques. One of those techniques is the K-Nearest Neighbor (KNN).

The main idea of KNN is to compare the new unclassified data to the original, categorized data. The comparison is the data with the least distance. We compare it with the data that is spaced around the uncategorized data until the K is complete. Then we use the result to categorize our new information.



Data communication

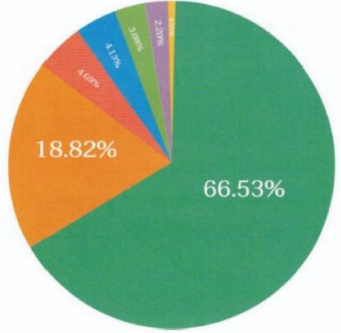
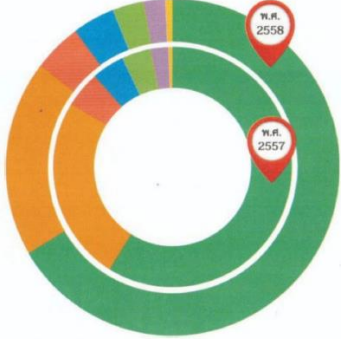
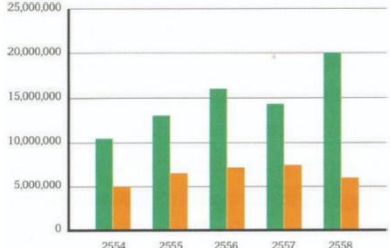
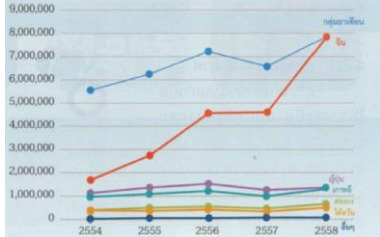
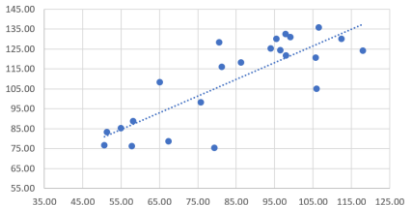
Data communication is the communication of information from the source (speaker) to the receiver (listener). The information has to be presented in a way that will help make decisions, answer questions, find patterns...

Data visualization

Data visualization is presenting numerical data in a more understandable way like graphs or charts.

Type of chart	Purpose of presentation
Pie chart/Donut chart	Shows the proportion of data with a small number of different data
Bar chart	Compare individual data sets
Line chart	Shows changes in continuous data
Scatter plot	Shows the relationship between two data sets

Graphs

Name	Explanation	Visual chart
<p>Pie chart</p>	<p>A pie chart is created by taking a circle and dividing it into proportions according to the amount of information. It should be used for data that has a small number of groups. It can show the information in percentage which can help us compare groups.</p>	
<p>Donut chart</p>	<p>It works in the same way as a pie chart but the middle of the circle is empty. It can be used to compare two similar datasets. It can be used to compare two similar datasets.</p>	
<p>Bar chart</p>	<p>It is used to show the difference in volume between data sets. It sorts the data from left to right and is usually arranged horizontally.</p>	
<p>Line graph</p>	<p>A line graph is used to show changes in individual data and compare those changes with other data. Because of its design, it will take less space than other graphs.</p>	
<p>Scatter plot chart</p>	<p>It is used to show the distribution of the data. It can be used to compare the data also.</p>	

ATTENTION: Each group will have one practical task in the exam. It will be either Finding a relationship, Numerical prediction, or Category prediction, depending on the group. So please prepare yourself for it. Any questions before the exam are welcome. Questions during the exam maybe will not be answered!