**Final Project**

1. *Requirement*: Consider a data set of your choice: e.g. from Kaggle or from any other source (e.g. job or your master thesis).

For this task I have chosen the “All laptops with Specifications Dataset” (<https://www.kaggle.com/datasets/durgeshrao9993/laptop-specification-dataset>) containing several information about the hardware specification of various laptops on the market.

***Motivation***

Buying a new laptop can be an important financial and practical decision due to its many personal and job related use cases in our daily life. But the reality is that most people looking to buy a new laptop are not familiar enough with all the hardware details that can make a big difference between a successful acquisition that can last years, and a disappointment. As such, I think that by using FCA on this dataset about laptops we might get a stronger understanding about how various hardware components are related, which in turn should make the acquisition process easier.

1. *Requirement*: If necessary, apply some AI, KD and Data Mining algorithms in order to extract a first layer of knowledge. Describe the results you have obtained and the knowledge you have extracted.

The raw version of the dataset contains several columns, each with a large variety of values, therefore, it needs some cleaning. I have chosen some representative columns such as **company, CPU, RAM, Memory and GPU**. In my opinion, these hardware pieces along with the brand should be the main factors to take into consideration when choosing a laptop.

First of all, for each field, I choose the most representative values, discarding the ones that are not frequent enough.

The company field has the following values, along with their frequencies: Acer (23), Apple (5), Asus (33), Chuwi (2), Dell (29), Fujitsu (3), Google (2), HP (36), Huawei (1), LG (2), Lenovo (36), MSI (5), Mediacom (4), Microsoft (3), Razer (4), Samsung (6), Toshiba (9), Vero (2), Xiaomi (2).

The CPU field has the following values: amd (27), intel (180).

The RAM field has the following values (GB): 2 (9), 4 (47), 6 (21), 8 (64), 12 (16), 16 (37), 24 (3), 32 (9), 64 (1).

The memory type field has the following values: HDD (63), SSD (85), SSD + HDD (36), flash (23).

The GPU field has the following values: AMD (49), Intel (88), Nvidia (70).

In the end, the processed dataset has **207** objects.

1. *Requirement*: Use either the result of this analysis or any other many-valued context to build a ToscanaJ system. Describe in detail the scales you have built and the amount of knowledge you was able to extract here from. Be inventive and creative in the scale building, usually the knowledge gems are there, but you need to dig after them. For ToscanaJ you will need an older version of Java to make it work.

For this task I have used both **ELBA and ToscanaJ** in order to extract knowledge. ELBA was useful in designing the contexts and scales by using **attribute list, nominal and ordinal scales**. This process was eased by the capacity of ELBA to connect and read my custom dataset from an sql file which I have generated using an online csv to sql tool (<https://www.convertcsv.com/csv-to-sql.htm?fbclid=IwAR1QEPb9QfeDWG4APvfzit2MSq6agpxtfQe_6mn0eEUa87psX3_jL3rBUdg>). The output sql looks like this:

CREATE TABLE mytable(

id INTEGER NOT NULL PRIMARY KEY

,company VARCHAR(9) NOT NULL

,cpu VARCHAR(5) NOT NULL

,ram INTEGER NOT NULL

,memory VARCHAR(9) NOT NULL

,gpu VARCHAR(6) NOT NULL

);

INSERT INTO mytable(id,company,cpu,ram,memory,gpu) VALUES (0,'Apple','intel',8,'SSD','Intel');

INSERT INTO mytable(id,company,cpu,ram,memory,gpu) VALUES (1,'Apple','intel',8,'flash','Intel');

…

After loading the data into ELBA, I was able to generate conceptual schema (csx) files which are in turn used by ToscanaJ to visualize the concept lattices, with more details about where exactly the objects are located. In ToscanaJ I used the “**Show exact matches**” option from the view panel.

The next step is the knowledge extraction. In order to have a better understanding of the dataset, I have created several scales and diagrams as follows.

First of all, I wanted to get a better understanding of the dataset proportions with respect to each column. I started with RAM and I used the ordinal scale option:

A screenshot of a computer

Description automatically generated with medium confidence

Afterwards, I visualized it in ToscanaJ:

A picture containing line, screenshot

Description automatically generated

It seems clear that most laptops (68.12%) have less than 8GB of RAM which is nowadays enough for usual activities such as browsing the internet or using Microsoft Office, but not quite enough for more demanding activities such as playing games or training machine learning models. This is reflected by the low proportions of laptops which have RAM ranging from 8 to 16 GB (25.6%) and even lower proportion of laptops with a lot of RAM (6.28%). Considering what most people need from a laptop, these proportions make sense.

The next element from which I tried to extract knowledge is the CPU:

A picture containing diagram, screenshot, line, circle

Description automatically generated

From the diagram it is clear that the most popular option are the CPUs from Intel which is a well established producer in the market.

The next component might not be that important for basic usage of the latop. The GPU is usually important to those that need heavy algebraic computations. This is usually needed for computer games or machine learning. The diagram is the following:

A picture containing line, diagram, circle

Description automatically generated

Here Intel is also leading in popularity, probably because it is a cheaper option, suitable for basic usage of the laptop. Following closely, there is Nvidia with a reasonably high percentage of 33.82%. This is the kind of GPU needed for machine learning and game playing.

I also analyzed the memory component. This is important because it decides how fast applications are loading on the laptop:

A picture containing line, diagram, circle

Description automatically generated

We can see that there are mostly two types: SSD and HDD, with the latter being an older technology, but cheaper. Laptops also may have both an SSD and an HDD. Flash memory is similar to SSD but smaller in size, which is why it is not that popular.

A diagram of a network

Description automatically generated with low confidenceThe next component is not related to hardware, but it is the company producing the laptop. From this diagram I can tell the variety of options offered by each brand:

Now that I have covered the basic components, I want to see what knowledge I can extract such that I can better understand laptops suitable for machine learning.

A picture containing line, diagram

Description automatically generatedFirst of all, it needs to have a Nvidia GPU. In the next diagram I have illustrated some of the most popular brands and whether they have a Nvidia GPU or not:

It seems that there are no Apple laptops with Nvidia GPU, therefore they are not suitable for this. Also, Asus seems to lead with 6.36% of laptops with Nvidia GPU, which means it offers more variety with a Nvidia GPU.

Finally, I wanted to have the bigger picture. I want to understand all components together. This ideally will help in deciding which components are worth it, which is especially important when buying a laptop on a budget.

Therefore, I have chosen the following attributes: if it has SSD, Intel CPU and Nvidia GPU. Also, I differentiated between RAM larger than 8GB, 16GB and 32GB.

A picture containing diagram, line

Description automatically generated

Also, this is the context:

A screenshot of a computer

Description automatically generated with medium confidence

And here we have the same diagram, but showing the list of objects for each concept:

A picture containing diagram, plan, line, technical drawing

Description automatically generated

Finally, it seems that the laptop brands having the best components for machine learning are: Asus, Dell, Lenovo, MSI and Razer. But depending on the budget, it seems that there are options for all combinations of hardware components.