Lazy FCA. Chepurchenko Alexey Sergeevich.

Github repository: https://github.com/kikendopalo/lazy_fca_project

I used one of the datasets from “Kaggle” that is named “The Estonia Disaster Passenger List”

The dataset contains the name, age, sex, category and fate of the 989 passengers aboard the MS Estonia on the night of the sinking.

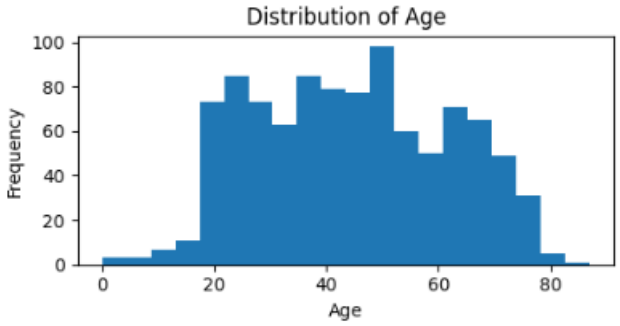
https://www.kaggle.com/datasets/christianlillelund/passenger-list-for-the-estonia-ferry-disaster/data

For column ”Country” i used nominal scale because there are 16 unique values.

For columns ”Sex” and ”Category” i used dichotomic scales, because these are binary features.

I deleted columns ”PassengerId”, ”Firstname” and ”Lastname”, because these columns only have identificational information that can’t affect object’s class.

For column ”Age” i used ordinal scale ($\geq 0, \geq 26, \geq 35, \geq 44, \geq 52, \geq 61, \geq 70$), because these feature is numerical and it’s distribution looks like this:



And 0, 26, 35, 44, 52, 61, 70 are the edges of buckets.

As we can see, on such a complex data when classes are hard to distinguish, Lazy FCA model gives scores that are less than constant model can give (the situation is exacerbated by the imbalance of classes).

But in comparison to other models it is a good performing model. Seems that our LazyFCA model performs better than every other model in terms of f1-score, but for example in terms of accuracy it is one of the worst models (only Naive Bayes is worse).