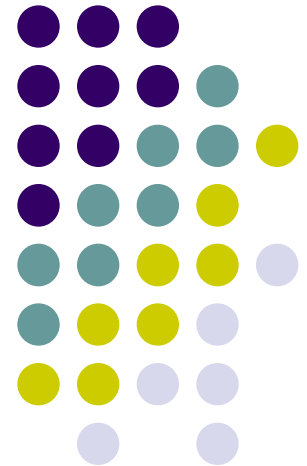
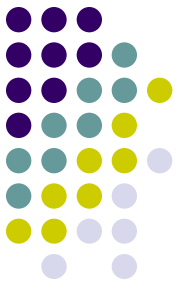


# Polimetrics

## Seventh Assignment

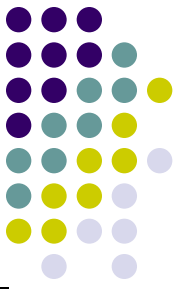


# Deadline: 19 November 2019



- Use `train_review.csv` and `test_review.csv`. The texts included in such two database refer to actual reviews of movies classified as either positive or negative (therefore, REMEMBER: **here your output variable assumes in this case only 2 values!**)
- Use as your training-set the reviews included in `train_review.csv`
- Apply the three ML algorithms as well as iSAX to classify the reviews included in `test_review.csv`
- Compare the results you get

# Deadline: 19 November 2019



- HINT: when creating the dfm for your training and test-set, trim both dfm by keeping only those features that appear in at least 5% of the texts
- Otherwise your dfm will be too large, and when you run a ML algorithm you could get this error: «Error: protect(): protection stack overflow»
- It means that given that you use a VERY large number of variables in the dataframe, R refuses to go on with the analysis given the power of your laptop