



CLARANS

MAD III

Matej Kubinec (KUB0462)

VŠB-TUO



CLARANS

(Clustering Large Applications based on RANdomized Search)

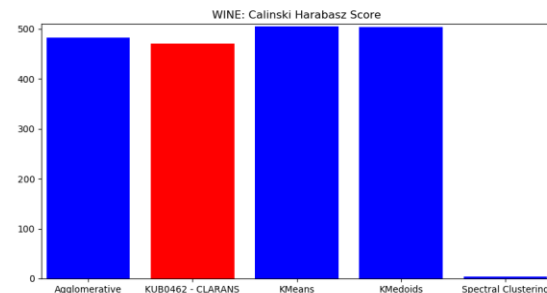
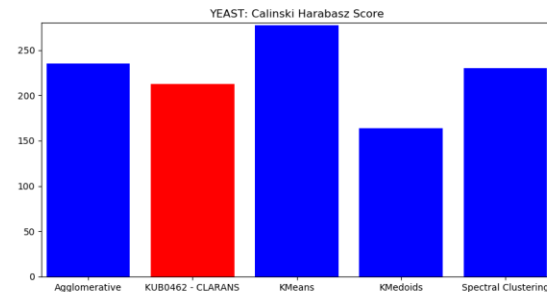
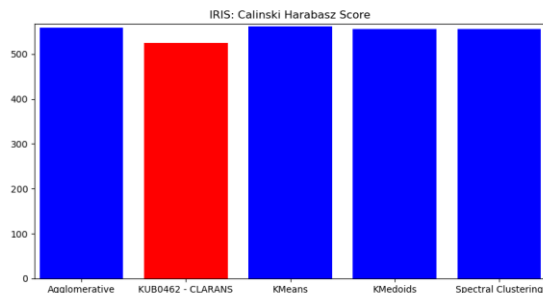
- Vylepšenie algoritmu CLARA
- Prezентuje trade-off medzi výpočelným výkonom a efektívnosťou použitím vzorkovania na vytvorenie zhlučkov

• Parametre

- **K** – počet zhlučkov
- **NumLocal** – počet iterácií
- **MaxNeighbour** – maximálny počet susedov, ktorý sa vymenia

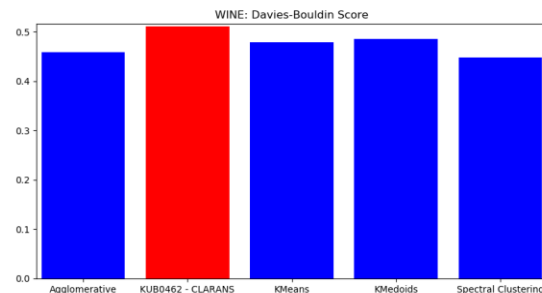
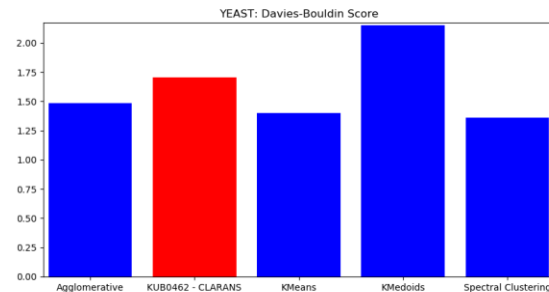
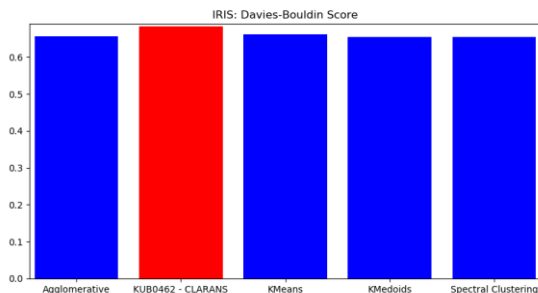
Calinski Harabasz

- Zname aj ako Variance Ratio Criterion
- Pomer medzi rozptylom vnútri zhlukov a medzi zhlukmi navzájom



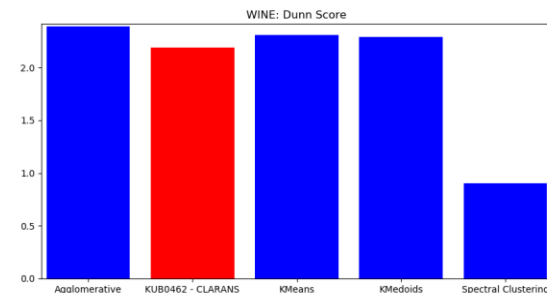
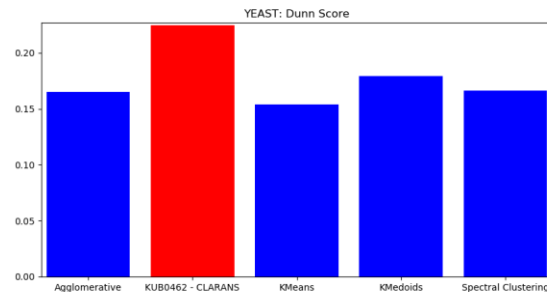
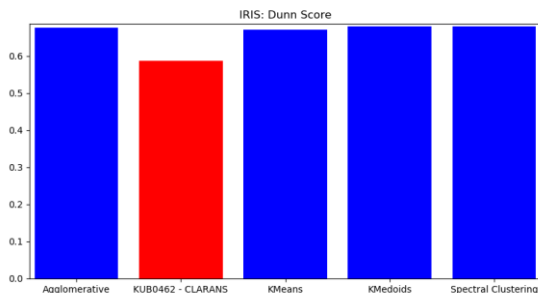
Davies-Bouldin

- Vnútročné ohodnotenie
- Pomer medzi rozmiestnením vnútri zhluku a zhlukmi navzájom
- Nižšie skóre lepší výsledok



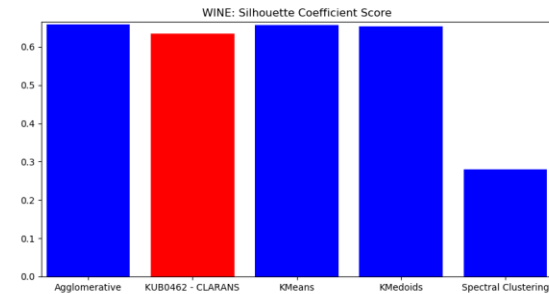
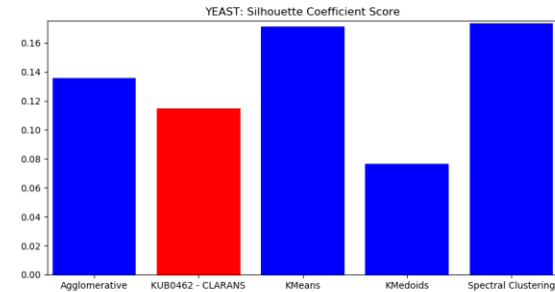
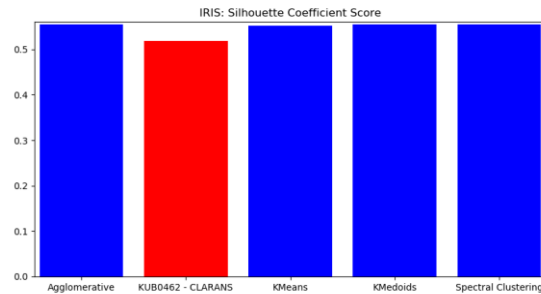
Dunn Index

- Skóre závisí na použitej metrike (Euklidovská)
- Pomer medzi minimálnou vzdialenosťou medzi zhlukmi a maximálnou vnútri zhluku



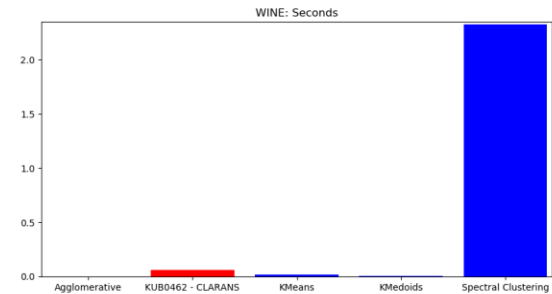
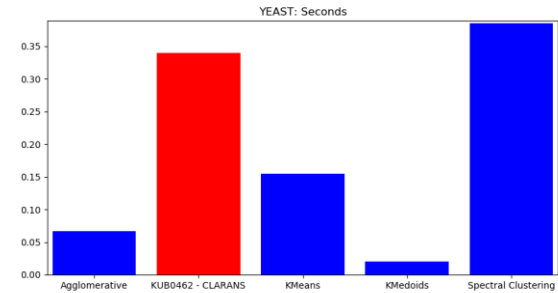
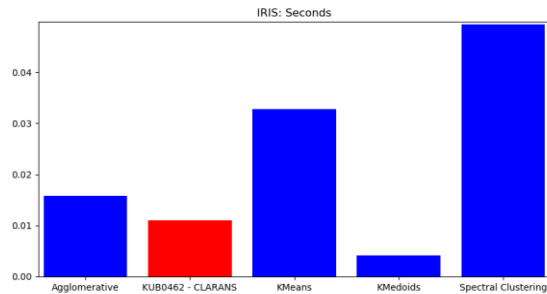
Silhouette Coefficient

- Podobnosť objektov k ostatným v zhlukoch



Výkon

- Počet sekund behu programu.

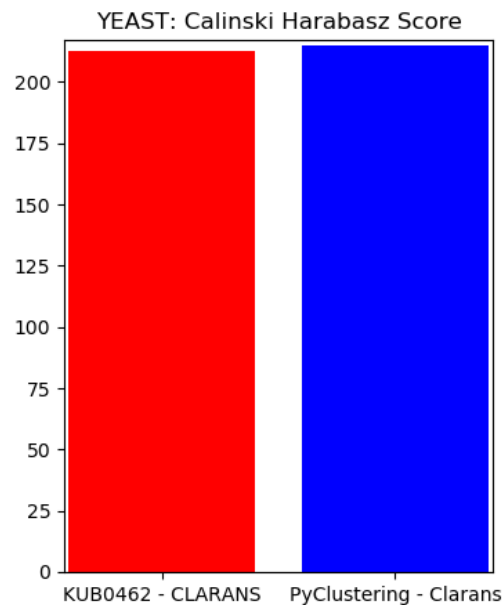


Porovnanie s PyClustering (YEAST)

Calinski Harabasz

–CLARANS – 212.65

–PyClustering – 215.03

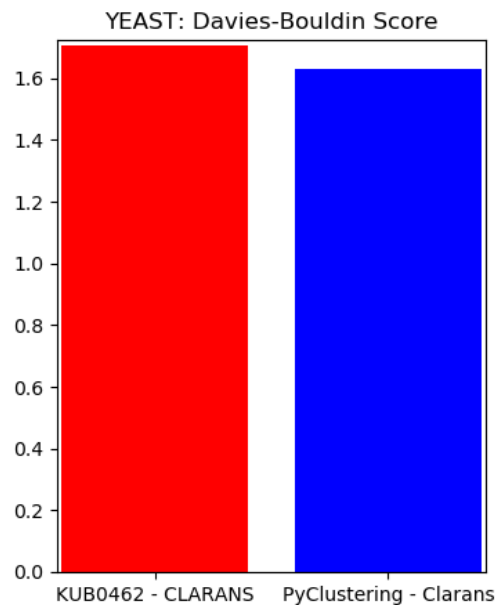


Porovnanie s PyClustering (YEAST)

Davies-Bouldin

–CLARANS – 1.71

–PyClustering – 1.63

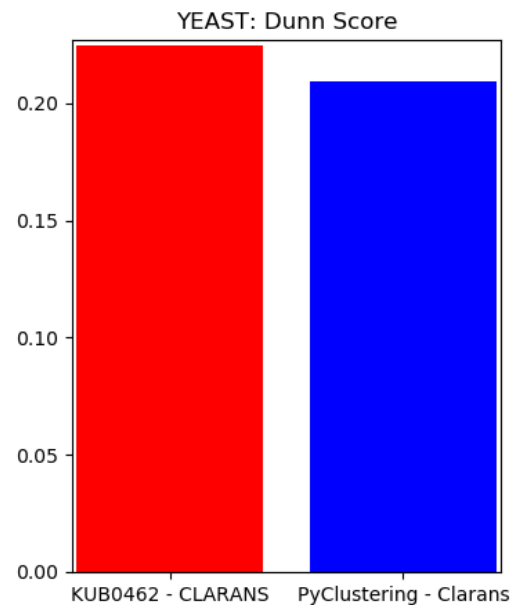


Porovnanie s PyClustering (YEAST)

Dunn

–CLARANS – 0.22

–PyClustering – 0.20

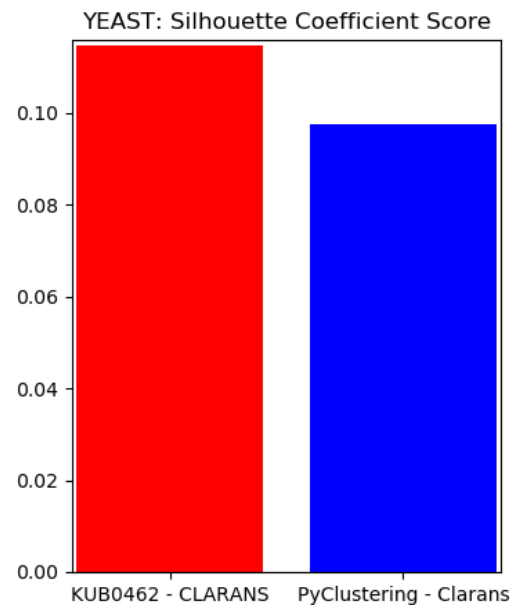


Porovnanie s PyClustering (YEAST)

Silhouette Coefficient

–CLARANS – 0.11

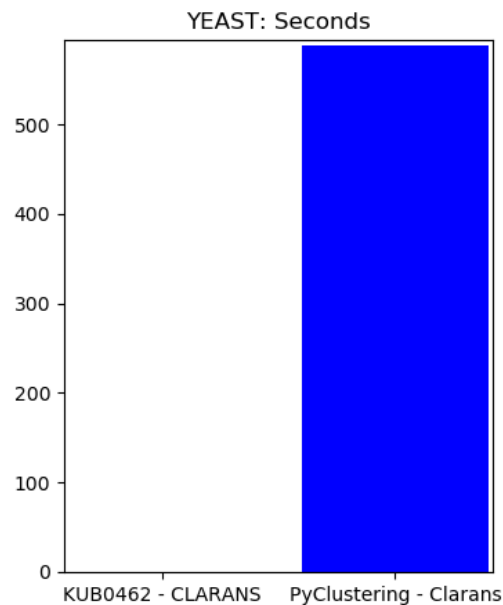
–PyClustering – 0.09



Porovnanie s PyClustering (YEAST)

Výkon

- **CLARANS** – 0.34 sekúnd
- **PyClustering** – 588.4 sekúnd





Ďakujem za pozornosť 