

1 Autorzy projektu

- Piotr Faron
- Artur Górak
- Kamil Harlacz
- Yurii Titov

2 Temat Projektu

Generycznie algorytmy z dokowaniem

3 Link do repozytorium

<https://github.com/YuriiBatkovich/Genetic-algorithms-with-docking>

4 Dodatkowe paczki danych

4.1 Pobieranie i obróbka danych

- Pandas v1.5.3
- SDF v0.3.5
- rdkit v2022.9.5
- numpy v1.24.2

4.2 Wizualizcja danych

- seaborn v0.12.0
- matplotlib v3.7.1.

4.3 Nauczanie Maszynowe

- pytorch/ torch_geometric
- sklearn
- tqdm

4.4 Możliwe algorytmy dokowania

- Autodock
- DOCK
- SwissDock
- PatchDock

5 Źródła danych

- baza danych ChEMBL - dane aktywnościowe
- przesłane przez Panią Doktor pliki PDB - przeprowadzenie procedury dokowania

6 Możliwe problemy

- zasobożerność dokowania molekularnego - możliwość braku mocy obliczeniowej
- trudności implementacyjne - w czasie pisania algorytmu może dojść do wielu niespodziewanych wcześniej problemów np.
 - niekompatybilność bibliotek
 - niedostateczna lub brak dokumentacji
 - problemy z optymalizacją
- ograniczenia płynące z zastosowanego konkretnego narzędzia dokowania molekularnego. W zależności od stopnia zaawansowania, może ono nie przewidywać innych czynników wpływających na interakcje związku chemicznego z białkiem
- algorytm genetyczny wygeneruje bardzo podobne związki - będziemy musieli sprawić, aby populacja zawierała różnorodne związki chemiczne
- algorytm genetyczny wymaga odpowiedniej kalibracji i doboru hiperparametrów - co w połączeniu z naszym brakiem doświadczenia może przynieść słabe rezultaty
- problemy z odczytem i obróbką danych - w grupie mamy 4 koderów z znikomą wiedzą chemiczną, przez co wydobycie przez nas istotnych informacji z danych będzie utrudnione
- molekuły związane z danym białkiem muszą mieć określone właściwości chemiczne, takie jak rozmiar, kształt i ładunek, aby mogły dokować poprawnie. Związki chemiczne wygenerowane przez algorytm genetyczny mogą nie spełniać tych wymagań, co może prowadzić do wyników o słabej jakości