**Projekt:**

Podstawią zaliczenia przedmiotu jest zaprojektowanie i przygotowanie programu, który obejmuje poznane na zajęciach metody i struktury danych. Program powinien zawierać ~200 linii kodu, w związku z tym należy przemyśleć projekt elementów składowych. Program powinien dotyczyć zagadnienia podejmowanego w obrębie własnej pracy badawczej np. analiza danych, wizualizacja, metoda obliczeniowa, automatyzacja zadań/obliczeń. W przypadku braku takiej możliwości należy zaproponować własny temat, nie związany z pracą magisterską ale nawiązujący do kierunku studiów.

Praca w zespołach 2-3 osobowych. Skład grupy, tematyka projektu i jego zakres musi być **omówiony z prowadzącym i zaakceptowany** do **końca kwietnia 2020**.

***Forma zaliczenia:***

Prezentacja elementów programu i jego możliwości (10-15 minut) na forum grupy oraz kod razem z opisem w formacie .py lub jupiter notebook dostarczony do prowadzącego.

Ostateczny termin zakończenia projektu : 16.06.2020

W ramach ćwiczeń odbędą się zajęcia konsultacyjne 1-2 tyg. przed zaliczeniem na których przewidziana jest praca w grupach nad własnymi projektami.

***Przykład projektu:***

**Źródło: http://davinci.mimuw.edu.pl/biologia/fasta-kombajn**

Program, który wczyta plik w formacie i go przeanalizuje, tj.:

* Obliczy GC-content
* Znajdzie wszystkie potencjalne kodony start (ATG), tj. pozycje pierwszych nukleotydów w kodonie,
* Znajdzie wszystkie potencjalne TATA-box’y (sekwencje ‘TATAAA’), tj. pozycje pierwszej tyminy w box’ie,
* Znajdzie wszystkie wystąpienia sekwencji ANTTNCCNA, gdzie N to dowolny nukleotyd (może być nieokreślony), tj. pozycje pierwszej adeniny w tej sekwencji,
* Znajdzie pozycje wszystkich niedokładnie określonych nukleotydów,
* Poda ilość nukleotydów w sekwencji.

Wyniki analizy powinny być zwrócone w czytelnej formie w pliku tekstowym.