

Bioinformatics Competition

Competitors:

- ★ Mai
- ★ Shima
- ★ Nehal
- ★ Menna

Do:

- Use any suitable programming language for the challenges
- Make comments on lines for clarification
- Make a google document including a table with the code and output

Ex:

| SQL command | Output | | | | | | | | | | | | | | | | |
|--|---|---------|--------------|---------|-----|---------|-------|--|--|--|--|------------------|-----------|---------|--|--|--|
| DDL: CREATE TABLE Patients (ID INT PRIMARY KEY, Patient_name VARCHAR(50), Gender CHAR(5), DOB DATE , Disease VARCHAR(20), Examination_room VARCHAR(3), Physician VARCHAR(10), Payment VARCHAR(10)); | <div>Patients</div> <table><tr><th>ID</th><th>Patient_name</th><th>Gender</th><th>DOB</th><th>Disease</th></tr><tr><td colspan="5">empty</td></tr></table> <div><table><tr><th>Examination_room</th><th>Physician</th><th>Payment</th></tr><tr><td colspan="3"></td></tr></table></div> | ID | Patient_name | Gender | DOB | Disease | empty | | | | | Examination_room | Physician | Payment | | | |
| ID | Patient_name | Gender | DOB | Disease | | | | | | | | | | | | | |
| empty | | | | | | | | | | | | | | | | | |
| Examination_room | Physician | Payment | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

Don't:

- Copy lines of code without understanding them (You can copy lines but do understand them for a final small discussion 😊)

Challenges:

1. Sequence alignment:

- Sequence alignment is the process of comparing two or more sequences of nucleotides or amino acids to identify regions of similarity and difference. It is an important tool in bioinformatics and is used to compare DNA or protein sequences to identify evolutionary relationships, functional domains, mutations, and other features.
- There are two main types of sequence alignment: global alignment and local alignment. In global alignment, the entire length of both sequences is aligned from the beginning to the end, whereas in local alignment, only the most similar regions of the sequences are aligned.

Target:

- Write a code that takes 2 sequences as an input and performs sequence alignment

Global alignment: 5 points

Local Alignment: 10 points

Bonus: 5 points (Use the same 2 sequences and compare your code output with sequence alignment available tools as BLAST)

N.B. Using more tools, you can earn more points 100

2. Clustering:

- Clustering is a technique used to group similar objects or data points together based on their similarity. In bioinformatics, clustering is often used to group genes or proteins based on their expression profiles, sequence similarity, or other features. Clustering algorithms commonly used in bioinformatics include hierarchical clustering, K-means clustering, and self-organizing maps.

Target:

- Write a code using bioinformatics data to perform k-means clustering

Data loading: 5 points

Clustering: 5 points

Bonus: 5 points (Data cleaning and normalization)

Bonus: 5 points (Data Visualization)

3. Classification:

- Classification is a technique used to assign objects or data points to predefined categories or classes. In bioinformatics, classification is often used to predict the function of genes or proteins based on their sequence or expression data. Classification algorithms commonly used in bioinformatics include decision trees, random forests, support vector machines, and neural networks.

Target:

- Write a code using bioinformatics data to perform k-means clustering

Data loading: 5 points

Classification: 5 points

Bonus: 5 points (Data cleaning and normalization)

Bonus: 5 points (Data Visualization)

4. Regression:

- Regression is a technique used to predict a continuous variable based on one or more input variables. In bioinformatics, regression is often used to predict gene expression levels or protein interactions based on sequence or other features. Regression algorithms commonly used in bioinformatics include linear regression, logistic regression, and ridge regression.

Target:

- Write a code using bioinformatics data to perform k-means clustering

Data loading: 5 points

Regression: 5 points

Bonus: 5 points (Data cleaning and normalization)

Bonus: 5 points (Data Visualization)