**Data cleaning:**

* Remove irrelevant columns from the dataset

**Mutation data:** Deals with change of DNA sequence: deletion, insertion,