<https://www.kaggle.com/competitions/3722genomics/rules>

Это ссылка на соревнование, необходимо ознакомиться с кодом в google colab (прикрепляю его к заказу) и правилами, модифицировать код в google colab для достижения наибольшего результата в leadboard (желательно получить больше 0,98). Начинать участие в соревнование не нужно (и загружать что-то), все черновые варианты необходимо отправлять мне и проверять через мой аккаунт.

Colab, который производит результат 0.9423, в архиве:

Your goal is to classify two bacteria based on their DNA sequences. See Rules tab for guidelines, grading, submission details, and the starter Colab notebook. You are encouraged to learn from other competitions in Kaggle or elsewhere, but keep your modeling constrained by rules of this competition.

**ALLOWED MODELS**

1. **Ok to use these models:**

[SVM](https://scikit-learn.org/stable/modules/classes.html" \l "module-sklearn.svm) flavors only

1. You can also use clustering algorithms and dimensions reduction algorithms (stats::pca() in R, PCA() in SKL, etc.)
2. Ok to engineer features, tune hyperparameters/optimizers, sub/oversample, regularize, do stepwise selection, reduce dimensions, use CPU/GPU/TPU.

The evaluation metric is [accuracy score](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html), which is the default classification accuracy in sklearn library.

Use \*sampleSubmission.csv file template to populate your submission labels or you can generate your submission file from Pandas in the following format:

id,y

0,0

1,0