

# Разработка методов для точного предсказания промоторов на геноме

Руководитель: Татьяна Татаринова

Команда №3:

Мария Фирулёва

Полина Пчелинцева

Степан Пачганов

Аделия Попова

Иван Климук

Егор Бондаренко

# Цель

- На основе данных нуклеотидных последовательностей *Oryza sativa* разработать метод, предсказывающий положение TSS (*Transcriptional Start Site*).

# Решения

```
graph TD; A[Решения] --> B[Классический подход]; A --> C[Machine learning]; A --> D[Deep learning];
```

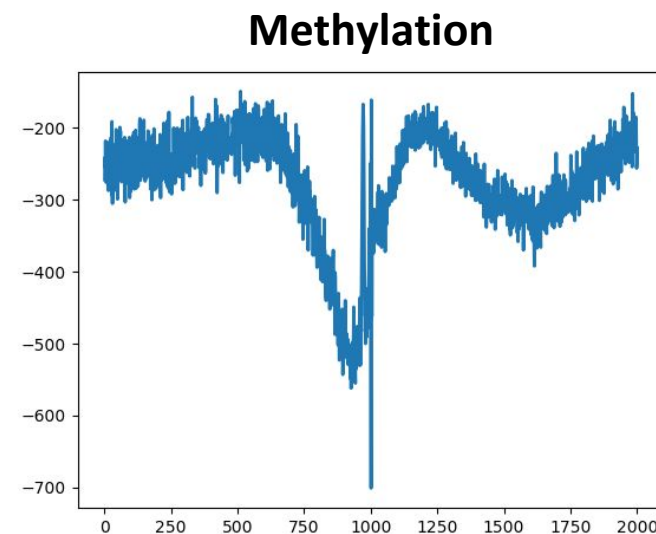
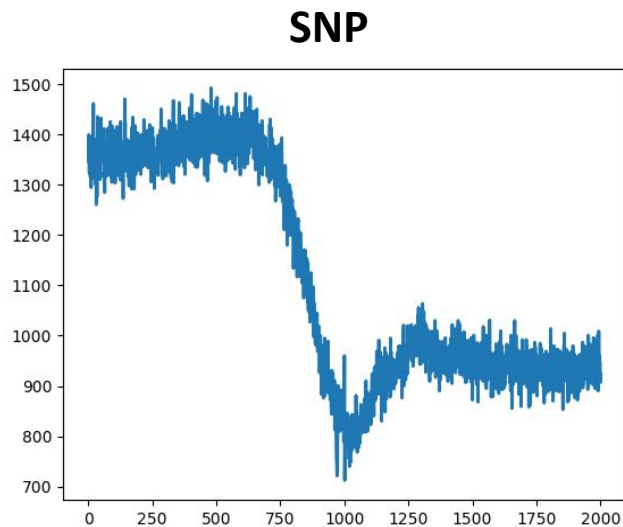
Классический  
подход

Machine  
learning

Deep  
learning

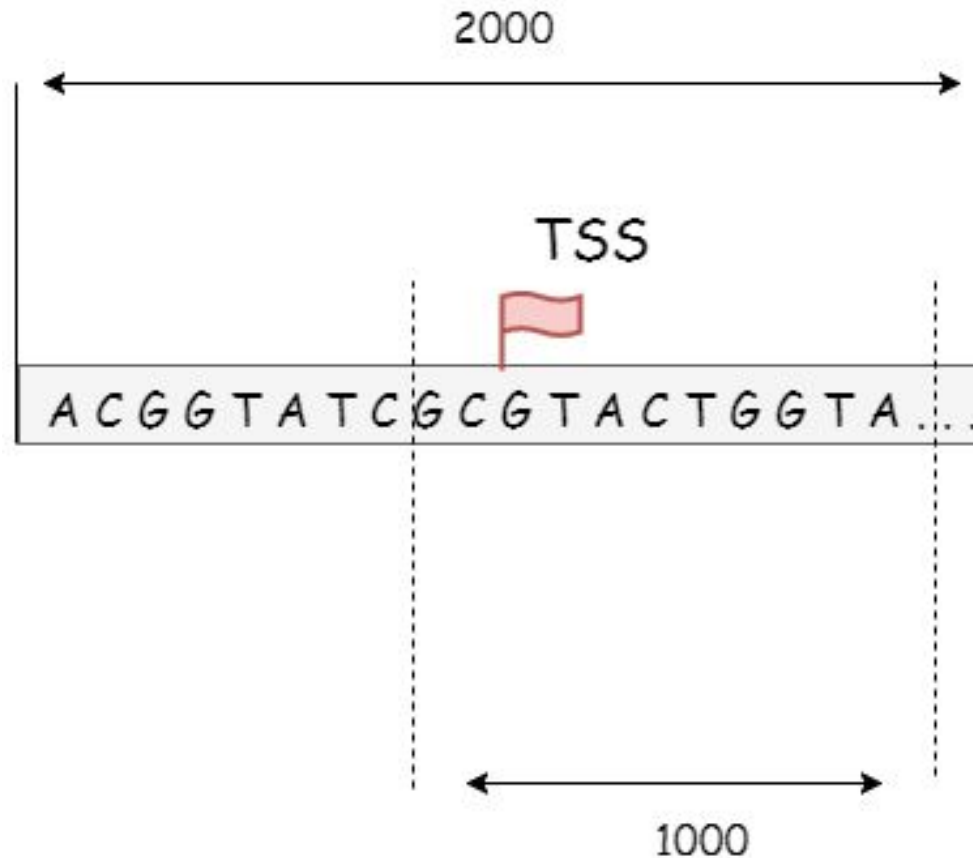
# Основные проблемы ML/DL

- Неинформативная выборка (в каждой последовательности *TSS* начинается с позиции 1000) => переобучение
- Представление признаков одним вектором для каждой последовательности

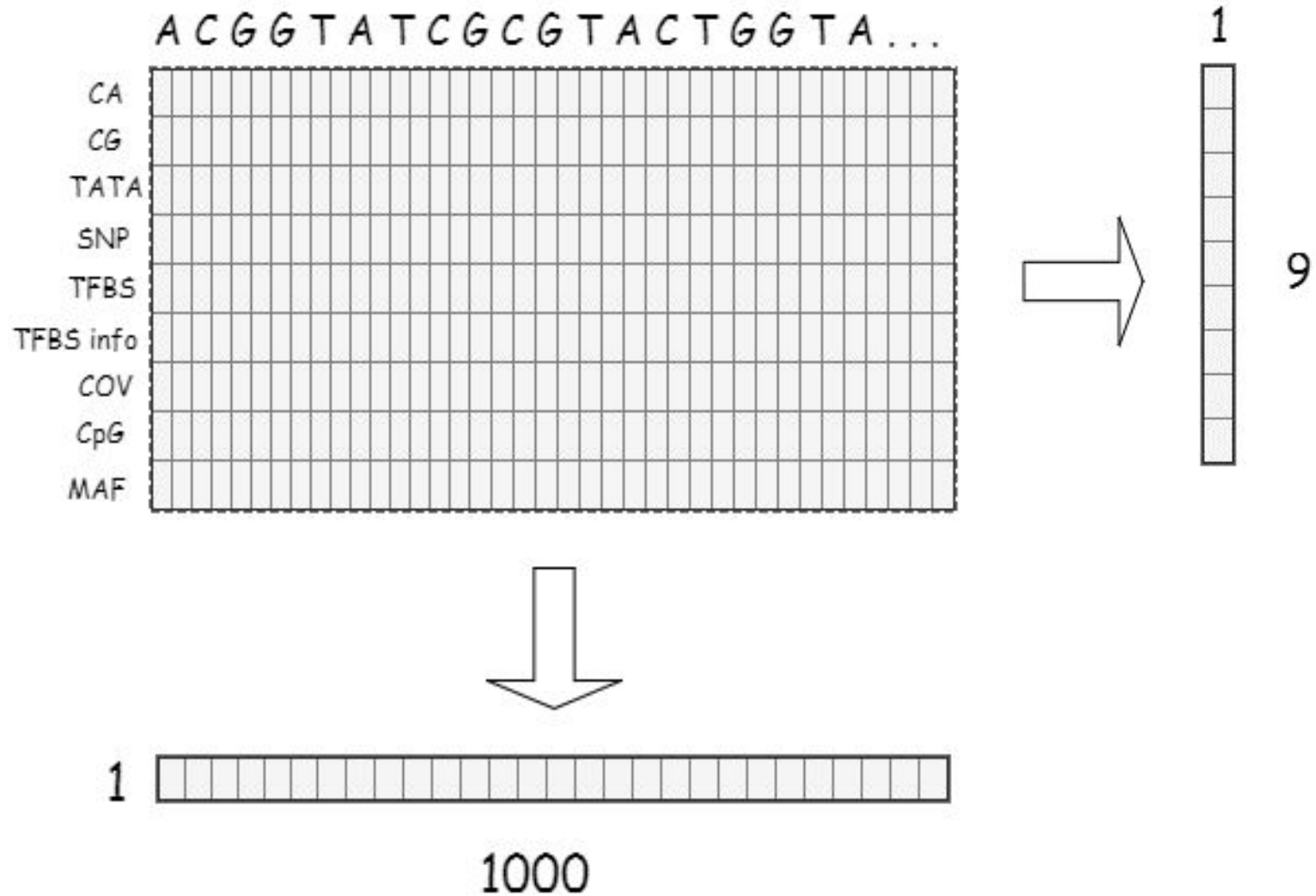


# Обработка данных

- Из каждой последовательности длиной 2000 вырезан случайный участок длиной 1000

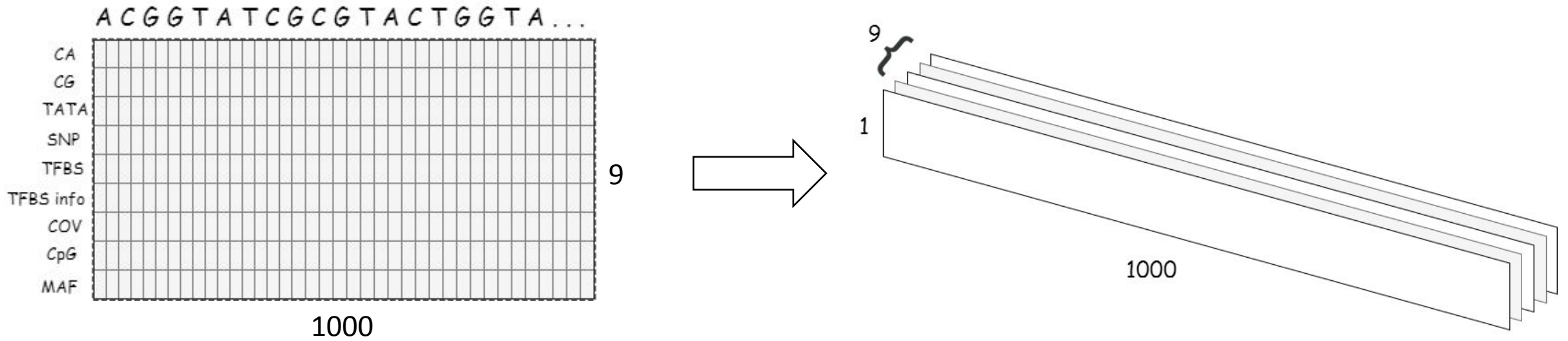


# Machine learning



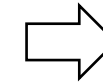
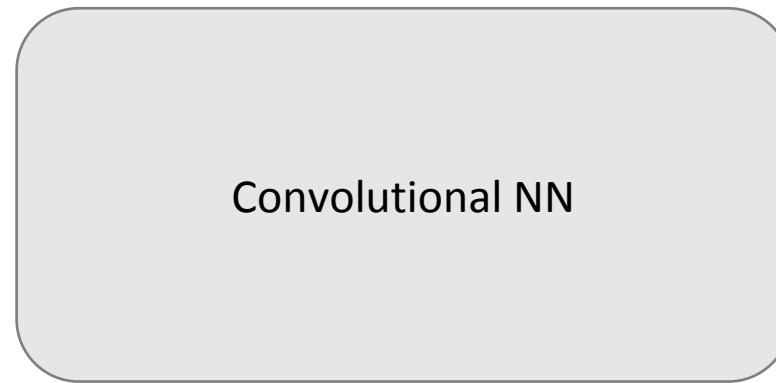
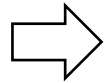
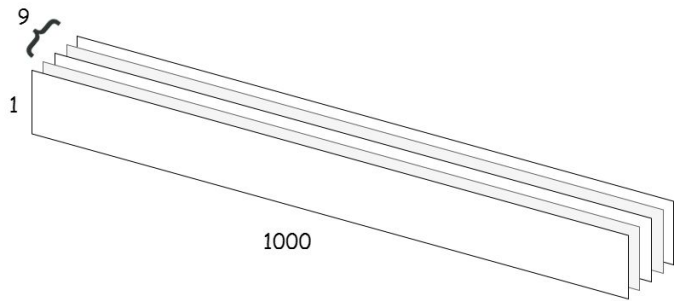
# Deep learning

- Конволюционная нейронная сеть



# Deep learning

- Конволюционная нейронная сеть

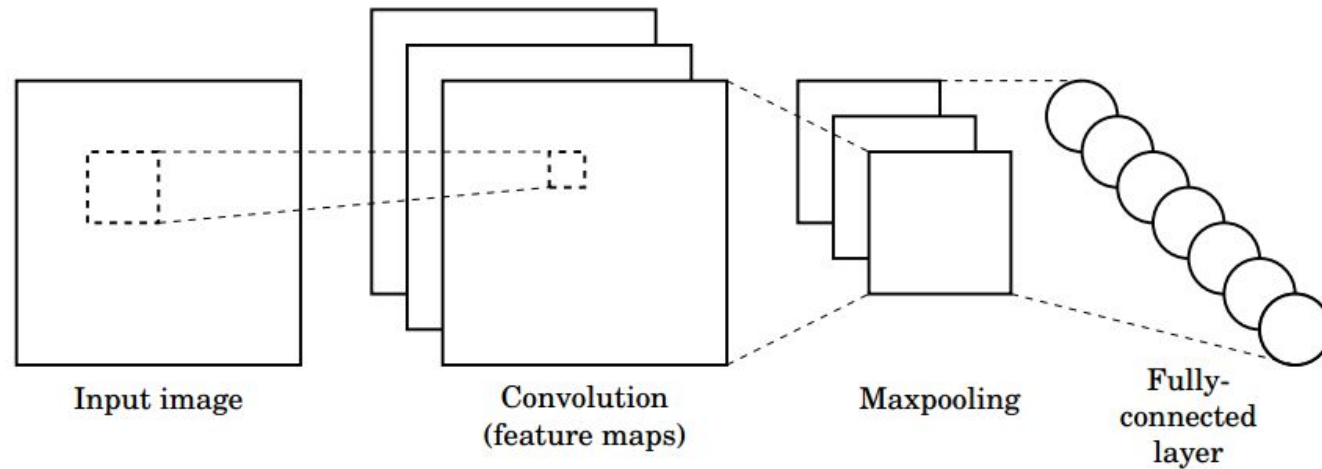


TSS position prediction



# Deep learning

- Разработка архитектуры сети



- 3 слоя Convolution – MaxPooling
- 3 FC слоя

# Результаты и выводы

- Разработали сверточную нейронную сеть
- На имеющихся данных обучили сеть предсказывать положение TSS точностью  $\pm 86$  нуклетоидов для последовательности длиной 1000
- В дальнейшем можно продолжить усложнять архитектуру данной сети для повышения точности ее предсказаний

**Спасибо за внимание!**