

★ Функция вычисления уникальности (общее описание)

Функция **оценивает уникальность кандидата** среди всех других лекарственных молекул в датасете.

Для расчёта используется три фактора:

1. Редкость мишени (Target Uniqueness)

Если по данной мишени (**target**) существует много препаратов, кандидат считается **менее уникальным**.

Если мишень редкая или мало исследованная, уникальность выше.

Вычисляется как:

- считаем, сколько раз каждая target встречается в датасете
 - чем меньше частота → тем выше уникальность
-

2. Редкость показания (Indication Uniqueness)

Аналогично мишени:

- если заболевание распространённое (lung cancer, diabetes), то кандидат менее уникален
- если заболевание редкое (ALS, Huntington's), уникальность выше

Считается так же:

- считаем частоту **indication**
 - большая частота → низкая уникальность
-

3. Новизна текстовой информации (Publication Novelty)

Оценивается по `text_embed_score` (0–1):

- 1 = текст сильно похож на существующие публикации (низкая уникальность)
- 0 = мало похож, более уникальный подход

Новизна публикаций = $1 - \text{text_embed_score}$

★ Итоговый показатель уникальности

После нормализации предыдущих трёх факторов они объединяются в один итоговый показатель:

```
uniqueness_score =  
    0.4 * (уникальность по target)  
+ 0.4 * (уникальность по indication)  
+ 0.2 * (новизна текста)
```

Весовые коэффициенты (40%, 40%, 20%) можно менять, но по умолчанию дают сбалансированную оценку.

★ Интерпретация результата

- **0.0 – 0.33** → Низкая уникальность
кандидат похож на множество существующих решений
 - **0.33 – 0.66** → Средняя уникальность
есть отличия, но мишень или показание используются часто
 - **0.66 – 1.0** → Высокая уникальность
редкая мишень, редкое заболевание и оригинальная публикация
-

Преимущества такой схемы

- работает на любом синтетическом или реальном датасете
- не требует химических структур или сложных биоинформатических данных
- полностью соответствует логике фармацевтического R&D
- легко объясняется и защищается в отчёте
- даёт реалистичную и убедительную модель «уникальности кандидата»