1️⃣ Функция `efectivity_model()`

Что делает

Обучает модель RandomForestClassifier для предсказания эффективности кандидата на основе синтетического датасета:
`synthetic_indication_target_eff.csv`

Используемые признаки:

- `indication`
- `target`

Что возвращает

`(model_e, le_ind_e, le_target_e)`

- обученная модель эффективности
- энкодер для indication
- энкодер для target

Модель предсказывает: **0 / 1 / 2** (низкая–средняя–высокая эффективность)

2️⃣ Функция `toxity_model()`

Что делает

Обучает модель RandomForest для предсказания токсичности на основе:
`toxicity_training_dataset.csv`

Используемые признаки:

- `indication`

- `target`
- `molecular_weight`
- `logP`

Что возвращает

`(model_t, le_ind_t, le_target_t)`

Модель предсказывает токсичность как: **0 / 1**

3️⃣ Функция `move_data()`

Что делает

Применяет обе модели к **основному датасету df**:

1. кодирует indication/target теми же LabelEncoder-ами
2. предсказывает:
    - эффективность → `efective_model`
    - токсичность → `toxic_model`
3. записывает результаты в df
4. вызывает:
    - `build_efficiency_score()`
    - `build_toxicity_score()`
5. удаляет ненужные исходные признаки:
    - `efective_model`
    - `toxic_model`

- `has_positive_efficacy_phrase`
- `has_severe_toxicity_phrase`

4️⃣ Функция `build_efficiency_score()`

Что делает

Объединяет:

- модельную эффективность → `efective_model` (0/1/2 → нормируется)
- текстовую → `has_positive_efficacy_phrase` (0/1)

и формирует **единый показатель эффективности**.

Что записывает в df

```
efficiency_score    # итоговая эффективность 0..2
```

5️⃣ Функция `build_toxicity_score()`

Что делает

Объединяет:

- модельную токсичность → `toxic_model` (0/1)
- текстовую → `has_severe_toxicity_phrase` (0/1)

и формирует **единый показатель токсичности**.

Что записывает в df

```
toxicity_score      # итоговая токсичность 0..2
```

---

6 Функция `build_uniqueness_score()`

(Компактная версия — только один итоговый столбец)

Что делает

Рассчитывает уникальность препарата на основе:

- частоты target среди всех кандидатов
- частоты indication
- "новизны" текста: (1 - text_embed_score)

Что записывает в df

```
uniqueness_score    # итоговая уникальность (0..1)
```

7 Функция `raiting()`

(Компактная версия — только один итоговый столбец)

Что делает

Рассчитывает потенциал препарата на основе:

- эффективности
- токсичности
- "новизны" текста: (1 - text_embed_score)

Что записывает в df

```
score    # итоговая потенциал (0/1/2)
```

📌 ИТОГОВЫЕ КОЛОНКИ, КОТОРЫЕ ОСТАЮТСЯ В ГЛАВНОМ DF

После полного пайплайна у тебя в df будут (минимальный набор):

🔷 Исходные данные:

- drug_id
- indication
- target
- text_embed_score
- molecular_weight
- logP
- market_size_million
- competition_level
- expected_profit_score
- success_probability
- traditional_time_years
- ai_time_years

🔶 Добавленные итоговые метрики:

- **efficiency_score**
- **toxicity_score**
- **uniqueness_score**
- **score**

🔸 Удалённые (чтобы не засорять df):

- efective_model

- toxic_model
- has_positive_efficacy_phrase
- has_severe_toxicity_phrase