*Построение обучаемого алгоритма для анализа формулировок образовательных и карьерных целей человека на предмет соответствия критериям целеполагания, степени конкретности и тематической области цели*

**О задаче**

В основе датасета - результаты открытого онлайн-опроса людей об их образовательных и карьерных целях, содержащие числовую и текстовую информацию. Об участниках опроса известны:

* пол и возраст, образование, город проживания (информация о человеке);
* формулировка цели, оценка человеком тематических категорий цели, оценка предполагаемого времени достижения и первых шагов (информация о цели).

Часть датасета была размечена профессиональными коучами для уточнения соответствия заявленных человеком целей критериям [SMART](https://ru.wikipedia.org/wiki/SMART), степени абстрактности и тематической области. Другая часть датасета подобной разметки не имеет.

Задача участников хакатона: построить алгоритмизированный механизм оценки текстовых формулировок цели человека на предмет соответствия критериям SMART, областям образования или карьеры, уровню абстрактности и тематической области.

Примеры разметки и списки полей с пояснением выдаются участникам дополнительно.

**Что ожидается**

Решение представляет собой обучаемый алгоритм, способный:

* выявлять “шум” в формулировках (случайные слова, слишком короткие слова, тестовые прохождения и т.п.);
* определять, насколько заявленная человеком цель соответствует имеющимся критериям SMART;
* определять, относится ли заявленная человеком цель к образованию или карьере;
* определять, насколько заявленная человеком цель является абстрактной;
* определять, к какой тематической области относится заявленная человеком цель.

Также предоставляются оценки надёжности предлагаемого алгоритма (precision, recall, f-мера…).

**Описание датасета**

Датасет содержит текстовые значения свободных формулировок целей, времени достижения и предполагаемых первых шагов и преград. Остальные значения представляют собой числовые значения номинальных или порядковых переменных.