**Задание**

**Описание**: К нам пришли из смежной дирекции, с просьбой сделать алгоритм, который будет выполнять “разметку” документов на базе произвольной таксономии. Раньше такую разметку делали в полуручном режиме при помощи регулярных выражений, но для составления запросов к новым таксономиям требуется слишком много человеко-часов. Процесс “разметки“ можно представить как задачу классификации произвольного текста, однако основная сложность состоит в том, что классы заранее неизвестны, при этом у них есть название на естественном языке, отражающее смысл того, что необходимо искать в тексте. Кроме того, таксономия обычно состоит из нескольких уровней, которые могут содержать в себе контекст для уточнения дочерних или родительских узлов. Дополнительную сложность вносит то, что поскольку дирекция отказалась от ручной разметки, можно считать, что обучающие данные отсутствуют.

**Входные данные:**

[Тестовое задание](https://docs.google.com/spreadsheets/d/1eLGSfE_Ne4YqAp5oVRRLbRcKCglhqNOwDZHnU5pyTR4/edit?usp=sharing)

На листе “Таксономия” содержится кусок одной из таксономий, по которой необходимо размечать тексты. В ней описываются задачи выполняемые пилотом БАС. Таксономия состоит из корневого уровня (Level\_0) и 3-х уровней обобщения задач. Листовыми узлами считаются те, у которых нет дочерних узлов.

На листе “Тестовые данные” находятся тексты для тестирования подходов. Соответствие узлам таксономии выполнялось при помощи одной из моделей (колонка Label), после чего вручную был проставлен маркер Cls (=1 - правильное сопоставление, =2 - “частично” правильное, =0 - неправильное). Таблица содержит тексты не для всех узлов, но может позволить лучше понять суть задачи и попытаться частично оценить/сравнить различные подходы.

**Задачи**:

1. Предложи (в текстовом виде) несколько подходов плоской классификации (используя только “листья” таксономии);
2. Реализуй один из предложенных вариантов (желательно тот, который кажется перспективным, а не наиболее простой);
3. Оцени точность классификации на тестовых данных;
4. Предложи подход(-ы) учета иерархиичной структуры таксономии для улучшения качества предсказаний.