



## Задача «Разработка алгоритма по прогнозированию карьерной траектории сотрудника»

### Введение

Технологии искусственного интеллекта трансформируют сферу управления персоналом, позволяя решать практические HR-задачи. Уже сейчас умные алгоритмы могут подобрать кадры, спланировать размер фонда заработной платы, выбрать образовательные курсы, оценить мотивацию, эффективность труда и даже спрогнозировать карьерное развитие специалиста.

Аналитика по определению карьерной траектории строится, в том числе, на основе данных по использованию различных корпоративных информационных систем (система мониторинга рабочего времени, сервисы видеоконференцсвязи, IP-телефонии, СЭД, СКУД, электронная почта). Таким образом, для некоторых профессий, связанных с активным использованием электронных устройств (программист, бухгалтер, маркетолог и т.д.), можно выделить цифровой профиль успешных специалистов, выполняющих задачи своевременно и качественно и получающих новые должности.

Имея на руках данные о повышении работников и информацию об их активности, истории посещения офиса и успеваемости по задачам, можно получить прогнозную модель, которая будет предсказывать изменение должности сотрудников автоматически.

### Условие задачи

На основе метаданных работников разработайте прогнозную модель, которая будет предсказывать, какой тип повышения получит сотрудник в будущем.

### Описание входных значений

**Calls.csv** — файл, содержащий данные о звонках

- Date: дата звонка;

- Calltime: длительность звонка в часах;
- NumberOfCalls: количество звонков;
- Вид учета времени: будни/выходные;
- InOut: входящий или исходящий звонок (ToUser: входящий, FromUser: исходящий);
- ID: уникальный номер сотрудника.

**ConnectionTime.csv** — определение опозданий по данным

- dateNum: дата;
- maxLogOff: последний выход из сети;
- Нормативное время начала раб.дня: время начала рабочего дня согласно должности и часового пояса;
- Фактич. время начала раб.дня: фактическое время начала рабочего дня
- Время опоздания: длительность опоздания;
- Признак опоздания: опоздание или нет;
- Вых/Будни: выходной или будний день;
- ID: уникальный номер сотрудника.

**Education.csv** — файл, содержащий данные об образовании сотрудников:

- Табельный номер руководителя;
- Табельный номер сотрудника;
- ID: уникальный номер сотрудника.
- Специальность: название специальности.

**SKUD.csv** — файл, содержащий данные прихода/ухода сотрудников на рабочее место

- Дата: дата входа/выхода в офис;
- Приход.1: первый вход;
- Уход.1: последний выход;
- Длительность общая: длительность пребывания в офисе;
- Длительность раб.дня без обеда: длительность пребывания в офисе без обеда;
- Вых/Будни: выходной или будний день;
- ID: уникальный номер сотрудника.

**Tasks.csv** — файл, содержащий данные о исполнении задач:

- Статус по просрочке: без нарушения срока или с нарушением;
- Срок плановый: плановая дата завершения;
- Просрочено, дней: на сколько дней просрочено завершение задания;
- Длительность просрочки: без нарушения срока, более 30 дней, менее 30 дней, менее 7 дней;
- Вид документа: виды документа, связанного с заданием (СЗ, договор и т.п.);

- Дата старта задания;
- Дата завершения задания плановая;
- Дата завершения задания фактическая;
- Состояние задания:
  - завершено – завершенное задание,
  - на приёме, на проверке – завершенное, но еще не принятое контролёром задание,
  - делегировано – передано другому сотруднику,
  - все остальные состояния означают, что задание не завершено.
- ID: уникальный номер сотрудника.

**TimenNetwork.csv** — интервалы времени в сети в системе учета рабочего времени сотрудника за компьютером

- Вых/Будни: выходной или будний день;
- monitorTime: общее время активности пользователя в сети;
- startTime: дата старта активности;
- ID: уникальный номер сотрудника.

**WorkingDay.csv** — файл, содержащий интервалы времени активности система учета рабочего времени сотрудника за компьютером

- startTime: дата старта активности;
- activeTime: активное время (пользователь активно пользуется мышью и клавиатурой);
- Вых/Будни: выходной или будний день;
- monitorTime: общее время активности пользователя;
- ID: уникальный номер сотрудника.

**train.csv** — файл, содержащий данные сотрудников с типом повышения для тренировки;

- "Перевод по должности": 0,
- "Перевод по подразделению с сохранением должности": 1,
- "Перевод и по подразделению и по должности": 2,
- "Перевод с сохранением должности и подразделения": 3

**test.csv** — пример файла для отправки, содержит id работников, для которых требуется предсказать тип повышения.

## На что стоит обратить внимание

Вам дается излишне много данных для анализа, размещенных в разных таблицах. По этому не стоит их пугаться, сначала попробуйте предсказать

повышение на основе небольшого числа наиболее значимых действий, после — расширяйте список признаков.

## Метрика

В качестве метрики выступает Recall по столбцу **type**.

Recall считается как:

$$recall = \frac{TP}{TP + FN}$$

где:

TP (True Positive) — количество верно угаданных значений одного класса

FN (False Negative) — количество неправильно угаданных значений класса

## Правила чемпионата:

1. С момента открытия датасета до момента завершения приема решений репозиторий участника, в котором он ведет разработку по задаче текущего чемпионата, должен оставаться закрытым.
2. Участник обязан открыть доступ к репозиторию на чтение по ссылке (которая была прикреплена в ЛК в поле «Ссылка на код (гитхаб)») не позднее чем в течение 12 часов с момента окончания дедлайна отправки решений на региональном чемпионате.
3. Согласно п. 5.8 Положения в процессе верификации решений организаторы и технические эксперты, проверяющие решения участников, в праве назначить интервью с участниками чемпионата. Участник получит приглашение и ссылку на интервью не позднее чем за 12 часов до публикации итогового лидерборда. Пропуск интервью участником является поводом для дисквалификации.
4. Организаторы вправе исключить участника из призовых позиций лидерборда за непредоставление одного из артефактов решения задачи: тизера, скринкаста, презентации, ссылки на репозиторий.
5. Организаторы вправе дисквалифицировать участника в случае выявления плагиата кода или несоблюдения Положения конкурса.

6. Участник, получивший 2 дисквалификации за сезон конкурса, попадает в чёрный список с дальнейшим отстранением от участия в чемпионатах до конца сезона.