

## Кейс “Классификация позитивных комментариев по работе управляющей компании”

<b>Постановщик задачи</b>	Компания «Самолет»
<b>Название кейсового задания</b>	Модель склонности клиента к приобретению машиноместа
<b>Краткое описание кейсового задания</b>	Классификация позитивных комментариев по работе управляющей компании
<b>Полное описание кейсового задания</b>	<p>Предварительно требуется подготовить исходные данные</p> <p>Разработать модель классификации комментариев по работе управляющей компании на категории:</p> <ul style="list-style-type: none"><li>Нравится скорость отработки заявок</li><li>Нравится качество выполнения заявки</li><li>Нравится качество работы сотрудников</li><li>Понравилось выполнение заявки</li><li>Вопрос решен</li></ul> <p>Входные данные представляют собой набор комментариев с отметками 1 в колонке соответствующей категории</p> <p>Оцениваться результат будет метрикой ROC-AUC</p>
<b>Образ результата</b>	<p>Ожидаемый результат: подготовленные данные, разработанная модель машинного или глубокого обучения, позволяющая на основе имеющихся данных проводить классификацию комментариев по нужным категориям</p> <p>Подготовленная презентация разработанной модели, содержащая основные аспекты:</p> <ul style="list-style-type: none"><li>● подход к решению бизнес-задачи</li><li>● подробно изложенный процесс моделирования</li><li>● полученные результаты</li><li>● раскрытые инсайды если есть.</li></ul>
<b>Стек технологий, рекомендуемых к использованию</b>	Jupyter notebook, Python, scikit-learn, pandas, Data Science, Машинное обучение
<b>Мощности</b>	Google Colab (бесплатный доступ)

<p><b>Необходимые данные, дополнения/пояснения/уточнения</b>  <b>Информация для участников интенсива, которая поможет разобраться в проблематике и предложить качественные решения кейсового задания</b></p>	<p>Данные содержат дубли и, возможно, ошибки разметки. Число примеров по классам неравномерно и примерно соответствует реальному распределению комментариев.</p> <p>При моделировании рекомендуется использовать кросс-валидацию.</p> <p>При оценке будут учитываться точность модели, использованные подходы к предобработке данных, отбора признаков, выбранный алгоритм построения модели.</p> <p>Будет плюсом, если презентация будет отражать следующую информацию:</p> <ul style="list-style-type: none"> <li>- Какие методы использовались при подготовке данных?</li> <li>- Какой предварительный анализ данных проводился? Есть ли какие-то интересные выводы?</li> <li>- Какие подходы к отбору признаков пробовали? Какой в итоге выбрали и почему?</li> <li>- Добавляли ли новые фичи?</li> <li>- Какую валидационную схему использовали?</li> <li>- Как подбирали гиперпараметры? Какие параметры сильно влияют на метрики?</li> <li>- Какие алгоритмы машинного обучения применяли и какой выбрали для итоговой модели?</li> <li>- Работает ли модель стабильно? Наблюдалось ли переобучение? Как с ним боролись?</li> <li>- С какими сложностями столкнулись?</li> </ul>
<p><b>Дополнительно</b></p>	<p>Есть несколько причин, почему этот кейс может быть привлекательным:</p> <p>Реальная бизнес-задача с большим объемом данных от крупной компании. Это позволяет участникам потренировать навыки решения задач ML в промышленных масштабах и реальных условиях.</p> <p>Успешное решение кейса даст участникам хороший пример для портфолио и возможность выделиться при трудоустройстве в крупные компании с опытом решения реальной боевой задачи.</p>