Финальный проект курса “Специализация Data Science” (SkillFactory)

Определение территории дистрибьютора, на которой находится торговая точка. Создание системы для подсказки оператору

Сергей Пигида

2020

# 

# Бизнес-кейс

В отделе продаж имеется периодический бизнес процесс выплат дистрибьюторам, которые осуществляют продажу продукции компании в определенной географической области. Для подсчета размера выплат используются данные по продажам торговых точек, которые должны быть отнесены к тому или иному дистрибьютору. Сложность заключается в том, что в одну торговую точку могут продавать несколько дистрибьюторов, при этом необходимо точку однозначно отнести точку к конкретному дистрибьютору.

Возможность различать торговые точки по географическому принципу затруднена: неявные границы территорий внутри больших городов, пограничные точки на границах областей, дистрибьюторы работают в одной области, но в разных форматах торговых точек и т.п..

Для обеспечения процесса в каждом регионе имеется оператор, который вручную назначает торговым точкам правильный филиал нужного дистрибьютора. Чтобы снизить человеческий фактор и ускорить процесс формирования выплат было принято решение внедрить систему для формирования подсказки оператору о том, к какому филиалу дистрибьютора относится новая точка.

# Постановка задачи

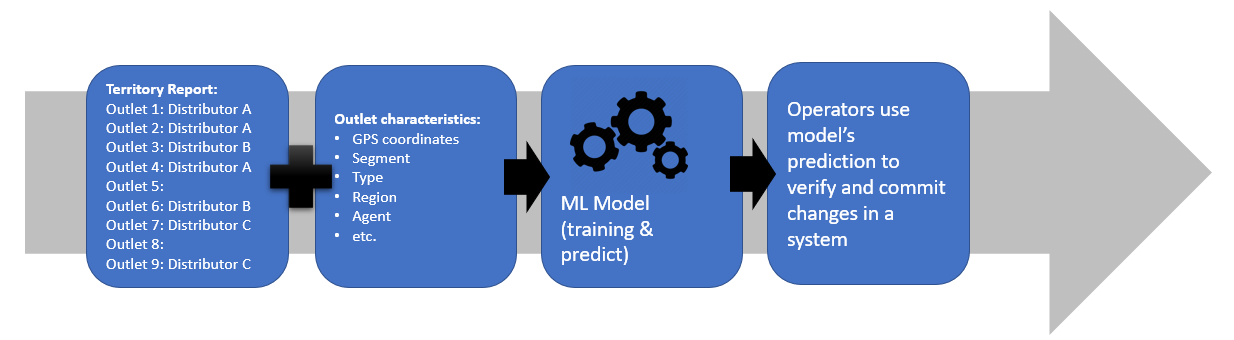
Необходимо в имеющийся отчет Territory Management Report, с которым работают операторы, добавить новые колонки в которых только для новых торговых точек показывать ТОП-3 филиалов дистрибьюторов, к которым относится данная торговая точка. Предсказание выполнять на основе доступной информации о торговой точке: регион, сеть, формат магазина, координаты и т.п.

# Описание решения

Поскольку бизнес-процесс имеет особый легальный статус, необходимо минимальное вмешательство в бизнес-процесс. Для решения задачи будет разработана программа, которая в качестве входной информации будет использовать два Excel-файла: исходный отчет Territory Management Report и файл с координатами торговых точек. На выходе будет формироваться новый файл Territory Management Report Updated с дополнительными колонками информации.

В процессе выполнения программы поступившая информация о точках будет очищена и подготовлена для обучения в модели машинного обучения. Обучение необходимо выполнять при каждом запуске отчета из-за периодических изменений территорий дистрибьюторов. На выходе модели будет оцениваться качество работы модели, а затем запускаться предсказание для новых торговых точек.

Дополнительная информация в обновленном отчете будет использоваться операторами для проверки и утверждения. По завершению работы оператор будет вручную удалять дополнительную информацию, поскольку отчет в неизменном виде участвует в дальнейшем процессе, в том числе используется для загрузки в систему для дальнейшего выполнения расчета.



# Реализация и продукционирование решения

Для реализации решения был выбран язык Python, поскольку имеет библиотеку для работы с Excel-файлами (Openpyxl), а также библиотеку для работы с машинным обучением (Scikit-Learn). Для продукционирования и передачи программы конечному пользователю были принято решение скомпилировать программу в выполняемый файл (библиотека Pyinstaller) для запуска на корпоративных компьютерах с операционной системой Windows 10 (решение продиктовано необходимостью быстрого внедрения прототипа, а также малым количеством пользователей).

## Git-репозиторий

<https://github.com/pigidser/Territory-Prediction>