**Из IESE City in Motion 2020**

Седьмое исследование

В результате Ковид-кризиса появились особенности, о которых раньше никто не задумывался.

Многие рекомендации сити-менеджеров, которые давались последние годы, перестали быть нужны. Стремление к более высокому показателю плотности населения заменилось на уровень социального дистанцирования; потребность в общественном транспорте снизилась, среда для личного автотранспорта пришла взамен; вместо людных общественных событий – укрытие внутри домов. Само название инициативы Citi in Motion в 2020 году стало желанием в будущем, а не реальностью.

Таким образом, пришлось оценивать способность городов пережить сложные «ковидные» обстоятельства.

Каждый год авторы пытаются улучшить индекс, т.е. набор показателей разный, методы расчета меняются (проверить!).

В 2020 году – 101 индикатор (на 5 больше)

Сравнение рейтингов разных лет происходит на разной базе индикаторов (количество, разные способы измерения, разные источники) и с изменяющимися методиками расчета рейтинга. То есть, сравнить один и тот же индекс в разные годы нельзя.

Также добавляются новые города в рейтинг. В каждом исследовании приводится сравнение рейтинга за 3 года, чтобы понять тенденции изменения. То есть, изменения рейтинг за большее, чем 3 года, время, по мнению авторов, дает меньше объективности в оценке тенденции из-за разного количества городов, индикаторов и методики.

Два измерения в модели, которые принимают в расчет 10 ключевых измерений (2020):

1. human capital
2. social cohesion
3. the economy
4. public management
5. governance (вместе)
6. the environment
7. mobility and transportation
8. urban planning
9. international projection
10. technology

Итого: 9 категорий; 174 страны.

Измерения в модели 2015:

1. human capital
2. social cohesion
3. the economy
4. public management
5. governance (отдельно)
6. mobility and transportation
7. the environment
8. urban planning
9. international projection
10. technology

Сегодня у нас много «городских» индикаторов, но большинство из них не стандартизованы и не постоянны, поэтому их нельзя использовать для оценки всех городов. Пока не удается создать всем понятные стандартные индикаторы, которые будут доступы все заинтересованным людям).

Модель предполагает набор шагов, которые включают диагностику ситуации, развитие стратегии и ее реализацию. Первый шаг, чтобы сделать хороший диагноз – проанализировать статус ключевых индикаторов.

Нужно брать рейтинг 2020, 2019 (как пограничные годы до ковида и после) и смотреть:

Изменение рейтинга готова до и после

Искать общее изменение показателей / секторов в каждом городе (будет ли связь: что в 2020 г. Победили те, у кого лучше медицина, или что-то еще, а проиграли те, у кого были лучше социальные активности в 2019? Можно ли найти какие-то тенденции между показателями и рейтингами до и после?

Плюс берем рейтинги всех лет, с 2013 г, если найдем (или берем все, какие найдем).

Только общий рейтинг с местом.

В отчете 2020 есть раздел, в котором представлено сравнение топ 10 городов из разных индексов – см. файл отдельный.

Сделать по нему вывод: …

По частоте встречаемости в топ 10 сделать рейтинг

В каждом годовом отчете дается изменение рейтинга топ 50 городов за 3 года.

Так в отчете 2020 г. Есть рейтинги за 2019, 2018, 2017, а в отчете 2019 г. Рейтинги 2018, 2017, 2016 и т.д. Так вот, рейтинг 2018 г. В отчете 2019 не равен рейтингу 2018 в отчете 2020 г. Единственное, почему это может быть: к оценкам за три года в каждом отчете применяют методологию и индексы, использующиеся только в отчете этого года, таким образом рейтинги 2018 г. Считаются в 2020 г. Не по тому же набору индексов и их весов, что считались в 2019 г. Таким образом, составить картину рейтингов всех лет с одинаковым подходом к набору показателей и весовым коэф-там, нельзя. То есть мы можем составить:

1. Изменение рейтинга 2017, 2018, 2019 – по одинаковой методике 2020 г.
2. Изменение рейтинга 2013 – 2019 каждый по своей методике, т.е. мы сравниваем, какие места достались городам при подведении итогов с разными подходами к индексам, весам и особенностям расчета. С позиции общественного мнения такой подход имеет ценность – мы смотрим, как котируются для общественного сознания города мира (если принять, что подведенные рейтинги отражают положение дел и влияют на общественное сознание).

**From McKinsey Global Institute.**

SMART CITIES: DIGITAL SOLUTIONS FOR A MORE LIVABLE FUTURE, June 2018

В городах живет более половины мирового населения, к 2050 году будет жить 2,5 млрд. человек. Всем видны экологические и инфраструктурные проблемы городов, и в целом запрос от жителей на улучшение качества жизни.

Умные технологии помогают городам справиться с вызовом, фактически это общественная инвестиция. Все умные технологии начинаются с данных. Жизнь в городе можно описать очень большим количеством данных. Чем умнее становятся города, тем они более жизнеспособные и более отзывчивые.

Цифровые технологии дают наиболее мощные и экономически эффективные дополнения к качеству жизни в городах. Они добавляют 10-30%. Но даже самые технологичные из городов в мире еще находятся в начале пути по улучшению качества жизни.

Можно спасать жизни, предотвращать преступления, уменьшать отходы, экономить время, расширять социальные связи и т.п.

Стать умным городом – это цель сама по себе, но по сути технологии дают способы оптимизировать инфраструктуру, ресурсы, пространства.

Умные города используют данные и технологии для принятия лучших решений.

Более эффективный, удобный и стабильный город, который дает лучшие решения для живущих в нем людей.

Было выбрано 50 городов для оценки. Оценка проводилась онлайн.

В обзоре McKinsey 6 показателей:

* Time and convenience
* Safety
* Health
* Cost of Living
* Environment
* Jobs
* Social connectedness

и 8 областей (сфер):

* Mobility
* Security
* Healthcare
* Energy
* Water
* Waste
* Economic Development and Housing
* Engagement and Community

Но количественные оценки даны по показателям:

* Technology: Sensor, Communication, Open data portal. 37 points max.
* Applications: Mobility, Security, Utilities, Healthcare, Economic development, housing and community. 55 points max.
* Citizen experience: Awareness, Usage, Satistaction. 30 points max.

То есть вольная интерпретация по сравнению с описанием. Как таковой общий рейтинг не дается. Рейтинг по каждому из трех показателей составляется легко, но объединить его в итоговый рейтинг города – как? Простой суммой трех оценок – неверно; взять % от максимальной оценки в каждом рейтинге и взять среднеарифметический % от макс. оценки – притянуто. Хочется сравнить рейтинги городов в разных оценках, но если у них принципиально разные подходы к оценке, то сравнение будет некорректным. Как поступить?

**2020/21 TOP 50 SMART CITY GOVERNMENT RANKINGS**

Smart City innovations offer mayors across the world a promising new way of engaging citizens and increasing quality of life. To offer city leaders guidance on assessing the readiness of their governments to develop, facilitate, or track their smart city initiatives, Eden Strategy Institute developed an independent ranking of the Top 50 Smart City Governments.

<https://www.smartcitygovt.com/>

50 городов

Критерии:

VISION. A clear and well-defined strategy to develop a “smart city”

LEADERSHIP. Dedicated City leadership that steers smart city projects

BUDGET. Sufficient funding for smart city projects

FINANCIAL. Financial incentives to effectively encourage private sector participation (e.g. grants, rebates, subsidies, competitions)

SUPPORT PROGRAMMES. In-kind programmes to encourage private actors to participate (e.g. incubators, events, networks)

POLICIES. A conducive policy environment for smart city development (e.g. data governance, IP protection, urban design)

ECOSYSTEMS. A comprehensive range of engaged stakeholders to sustain innovation

PEOPLE-CENTRICITY. A sincere, people-first design of the future city

TALENT-READINESS. Programmes to equip the city’s talent with smart skills

TRACK RECORD. The government’s experience in catalyzing successful smart city initiatives

Оценки от 1 до 4 в каждой категории, чем выше, тем лучше.

Затем просуммированы все 10 оценок критериев и ранжированы по сумме.

EDA – exploratory data analysis, исследовательский/разведывательный анализ данных.

Анализ основных свойств данных, нахождение в них общих закономерностей, распределений и аномалий, построение начальных моделей, зачастую с использованием инструментов визуализации.

цели такого анализа следующим образом:

* максимальное «проникновение» в данные,
* выявление основных структур,
* выбор наиболее важных переменных,
* обнаружение отклонений и аномалий,
* проверка основных гипотез,
* разработка начальных моделей.

Основные средства разведочного анализа — изучение вероятностных распределений переменных, построение и анализ корреляционных матриц, факторный анализ, дискриминантный анализ, многомерное шкалирование.

Какие из трех классических задач буду решать:

*Классификация - Определять тип (Мужчина/Женщина) Обучение с учителем*

*● Возьмет клиент кредит или нет*

*● Что изображено на картинке*

*● Отзыв положительный или негативный*

*Регрессия - Прогнозировать значения для объектов (Возраст, Доход, Рост) Обучение с учителем*

*● Предсказание погоды*

*● Прогноз цены акций*

*● Прогноз спроса*

*Кластеризация - Группировать (Школьники, Бизнесмены, Политики, Любители Чая) Обучение без учителя*

● Какие основные темы обращений клиентов?

● Какие группы пользователей у нас есть?

Общая схема

1. Получить данные

2. Подготовить объекты и признаки

Данные нужно привести в порядок: привести к одной размерности, проверить на пропуски, на выбросы и шумы

3. Разделить данные на обучающую и тестовую выборку при необходимости

4. Выбрать алгоритм машинного обучения

Алгоритм зависит от:

● Задачи (классификация/регрессия/кластеризация)

● Структуры и особенностей данных

● … <- работа data science

Классификация – деревья решения. Важно построить дерево так, чтобы объекты обучающей выборки классифицировались максимально правильно

Классификация – метод ближайшего соседа

Классификация – метод опорных векторов

Регрессия – линейная

Кластеризация – Kmeans

Кластеризация - иерархическая

5. Обучить модель на обучающей выборке

6. Оценить качество на тестовой выборке

Виды аналитики

Описательная аналитика использует данные для понимания прошлого и настоящего.

Отвечает на вопрос «Что случилось?»

Диагностическая аналитика анализирует информацию, чтобы ответить на вопрос «Почему это случилось?»

Используются методы анализа данных с целью выявления причин того или иного события.

Прогнозная аналитика прогнозирует события в будущем, отвечая на вопрос «Что

может случиться?» на основе накопленной информации.

Активно использует статистику и машинное обучение.

Предписательная аналитика отвечает на вопрос «что делать?»

Также использует статистику, машинное обучение и другие методы анализа данных.

Оценка качества данных и исправление

Проблемы:

● Дублирование (один и тот же город/страна написаны по-разному)

● Противоречия (в моем случае – в данных из одного источника нет противоречий, они уже обработаны на противоречивость авторами исследований; в моем случае противоречивость может быть только в интерпретации данных и применении к ним верного метода анализа)

● Неоднозначности (нет уверенности в правильной идентификации городов из-за особенностей написания, некоторые города в Китае, Корее пишутся почти одинаково, но в таблице рейтингов таких городов в конце концов не оказалось, было опасение, что такие города могут быть и неправильно прочитаться)

● Полнота (все ли данные по численности, ВВП нашлись?)

Показатели качества данных:

Корректность / Точность – да

Согласованность – да

Полнота – да, почти, с учетом допущения, что временное отличие в один год не существенно (население – за … год, ВВП за … год, CIMI – за 2019 г.)

Своевременность – не применимо

Метаданные - ???

Были проведены проверки данных:

- приведение к одному справочнику названий городов и стран

- создание справочника категории по «высоте» рейтинга

- пустых значений нет

- числовые значения приведены в один формат (разделитель – точка)