

Космонавты (Data mining/economic modeling)

Вы - консультант крупной консалтинговой компании и известный аналитик.

К вам пришел основатель компании X, частной российской компании, которая хочет развивать спутниковый интернет (аналог Старлинк). В штате компании уже сейчас 1000 человек, 750 из которых являются инженерами. Компания уже провела два успешных пуска и испытания собственных спутников и смотрит в будущее очень оптимистично.

Главной проблемой для дальнейшего развития стали кадры.

Оказалось, что в России очень мало специалистов, способных разрабатывать современные космические технологии и воплощать их в железе.

Как смелый покоритель космоса, его не остановит никакое препятствие и он решил сам построить инженерную школу по космическим технологиям.

Для старта этого проекта ему не хватает вводных данных: сколько потребуется людей и сколько стоит их обучить?

Направления, где нужны специалисты:

- Инфокоммуникационные технологии и системы связи
- Конструирование
- Технологии электронных средств
- Радиоэлектронные системы и комплексы
- Приборостроение
- Лазерная техника и лазерные технологии
- Радиофизика
- Системы управления движением и навигация
- Информатика и вычислительная техника
- Прикладная математика и физика
- Высокотехнологичные плазменные и энергетические установки

Ожидаемый вид решения:

Пользуясь открытыми данными (поиск и оценка релевантности данных - часть задачи) и информацией о текущей стадии проекта, постройте аналитическую модель, которая оценивает потребность в специалистах каждого профиля.

Пользуясь открытыми данными оцените стоимость обучения специалиста по каждому направлению.

Идеальный вид (Data mining/Data analytics)

Мистер Чен обожает смотреть на волны.

Он всегда мечтал о доме на скале с видом на океан. В своих мечтах он даже представлял, как в один прекрасный день он увидит цунами или на худой конец какое-нибудь другое стихийное бедствие.

Скопив большое состояние, мистер Чен решил купить дом своей мечты. Однако являясь очень осторожным человеком, он совершенно не намерен подвергать себя или свою собственность риску. Поэтому при выборе дома он хоть и желает быть иметь прекрасный вид из окна, желательно с возможностью лицезреть стихийные бедствия, сам дом должен оставаться в настолько безопасном месте, насколько это возможно.

При этом мистер Чен (гражданин Китая) настаивает, чтобы страна, в которой располагался дом, позволяла китайцам свободно приобретать недвижимость.

Для выбора места для дома мистер Чен обратился за помощью к вам, известному специалисту по анализу данных. Польщенные оказанной честью, вы согласились помочь.

Ожидаемый вид решения:

Пользуясь открытыми данными (поиск и оценка релевантности данных - часть задачи) предложите подходящее место для дома (и метрику для оценивания дома).

Критерии:

- Дом имеет вид на океан или открытое море.
- Страна позволяет приобретать недвижимость в собственность гражданам КНР.
- Расположение безопасно.
- Желательно, чтобы в пределах видимости была вероятность наблюдать цунами или другой природный катаклизм.

Подсказка: для анализа могут оказаться полезными датасеты вот такого типа

<https://www.kaggle.com/datasets/stealthtechnologies/earthquakes-dataset>

Анализ стратегий управления буфером (algorithms)

Как вы знаете, в стандартной комплектации компьютера есть два типа памяти: оперативная память и жесткий диск. Доступ к оперативной памяти намного быстрее, чем к диску, однако ее объем существенно меньше. Поэтому важно уметь правильно управлять данными, чтобы как можно чаще обращаться именно к оперативной памяти, а не к диску. Эта задача актуальна не только для вашего персонального компьютера, но и для огромных промышленных систем. В рамках этой задачи мы рассмотрим как это делается в Базах Данных (БД).

Менеджер буфер памяти в базах данных, как и в задаче кеширования, призван улучшить производительность за счет правильного управления хранением данных.

Для знакомства с его устройством и работой мы рекомендуем вот эту презентацию:

https://ethz.ch/content/dam/ethz/special-interest/infk/inst-cp/inst-cp-dam/education/courses/2020-fall/data-management-systems/slides/DMS-HS20-Storage_Buffer_Cache.pdf

Как указано в презентации, в промышленных БД используются следующие алгоритмы:

- Touch Count (Hot/Cold)
- Clock Sweep
- 2Q
- LRU-K/2

В этой задаче от Вас требуется сравнить эффективность указанных алгоритмов по двум показателям:

- [Competitiveness Ratio](#)
- [Regret](#)

Первый подход оценивает отношение целевой метрики (в нашем случае это hit rate) при использовании алгоритма с целевой метрикой, если бы мы знали оптимальную стратегию управления буфером.

Второй подход оценивает разность целевой метрики при использовании алгоритма с значением целевой метрики при использовании оптимального буфера.

Ожидаемый вид решения:

Оцените Competitiveness ratio для указанных алгоритмов.

Оцените regret для указанных алгоритмов в заданной среде.

- Для случая случайного выбора следующей страницы для подгрузки. Т.е. у каждой страницы есть вероятность вызова, которая не меняется со временем (но неизвестна вам) и одинакова для любого состояния буфера.
- Для марковой среды. Т.е. у каждой страницы есть вероятность вызова, которая не меняется со временем (но неизвестна вам), но может зависеть от состояния буфера.
- Для противодействующей среды. Т.е. среда действует наихудшим для алгоритма способом.

Блуждая по решетке (statistics/stochastics)

Саудовская Аравия известна своим желанием строить чудеса света (архитектурные). К 2049-му году все необычные идеи для чудес света закончились и саудиты решили построить улучшенную версию уже имеющихся чудес. Так, в 2055 году была построена Очень Великая Стена. Хотя по всем параметрам ребра между вышками были как у Великой стены в Китае, сама стена имела форму правильной решетки.

Как-то раз один турист решил походить по этой стене. Так как все ее звенья были абсолютно одинаковы и вид открывался только на пески, турист ходил бесцельно, случайно выбирая направления в точках пересечения и считая звенья.

На счете 96 он опять оказался у входа (уже в 5-й раз) и решил уйти.

Как Вы думаете, какую форму имеет элементарная ячейка стены (треугольник, квадрат и т.д.)