



IT RECRUITER

Информационная карта № 1

Платформа для создания интеллектуальных объектов на основе больших данных

Задача №1

Обработка комментарием специализированных youtube каналов с целью идентификации мнения аудитории обсуждаемых вопросов (проблем, мнений) на примере автомобильной области.

Краткое описание:

На вход подается любой url видео с наличием соответствующих условий (обзор из автомобильной области, в обзоре есть явное или неявное обращение за комментариями к зрителям), из которого необходимо выделить все комментарии, провести анализ интенгов, выделить качественные и количественные характеристики, мнения, классифицировать их по автомобильным тематикам обсуждения.

Требования:

1. Представить в виде агрегированной статистики: выявленные темы, распределение людей по темам, распределением мнений по темам (предложить свои критерии).
2. Отдельно в статистике необходимо выделить положительные и отрицательные мнения.
3. Должен быть реализован полный пайплайн обратки в виде jupyter notebook-ов, включающие: краулинг нужной страницы, преобразование, очистку, нормализацию, обработку тем, выделение классов обсуждаемых тем, выделение мнений и распределение среди участников.

Задача №2

Построение модели рекомендации покупки недвижимости на основе открытых данных Интернет.

Краткое описание:

На вход подаются критерии ожидания пользователя, возможно не в полной форме (формат критериев может быть предложен, как пример, возраст, цели приобретения, интересы). Модель рекомендации предлагает определенные варианты с учетом критериев с фактическим обоснованием выбора, путем представления статистических и семантических оценок и т.д.



Требования

1. Представить в виде агрегированной статистики: предложения по объектам покупки как точки на домах на карте без ссылки на конкретное предложение, выбор каждой точки должен быть обоснован установленными критериями (рентабельность, благополучие, инфраструктура, образование, близость к точкам притяжения, экология, социальный уровень и т.д.).
2. Отдельно в статистике необходимо выделить положительные и отрицательные мнения, полученные из отзывов жителей города на примере СПб.
3. Должен быть реализован полный пайплайн обработки в виде jupyter notebook-ов, включающие: обрабатываемые датасеты, операции преобразования, очистки, нормализации данных, выстраивание и формирование критериев, выделение подклассов критериев, выделение положительных/отрицательных мнений, построение общей оценки.



IT RECRUITER

Информационная карта № 2

Платформа для интеллектуализации человеко-компьютерного взаимодействия

Задача №1

Разработка событийного агрегатора по интересам

Краткое описание:

Сервис агрегации событий ИТМО должен предоставлять пользователю системы возможность создания событий/встреч (спортивных, музыкальных и других) для потенциально заинтересованных пользователей и формирования групп/команд по интересам и формам проведения досуга. Кроме того, в пассивном режиме сервис должен представлять собой рекомендательную систему событий и мест досуга по интересам из уже созданных событий и встреч.

Требуемые навыки:

Python 3.6, NLP, Git, HTTP, BaseAuth, OAuth

Задача №2

Разработка сервис-агрегатора студенческих проектов

Краткое описание:

Сервис агрегации проектов должен предоставлять пользователю возможность поиска исполнителей научного (студенческого) проекта с заданными компетенциями по ключевым навыкам/ключевым словам и тематикам, а также поиска уже созданных проектов исходя из своих компетенций и подачи заявки на участие в них.

Требуемые навыки:

Python 3.6, NLP, Git, HTTP, BaseAuth, OAuth

Задача №3

Разработка вопросно-ответной системы ITMOOverflow

Краткое описание:

Сервис должен представлять собой агрегатор вопросов и ответов на различные темы, связанные с обучением, работой и другими видами активной деятельности в Университете ИТМО. У пользователя должна быть возможность самому задать вопрос экспертному сообществу, ответить на выбранный вопрос в рамках своей компетенции, а также произвести



поиск вопросов и ответов по темам и ключевым словам.

Требуемые навыки:

Python 3.6, NLP, Git, HTTP, BaseAuth, OAuth

Задача №4

Разработка сервиса оценки научного руководителя и преподавателей вопросно-ответной системы ITMOOverflow

Краткое описание:

Сервис, реализующий проведение опросов студентов ИТМО (бакалавров и магистров) с целью оценки удовлетворённости научным руководителем, преподавателем. Необходимо реализовать опросную систему минимум из 10 вопросов (параметров), с возможностью выставления оценки по пятибальной шкале (1 - «ужасно», 2 - «неудовлетворительно», 3 - «удовлетворительно», 4 - «хорошо», 5 - «отлично»). Сервис должен агрегировать выставленные оценки по всем пользователям и сохранять текущий средний балл преподавателей.

Требуемые навыки:

Python 3.6, NLP, Git, HTTP, BaseAuth, OAuth

Задача №5

Разработка системы check-in для образовательных и внеучебных мероприятий

Краткое описание:

Сервис должен представлять собой приложение, предоставляющее пользователю возможность автоматической регистрации факта присутствия на различных образовательных и внеучебных мероприятиях ИТМО для дальнейшей агрегации в профиле его цифрового аватара. Система check-in может быть реализована посредством QR кодов или checkbox-ов.

Требуемые навыки:

Python 3.6, NLP, Git, HTTP, BaseAuth, OAuth



IT RECRUITER

Информационная карта № 3

Платформа для управления компьютерными моделями

Задача №1

Разработка систем поддержки принятия решений на основе потоковых данных для технических объектов с применением методов машинного обучения

Краткое описание:

Сервис должен осуществлять сбор и мониторинг данных технологического или бизнес-процесса организации с целью поддержки принятия решений. Примерами результатов мониторинга являются выявление выбросов, отклонение от плана, оповещение о событиях, краткосрочное прогнозирование и т.п.

Данные могут собираться из следующих источников:

- информационные системы организации,
- сервисы сети Интернет,
- датчики реального объекта, оборудования или устройства.

Данные должны иметь возможность представления в форме временных рядов, сохраняться внутри платформы и быть извлекаемыми для анализа с использованием методов машинного обучения. Допускается реализация модели бизнес- или технологических процессов внутри платформы с возможностью оперативного сбора информации из внутреннего контура организации. Для недостающих источников данных допустимо использование генеративных моделей для получения потока данных.

Технические требования:

- Бизнес- или технологический процесс должен быть представлен в форме потока задач, между которыми осуществляется обмен информацией с использованием инструментов платформы. Количество блоков должно быть не менее трех.
- Количество входных параметров для анализа должно быть не менее 10.
- Блоки (хотя бы частично) должны использовать инструменты платформы по машинному обучению и визуализации.
- Каждый блок процесса должен быть визуализирован средствами платформы, включая динамическую визуализацию.
- Допускается интеграция новых пользовательских блоков в состав платформы.
- Внешние данные могут собираться через WEB-сервисы, сокет или другие коммуникационные инструменты.
- Допустимо использовать предыдущие наработки и программные коды Исполнителя для интеграции в состав проекта.
- Допускается обфускация части данных в случае необходимости использования конфиденциальной информации.

Требуемые навыки:



Python 3.6+, Git, HTTP

Примечание: для уточнения технических возможностей платформы допустимы прямые контакты с разработчиками. Допускается запрос на реализацию дополнительных функциональностей платформы.

Задача №2

Реализация модуля предобработки данных в рамках фреймворка FEDOT (<https://github.com/nccr-itmo/FEDOT>).

Технические требования:

- Модуль предобработки должен предоставлять возможности преобразования исходного датасета (включая временные ряды) с использованием методов поиска пропусков и выбросов, процедур стандартизации и нормализации переменных, выбора наиболее важных предикторов (в случае использования существующих библиотек необходимо обосновать их применения и внести в документацию описание лежащих в их основе принципов).
- Модуль предобработки данных должен содержать инструменты для преобразования исходного пространства переменных, на основе методов понижения размерности и ортогонализации переменных, спектральных и вейвлет преобразований, процедур вычисления агрегирующих параметров, методов автоматического извлечения признаков (в т.ч. нейросетевых, например – с помощью автоэнкодеров).
- Модуль предобработки данных должен позволять интегрировать аналогичные open-source решения (блоки обработки данных из TPOT, tsfresh и т.п.)
- Для обоснования выбора перечня подходов, наиболее подходящих для задачи предобработки данных в рамках фреймворка, должен быть проведен аналитический обзор и сравнительный анализ не менее 10 перспективных подходов.
- Для выполненной реализации должна быть реализован набор примеров и тестов, а также подготовлена документация для <https://itmo-nss-team.github.io/FEDOT.Docs/>

Задача №3

Реализация модуля настройки гипер-параметров сложных моделей в рамках фреймворка FEDOT (<https://github.com/nccr-itmo/FEDOT>).

Технические требования:

- Модуль настройки гипер-параметров должен включать алгоритмы байесовской оптимизации, эволюционные алгоритмы, роевые алгоритмы и др. При настройке моделей должна учитываться предыстория настройки аналогичных или близких моделей из предыдущих запусков. Должна осуществляться настройка как моделей машинного обучения, так и статистических и нейросетевых моделей. Для ускорения процесса настройки должны применяться вспомогательные (в т.ч. суррогатные) модели.
- Должна быть реализовано возможность автоматической оценки значимости параметров в ходе настройки.
- Для обоснования выбора перечня алгоритмов, наиболее подходящих для задачи настройки гипер-параметров атомарных моделей, должен быть проведен аналитический



обзор и сравнительный анализ не менее 15 перспективных алгоритмов и подходов.

— Модуль настройки гипер-параметров должен позволять интегрировать существующие open-source решения для решения данной задачи (hyperopt-sklearn, keras-tuner etc)

— Для выполненной реализации должна быть реализован набор примеров и тестов, а также подготовлена документация для <https://itmo-nss-team.github.io/FEDOT.Docs/>

— Должна быть продемонстрирована эффективность и устойчивость выполненной реализации на наборах тестовых задач (общего назначения или специализированных) для классификации, регрессии и прогнозирования временных рядов.

Задача №4

Реализация модуля для упаковки и отчуждения готовой модели, созданной средствами фреймворка FEDOT (<https://github.com/nccr-itmo/FEDOT>), с кодогенератором, выполняющим автоматическое создание скриптов на Python, для построения композитных моделей, идентичных заданной цепочке, но реализованных без использования методов и объектов ядра фреймворка (т.е. обеспечение отчуждаемости композитных моделей).

Технические требования:

— Экспорт композитной модели должен осуществляться в виде скрипта на python, а также в виде JSON-описания, включающего описание гиперпараметров, структуры.

— Должен выполняться экспорт файлов обученных моделей (входящих в состав композитной модели) и реализована возможность их подключения к результатам экспорта из п. 1.

— Реализованный модуль должен реализовывать наглядную визуализацию структуры и параметров экспортированной модели.

Задача №5

Программа для визуализации и отладки процесса обучения (средствами фреймворка FEDOT <https://github.com/nccr-itmo/FEDOT>) структуры композитной (мета-) модели и ее результатов.

Технические требования:

— Программа для визуализации и отладки композитных моделей должна быть снабжена интерактивным графическим веб-интерфейсом.

— Программа для визуализации и отладки должна содержать редактор композитной модели с обратной связью для online-коррекции процесса сборки и обучения структуры модели через GUI (динамическое изменение структуры модели, её гиперпараметров, изменения ограничений, накладываемых на модель). Также должна осуществляться визуализация хода работы методов поиска оптимальной структуры модели.

— Для выполненной реализации должна быть реализован набор примеров и тестов, а также подготовлена документация для <https://itmo-nss-team.github.io/FEDOT.Docs/>



Задача №6

Тема на выбор участников. Предложите, как можно расширить текущую функциональность FEDOT, повысив его эффективность, производительность и т.д. Заявка будет оцениваться по степени масштабности, полезности и реализуемости предложенных изменений.

Общие требования ко всем задачам:

- Выполненная реализация не должна дублировать существующую функциональность фреймворка FEDOT.
- Реализация должна соответствовать существующим архитектурным решениям. В случае необходимости их изменения, предложенные модификации должны быть согласованы с разработчиками фреймворка.
- Выполненная реализация должна соответствовать принятым практикам разработки на Python (<https://itmo-nss-team.github.io/FEDOT.Docs/contribution/guidelines>), быть понятной, документированной, покрытой автоматическими тестами и проиллюстрированной примерами.
- Использование сторонних библиотек допустимо (если это допускает лицензионное соглашение библиотеки, которое должно быть совместимо с лицензией фреймворка FEDOT – BSD3 <https://github.com/nccr-itmo/FEDOT/blob/master/LICENSE.md>), однако применение каждой из них должно быть обосновано. Не должно возникать дублирования со схожими по функциональности библиотеками, подключенными ранее.



IT RECRUITER

Информационная карта №4

Комплекс решений для умного города

Задача Участников — предложить и в дальнейшем реализовать новые методы сбора и подготовки городских данных для обогащения данных в платформе городского акселератора и решения новых актуальных городских задач. Предпочтение должно отдаваться методам, позволяющим получить оперативные данные и основанным на применении машинного обучения и искусственного интеллекта.

Общей целью перечисленных задач является получение возможности городскому акселератору обрабатывать уникальные (отсутствующие у других) городские данные.

Задача №1

Выявление поведения горожан в городе на основе публикаций в социальных сетях.

Описание задачи:

Поведение описывается конкретными действиями, совершаемыми горожанами в городе. Например, отдых, выезд на шашлыки, распитие спиртных напитков и прочее. Необходимо предложить и реализовать метод, позволяющий выявить такого рода активности из публикаций пользователей социальных сетей. Участникам необходимо самостоятельно выбрать социальную сеть (одну или несколько), выбрать сервис или источник данных, для фрагмента данных необходимо предложить алгоритм идентификации поведения горожан применительно к конкретным местам в городе.

Разработанный алгоритм необходимо оценить на предмет точности.

Задача №2

Определение точек притяжения людей в городе на основе публикаций в социальных сетях и определение тематики активностей в этих точках.

Описание задачи:

Точки притяжения – это места в городе, которые притягивают горожан, и туристов чем-то интересным. Тематика точек притяжения может быть разная – туристический центр, культурный объект, место сбора неформальных сообществ. Участникам необходимо самостоятельно выбрать социальную сеть (одну или несколько), выбрать сервис или источник данных, для фрагмента данных необходимо предложить алгоритм выявления таких точек притяжения.

Разработанный алгоритм необходимо оценить на предмет точности.



Задача №3

Определение точек негативного отношения людей в городе на основе публикаций в социальных сетях и определение причины негативного отношения.

Описание задачи:

Точки негативного отношения людей – это места в городе, которые вызывают споры и недовольство горожан происходящими там событиями. Это могут быть заброшенные корпуса заводов или места сбора людей, которые мешают окружающим жителям. Участникам необходимо самостоятельно выбрать социальную сеть (одну или несколько), выбрать сервис или источник данных, для фрагмента данных необходимо предложить алгоритм выявления таких точек притяжения.

Разработанный алгоритм необходимо оценить на предмет точности.

Задача №4

Разработка инструмента мониторинга экологической ситуации в городе на основе разбора информации от станций измерения уровня загрязнения воздуха и воды.

Описание задачи:

Мониторинг экологической ситуации в городе производится на основе замеров показателей состояния воздуха, воды, радиации на стационарных станциях измерения. Результаты замеров публикуются в открытом доступе раз в сутки в формате MS Word (<http://www.infoeco.ru/index.php?id=8880>). Участникам необходимо разработать метод автоматического извлечения и распознавания данных автоматических замеров с последующей обработкой. Обработка должна включать приведение результатов замеров к универсальной шкале оценки экологической ситуации в районе замеров.

Задача №5

Определение зон криминальной активности на основе анализа публикаций пресслужбы МВД.

Описание задачи:

Сведения о преступности в Санкт-Петербурге публикуются на нескольких профильных порталах города. Среди них официальный сайт МВД РФ (<https://mvdrus.ru/region/91.html>), группа в ВК (<https://vk.com/public151203927>), группа ДТП и ЧП (https://vk.com/spb_today) и ряд других. Участникам необходимо разработать метод автоматического извлечения и распознавания данных о преступности и определения из этих данных (а) места и (б) типа происшествия. Результат должен быть представлен в виде гео-слоя.



Задача №6

Формирование полного набора данных по жилым домам на основе нескольких источников данных.

Описание задачи:

Сведения о жилых домах Санкт-Петербурга размещены на трех ресурсах: OSM, портал открытых данных СПб и портал жилищной инспекции. На всех этих ресурсах данные не являются полными в том смысле, что перечень домов и их атрибуты отличаются. Необходимо разработать метод, позволяющий автоматизировано объединять указанные наборы данных и формировать на их основе гео-слой. Дополнительно необходимо предложить порядок хранения адресов жилых домов.

Задача №7

Разработка метода оптимального планирования развития системы освещения города.

Описание задачи:

Система освещения города развивается на основе адресных программ, формируемых на несколько лет и включающих адреса улиц, парков, зданий и кварталов, подлежащих освещению или ремонту. Необходимо предложить алгоритм формирования оптимального плана освещения для Санкт-Петербурга при ограниченном финансировании. Участникам представляется информация о состоянии системы освещения и методах оценки стоимости объектов освещения.

Задача №8

Оценка озелененности города на основе данных изображений улиц.

Описание задачи:

Реальная озелененность улиц отличается от документированной по той причине, что не все зеленые зоны отражены в документах типа ЗНОП. Задача состоит в том, чтобы разработать метод оценки реальной озелененности улиц на основе анализа панорам изображений.

Задача №9

Оценка озелененности города на основе данных изображений со спутника.

Описание задачи:

Реальная озелененность улиц отличается от документированной по той причине, что не все зеленые зоны отражены в документах типа ЗНОП. Задача состоит в том, чтобы разработать метод оценки реальной озелененности территорий города на основе анализа изображений со спутника.



Задача №10

Формирование сети кварталов на основе данных изображений со спутника.

Описание задачи:

Кварталом в общем смысле является территория города, огороженная дорогами или другими препятствиями, например водой. При наличии данных о сети улиц, выделение кварталов может быть произведено геометрически. При отсутствии таких данных кварталы могут выделяться на основе распознавания изображений со спутника. Участникам необходимо предложить алгоритм формирования гео-слоя кварталов как при наличии так и при отсутствии информации о сети улиц.

Задача №11

Определение типологии застройки городских кварталов на основе нескольких источников данных.

Описание задачи:

Типология застройки определяется наличием и способом размещения зданий на территории кварталов (города). Выделяют жилую застройку, свободную жилую застройку микрорайонного типа, территории озеленения, промышленные территории, водные объекты, историческую застройку и прочие. Участникам необходимо предложить алгоритм формирования гео-слоя кварталов с определением типа застройки квартала на основе открытых данных.