**Проектное предложение**

|  |  |
| --- | --- |
| Тип элемента практической подготовки | Проект |
| Если проект, тип проекта | Исследовательский |
| Наименование проекта | Применение методов обработки естественного языка в социальных науках |
| Подразделение инициатор проекта | Департамент социологии |
| Руководитель проекта | Александрова Марина Юрьевна |
| Основное место работы руководителя проекта в НИУ ВШЭ | Международная лаборатория исследований социальной интеграции |
| Контакты руководителя (адрес эл. почты) | myaleksandrova@hse.ru |
| Соруководители проекта от НИУ ВШЭ (если имеются) | - |
| Контакты соруководителей от НИУ ВШЭ (адрес эл. почты) | - |
| Основная проектная идея / описание решаемой проблемы | Проект предполагает анализ взаимосвязи особенностей формулировок вопросов в анкетах социологических исследований и пропусков в опросных данных. Такое исследование позволит выявить закономерности в поведении респондентов во время участия в массовых опросах. |
| Цель и задачи проекта | Цель проекта – разработка модели предсказания возникновения пропусков в опросных данных в зависимости от формулировок, используемых в жесткой социологической анкете, а также накопление данных для развития методов текст-майнинга для целей социальных наук. |
| Проектное задание | Каждому из участников проектной команды предлагается корпус социологических анкет (одна анкета ≈70-100 страниц машиночитаемого файла формата .pdf).  На первом этапе, после вводного инструктажа, на котором объясняются основные цели и методика исследования, участники структурируют и аннотируют материал по определенным правилам и визуализируют данные, подготовленные по итогам работы (в виде диаграмм и графиков), структурированные текстовые данные загружают в соответствующий файл, размещенный на гугл-диске.  На втором этапе исследования, после инструктажа, предполагается предварительный анализ полученных структурированных текстовых данных в команде(участники познакомятся с методами предварительной обработки текстовых данных и первичного их анализа, а также подготовят краткий отчет по проведенной работе по заранее подготовленному шаблону). |
| Планируемые результаты проекта, специальные или функциональные требования к результату | База данных (размеченные социологические анкеты в электронном формате) и групповой отчет о проделанной работе по заранее имеющемуся шаблону. |
| Дата начала проекта | 15.12.2021 |
| Дата окончания проекта | 04.02.2022 |
| Трудоемкость (часы в неделю) на одного участника | 10 |
| Предполагаемое количество участников (вакантных мест) в проектной команде | 5 |
| Названия вакансий (ролей), краткое описание задач по каждой вакансии, количество кредитов и критерии отбора для участников проекта (если характер работ для всех участников совпадает, описывается одна вакансия)  Кредиты на 1 участника рассчитываются по формуле: продолжительность в неделях \* трудоемкость проекта в часах / 25 | Вакансия: Стажер/аналитик  Задачи: подготовка текстовых данных к анализу, предварительный анализ текстовых данных (навыки обработки естественного языка не требуются, но приветствуются).  Количество кредитов: 3  Критерии отбора на вакансию: Понимание структуры анкет, используемых для массовых опросов, готовность работать с данными, наличие базовых навыков работы с данными, внимательность, ответственность. Навыки работы в Python не являются обязательными, но приветствуются. |
| Общее количество кредитов | 15 |
| Форма итогового контроля | Экзамен |
| Формат представления результатов, который подлежит оцениванию | База данных (размеченные социологические анкеты в электронном формате) и групповой отчет о проделанной работе по заранее имеющемуся шаблону. |
| Формула оценки результатов, возможные критерии оценивания результатов с указанием всех требований и параметров | Оценивается чистота разметки текста социологических анкет на русском и на английском языках, навыки работы с данными. |
| Возможность пересдач при получении неудовлетворительной оценки | Да |
| Ожидаемые образовательные результаты проекта | Участники проекта приобретут навыки компьютерной обработки текстовых данных и интерпретации полученных результатов, внесут свой вклад в изучение возможностей применения методов текст-майнинга для социальных наук. |
| Особенности реализации проекта: территория, время, информационные ресурсы и т.п. | Мск, СПб, удаленная работа  Гибкий график |
| Рекомендуемые образовательные программы | Бакалавриат Социология (Мск), Магистратура: Комплексный социальный анализ (Мск), Прикладные методы социального анализа рынков (Мск), Современный социальный анализ (СПб), Социология публичной и деловой сферы (Мск), Сравнительные социальные исследования (Мск) |
| Требуется резюме студента | Да |
| Требуется мотивированное письмо студента | Нет |