**Научно-исследовательский семинар**

**«Аналитика больших данных»**

**Аннотация**

|  |
| --- |
| Основные задачи освоения дисциплины:   * сформировать у студента представление об основных методах хранения и анализа данных, в том числе больших данных; * сформировать у студента представление о типичных задачах HR-аналитики и подходах к их решению; * сформировать у студента умение формулировать гипотезы исследования, правильно выбирать метод в соответствии с целями, задачами исследования; * выработать навыки практического применения некоторых методов сбора и анализа данных.   В рамках курса проводится обучение основным методам анализа данных, техникам визуализации данных. |

**Цель освоения дисциплины**

|  |  |
| --- | --- |
| **1.** | Целью освоения студентами дисциплины является получение навыков всестороннего исследования всех доступных данных и поиска скрытых факторов в рамках конкретной решаемой аналитической задачи, формирование ясного и доступного формата изложения полученных выводов. |

**Планируемые результаты обучения**

|  |  |
| --- | --- |
| **–** | знание типичных задач анализа данных для нужд HR, методов хранения и обработки данных, понимание уровней зрелости аналитической функции в организации; |
| **–** | знание баз данных, их возможностей и места в современных процессах и ИТ-системах, представление о работе HR-систем, понимание природы транзакционных и аналитических данных в HR; |
| **–** | знание инструментов анализа данных и систем контроля версий: Python, GIT; |
| **–** | знание синтаксиса SQL, агрегационных функций, журналирование; |
| **–** | навыки выгрузки результатов запросов и их визуальной интерпретации; |
| **–** | навыки работы со статистическими критериями; |
| **–** | знание возможностей применения статистических тестов для решения практических задач. |

**Содержание учебной дисциплины**

|  |  |
| --- | --- |
| **1.** | **Тема 1. HR-аналитика в современной организации**  Значение функции HR-аналитики в современной организации. Уровни зрелости аналитической функции в организации. Инструменты и процессы, характерные для разных уровней зрелости. HR-системы. |
| **2.** | **Тема 2. Cтатистические гипотезы и статистический вывод для задач HR-аналитики**  Статистическая гипотеза и статистический вывод. Анализ среднего и доверительный интервал. Проверка распределений данных при помощи тестов. Построение выводов на основе данных. |
| **3.** | **Тема 3. Продвинутая работа с базами данных**  Реляционное представление данных. Нормализация данных и нормальные формы данных. Транзакционные и аналитические данные в HR. Ключевые сложности и типичные задачи HR-аналитики. Язык запросов SQL, агрегационные функции, журналирование. |
| **4.** | **Тема 4. Перспективные инструменты анализа и визуализации данных**  Задачи анализа данных в функции HR-аналитики в организации. Перспективные инструменты работы с данными. Визуализация данных в Jupyter Notebook и в BI-системах. Системы контроля версий на примере Git. Инструменты автоматической поставки данных. Метрики в HR. |
| **5.** | **Тема 5. Подготовка к итоговому проекту**  Типичные проекты аналитики для нужд HR. Распределение задач, работа над проектами. Примеры отчетов. Защита, подведение итогов. |

**Технологии реализации дисциплины**

|  |  |
| --- | --- |
| **Охват аудитории** | Московский кампус, ОП «HR-аналитика» |
| **Лекции** | 0 |
| **Практические занятия** | 32 |

**Текущий контроль**

|  |  |
| --- | --- |
| **–** | **Выполнение контрольных работ по итогам изучения каждой темы**  Критерии оценивания  Оценка за контрольную работу будет складываться на основе трех критериев:  (1) верно сформулированного подхода к решению задачи,  (2) верного и своевременного исполнения задачи,  (3) понятного визуального представления результатов задачи.  Оценка за контрольную работу рассчитывается как отношение верно решенных заданий к их общему числу. Это отношение нормируется на 10-ти бальную шкалу. Общая оценка по текущему контролю рассчитывается как среднее арифметическое оценок всех контрольных работ.  Комментарий  В случае болезни или отсутствия студента на контрольной работе по уважительной причине, студенту нужно предупредить преподавателя, проводящего контрольную работу, до срока её сдачи. Повторное проведение контрольной работы для студентов, пропустивших ее по уважительной причине, не производится, но вес контрольной работы переносится на решение итоговой контрольной работы с включением дополнительного задания в итоговую контрольную работу. В случае пропуска контрольной работы по неуважительной причине и/или отсутствия соответствующих документов повторное проведение контрольной работы не производится и в качестве оценки выставляется 0 баллов. |

**Итоговый контроль**

|  |  |
| --- | --- |
| **–** | **Итоговая контрольная работа (экзамен)**  Критерии оценивания  Оценка за итоговую контрольную работу будет складываться на основе трех критериев:  (1) верно сформулированного подхода к решению задачи,  (2) верного и своевременного исполнения задачи,  (3) понятного визуального представления результатов работы.  Оценка за контрольную работу рассчитывается как отношение верно решенных заданий к их общему числу. Это отношение нормируется на 10-ти бальную шкалу. |
| **–** | **Выполнение индивидуального проекта по анализу HR-данных по индивидуальной теме**  Примеры тем:   * Модель оттока сотрудников для задач сокращение издержек организации * Оценка причин увольнения сотрудников * Модель оценки абсентеизма сотрудников   Требования к формату и объему работы: работа сдается в электронном виде в объеме 5-10 страниц А4.  **Внимание!** Требования к срокам предоставления индивидуального проекта: не позднее чем за 7 дней до даты итоговой аттестации (экзамена).  Критерии оценивания  Оценка за индивидуальный проект будет складываться на основе трех критериев:  (1) верно сформулированного подхода к решению поставленной в проекте цели,  (2) охвата сопутствующих (не явно выраженных) данных, поиска и сравнения похожих кейсов из открытых источников,  (3) достижения поставленной в работе задачи, получения полных данных, адекватности выбранного метода для выполнения итоговой работы. |

**Итоговая аттестация**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***%*** | **Итоговая аттестация (2 модуль)**   |  |  | | --- | --- | | **0.200** | выполнение индивидуального проекта по анализу HR-данных | | **0.300** | контрольные работы после изучения каждой темы | | **0.500** | итоговая контрольная работа (экзамен) |   **Правила округления**  Оценки за контрольные и экзамен выставляются по 10-ти балльной шкале с округлением до ближайшего целого числа арифметическим способом (например, 3,4 округляется до 3, а 3,5 до 4). При расчете общей оценки за дисциплину по формуле с коэффициентами ее элементы не округляются. |

**Литература**

|  |  |
| --- | --- |
| **Рекомендуемая основная литература** | |
|  | Boslaugh S. (2012) Statistics in a Nutshell. 2nd edition. O’Reilly Media. – *Электронные ресурсы библиотеки НИУ ВШЭ*: <http://web.a.ebscohost.com/pfi/results?vid=2&sid=d42b6951-9e77-4040-9d95-2ffba52c6cb2%40sessionmgr4008&bquery=Statistics+in+a+Nutshell%2c+2nd+Edition&bdata=JmRiPWVkc3B1YiZ0eXBlPTQ0JnNlYXJjaE1vZGU9QW5kJnNpdGU9cGZpLWxpdmU%3d>  Géron A. (2019) Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow : Concepts, Tools, and Techniques to Build Intelligent Systems. Second edition. O’Reilly Media. |
|  | Glantz S.A. (2012) Primer of Biostatistics, Seventh Edition. McGraw-Hill Education / Medical. |
|  | Géron A. (2017) Hands-On Machine Learning with Scikit-Learn and TensorFlow : Concepts, Tools, and Techniques to Build Intelligent Systems (Vol. First edition). Sebastopol, CA: O’Reilly Media. – http://search.ebscohost.com/login.aspx?direct=true&site=eds-live&db=nlebk&AN=1486117 |
| **Рекомендуемая дополнительная литература** | |
|  | Бослаф С. (2017) Статистика для Всех. – М.: ДМК-Пресс.  Goodfellow I., Bengio Y., Courville A. (2016) Deep Learning. The MIT Press. |
|  | Lutz M. (2009) Learning Python: Powerful Object-Oriented Programming. 4th ed. O’Reilly Media. |

**Программное обеспечение**

|  |  |
| --- | --- |
| **–** | Microsoft Office Professional Plus 2010 / Из внутренней сети Университета |
| **–** | Microsoft Windows 7 Professional RUS / Из внутренней сети Университета |
| **–** | Anaconda Jupyter |
| **–** | PostgreSQL administration tool (pgAdmin) |

**Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (Электронные образовательные ресурсы)**

|  |  |
| --- | --- |
| **–** | Consultant Plus  Legal reference system, containing a complete database of legal information and brand-updated explanations |
| **–** | Garant system  Reference legal system on the legislation of the Russian Federation |