Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение

высшего образования «Новосибирский национальный исследовательский

государственный университет» (Новосибирский государственный университет, НГУ)

**Факультет информационных технологий**

СОГЛАСОВАНО

Декан ФИТ НГУ

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ М.М. Лаврентьев

«03» июля 2019 г.

РАБОЧАЯ ПРОГРАММА Дисциплины

|  |
| --- |
| **Аналитика больших массивов данных** |

Направление подготовки: 09.04.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА

Направленность (профиль): Технология разработки программных систем

Форма обучения: очная

Год обучения: 1, семестр: 1

|  |  |  |
| --- | --- | --- |
| **№** | **Вид деятельности** | **Семестр** |
| **1** |
| **1** | Лекции, час. | 16 |
| **2** | Практические занятия, час. | 16 |
| **3** | Лабораторные занятия, час. |  |
| **4** | Занятий в контактной форме без учета промежуточной аттестации, час, из них | 34 |
| **5** | в электронной форме, час. |  |
| **6** | из них аудиторных занятий, час. | 32 |
| **7** | из них в активной и интерактивной форме, час. | 24 |
| **8** | консультаций, час. | 2 |
| **9** | Самостоятельная работа, час. | 72 |
| **10** | в том числе на выполнение письменных работ, час |  |
| **11** | Форма аттестации (экзамен), час | Э 2 |
| **12** | Всего зачетных единиц[[1]](#footnote-1) | 3 |

Новосибирск 2019

Рабочая программа дисциплины составлена на основании федерального государственного образовательного стандарта (ФГОС) высшего образования - магистратура по направлению подготовки 09.04.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА.

Федеральный государственный образовательный стандарт (ФГОС) высшего образования - магистратура по направлению подготовки 09.04.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА введен в действие приказом Минобрнауки от 19.09.2017 № 918.

Место дисциплины в структуре учебного плана: Блок 1 Дисциплины (модули), обязательная часть, обязательная дисциплина.

Рабочая программа дисциплины утверждена решением Ученого совета факультета информационных технологий от 02.07.2019, протокол № 75.

Программу разработали:

Доцент кафедры вычислительных систем ММФ,

кандидат физико-математических наук Е.Н. Павловский

Заведующий кафедрой общей информатики ФИТ,

доктор физико-математических наук Д.Е. Пальчунов

Ответственный за образовательную программу:

заведующий кафедрой систем информатики ФИТ,

доктор физико-математических наук М.М. Лаврентьев

Согласовано:

Директор ООО «Диалоговые системы» Д.А. Луппов

**Аннотация к рабочей программе дисциплины**

**«Аналитика больших массивов данных»**

Дисциплина «Аналитика больших массивов данных» реализуется в рамках образовательной программы высшего образования – программы магистратуры 09.04.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА, направленность (профиль): Технология разработки программных систем по очной форме обучения на русском языке.

**Место в образовательной программе:**

Дисциплина «Аналитика больших массивов данных» является базовой для прохождения производственной практики и написания выпускной квалификационной работы.

Дисциплина «Аналитика больших массивов данных» реализуется в 1 семестре в рамках базовой части дисциплин (модулей) Блока 1 и является обязательной дисциплиной.

Дисциплина «Аналитика больших массивов данных» направлена на формирование компетенций:

**Способен применять на практике новые научные принципы и методы исследований (ОПК-4) , в части следующих индикаторов достижения компетенции:**

ОПК-4.1. Знать: новые научные принципы и методы исследований

ОПК-4.2. Умеет: применять на практике новые научные принципы и методы исследований

ОПК-4.3. Владеть: навыками применения новых научных принципов и методов исследования для решения профессиональных задач.

**Перечень основных разделов дисциплины:** Введение в большие данные. Жизненный цикл анализа больших данных. Регрессия. Методы и задачи классификации. Метрики качества для задач регрессии и классификации. Язык Python для анализа данных. Измерительные шкалы. Нейронные сети. Проблема переобучения. Регуляризация.

При освоении дисциплины студенты выполняют следующие виды учебной работы: лекции, практические занятия, консультации, самостоятельная работа. В учебном процессе предусматривается использование интерактивных форм проведения занятий*.* В том числе, предполагаются демонстрации работы известных алгоритмов.

Самостоятельная работа включает: подготовку к практическим занятиям по разделам дисциплины, выполнение заданий и подготовку к экзамену.

Общий объем дисциплины – 3 зачетных единицы (108 часов).

**Правила аттестации по дисциплине.** Текущий контроль по дисциплине «Аналитика больших массивов данных» осуществляется пятиминутными тестами вначале каждого третьего занятия на знание предыдущего материала. Также выдаются практические задания, которые входят в структуру портфолио, по результатам которых выставляется оценка «зачтено» или «не зачтено». Оценка «зачтено» по результатам выполнения задания является одним из условий успешного прохождения промежуточной аттестации.

Промежуточная аттестация по дисциплине «Аналитика больших массивов данных» проводится по завершению семестра в форме экзамена. Экзамен проводится в три этапа: (1) тест на знание теоретического материала, (2) тест на знание алгоритмов, (3) беседа по открытым темам. Результаты промежуточной аттестации по дисциплине оцениваются по шкале «неудовлетворительно», «удовлетворительно», «хорошо», «отлично». Оценки «отлично», «хорошо», «удовлетворительно», означают успешное прохождение промежуточной аттестации.

Оценка «отлично» соответствует продвинутому уровню сформированности компетенции.

Оценка «хорошо» соответствует базовому уровню сформированности компетенции.

Оценка «удовлетворительно» соответствует пороговому уровню сформированности компетенции.

**Учебно-методическое обеспечение дисциплины.**

1. Аналитика больших массивов данных [Электронный ресурс] : электронный учебно-методический комплекс / Е. Н. Павловский, Ю.А. Аникин, Н.Г. Загоруйко, И.А. Борисова, А.О. Зырянов; НОУ ИНТУИТ. - Москва, [2014]. - Режим доступа: <http://www.intuit.ru/studies/courses/12385/1181/info>. - Загл. с экрана.
2. Павловский Е. Н. Аналитика больших массивов данных [Электронный ресурс] : электронный учебно-методический комплекс / Е. Н. Павловский ; Новосиб. гос. ун-т. - Новосибирск, [2018]. - Режим доступа: https://bigdata.nsu.ru/ABMD (доступ только для домена g.nsu.ru). - Загл. с экрана.
3. **Внешние требования к дисциплине**

Таблица 1.1

|  |
| --- |
| **Компетенция ОПК-4 Способен применять на практике новые научные принципы и методы исследований, *в части следующих индикаторов достижения компетенции*:** |
| **ОПК-4.1** Знать: новые научные принципы и методы исследований |
| **ОПК-4.2** Умеет: применять на практике новые научные принципы и методы исследований |
| **ОПК-4.3** Владеть: навыками применения новых научных принципов и методов исследования для решения профессиональных задач |

1. **Требования к результатам освоения дисциплины**

Таблица 2.1

|  |  |  |  |
| --- | --- | --- | --- |
| **Результаты изучения дисциплины по уровням освоения (иметь представление, знать, уметь, владеть)** | **Формы организации занятий** | | |
| **Лекции** | **Практики** | **Самостоятельная работа** |
| **ОПК-4.1** Знать: новые научные принципы и методы исследований | | | |
| 1. Знать способы и методы применения алгоритмов анализа данных для решения научных и прикладных задач. | + | + | + |
| 1. Знать основные алгоритмы для решения задач классификации и регрессии. | + | + | + |
| **ОПК-4.2** Умеет: применять на практике новые научные принципы и методы исследований | | | |
| 1. Уметь формулировать бизнес-задачи в терминах анализа данных. |  | + | + |
| 1. Уметь решать задачи классификации и регрессии. |  | + | + |
| **ОПК-4.3** Владеть: навыками применения новых научных принципов и методов исследования для решения профессиональных задач | | | |
| 1. Владеть методами интеллектуального анализа данных, в т.ч. методами оценки качества моделей, алгоритмов, методами экспериментальной проверки гипотез, методами обоснования гипотез. |  | + | + |

**3. Содержание и структура учебной дисциплины**

Программа лекционных занятий. Таблица 3.1

|  |  |  |  |
| --- | --- | --- | --- |
| **Темы лекций** | **Активные формы, час. (входит в общее кол-во часов)** | **Часы** | **Ссылки на результаты обучения** |
| **Семестр: 1** | | | |
| Тема 1. Введение в большие данные | 0 | 2 | 1 |
| Тема 2. Жизненный цикл анализа больших данных | 0 | 2 | 1 |
| Тема 3. Регрессия. | 2 | 2 | 2 |
| Тема 4. Методы и задачи классификации | 2 | 2 | 2 |
| Тема 5. Измерительные шкалы | 0 | 2 | 1,2 |
| Тема 6. Системы хранения больших данных | 0 | 1 | 1,2 |
| Тема 7. Нейронные сети | 2 | 3 | 1,2 |
| Тема 8. Проблема переобучения. Регуляризация | 2 | 2 | 1,2 |
| **Итого:** | **8** | **16** |  |

Программа практических занятий. Таблица 3.2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Темы практических занятий** | **Активные формы, час.**  **(входит в общее кол-во часов)** | | **Часы** | **Ссылки на результаты обучения** | | **Учебная деятельность** |
| **Семестр: 1** | | | | | | |
| Тема 1. Оценка метрик. | 2 | 2 | | | 1,2,3 | Обучающиеся оценивают необходимые требования к качеству алгоритма исходя из бизнес-задачи. |
| Тема 2. Регрессия. | 4 | 4 | | | 1,4,5 | Обучающиеся решают аналитически и с помощью градиентного спуска задачу линейной регрессии. |
| Тема 3. Методы и задачи классификации | 4 | 4 | | | 1,4,5 | Обучающиеся разбирают алгоритмы классификации с примерами на языке Python. |
| Тема 4. Нейронные сети | 4 | 4 | | | 1,4,5 | Обучающиеся пишут программу на языке Python с реализацией нейронной сети. |
| Тема 5. Проблема переобучения. Регуляризация | 2 | 2 | | | 1,4,5 | Обучающиеся анализируют причины переобучения и предлагают подходы к регуляризации. |
| **Итого:** | **16** | **16** | | |  |  |

1. **Самостоятельная работа студентов**

Таблица 4.1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№** | **Виды самостоятельной работы** | **Ссылки на результаты обучения** | **Часы на выполнение** | **Часы на консультации** | | |
| **Семестр: 1** | | | | | | |
| 1 | Выполнение домашнего задания в рамках портфолио. | 1,2,3,4,5 | 48 | | |  |
| Обучающиеся решают практические задачи, входящие в портфолио. Методические рекомендации по выполнению домашнего задания представлены в приложении к рабочей программе дисциплины | | | | | |
| 2 | Подготовка к экзамену | 1,2,3,4,5 | 24 | | | 2 |
| Подготовка к экзамену по вопросам, представленным в фонде оценочных средств, являющихся приложением к рабочей программе дисциплины. | | | | | |
|  | **Итого:** |  | **72** | | **2** | |

1. **Образовательные технологии**

В ходе реализации учебного процесса по дисциплине проводятся лекционные и семинарские занятия. Темы, рассматриваемые на лекциях и изучаемые самостоятельно, закрепляются на практических занятиях. А также применяются следующие интерактивные формы обучения (таблица 5.1).

Таблица 5.1

|  |  |  |
| --- | --- | --- |
| **1** | Портфолио | ОПК-4.1, ОПК-4.2, ОПК-4.3 |
| **Формируемые умения:** Уметь формулировать научные задачи исходя из технологических вызовов. Уметь формулировать бизнес-задачи в терминах анализа данных. Знать о критериях проверки аналитических гипотез. Уметь решать задачи классификации и регрессии. Владеть методами интеллектуального анализа данных, в т.ч. методами оценки качества моделей, алгоритмов, методами экспериментальной проверки гипотез, методами обоснования гипотез. | | |
| **Краткое описание применения:** студенты ведут портфолио (коллекцию работ), которое является основой для проведения аттестации по дисциплине. | | |

Для организации и контроля самостоятельной работы студентов, а также проведения консультаций применяются информационно-коммуникационные технологии (таблица 5.2).

Таблица 5.2

|  |  |
| --- | --- |
| Информирование | d.luppov@dsys.ai |
| Консультирование | d.luppov@dsys.ai |
| Контроль | d.luppov@dsys.ai |
| Размещение учебных материалов | <https://drive.google.com/drive/folders/1rtqiTivB7ED-X0Yeo9_4asOWvkEyDsOc> |

**6. Правила аттестации студентов по учебной дисциплине**

По дисциплине «Интеллектуальные системы» проводится текущая и промежуточная аттестация (итоговая по дисциплине).

**Текущая аттестация** по дисциплине «Интеллектуальные системы» осуществляется на практических занятиях и заключается в выполнение домашнего задания в рамках портфолио. По результатам текущей аттестации выставляется оценка «зачтено» или «не зачтено». Оценка «зачтено» по результатам защиты докладов является одним из условий успешного прохождения промежуточной аттестации.

**Промежуточная аттестация** (итоговая по дисциплине) проводится по завершению семестра в форме экзамена. Экзамен проводится в три этапа: (1) тест на знание теоретического материала, (2) тест на знание алгоритмов, (3) беседа по открытым темам. Результаты промежуточной аттестации по дисциплине оцениваются по шкале «неудовлетворительно», «удовлетворительно», «хорошо», «отлично». Оценки «отлично», «хорошо», «удовлетворительно», означают успешное прохождение промежуточной аттестации.

В таблице 6.1 представлено соответствие форм аттестации заявляемым требованиям к результатам освоения дисциплины.

Таблица 6.1

|  |  |  |  |
| --- | --- | --- | --- |
| **Коды компетенций ФГОС** | **Результаты обучения** | **Формы аттестации** | |
| **Портфолио** | **Экзамен** |
| **ОПК-4** | **ОПК-4.1** Знать: новые научные принципы и методы исследований | **+** | **+** |
| **ОПК-4.2** Умеет: применять на практике новые научные принципы и методы исследований | **+** | **+** |
| **ОПК-4.3**Владеть: навыками применения новых научных принципов и методов исследования для решения профессиональных задач |  | **+** |

Требования к структуре и содержанию портфолио, оценочные средства, а также критерии оценки сформированности компетенций и освоения дисциплины в целом, представлены в Фонде оценочных средств, являющемся приложением 1 к настоящей рабочей программе дисциплины.

**7. Литература**

1. Лбов, Геннадий Сергеевич Анализ данных и знаний : учебное пособие / Г.С. Лбов ; Федер. агентство по образованию, Новосиб. гос. ун-т, Мех.-мат. факНовосибирск : Новосибирский государственный университет, 2010107 с. ; 20 см.Библиогр.: с.107.ISBN 978-5-94356-907-4 (36 экз)

2. Крутиков, В.Н. Анализ данных : учебное пособие / В.Н. Крутиков, В.В. Мешечкин ; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Кемеровский государственный университет». - Кемерово : Кемеровский государственный университет, 2014. - 138 с. : ил. - Библиогр. в кн. - ISBN 978-5-8353-1770-7 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=278426>.

*Интернет-ресурсы*

Таблица 7.1

|  |  |  |
| --- | --- | --- |
| № п/п | Наименование Интернет-ресурса | Краткое описание |
| 1 | Воронцов К.В. Математические методы обучения по прецедентам [Электронный ресурс]. – Режим доступа:  [http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf](%20%20http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf) . – Загл. с экрана | Известный курс Воронцова К.В. (2016). |
| 2 | Trevor Hastie, Elements of statistical learning [Электронный ресурс]. – Режим доступа: <http://statweb.stanford.edu/~tibs/ElemStatLearn/> . – Загл. с экрана | Известная книга Тревора Хасти (2009). |
| 3 | Программирование на Python – Stepik [Электронный ресурс] : электронный учебно-методический комплекс / Институт биоинформатики // Веб-сайт Stepik.org. – Москва, [2018]. – Режим доступа: <https://stepik.org/course/67/>. – Загл. с экрана. | Онлайн курс на платформе Stepik. |
| 4 | Аналитика больших массивов данных [Электронный ресурс] : электронный учебно-методический комплекс / Е. Н. Павловский, Ю.А. Аникин, Н.Г. Загоруйко, И.А. Борисова, А.О. Зырянов; НОУ ИНТУИТ. - Москва, [2014]. - Режим доступа: <http://www.intuit.ru/studies/courses/12385/1181/info>. - Загл. с экрана. | Онлайн-курс на платформе ИНТУИТ (2014). |
| 5 | C. B. B. D. Manyika, Big Data: The Next Frontier for Innovation, Competition, and Productivity [Electronic resource] // McKinsey Global Institute, 2011. – URL:  [http://www.mckinsey.com/~/media/McKinsey/dotcom/ Insights%20and%20pubs/MGI/Research/](http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx)  [Technology%20and%20Innovation/Big%20Data/](http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx)  [MGI\_big\_data\_full\_report.ashx](http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx) | Отчёт МакКинзи о Больших Данных (2011). |
| 6 | Frontiers in Massive Data Analysis [Electornic resource] / National Research Council, 2013. – URL: <http://www.nap.edu/catalog/18374/frontiers-in-massive-data-analysis>. | Отчёт трёх национальных академий наук США о научных задачах в сфере больших данных (2013). |
| 7 | Павловский Е.Н. Научные вызовы Больших Данных // Тезисы докладов конференции «Большие Данные в национальной экономике» (Москва, 21 октября 2014 г.). / [Под ред. Дубовой Н.А.]. — М.: «Открытые системы». — 2014. С.12-14. URL: <https://www.osp.ru/netcat_files/18/10/06_2014_Nau4nye.pdf> | Статья о научных вызовах Больших данных (2014). |

**8. Учебно-методическое и программное обеспечение дисциплины**

**8.1. Учебно-методическое обеспечение**

1. Аналитика больших массивов данных [Электронный ресурс] : электронный учебно-методический комплекс / Е. Н. Павловский, Ю.А. Аникин, Н.Г. Загоруйко, И.А. Борисова, А.О. Зырянов; НОУ ИНТУИТ. - Москва, [2014]. - Режим доступа: <http://www.intuit.ru/studies/courses/12385/1181/info>. - Загл. с экрана.
2. Павловский Е. Н. Аналитика больших массивов данных [Электронный ресурс] : электронный учебно-методический комплекс / Е. Н. Павловский ; Новосиб. гос. ун-т. - Новосибирск, [2018]. - Режим доступа: <https://bigdata.nsu.ru/ABMD/> (доступ только для домена g.nsu.ru). - Загл. с экрана.

**8.2. Программное обеспечение**

Для обеспечения реализации дисциплины используется стандартный комплект программного обеспечения (ПО), включающий регулярно обновляемое лицензионное ПО Windows и MS Office

Специализированное программное обеспечение не требуется.

**9. Профессиональные базы данных и информационные справочные системы**

1. Полнотекстовые журналы Springer Journals за 1997-2015 г., электронные книги (2005-2016 гг.), коллекция научных биомедицинских и биологических протоколов SpringerProtocols, коллекция научных материалов в области физических наук и инжиниринга SpringerMaterials, реферативная БД по чистой и прикладной математике zbMATH.
2. Электронные ресурсы Web of Science Core Collection (Thomson Reuters Scientific LLC.), Journal Citation Reports + ESI
3. БД Scopus (Elsevier)
4. Лицензионные материалы на сайте eLibrary.ru

**10. Материально-техническое обеспечение**

Таблица 10.1

|  |  |  |
| --- | --- | --- |
| **№** | **Наименование** | **Назначение** |
| 1 | Презентационное оборудование (мультимедиа-проектор, экран, компьютер для управления) | Для проведения лекционных и практических занятий |
| 2 | Компьютерный класс (с выходом в Internet) | Для организации самостоятельной работы обучающихся |

Материально-техническое обеспечение образовательного процесса по дисциплине для обучающихся из числа лиц с ограниченными возможностями здоровья осуществляется согласно «Порядку организации и осуществления образовательной деятельности по образовательным программам для инвалидов и лиц с ограниченными возможностями здоровья в Новосибирском государственном университете».

**Лист актуализации рабочей программы дисциплины  
«Аналитика больших массивов данных»**

|  |  |  |  |
| --- | --- | --- | --- |
| № | Характеристика внесенных изменений (с указанием пунктов документа) | Дата и №  протокола Ученого совета ФИТ | Подпись  ответственного |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

1. С учетом выделенных часов на промежуточную аттестацию [↑](#footnote-ref-1)