

МИНОБРНАУКИ РОССИИ
Федеральное государственное автономное образовательное
учреждение высшего образования "Пермский
государственный национальный исследовательский
университет"

Авторы-составители: **Шишкин Владимир Андреевич**
Радионова Марина Владимировна

Рабочая программа дисциплины
MACHINE LEARNING FOR BIG DATA
Код УМК 97817

Утверждено
Протокол №10
от «09» июня 2021 г.

Пермь, 2021

1. Наименование дисциплины

Machine Learning for Big Data

2. Место дисциплины в структуре образовательной программы

Дисциплина входит в обязательную часть Блока « Б.1 » образовательной программы по направлениям подготовки (специальностям):

Направление: **38.03.05** Бизнес-информатика

направленность Информационные системы и большие данные

3. Планируемые результаты обучения по дисциплине

В результате освоения дисциплины **Machine Learning for Big Data** у обучающегося должны быть сформированы следующие компетенции:

38.03.05 Бизнес-информатика (направленность : Информационные системы и большие данные)

ОПК.9 Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений

Индикаторы

ОПК.9.1 Обоснованно использует информацию для информационно-аналитической поддержки принятия управленческих решений

ПК.3 Способен к планированию и организации аналитических работ, в том числе, с использованием технологий больших данных

Индикаторы

ПК.3.1 Разрабатывает, обсуждает и утверждает содержание аналитических работ, в том числе, с использованием технологий больших данных

4. Объем и содержание дисциплины

Направления подготовки	38.03.05 Бизнес-информатика (направленность: Информационные системы и большие данные)
форма обучения	очная
№№ триместров, выделенных для изучения дисциплины	11
Объем дисциплины (з.е.)	3
Объем дисциплины (ак.час.)	108
Контактная работа с преподавателем (ак.час.), в том числе:	42
Проведение лекционных занятий	14
Проведение лабораторных работ, занятий по иностранному языку	28
Самостоятельная работа (ак.час.)	66
Формы текущего контроля	Входное тестирование (1) Итоговое контрольное мероприятие (1) Письменное контрольное мероприятие (2)
Формы промежуточной аттестации	Экзамен (11 триместр)

5. Аннотированное описание содержания разделов и тем дисциплины

Machine Learning for Big Data

Input control

Introduction

The concept of big data. Processing and preparing data. The ecosystem of big data.

1 Regression Analysis

Models of linear and non-linear regression. Estimation of model coefficients. Model testing.

2 Classification

Naïve Bayes and classification trees. Binary choice models. Linear and non-linear discriminant functions.

Test 1

3 Cluster analysis

The concept of closeness. Hierarchical and non-hierarchical clustering. Methods of k-means, fuzzy k-means, Gustafson-Kessel method. Clustering based on the fuzzy alpha-quasy-equivalence relation.

4 Factor analysis

The principal component method. The singular decomposition.

Test 2

5 Association rules

Associative and sequential rules. A priori algorithm.

Final control activity

6. Методические указания для обучающихся по освоению дисциплины

Освоение дисциплины требует систематического изучения всех тем в той последовательности, в какой они указаны в рабочей программе.

Основными видами учебной работы являются аудиторные занятия. Их цель - расширить базовые знания обучающихся по осваиваемой дисциплине и систему теоретических ориентиров для последующего более глубокого освоения программного материала в ходе самостоятельной работы. Обучающемуся важно помнить, что контактная работа с преподавателем эффективно помогает ему овладеть программным материалом благодаря расстановке необходимых акцентов и удержанию внимания интонационными модуляциями голоса, а также подключением аудио-визуального механизма восприятия информации.

Самостоятельная работа преследует следующие цели:

- закрепление и совершенствование теоретических знаний, полученных на лекционных занятиях;
- формирование навыков подготовки текстовой составляющей информации учебного и научного назначения для размещения в различных информационных системах;
- совершенствование навыков поиска научных публикаций и образовательных ресурсов, размещенных в сети Интернет;
- самоконтроль освоения программного материала.

Обучающемуся необходимо помнить, что результаты самостоятельной работы контролируются преподавателем во время проведения мероприятий текущего контроля и учитываются при промежуточной аттестации.

Обучающимся с ОВЗ и инвалидов предоставляется возможность выбора форм проведения мероприятий текущего контроля, альтернативных формам, предусмотренным рабочей программой дисциплины. Предусматривается возможность увеличения в пределах 1 академического часа времени, отводимого на выполнение контрольных мероприятий.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации.

При проведении текущего контроля применяются оценочные средства, обеспечивающие передачу информации, от обучающегося к преподавателю, с учетом психофизиологических особенностей здоровья обучающихся.

7. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

При самостоятельной работе обучающимся следует использовать:

- конспекты лекций;
- литературу из перечня основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля);
- текст лекций на электронных носителях;
- ресурсы информационно-телекоммуникационной сети "Интернет", необходимые для освоения дисциплины;
- лицензионное и свободно распространяемое программное обеспечение из перечня информационных технологий, используемых при осуществлении образовательного процесса по дисциплине;
- методические указания для обучающихся по освоению дисциплины.

8. Перечень основной и дополнительной учебной литературы

Основная:

1. Valentina Janev. Knowledge Graphs and Big Data Processing / Valentina Janev, Damien Graux, Hajira Jabeen, Emanuel Sallinger // — 2020. — 209 p. — ISBN978-3-030-53199-7. [Электронный ресурс]. <https://link.springer.com/book/10.1007/978-3-030-53199-7>

Дополнительная:

1. Holdaway K. R. Harness oil and gas big data with analytics:optimize exploration and production with data driven models/Keith R. Holdaway.-Hoboken:John Wiley and sons,2014, ISBN 9781118779316.-364.-

Библиогр. в конце глав. - Указ.: с. 351

2. Grant Humphries, Machine Learning for Ecology and Sustainable Natural Resource Management / Grant Humphries, Dawn R. Magness, Falk Huettmann // Publisher Name: Springer, Cham, 2018. — 441 p. — ISBN 978-3-319-96978-7. — Текст : электронный // Электронно-библиотечная система SpringerLink : [сайт]. <https://link.springer.com/book/10.1007/978-3-319-96978-7>

9. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины

При освоении дисциплины использование ресурсов сети Интернет не предусмотрено.

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

Образовательный процесс по дисциплине **Machine Learning for Big Data** предполагает использование следующего программного обеспечения и информационных справочных систем:

- 1) presentation materials (slides on the topics of lectures);
- 2) on-line access to the Electronic library system (ELS);
- 3) access to the electronic information and educational environment of the University; 4) Internet services and electronic resources.

Free software: WPS Office - Office automation system

При освоении материала и выполнения заданий по дисциплине рекомендуется использование материалов, размещенных в Личных кабинетах обучающихся ЕТИС ПГНИУ (**student.psu.ru**).

При организации дистанционной работы и проведении занятий в режиме онлайн могут использоваться:

система видеоконференцсвязи на основе платформы BigBlueButton (<https://bigbluebutton.org/>).

система LMS Moodle (<http://e-learn.psu.ru/>), которая поддерживает возможность использования текстовых материалов и презентаций, аудио- и видеоконтент, а так же тесты, проверяемые задания, задания для совместной работы.

система тестирования Indigo (<https://indigotech.ru/>).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

1. For lectures - a classroom with the presentation equipment (projector, screen, computer/laptop) and the suitable software, chalkboard (and) or whiteboard.
2. Practices - a computer class equipped with personal computers and appropriate software. The composition of the equipment is defined in the Passport of the computer class.
3. For self-directed study – a classroom for independent work that is equipped with computer hardware and access to the Internet and thereby to the electronic educational environment of the university. Halls of PSU Scientific Library.
4. For the current or interim knowledge assessment – a classroom with the presentation equipment (projector, screen, computer/laptop) and the suitable software, chalkboard (and) or whiteboard

Помещения научной библиотеки ПГНИУ для обеспечения самостоятельной работы обучающихся:

1. Научно-библиографический отдел, корп.1, ауд. 142. Оборудован 3 персональными компьютера с доступом к локальной и глобальной компьютерным сетям.
2. Читальный зал гуманитарной литературы, корп. 2, ауд. 418. Оборудован 7 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.
3. Читальный зал естественной литературы, корп.6, ауд. 107а. Оборудован 5 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.
4. Отдел иностранной литературы, корп.2 ауд. 207. Оборудован 1 персональным компьютером с доступом к локальной и глобальной компьютерным сетям.
5. Библиотека юридического факультета, корп.9, ауд. 4. Оборудована 11 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

6. Читальный зал географического факультета, корп.8, ауд. 419. Оборудован 6 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

Все компьютеры, установленные в помещениях научной библиотеки, оснащены следующим программным обеспечением:

Операционная система ALT Linux;

Офисный пакет Libreoffice.

Справочно-правовая система «КонсультантПлюс»

**Фонды оценочных средств для аттестации по дисциплине
Machine Learning for Big Data**

**Планируемые результаты обучения по дисциплине для формирования компетенции.
Индикаторы и критерии их оценивания**

ОПК.9

Способен использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
ОПК.9.1 Обоснованно использует информацию для информационно- аналитической поддержки принятия управленческих решений	Knows the methods of information and analytical support for managerial decision-making. Knows how to reasonably use information for information and analytical support for managerial decision-making. Owns the methods of information- analytical support for managerial decision-making.	Неудовлетворител Does not know the methods of information and analytical support for managerial decision- making. Does not know how to reasonably use information for information and analytical support for managerial decision-making. Does not own the methods of information and analytical support for managerial decision- making.
		Удовлетворительн Poorly knows the methods of information and analytical support for managerial decision- making. Makes gross mistakes when using information for information and analytical support for managerial decision-making. Owns at a satisfactory level the methods of information and analytical support for managerial decision-making.
		Хорошо Knows the methods of information and analytical support for managerial decision- making. Makes minor mistakes when using information for information and analytical support for managerial decision-making. He has a good command of the methods of information and analytical support for managerial decision-making.
		Отлично Knows at a high level the methods of

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
		<p>Отлично</p> <p>information and analytical support for managerial decision-making. Knows how to reasonably use information for information and analytical support for managerial decision-making. He is fluent in the methods of information and analytical support for managerial decision-making.</p>

ПК.3

Способен к планированию и организации аналитических работ, в том числе, с использованием технологий больших данных

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
<p>ПК.3.1</p> <p>Разрабатывает, обсуждает и утверждает содержание аналитических работ, в том числе, с использованием технологий больших данных</p>	<p>Knows the content of analytical work.</p> <p>Able to develop, discuss and approve the content of analytical work.</p> <p>Owens methods of analytical work using big data technology.</p>	<p>Неудовлетворител</p> <p>Does not know the content of analytical works. Does not know how to develop, discuss and approve the content of analytical works. Does not own methods of analytical work using big data technology.</p> <p>Удовлетворительн</p> <p>Poor knowledge of the content of analytical works. Makes gross mistakes in the process of developing, discussing or approving the content of analytical works. Has a satisfactory level of analytical skills using big data technology.</p> <p>Хорошо</p> <p>Knows the content of analytical work. Makes minor mistakes in the process of developing, discussing or approving the content of analytical papers. In general, he owns the methods of analytical work using big data technology.</p> <p>Отлично</p> <p>Knows the content of analytical works at a high level. Able to develop, discuss and approve the content of analytical work.</p>

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
		Отлично He is fluent in the methods of analytical work using big data technology.

Оценочные средства текущего контроля и промежуточной аттестации

Схема доставки : Базовая

Вид мероприятия промежуточной аттестации : Экзамен

Способ проведения мероприятия промежуточной аттестации : Оценка по дисциплине в рамках промежуточной аттестации определяется на основе баллов, набранных обучающимся на контрольных мероприятиях, проводимых в течение учебного периода.

Максимальное количество баллов : 100

Конвертация баллов в отметки

«отлично» - от 81 до 100

«хорошо» - от 61 до 80

«удовлетворительно» - от 43 до 60

«неудовлетворительно» / «незачтено» менее 43 балла

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
Входной контроль	Input control Входное тестирование	Knowledge of mathematics, probability theory and mathematical statistics is tested.
ПК.3.1 Разрабатывает, обсуждает и утверждает содержание аналитических работ, в том числе, с использованием технологий больших данных ОПК.9.1 Обоснованно использует информацию для информационно-аналитической поддержки принятия управленческих решений	Test 1 Письменное контрольное мероприятие	It tests knowledge of the basic concepts related to machine learning and big data technology. Knowledge of regression and classification data models is tested.
ПК.3.1 Разрабатывает, обсуждает и утверждает содержание аналитических работ, в том числе, с использованием технологий больших данных ОПК.9.1 Обоснованно использует информацию для информационно-аналитической поддержки принятия управленческих решений	Test 2 Письменное контрольное мероприятие	Knowledge and skills related to working with big data are tested. Knowledge of methods of cluster analysis and factor analysis is tested.

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
ПК.3.1 Разрабатывает, обсуждает и утверждает содержание аналитических работ, в том числе, с использованием технологий больших данных ОПК.9.1 Обоснованно использует информацию для информационно-аналитической поддержки принятия управленческих решений	Final control activity Итоговое контрольное мероприятие	The level of knowledge on all studied material is checked.

Спецификация мероприятий текущего контроля

Input control

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **0**

Проходной балл: **0**

Показатели оценивания	Баллы
There are basic knowledge of linear algebra, mathematical analysis, optimization theory.	4
Basic knowledge of the course of mathematical statistics.	3
There are basic knowledge of the course of probability theory.	3

Test 1

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **30**

Проходной балл: **13**

Показатели оценивания	Баллы
Knows the basic concepts related to machine learning on big data.	10
Knows and knows how to use methods for constructing classifiers	10
Knows and can use the methods of building regression models	10

Test 2

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **30**

Проходной балл: **13**

Показатели оценивания	Баллы
-----------------------	-------

Knows the basic concepts related to machine learning on big data.	10
Knows and can use the methods of factor analysis.	10
Knows and can use cluster analysis methods.	10

Final control activity

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **40**

Проходной балл: **17**

Показатели оценивания	Баллы
Knows the basic concepts related to machine learning on big data.	10
Knows and can use the methods of constructing associative and sequential rules.	10
Knows and can use the methods of cluster analysis and factor analysis.	10
Knows and can use the methods of building regression and classification data models.	10