

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тверской государственный технический университет»
(ТвГТУ)

УТВЕРЖДАЮ

Проректор
по учебной работе

_____ Э.Ю. Майкова
« _____ » _____ 20__г.

РАБОЧАЯ ПРОГРАММА

дисциплины части, формируемой участниками образовательных отношений
Блока 1 «Дисциплины (модули)»

«Анализ больших данных»

Направление подготовки бакалавров 09.03.04 Программная инженерия
Направленность (профиль) – разработка программно-информационных систем
Типы задач профессиональной деятельности – производственно-технологический

Форма обучения - очная
Факультет информационных технологий
Кафедра Программное обеспечение

Тверь 20__

Рабочая программа дисциплины соответствует ОХОП подготовки бакалавров в части требований к результатам обучения по дисциплине и учебному плану.

Разработчик программы: доцент кафедры ПО

А.Н. Прохныч

Программа рассмотрена и одобрена на заседании кафедры ПО
« ___ » _____ 20____ г., протокол № ____.

Заведующий кафедрой

А.Л. Калабин

Согласовано
Начальник учебно-методического
отдела УМУ

Д.А. Барчуков

Начальник отдела
комплектования
зональной научной библиотеки

О.Ф. Жмыхова

1. Цели и задачи дисциплины.

Целью изучения дисциплины является: изучение математических методов и подходов, используемых в программных системах обработки и анализа больших данных компании SAS - мирового лидера в области разработки и внедрения IT решений и услуг в бизнес-аналитике, в том числе в финансовой и банковской сфере; развитие профессиональных навыков учащихся за счет получения практического опыта работы с IT решениями компании SAS в части обработки и анализа больших данных.

Достижение этих целей позволит подготовить высококвалифицированных специалистов, обладающих уникальной комбинацией компетенций и определенным практическим в передовых областях развития информационных технологий и математики, применяемых в настоящее время для решения задач анализа больших данных в ведущих финансовых организациях мира.

Задачами дисциплины являются:

- формирование базовых знаний в области анализа данных;
- изучение современных методов и алгоритмов баз данных;
- работа и настройка базы данных для производства анализа знаний из данных.

2. Место дисциплины в структуре ОП.

Дисциплина относится к части, формируемой участниками образовательных отношений Блока 1 ОП ВО. Для изучения курса требуются знания дисциплин «Основы программирования», «Разработки и анализ требований», «Тестирование программного обеспечения».

Приобретенные знания в рамках данной дисциплины необходимы в дальнейшем для прохождения производственной практики и при выполнении выпускной квалификационной работы.

3. Планируемые результаты обучения по дисциплине.

3.1. Планируемые результаты обучения по дисциплине.

Компетенция, закрепленная за дисциплиной в ОХОП:

ПК-2. *Способен разрабатывать, вносить изменения в процедуры сборки модулей и компонент, миграции и преобразования (конвертации) данных, процедур развертывания и обновления программного обеспечения.*

Индикаторы компетенции, закреплённых за дисциплиной в ОХОП:

ИПК-2.3 Выбирает стандарты и модели процедуры миграции и преобразования (конвертации) данных.

Показатели оценивания индикаторов достижения компетенций

Знать:

З1. Основные понятия и терминологию в области технологий обработки и анализа больших данных, понятие модели параллельной обработки данных MapReduce и ее практическую реализацию в Hadoop, основные API и типовые примеры программирования в Hadoop, стек Hadoop-технологий для распределенного хранения и обработки данных.

32. Языки программирования для обработки данных SAS Base и SAS Data Step 2, а также методы организации взаимодействия аналитической платформы SAS с распределенными хранилищами информации на основе Hadoop-стека технологий.

Уметь:

У1. Строить и применять на практике описательные и прогнозные модели интеллектуального анализа больших данных и машинного обучения с использованием технологий компании SAS, в том числе разнородных сложно-структурированных данных большого объема.

У2. Использовать программные средства визуализации и интерактивного исследования больших данных.

Иметь опыт практической подготовки:

Реализация стандартных методов обработки и анализа данных.

3.2. Технологии, обеспечивающие формирование компетенций

Проведение лекционных занятий, выполнение лабораторных работ, самостоятельная работа студента.

4. Трудоемкость дисциплины и виды учебной работы.

Таблица 1. Распределение трудоемкости дисциплины по видам учебной работы

Вид учебной работы	Зачетных единиц	Академических часов
Общая трудоемкость дисциплины	3	108
Аудиторные занятия (всего)		65
В том числе:		
Лекции		26
Практические занятия (ПЗ)		не предусмотрены
Лабораторные работы (ЛР)		39
Самостоятельная работа (всего)		43
В том числе:		
Курсовая работа		не предусмотрена
Курсовой проект		не предусмотрен
Расчетно-графические работы		не предусмотрены
Реферат		не предусмотрен
Другие виды самостоятельной работы: - подготовка к лабораторному практикуму - подготовка к практическим занятиям		43
Контроль текущий и промежуточный (балльно-рейтинговый, зачет)		не предусмотрен
Практическая подготовка при реализации дисциплины (всего)		39
В том числе:		
Практические занятия (ПЗ)		не предусмотрены
Лабораторные работы (ЛР)		39
Курсовая работа		не предусмотрена
Курсовой проект		не предусмотрен

5. Структура и содержание дисциплины.

Структура и содержание дисциплины построены по модульно-блочному принципу. Под модулем дисциплины понимается укрупненная логико-понятийная тема, характеризующаяся общностью использованного понятийно-терминологического аппарата.

5.1. Структура дисциплины.

Таблица 2. Модули дисциплины, трудоемкость в часах и виды учебной работы.

№	Наименование модуля	Труд-ть часы	Лекции	Лаб. практика	Сам. работа
1	Введение в большие данные.	32	6	13	6
2	Изучение алгоритмов и методов работы.	33	10	13	15
3	Практическое применение узконаправленных методов и логики на получение знаний из баз данных.	43	10	13	22
Всего на дисциплину		108	26	39	43

5.2. Содержание дисциплины.

Модуль 1. «Введение в большие данные.»

Терминология, история появления. Технические сложности работы с большими данными. Распределенная файловая система HDFS. Базовая идея модели MapReduce, пример использования MapReduce.

Модуль 2. «Изучение алгоритмов и методов работы.»

Современные и полезные алгоритмы: наивный байесс, работа классификатора, кластеризация, априори, методы k, c- средних, методы и алгоритмы снижения размерности баз данных и работа с полученными знаниями.

Модуль 3. «Практическое применение узконаправленных методов и логики на получение знаний из баз данных.»

Применения технологий текстовой аналитики: обогащение информации по клиентам компаний, мониторинг потоков сообщений. Принципы статистического анализа текста (Text Mining): лингвистическая предобработка текста, статистическая фильтрация текста, автоматическое выявление тематик, кластеризация текстов.

5.3. Лабораторные работы

Таблица 3. Лабораторный практикум и его трудоемкость.

№ пп.	Учебно - образовательный модуль. Цели лабораторного практикума	Тематика лабораторного практикума	Трудоемкость в часах
1.	Модуль 1 Цель: Введение в большие данные.	1. Изучение и введение в большие данные.	1
		2. Виды больших данных.	2
		3. Работа с большими данными.	2
		4. Обработка больших данных.	2
		5. Методы и алгоритмы обработки.	2
		6. Среда разработки и выбор программных средств.	2
2	Модуль 2 Цель: Изучение алгоритмов и методов работы.	7. Применение на практике основных и базовых методов для работы с большими данными.	2
3	Модуль 3 Цель: Практическое применение узконаправленных методов и логики на получение знаний из баз данных.	8. Изучение малоизвестных методов.	2
		9. Выбор данных для работы.	2
		10. Получение знаний из данных.	2
		11. Основные принципы получения знаний по узконаправленным методам.	3
		12. Практическое изучение методов.	2
		13. Создание собственных данных.	2

5.4. Практические занятия.

Учебным планом практические занятия не предусмотрены.

6. Самостоятельная работа обучающихся и текущий контроль успеваемости.

6.1. Цели самостоятельной работы

Формирование способностей к самостоятельному познанию и обучению, поиску литературы, обобщению, оформлению и представлению полученных результатов, их критическому анализу, поиску новых и неординарных решений, аргументированному отстаиванию своих предложений, умений подготовки выступлений и ведения дискуссий.

6.2. Организация и содержание самостоятельной работы

Самостоятельная работа заключается в изучении отдельных тем курса по заданию преподавателя по рекомендуемой им учебной литературе, в подготовке к лабораторным работам, к текущему контролю успеваемости, экзамену, зачету и выполнению курсовой работы.

После вводных лекций, в которых обозначается содержание дисциплины, ее проблематика и практическая значимость, студентам выдаются темы курсовой работы, определяется порядок подготовки доклада и презентации для ее защиты.

В рамках дисциплины выполняются лабораторные работы, которые защищаются посредством тестирования или устным опросом (по желанию

обучающегося). Максимальная оценка за каждую выполненную лабораторную работу – 5 баллов, минимальная – 3 балла.

Выполнение всех лабораторных работ обязательно. В случае невыполнения лабораторной работы по уважительной причине студент имеет право выполнить письменный реферат, по согласованной с преподавателем теме по модулю, по которому пропущена лабораторная работа. Возможная тематическая направленность реферативной работы для каждого учебно-образовательного модуля представлена в следующей таблице:

Таблица 3. Темы рефератов

№ п/п	Модули	Возможная тематика самостоятельной реферативной работы
1.	Модуль 1	Big Data; Text mining;
2.	Модуль 2	Наивный Байесс; Метод Йео-Джонсона.
3.	Модуль 3	Классификация; Кластеризация.

Оценивание в этом случае осуществляется путем устного опроса проводится по содержанию и качеству выполненного реферата.

7. Учебно-методическое и информационное обеспечение дисциплины.

7.1. Основная литература

1. Липаев, В.В. Программная инженерия. Методологические основы [Текст]: учебник для вузов по напр. "Бизнес-информатика" (080700) / Гос. ун-т высш. шк. экономики - М.: ТЕИС, 2006. - 607 с. - (75271-1)
2. Грекул, В.И. Проектное управление в сфере информационных технологий [Электронный ресурс] / Грекул, В.И., Н.Л. Коровкина, Купричнов, Ю.Л. - Москва: БИНОМ. Лаборатория знаний, 2013. - ЭБС Лань. - (107032-1)
3. Адизес, И.К. Управляя изменениями [Текст]: пособие - СПб.: Питер, 2010. - 223 с. - (84163-2)
4. Мейер, Б. Agile! Прекрасный, ужасный, шумный [Текст]: пер. с англ. / пер. с англ. и ред. В.А.Биллига - Москва;Тверь: Интуит - ЗАО НИИ ЦПС, 2015. - 248 с. - (109888-1)
5. Сазерленд, Д. *SCRUM. Революционный метод управления проектами - Москва: Манн, Иванов и Фербер, 2016. - (114795-3)

7.2. Дополнительная литература

1. Чудов, С.А. Согласованное управление проектом создания системы электронного документооборота сетевой организации с использованием эвристической процедуры нечеткой оценки решений [Текст]: диссертация .. канд. техн. наук: спец. 05.13.01 - Системный анализ, управление и обработка информации, 05.13.10 - Управление в соц. и экон. системах - Тверь, 2013. - 190 с. - (102165-1)

2. Гусятников, В.Н. Стандартизация и разработка программных систем [Текст]: учеб. пособие для вузов по спец. 080801 "Прикл. информатика" и др. экон. спец. / Гусятников, В.Н., Безруков, А.И. - М.: Финансы и статистика : ИНФРА-М, 2010. - 286 с. - (82353-106)

3. Маккарти, Д. Правила разработки программного обеспечения [Текст]: [практ. рук.] / Маккарти, Д., Маккарти, М. - СПб. [и др.]: Питер, 2007. - 220 с. - (66402-3)

4. Пилон, Д. Управление разработкой ПО [Текст] / Пилон, Д., Майлз, Р. - СПб.: Питер, 2011. - 459 с. - (93822-2)

5. Мартин, Р.С. Быстрая разработка программ: принципы, примеры, практика / Мартин, Р.С., Ньюкирк, Д.В., Косс, Р.С. - М. [и др.]: Вильямс, 2004. - 744 с. - (16348-1)

6. Ильина, Н.С. Календарное планирование в системах Time Line и Microsoft Project [Электронный ресурс]: учеб. пособие / Ильина, Н.С., Лобанова, Е.В., Стрелец, К.И. ; Санкт-Петербургский гос. политехн. ун-т - СПб.: Санкт-Петербургский гос. политехн. ун-т, 2005. - Внешний сервер. - (78733-1)

7. Белладжио, Д. Стратегия управления конфигурацией программного обеспечения с использованием IBM Rational ClearCase [Электронный ресурс] / Белладжио, Д., Миллиган, Т. ; пер. с англ. Н. Мухина - М.: ДМК Пресс, 2009. - ЭБС Лань. - (110620-1)

8. Мороз, О.А. Управление проектами в ProjectLibre [Текст]: [учеб. пособие для вузов] - Ростов-на-Дону: Феникс, 2015. - 254 с. - (83492-2)

7.3. Методические материалы

1. Учебно-методический комплекс дисциплины вариативной части Блока 1 "Анализ больших данных". Направление подготовки бакалавров 09.03.04 Программная инженерия Направленность (профиль) – разработка программно-информационных систем : ФГОС 3++ / Тверской гос. техн. ун-т, Каф. ПО ; сост. А.Н. Прохныч. - 2022. - (УМК). - Текст : электронный. - URL: <https://elib.tstu.tver.ru/MegaPro/GetDoc/Megapro/153318> . - (ID=153318-0)

7.4. Программное обеспечение по дисциплине

1. Операционная система Microsoft Windows: лицензии № ICM-176609 и № ICM-176613 (Azure Dev Tools for Teaching).

2. Microsoft Office 2007 Russian Academic: OPEN No Level: лицензия № 41902814.

7.5. Специализированные базы данных, справочные системы, электронно-библиотечные системы, профессиональные порталы в Интернет

ЭБС и лицензионные ресурсы размещены: <http://lib.tstu.tver.ru/header/obr-res>

1. ЭБС "Лань": <https://e.lanbook.com/>

2. ЭБС "Университетская библиотека онлайн": <http://www.biblioclub.ru/>

3. ЭБС «IPRBooks»: <http://www.iprbookshop.ru/>

4. Электронная образовательная платформа "Юрайт" (ЭБС «Юрайт»): <http://urait.ru/>

5. Научная электронная библиотека eLIBRARY: <http://elibrary.ru/>

УМК размещен: <https://elib.tstu.tver.ru/MegaPro/GetDoc/Megapro/153318>

8. Материально-техническое обеспечение.

При изучении дисциплины используются современные средства обучения: наглядные пособия, диаграммы, схемы.

Возможна демонстрация лекционного материала с помощью оверхедпроектора (кодоскопа) и мультипроектора.

Вуз имеет лабораторию для реализации лабораторного практикума; учебный класс для проведения самостоятельной работы, оснащенный современной компьютерной и офисной техникой, необходимым программным обеспечением, электронными учебными пособиями, имеющий безлимитный выход в глобальную сеть; аудиторию для проведения семинарских занятий, практикумов и презентаций студенческих работ, оснащенную аудиовизуальной техникой.

Таблица 4. Материально-техническое обеспечение.

№№	Компьютерное и программное обеспечение
1	Компьютерный класс, имеющий безлимитный выход в глобальную сеть, оснащенный операционной системой семейства Windows (не ниже Windows XP), программным обеспечением MS Office 2003.
2	Аудиторию для проведения лекционных занятий и презентаций студенческих работ, оснащенную аудиовизуальной техникой.

9. Оценочные средства для проведения промежуточной аттестации

9.1. Оценочные средства для проведения промежуточной аттестации в форме экзамена

Учебным планом экзамен по дисциплине не предусмотрен.

9.2. Оценочные средства для проведения промежуточной аттестации в форме зачета

1. Вид промежуточной аттестации в форме зачета.

Вид промежуточной аттестации:

по результатам текущего контроля знаний и умений обучающегося без дополнительных контрольных испытаний;

2. При промежуточной аттестации без выполнения дополнительного итогового контрольного испытания студенту в обязательном порядке описываются критерии проставления зачёта:

«зачтено» - выставляется обучающемуся при условии выполнения им всех контрольных мероприятий: посещение лекций в объеме не менее 80% контактной работы с преподавателем, выполнения практических и защиты лабораторных работ, курсовой работы.

9.3. Оценочные средства для проведения промежуточной аттестации в форме курсового проекта или курсовой работы

Учебным планом курсовая работа и курсовой проект по дисциплине не предусмотрены.

10. Методические рекомендации по организации изучения дисциплины.

Студенты перед началом изучения дисциплины ознакомлены с системами кредитных единиц и балльно-рейтинговой оценки, которые опубликованы и размещены на сайте вуза или кафедры.

В учебном процесс внедрена субъект-субъектная педагогическая технология, при которой в расписании каждого преподавателя определяется время консультаций студентов по закрепленному за ним модулю дисциплины.

Студентов, изучающих дисциплину, обеспечены электронными учебниками, учебно-методическим комплексом по дисциплине.

11. Внесение изменений и дополнений в рабочую программу дисциплин.

Кафедра ежегодно обновляет содержание рабочих программ дисциплин, которые оформляются протоколами заседаний дисциплин, форма которых утверждена Положением о рабочих программах дисциплин, соответствующих ФГОС ВО.