



Федеральное агентство по рыболовству
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Калининградский государственный технический университет»
(ФГБОУ ВО «КГТУ»)

Институт цифровых технологий

УТВЕРЖДАЮ:
Первый проректор
О.Г. Огий

ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА
(программа повышения квалификации)
«ТЕХНОЛОГИИ DATA MINING»

Трудоемкость – 72 ч

Разработчик: *кафедра прикладной математики и информационных технологий*

Автор: *ст. преподаватель Романов Михаил Андреевич*

г. Калининград
2024

СОДЕРЖАНИЕ

1.	ПОЯСНИТЕЛЬНАЯ ЗАПИСКА	3
2.	УЧЕБНЫЙ ПЛАН И КАЛЕНДАРНЫЙ УЧЕБНЫЙ ПЛАН	4
3.	РАБОЧИЕ ПРОГРАММЫ ПРЕДМЕТОВ, КУРСОВ, ДИСЦИПЛИН (МОДУЛЕЙ) ПРОГРАММЫ	5
4	ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ	9
4.1	Материально-техническое обеспечение учебного процесса	9
4.2	Организация образовательного процесса	9
4.3	Кадровое обеспечение	9
4.4	Методические рекомендации по реализации программы	10
5	ИТОГОВАЯ АТТЕСТАЦИЯ ПО ПРОГРАММЕ	10

1. ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Дополнительная профессиональная программа (повышение квалификации) (далее - ДПП), реализуется в соответствии с Федеральным законом от 29.12.2012 г. № 273-ФЗ "Об образовании в Российской Федерации", Приказом Минобрнауки России от 01.07.2013 № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам», «Стратегией в области цифровой трансформации отраслей экономики, социальной сферы и государственного управления в РФ», а также в соответствии с современными тенденциями в IT-сфере.

Программа курса «Технологии Data Mining» раскрывает вопросы современных технологий, применяемых для сбора, обработки и анализа больших данных, а также поиск новых, нетривиальных, практически полезных знаний. Слушатель познакомится с основными алгоритмами интеллектуального анализа больших данных и получит навыки использования современных интеллектуальных средств.

Цель: сформировать у студентов целостное представление о технологии интеллектуального анализа больших наборов данных, а также знания и умения в предварительной обработке данных.

Задачи: ознакомиться с базовыми концепциями технологии и структурой отдельных алгоритмов, а также с алгоритмами обнаружения новых, нетривиальных, практически полезных знаний в больших наборах данных;
получить представление о процедуре сбора, транспортировки, хранения и анализа больших данных.

Категория слушателей. (требования к квалификации слушателей): Лица, имеющие высшее (бакалавриат, магистратура или специалитет) образование

Срок освоения: 72 ч.

Режим занятий: с отрывом от работы/ без отрыва от работы

Форма обучения: очная, применение электронного обучения

Планируемые результаты обучения. Компетентностный профиль программы.

Перечень компетенций, подлежащих совершенствованию, и (или) перечень новых компетенций, формирующихся в результате освоения.

ПК-1 – Способен разрабатывать и внедрять новые методы и технологии больших данных

ПК-1.2 – Проводит испытаний и разрабатывает рекомендации по внедрению и использованию усовершенствованных или разработанных новых методов, моделей, алгоритмов, технологий и инструментальных средств работы с большими данными

Профессиональный стандарт «Специалист по большим данным», утвержденный приказом Минтруда № 405н от 06.07.2020

ОТФ А.6: Анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры

ТФ А/04.6: Проведение аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика

- знания:**
- 1) Содержание и последовательность выполнения этапов аналитического проекта;
 - 2) Предметная область анализа;
 - 3) Теоретические и прикладные основы анализа больших данных;
 - 4) Технологии анализа данных;

- 5) Алгоритмы машинного обучения.
- умения:**
- 1) Планировать аналитические работы с использованием технологий больших данных;
 - 2) Проводить аналитические работы с использованием технологий больших данных, как индивидуально, так и, осуществляя руководство малыми аналитическими группами;
 - 3) Проводить сравнительный анализ методов и инструментальных средств анализа больших данных;
 - 4) Разрабатывать и оценивать модели больших данных;
 - 5) Решать задачи классификации, кластеризации, регрессии, прогнозирования, снижения размерности и ранжирования данных;
 - 6) Осуществлять поиск информации о новых и перспективных методах анализа больших данных, выполнять сравнительный анализ методов.
- трудовые действия:**
- 1) Выбор методов и инструментальных средств анализа больших данных для проведения аналитических работ;
 - 2) Разработка, проверка, оценка используемых моделей больших данных;
 - 3) Выбор средств представления результатов аналитики больших данных;
 - 4) Подготовка отчета по результатам аналитических работ с использованием технологий больших данных.

2. УЧЕБНЫЙ ПЛАН И КАЛЕНДАРНЫЙ УЧЕБНЫЙ ПЛАН

2.1 Учебный план

№ п/п	Наименование предметов, курсов, дисциплин (модулей)	Всего часов	В том числе			Форма контроля
			ЛК	ПЗ	СР	
1.	Модуль 1. Введение в Data Mining	10	4	4	2	Дискуссия
2.	Модуль 2. Предобработка данных	14	4	4	6	Дискуссия
3.	Модуль 3. Обучение с учителем	24	10	8	6	Дискуссия
4.	Модуль 4. Обучение без учителя	18	6	6	6	Дискуссия
5.	Итоговая аттестация	6	-	-	6	Тестирование
	Итого часов	72	24	22	26	-

Примечание: при необходимости количество часов по отдельным модулям программы может быть изменено

2.2 Календарный учебный график

№ п/п	Наименование предметов, курсов, дисциплин (модулей)	Номер дня 1-й учебной недели с начала обучения					Номер дня 2-й учебной недели с начала обучения				
		1	2	3	4	5	1	2	3	4	5
		часов в день					часов в день				
1	Модуль 1. Введение в Data Mining	Т	Т	Т	х	х	х	х	х	х	х
2	Модуль 2. Предобработка данных	х	х	Т	Т	х	х	х	х	х	х
3	Модуль 3. Обучение с учителем	х	х	х	Т	Т	х	х	х	х	х
4	Модуль 4. Обучение без учителя	х	х	х	х	Т	Т	х	х	х	х
5	Итоговая аттестация	х	х	х	х	х	х	И	х	х	х

¹Даты обучения будут определены в расписании занятий при наборе группы на обучение

□ – учебная неделя; Т – теоретическое обучение; С – стажировка; А – промежуточная аттестация; И – защита проекта; × – нет недели

3. РАБОЧИЕ ПРОГРАММЫ ПРЕДМЕТОВ, КУРСОВ, ДИСЦИПЛИН (МОДУЛЕЙ) ПРОГРАММЫ

3.1 Рабочая программа модуля «Введение в Data Mining»

3.1.1 Пояснительная записка

Цель:	определить цели и задачи дисциплины, планируемые результаты освоения дисциплины; иметь представление об основных этапах анализа данных.
В результате изучения слушатели должны:	
знать:	Цели и задачи анализа данных; Основные этапы анализа данных; Основные методы, применяемые в Data Mining
уметь:	Определить класс решаемой задачи и подобрать необходимый метод для работы.
владеть:	навыками планирования анализа больших данных; навыками работы с инструментальными программными средствами интеллектуального анализа данных

3.1.2 Учебно-тематический план

№ п/п	Наименование тем и разделов	Всего часов	в том числе			Проверка знаний
			ЛК	ПЗ	СР	
1.	Цели и задачи дисциплины.	4	2	2	0	Дискуссия
2.	Анализ данных.	6	2	2	2	Дискуссия
Итого:		10	4	4	2	

3.1.3 Содержание программы

Тема	Содержание темы
Цели и задачи дисциплины.	Цели и задачи дисциплины. Место дисциплины в структуре образовательной программы. Планируемые результаты освоения дисциплины.
Анализ данных.	Понятие анализа данных. Задачи анализа данных. Методы анализа данных. Принципы анализа данных. Инструменты анализа данных. Области применения анализа данных

3.1.4 Промежуточная аттестация по программе

Промежуточная и итоговая аттестация по модулю не предусмотрена.

3.1.5 Обеспеченность образовательного процесса учебной литературой и информационными ресурсами

Материалы для самостоятельной работы и подготовки итогового проекта размещаются на <http://eios.klgtu.ru/mod> ЭИОС КГТУ. Доступ к материалам осуществляется после регистрации на основании договора об оказании образовательных услуг по программе ДПО.

3.2 Рабочая программа модуля «Предобработка данных»

3.2.1 Пояснительная записка

Цель:	Получить сведения об этапах предварительной обработки данных
В результате изучения слушатели должны:	
знать:	Все этапы предварительной обработки данных;
уметь:	Собирать, хранить и обрабатывать данные;
владеть:	навыками обработки сырых данных

3.2.2 Учебно-тематический план

№ п/п	Наименование тем и разделов	Всего часов	в том числе			Проверка знаний
			ЛК	ПЗ	СР	
1.	Очистка данных. Преобразование данных. Интеграция данных.	8	4	2	2	Дискуссия
2.	Подготовка признаков для моделирования.	6	2	2	2	Дискуссия
Итого:		14	6	4	4	

3.2.3 Содержание программы

Тема	Содержание темы
Очистка данных. Преобразование данных. Интеграция данных.	Очистка данных. Преобразование данных. Интеграция данных. Разделение данных. Определение типов данных. Обработка отсутствующих значений. Обработка выбросов. Масштабирование данных. Нормализация данных. Кодирование категориальных признаков. Слияние данных. Удаление дубликатов. Индексирование данных.
Подготовка признаков для моделирования.	Разделение на обучающую и тестовую выборки. Балансировка классов. Подготовка признаков для моделирования. Уменьшение размерности данных. Визуализация данных. Оценка качества данных.

3.2.4 Промежуточная аттестация по программе

Промежуточная и итоговая аттестация по модулю не предусмотрена.

3.2.5 Обеспеченность образовательного процесса учебной литературой и информационными ресурсами

Материалы для самостоятельной работы и подготовки итогового проекта размещаются на <http://eios.klgtu.ru/mod> ЭИОС КГТУ. Доступ к материалам осуществляется после регистрации на основании договора об оказании образовательных услуг по программе ДПО.

3.3 Рабочая программа модуля «Обучение с учителем»

3.3.1 Пояснительная записка

Цель:	Изучить основные методы обучения без учителя
В результате изучения слушатели должны:	
знать:	Основные алгоритмы обучения с учителем;
уметь:	применять соответствующие алгоритмы для решения профессиональных задач
владеть:	Навыками интеллектуального анализа больших данных алгоритмами классификации и регрессионного анализа

3.3.2 Учебно-тематический план

№ п/п	Наименование тем и разделов	Всего часов	в том числе			Проверка знаний
			ЛК	ПЗ	СР	
1.	Линейная регрессия	8	4	2	2	Дискуссия
2.	Классификация	16	6	6	4	Дискуссия
Итого:		24	10	8	6	

3.3.3 Содержание программы

Тема	Содержание темы
Линейная регрессия	Линейная регрессия. Модель линейной регрессии. Коэффициенты линейной регрессии. Метод наименьших квадратов. Регрессионный анализ. Множественная линейная регрессия. Кросс-валидация. Применение линейной регрессии в реальных задачах.
Классификация	Классификация. Класс. Объект. Признак. Дерево решений. Нейронные сети. Метод опорных векторов. Метрики качества.

3.3.4 Промежуточная аттестация по программе

Промежуточная и итоговая аттестация по модулю не предусмотрена.

3.3.5 Обеспеченность образовательного процесса учебной литературой и информационными ресурсами

Материалы для самостоятельной работы и подготовки итогового проекта размещаются на <http://eios.klgtu.ru/mod> ЭИОС КГТУ. Доступ к материалам осуществляется после регистрации на основании договора об оказании образовательных услуг по программе ДПО.

3.4 Рабочая программа модуля «Обучение без учителя»

3.4.1 Пояснительная записка

Цель:	изучить основные методы обучения без учителя
В результате изучения слушатели должны:	
знать:	Основные алгоритмы обучения без учителя;
уметь:	применять соответствующие алгоритмы для решения профессиональных задач
владеть:	Навыками интеллектуального анализа больших данных алгоритмами кластеризации и ассоциативных правил

3.4.2 Учебно-тематический план

№ п/п	Наименование тем и разделов	Всего часов	в том числе			Проверка знаний
			ЛК	ПЗ	СР	
1.	Кластеризация	11	3	4	4	Дискуссия
2.	Ассоциативные правила	7	3	2	2	Дискуссия
Итого:		18	6	6	6	

3.4.3 Содержание программы

Тема	Содержание темы
Кластеризация	Кластеризация. Кластер. Разбиение. Иерархическая кластеризация. Алгоритм к-средних. EM-алгоритм. Алгоритм DBSCAN.
Ассоциативные правила	Ассоциативные правила. Поддержка и уверенность. Правила ассоциации первого порядка. Правила ассоциации высоких порядков. Методы извлечения правил ассоциации. Алгоритм Apriori. Алгоритм FP-growth.

3.2.4 Промежуточная аттестация по программе

Промежуточная и итоговая аттестация по модулю не предусмотрена.

3.2.5 Обеспеченность образовательного процесса учебной литературой и информационными ресурсами

Материалы для самостоятельной работы и подготовки итогового проекта размещаются на <http://eios.klgtu.ru/mod> ЭИОС КГТУ. Доступ к материалам осуществляется после регистрации на основании договора об оказании образовательных услуг по программе ДПО.

4 ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ

4.1 Материально-техническое обеспечение учебного процесса

В ходе освоения программы слушатели используют возможности интерактивной коммуникации со всеми участниками и заинтересованными сторонами образовательного процесса, ресурсы и информационные технологии посредством электронной информационной образовательной среды университета.

Перечень современных профессиональных баз данных и информационных справочных систем, к которым обучающимся по образовательной программе обеспечивается доступ (удаленный доступ) является ежегодно обновляемым приложением и размещается на официальном сайте в разделе «Образовательные программы высшего образования университета» и в ЭИОС.

При дистанционном обучении преподавателю обеспечивается доступ к платформе проведения вебинаров в соответствии с расписанием. Технические и программные средства обеспечиваются слушателем самостоятельно.

При смешанном обучении занятия проводятся в компьютерных классах и мультимедийных аудиториях, оборудованных техническими средствами для проведения презентаций:

- персональный компьютер с ОС Windows 10 – 11;
- проектор;
- программное обеспечение MSOffice версий 2007 и выше;
- программное обеспечение Loginom версии 7.0.1 и выше;
- программное обеспечение Python версии 3.6 и выше;
- доступ в сеть Интернет.

При всех формах реализации программы должны соблюдаться требования соответствующих СанПиН.

4.2 Организация образовательного процесса

Реализация программы осуществляется в соответствии с требованиями к организации образовательного процесса в университете, изложенными в локальных нормативных актах.

Приведенное выше распределение модулей и тем занятий по дням занятий может уточняться с учетом выбранной формы обучения (очной, очно-заочной, заочной с применением электронного обучения и дистанционных образовательных технологий).

Обучение осуществляется на образовательной площадке университета и носит непрерывный характер. Преподаватели консультируют слушателей как в очном режиме, так и в режиме онлайн.

Программа разработана на основе практико-ориентированного подхода. Её освоение позволит слушателям решать на современном уровне практические задачи, связанные с функциями по обеспечению требований охраны и безопасности труда в своих организациях.

4.3 Кадровое обеспечение

Реализация программы обеспечивается профессорско-преподавательским составом, отвечающим одному из следующих критериев:

- наличие профильного образования по направлению читаемых дисциплин;
- наличие опыта практической работы не менее 5 лет по направлению дисциплины и опыта преподавательской работы не менее 2 лет.

К реализации программы привлекаются как штатные преподаватели университета, так и сторонние специалисты по договорам гражданско-правового характера.

4.4 Методические рекомендации по реализации программы

В процессе преподавания используются следующие образовательные технологии:

- проведение лекций и практических занятий;
- использование возможностей дистанционного консультирования и обучения.

При изложении материала важно помнить, что почти половина информации на лекции передается через интонацию. Учитывать тот факт, что первый кризис внимания слушателей наступает на 15-20-й минутах, второй – на 30-35-й минутах

В основу дидактических требований к лекционному материалу должны быть положены следующие требования:

- изложение материала от простого к сложному, от известного к неизвестному;
- логичность, четкость и ясность в изложении материала;
- возможность проблемного изложения, дискуссии, диалога с целью активизации деятельности слушателей;
- опора смысловой части лекции на подлинные факты, события, явления, статистические данные;
- тесная связь с теоретических положений и выводов с практикой.

Каждый раздел лекции целесообразно завершать резюме и дискуссией.

Контролируемая самостоятельная работа направлена на углубление и закрепление знаний слушателя, развитие аналитических навыков по тематике курса. Подведение итогов и оценка результатов таких форм самостоятельной работы осуществляется во время аудиторных занятий с преподавателем.

5 ИТОГОВАЯ АТТЕСТАЦИЯ ПО ПРОГРАММЕ

Итоговая аттестация по программе проводится в виде итогового тестирования. Тестирования производиться через внутриуниверситетскую электронную систему.

Лицам, успешно освоившим программу и прошедшим итоговую аттестацию, выдается документ о повышении квалификации установленного образца.

ЛИСТ СОГЛАСОВАНИЯ

Дополнительная профессиональная программа (программа повышения квалификации) «**Технологии Data mining**» утверждена на заседании учебно-методической комиссии Института цифровых технологий.

Зам. директора Института
цифровых технологий по ДО и ПП



Е.В. Кривоускова

и.о. директор Института
цифровых технологий

М.В. Гоконаев