

Минобрнауки России
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Сыктывкарский государственный университет имени Питирима Сорокина»
(ФГБОУ ВО «СГУ им. Питирима Сорокина»)

УТВЕРЖДЕНО:
решением Ученого совета
от 31 августа 2020 г.
№ 4.5/21 (541)

Дополнительная профессиональная программа
повышения квалификации
«БОЛЬШИЕ ДАННЫЕ»

г. Сыктывкар

2020

ОБЩИЕ СВЕДЕНИЯ О ПРОГРАММЕ

1. Дополнительная профессиональная программа повышения квалификации разработана на основе следующих документов:

— Федеральный закон от 29 декабря 2012 г. № 273-ФЗ «Об образовании в Российской Федерации»;

— Приказ Минобрнауки России от 1 июля 2013 г. № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам»;

— Методические рекомендации-разъяснения по разработке дополнительных профессиональных программ на основе профессиональных стандартов (Письмо Минобрнауки России от 22.04.2015 № ВК-1032/06);

— «Методические рекомендации по разработке дополнительных профессиональных программ», утвержденные решением Совета по инновационной деятельности и информатизации ФГБОУ ВО «СГУ им. Питирима Сорокина» от 26 мая 2015 г.

— Положение по организации и осуществлению образовательной деятельности по дополнительным профессиональным программам ФГБОУ ВПО «СыктГУ», утвержденное приказом № 378-ОСД от 27 мая 2014 г.

— Профессиональный стандарт 06.001 «Программист», утв. приказом Минтруда России от 18.11.2013 № 679н;

— 08.022 «Статистик», утвержден приказом Министерства труда и социальной защиты Российской Федерации от 08.09.2015 № 605н;

— Спецификации стандартов Worldskills Russia.

2. Связь дополнительной профессиональной программы повышения квалификации с профессиональными и федеральными государственными образовательными стандартами ВО.

Программа предназначена для повышения квалификации лиц, имеющих среднее профессиональное или высшее образование, в части формирования теоретических знаний и практических навыков в сфере работы с большими массивами данных, повышения мотивации к подобной деятельности.

Программно-аппаратной основой обучения является оборудование и программное обеспечение мастерской «Машинное обучение и большие данные» ФГБОУ ВО «СГУ им. Питирима Сорокина».

Целевой аудиторией программы являются лица, имеющие среднее профессиональное или высшее образование, или имеющие необходимость приобрести или повысить свои теоретические знания и практические навыки при работе с большими массивами данных. В

силу этого освоение программы требует наличия базовых компетенций и трудовых функций, включающих знания, умения и навыки, необходимые для работы с компьютером на уровне пользователя, а также знания, умения и навыки, необходимые для установки, настройки, применения, разработки компьютерных программ, желательно наличие компетенций, включающих владение языками программирования высокого уровня, например, Python, знания основ математической статистики, математического анализа.

Указанные компетенции и трудовые функции составляют входные требования программы и отражены в разных вариациях в следующих образовательных и профессиональных стандартах (см. таблицу 1).

Таблица 1.

Образовательный стандарт	Профессиональный стандарт
<p>— Практически все образовательные стандарты, содержащие компетенции, связанные с умением применять информационные и информационно-коммуникационные технологии (в том числе в профессиональной деятельности), формируются дисциплинами «Информатика», «Информационно-коммуникационные технологии», «Информационные технологии в профессиональной деятельности» и т.п.;</p> <p>— Образовательные стандарты, содержащие компетенции, связанные с обучением проектированию и разработке программного обеспечения (группа 09.00.00). Например, 09.02.07 «Информационные системы и программирование», 09.03.03 «Прикладная информатика» (профиль – «Прикладная информатика в экономике»);</p> <p>— Техническое описание компетенций «Машинное обучение и большие данные» Worldskills Russia.</p>	<p>06.001 «Программист», утв. приказом Минтруда России от 18.11.2013 № 679н;</p> <p>08.022 «Статистик», утвержден приказом Министерства труда и социальной защиты Российской Федерации от 08.09.2015 № 605н</p>

1. ЦЕЛЬ РЕАЛИЗАЦИИ ПРОГРАММЫ

Цели реализации дополнительной профессиональной программы повышения квалификации:

- формирование теоретических знаний и практических навыков в сфере работы с большими массивами данных, в части анализа и обработки данных;
- формирование теоретических знаний и практических навыков в области машинного обучения: построение и оценка качества моделей машинного обучения.

Образовательные результаты

Имеющиеся компетенции

Минимальные требования: владение навыками работы с компьютером на уровне среднего пользователя, владение приемами работы с ресурсами и технологиями сети Интернет.

Рекомендуемые требования: владение навыками работы с компьютером на уровне опытного пользователя, владение приемами работы с ресурсами и технологиями сети Интернет, знания и навыки в области программирования, владение языками программирования высокого уровня (Python и т.д.), знание основ высшей математики (основы математической статистики, математического анализа).

Вид деятельности:

Загрузка, анализ, обработка массивов больших данных, построение моделей машинного обучения, проверка качества моделей.

Основные трудовые функции, знания и умения согласно указанным профессиональным стандартам:

06.001 «Программист», утв. Приказом Минтруда России от 18.11.2013 № 679н

1. Разработка и отладка программного кода:

Знания:

- методологии и технологии проектирования и использования баз данных;
- технологии программирования;
- методы и приемы отладки программного кода;

Умения:

- применять выбранные языки программирования для написания программного кода;
- использовать выбранную среду программирования и средства системы управления базами данных;
- использовать возможности имеющейся технической и/или программной архитектуры;

– выявлять ошибки в программном коде.

2. Проверка работоспособности и рефакторинг кода программного обеспечения:

Знания:

- языки программирования и среды разработки;
- методы и приемы отладки дефектного программного кода;

Умения:

- анализировать значения полученных характеристик программного обеспечения;
- применять методы и приемы отладки дефектного программного кода.

3. Разработка требований и проектирование программного обеспечения:

Знания:

- возможности современных и перспективных средств разработки программных продуктов, технических средств;
- методы и средства проектирования программного обеспечения;
- методы и средства проектирования баз данных.

08.022 «Статистик», утвержден приказом Министерства труда и социальной защиты Российской Федерации от 08.09.2015 № 605н

1. Разработка и совершенствование вероятностных статистических методов анализа массовых количественных данных. Анализ данных на основе методов математической статистики.

Необходимые умения

- Производить статистические расчеты на основе соответствующих математических и технических средств
- Применять статистические пакеты прикладных программ
- Работать с различными источниками статистической информации

Необходимые знания

- Методики осуществления статистических расчетов
- Статистические пакеты прикладных программ

Техническое описание компетенции «Машинное обучение и большие данные». Worldskills Russia (даны выборочно)

Специалист должен знать и понимать (по разделам):

1. Построение моделей машинного обучения

- различные методы и алгоритмы машинного обучения;
- Критерии качества моделей машинного обучения;
- Последовательность разработки моделей машинного обучения;

- Какие методы машинного обучения применять в зависимости от исходных данных;
- Как работать с различными выборками данных;
- Как использовать различные программные средства для разработки и улучшения моделей.

2. Разработка прикладных решений

- Подходы к разработке прикладных решений;
- архитектуры прикладных решений;
- способы разработки прикладных решений;
- принципы построения человеко-машинного взаимодействия.

3. Анализ и обработка данных

- разновидности анализа данных;
- приемы анализа данных;
- ставить и проверять гипотезы о зависимости данных;
- методы сбора данных;
- принципы работы с графическими данными;
- принципы работы с текстовыми данными;
- принципы работы с аудио данными;
- принципы работы с видео данными.

Специалист должен уметь (по разделам):

1. Построение моделей машинного обучения

применять классические алгоритмы машинного обучения:

- Обучение без учителя (уменьшение размерности, поиск правил, кластеризация);
- Обучение с учителем (регрессия, классификация);

применять алгоритмы обучения с подкреплением:

- Генетические алгоритмы;
- Алгоритмы Q-Learning;

применять ансамблевые методы машинного обучения:

- Стекинг;
- Беггинг;
- Бустинг;

применять методы глубокого обучения и нейронные сети:

- Перцептроны
- Сверточные нейросети;
- Рекуррентные сети;
- Автоэнкодеры.

2. Разработка прикладных решений

- владеть языками программирования высокого уровня;
- интерпретировать результаты решения;
- интерпретация больших данных и больших изображений;
- осуществлять разработку прикладных решений, использующих модели машинного обучения;
- демонстрировать работоспособность прикладных решений.

3. Анализ и обработка данных

- проводить анализ временных рядов;
- выявлять аномалии и отклонения в данных;
- структурировать данные;
- проводить нормализацию и подготовку данных;
- выделять признаки, свойства и характеристики объектов в данных;
- проводить кластеризацию данных;
- искать корреляцию данных;
- осуществлять операции с большими данными;
- выявлять закономерности в данных;
- проводить визуальный анализ данных;
- формировать шаблоны закономерностей;
- использовать программные средства для исследования и обработки данных.

Желаемые результаты обучения представлены в таблице 2.

Таблица 2.

Профессиональные компетенции	Практический опыт	Умения	Знания
Способность работать с большими массивами данных: проводить анализ и обработку.	Использование языка Python и библиотек для загрузки больших массивов данных	У1 — проводить первичный анализ данных; выявлять закономерности в данных; искать корреляцию данных; проводить визуальный	З1 - разновидности анализа данных; приемы анализа данных; методы сбора данных; принципы работы с текстовыми

	<ul style="list-style-type: none"> – анализа данных – обработки данных 	<p>анализ данных;</p> <p>У2 — Осуществлять подготовку данных (стандартизацию, нормализацию, отбор);</p> <p>У4 – составлять интерактивные отчеты <i>по анализу данных</i> в среде Jupyter Notebook;</p>	данными;
Способность строить модели для решения основных задач машинного обучения, проверять качество моделей машинного обучения.	Использование языка Python и библиотек для <ul style="list-style-type: none"> – построения и проверки качества моделей машинного обучения 	У3 — Применять классические алгоритмы машинного обучения (строить модели МО) и проверять их качество при помощи библиотек Python	З2 – разновидности задач машинного обучения, методы и алгоритмы машинного обучения; последовательность разработки моделей машинного обучения;
Способность разрабатывать прикладные решения.	Использование языка Python для составления интерактивных отчетов в среде Jupyter Notebook;	У5 - владеть языками программирования высокого уровня;	З4 – основы языка Python

2. СОДЕРЖАНИЕ ПРОГРАММЫ

2.1. Учебный план

Срок обучения: 18 час.

Форма обучения: очная, очно-заочная, заочная, с применением ДОТ.

Учебный план представлен в таблице 3.

Таблица 3.

Раздел	Часов						Форма контроля
	Всего	Лек.	Лаб.	В т. Ч. ДОТ	Сам. Раб.	Стажи-ровка	
Раздел 1. Основы синтаксиса Python	3	0	3	3			Отчет
Раздел 2. Первичный анализ больших данных с Pandas	3	0	3	3			Отчет
Раздел 3. Визуальный анализ больших данных с Seaborn	3	0	3	3			Отчет
Раздел 4. Задача классификации	3	0	3	3			Отчет
Раздел 5. Задача регрессии	3	0	3	3			Отчет
Раздел 6. Обучение без учителя	2	0	2	2			Отчет
Итоговая аттестация	1	0	1	1			Тест
Итого	18	0	18	18			

Занятия проводятся в электронной образовательной среде, где обучающимся открывается дистанционный доступ к учебным материалам и/или в виде вебинаров.

Обучающимся потребуется самостоятельно скачать с веб-ресурсов производителей и установить необходимое для работы программное обеспечение.

Курс состоит из шести разделов. Обучающиеся овладевают практическими навыками работы с большими массивами данных.

Содержание:

В процессе освоения программы обучающиеся знакомятся с основами синтаксиса языка Python, учатся создавать интерактивные отчеты в среде Jupyter Notebook, загружают данные, проводят их первичный и визуальный анализ, предварительную подготовку, выявляют закономерности, ищут корреляцию данных, строят модели машинного обучения и

оценивают качество построенных моделей.

Формой промежуточной аттестации является выполнение практических заданий и сдача отчетов.

Итоговая аттестация: курс считается пройденным, если пройдено итоговое тестирование.

2.2. Учебно-тематический план

Учебно-тематический план программы представлен в таблице 4.

Таблица 4.

Раздел	Часов						Форма контроля
	Всего	Лек.	Лаб.	В т. Ч. ДОТ	Сам. Раб.	Стажи-ровка	
Раздел 1. Основы синтаксиса Python							
1.1. Типы данных, операции, операторы, функции, структуры данных Python	3		3	3			
Раздел 2. Первичный анализ больших данных с Pandas							
2.1. Загрузка, подготовка, первичный анализ данных с Pandas	3		3	3			
Раздел 3. Визуальный анализ больших данных с Seaborn							
3.1. Визуальный анализ: функции plot, pairplot, distplot, jointplot, boxplot, heatmap, hist, corr_matrix, violin_plot, t-SNE-представление данных, библиотека plotly	3		3	3			
Раздел 4. Задача классификации							
4.1. Задача классификации, план решения задачи классификации, построение дерева решений	1		1	1			
4.2. Метод ближайших соседей, кросс-валидация	1		1	1			
4.3. Случайный лес, измерение качества модели классификации	1		1	1			
Раздел 5. Задача регрессии							
5.1. Задача регрессии, план решения задачи регрессии, линейная модель	1		1	1			
5.2. Модели регрессии с	1		1	1			

регуляризацией							
5.3. Измерение качества модели регрессии	1		1	1			
Раздел 6. Обучение без учителя							
6.1. Понижение размерности. Метод главных компонент	1		1	1			
6.2. Задача кластеризация. Метод k-means	1		1	1			
Итоговая аттестация	1		1	1			Тест
Итого	18		18	18			

2.3. Перечень лабораторных работ

Перечень лабораторных работ приведен в таблице 5.

Таблица 5.

Номер темы	Наименование лабораторной работы
1.1.	Основы синтаксиса Python
2.1.	Загрузка, подготовка, первичный анализ данных с Pandas
3.1.	Визуальный анализ данных
4.1.	Построение дерева решений
4.2.	Метод ближайших соседей, кросс-валидация
4.3.	Случайный лес, измерение качества модели классификации
5.1.	Линейная модель регрессии
5.2.	Модели регрессии с регуляризацией
5.3.	Измерение качества модели регрессии
6.1.	Понижение размерности. Метод главных компонент
6.2.	Задача кластеризация. Метод k-means

2.4. Программа учебного курса

Программа представлена в таблице 6.

Таблица 6.

Результаты (компетенции)	Должен уметь	Темы и виды занятий	Должен знать
Способность разрабатывать прикладные решения	У-5	Лабораторная 1.1	3-4
Способность работать с большими массивами данных: проводить анализ и обработку	У-1, У-2, У-4	Лабораторные разделы 2,3	3-1
Способность строить модели для решения основных задач машинного обучения, проверять качество моделей машинного обучения.	У-3	Лабораторные разделы 4 - 6	3-2, 3-3

2.5. Календарный учебный график

Календарный учебный график представляется в форме расписания при наборе группы на обучение в зависимости от вида обучения (с отрывом или без отрыва от исполнения обучающимися своих функциональных обязанностей).

3. ОЦЕНКА КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ

Оценка качества освоения программы включает два этапа:

1. Оценка качества выполнения работ при изучении разделов 1 - 6. Критерии положительной оценки: подготовка и анализ данных, а также построение моделей машинного обучения не должны содержать критических ошибок. В качестве критериев оценки могут быть выбраны отдельные критерии, предложенные в техническом описании компетенции «Машинное обучение и большие данные» стандартов Worldskills Russia.

2. Итоговая аттестация включает выполнение теста.

4. МАТЕРИАЛЬНО-ТЕХНИЧЕСКИЕ УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ

Занятия проводятся на материально-технической базе мастерской «Машинное обучение и большие данные» ФГБОУ ВО «СГУ им. Питирима Сорокина». Для проведения занятий имеется следующее оборудование.

Мастерская «Машинное обучение и большие данные»

Компьютер (1 место для обучающегося)

Процессор: Intel Core i5-8500, ОЗУ: 16 ГБ, накопитель: 240 ГБ SSD, 2 монитора: 23,8”.

Ноутбук

Lenovo V330-15IKB (процессор Intel Core i5-8250U, память 8 ГБ, накопитель 256 ГБ SSD, экран 15,6”, ОС Windows 10 Pro)

Программное обеспечение (на 1 компьютер)

- Операционная система: Microsoft Windows 10 Professional
- Adobe Acrobat Reader DC
- ПО для архивации: 7-Zip 19.00
- Web Browser: Firefox Developer Edition 69.0.3
- Web Browser: Chrome 77.0
- Офисный пакет: Microsoft Office 2019 Professional Plus
- Программное обеспечение Anaconda 2019.07 For Windows Python 3.6, включая следующие компоненты:
 - Jupyter Notebook;
 - библиотеки NumPy, pandas, Seaborn, Scikit-learn;
 - Buildozer;
 - PyQt;
 - Pillow;
 - pymysql.
- PyCharm Community Edition 2019.2.3
- Текстовый редактор: Notepad ++ 7.8
- ПО Git: Smartgit 19.1.4

Интерактивная панель

Интерактивная панель 65" на мобильной стойке TEACHTOUCH 3.5 65".

Доска

- Белая доска для маркеров.

Примечание. Версии и состав описанного аппаратного и программного обеспечения могут быть изменены, если это не повлияет на возможность реализации программы.

Дистанционная часть программы размещается в среде Moodle на площадке дистанционных курсов СГУ им. Питирима Сорокина (<http://mooc.syktu.ru>, раздел «Молодые профессионалы»).

5. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ПРОГРАММЫ

1. Крутиков, В.Н. Анализ данных / 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. [Электронный ресурс] – URL: <http://biblioclub.ru/index.php?page=book&id=278426>.

2. Замятин, А.В. Интеллектуальный анализ данных / А.В. Замятин ; Министерство образования и науки Российской Федерации, Национальный исследовательский Томский государственный университет. – Томск : Издательский Дом Томского государственного университета, 2016. – 119 с. : табл., граф., схем. [Электронный ресурс] – URL: <http://biblioclub.ru/index.php?page=book&id=464472>.

3. Кухаренко, Б.Г. Интеллектуальные системы и технологии / Б.Г. Кухаренко ; Министерство транспорта Российской Федерации, Московская государственная академия водного транспорта. – Москва : Альтаир : МГАВТ, 2015. – 115 с. : табл., граф., ил. [Электронный ресурс] – URL: <http://biblioclub.ru/index.php?page=book&id=429758>.

4. Гусаров, В.М. Общая теория статистики / В.М. Гусаров, С.М. Проява. – 2-е изд., перераб. и доп. – Москва : ЮНИТИ-ДАНА, 2012. – 207 с. : ил. [Электронный ресурс] – URL: <http://biblioclub.ru/index.php?page=book&id=447902>.

5. Чубукова, И.А. Data Mining / И.А. Чубукова. – 2-е изд., испр. – Москва : Интернет-Университет Информационных Технологий, 2008. – 383 с. [Электронный ресурс] – URL: <http://biblioclub.ru/index.php?page=book&id=233055>.

6. Серегин, М.Ю. Интеллектуальные информационные системы / М.Ю. Серегин, М.А. Ивановский, А.В. Яковлев ; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Тамбовский государственный технический университет». – Тамбов : Издательство ФГБОУ ВПО «ТГТУ», 2012. – 205 с. : ил. [Электронный ресурс] – URL: <http://biblioclub.ru/index.php?page=book&id=277790>.

Дополнительный перечень ресурсов информационно-телекоммуникационной сети «Интернет», полезных для освоения программы

1. Документация библиотеки scikit-learn // URL: <https://scikit-learn.org/stable/documentation.html>.

2. Руководство для начинающих датасайнтистов // Контексты по машинному обучению

и анализу данных. URL: <https://mlbootcamp.ru/article/tutorial/>.

3. Курс лекций К. В. Воронцова по машинному обучению // URL: <http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%BE>.

4. Открытый курс машинного обучения // Блог компании Open Data Science. URL: <https://habr.com/ru/company/ods/blog/322626/>.

5. Официальный сайт Python // URL: <https://www.python.org/>.

6. Сайт библиотеки seaborn // URL: <https://seaborn.pydata.org/>.

7. Сайт библиотеки pandas // URL: <https://pandas.pydata.org/>.

8. Система организации конкурсов по исследованию данных Kaggle // URL: <https://www.kaggle.com/>.

9. Курс Артема Шевлякова по машинному обучению на Stepik. // URL: <https://stepik.org/course/8057>.

6. КАДРОВОЕ ОБЕСПЕЧЕНИЕ ПРОГРАММЫ

1. Бабенко Виктор Васильевич, к. г.-м. н., доцент кафедры информационных систем.
2. Гольчевский Юрий Валентинович, к. ф.-м. н., доцент, заведующий кафедрой информационных систем.
3. Ермоленко Андрей Васильевич, к. ф.-м. н., доцент, заведующий кафедрой прикладной математики и информационных технологий в образовании.
4. Котелина Надежда Олеговна, к. ф.-м. н., доцент кафедры прикладной математики и информационных технологий в образовании.
5. Миронов Владимир Валериевич, к. ф.-м. н., доцент, директор института точных наук и информационных технологий.
6. Гуляева Сабина Тахировна, старший преподаватель кафедры информационных систем.

7. СОСТАВИТЕЛЬ ПРОГРАММЫ

Котелина Надежда Олеговна, к. ф.-м. н., доцент кафедры прикладной математики и ИТО, руководитель Мастерской «Программные решения для бизнеса».