

Автономная некоммерческая организация «Научно-производственный центр испытаний и компетенций в сфере развития беспилотных авиационных систем и средств защиты от их противоправного применения «Крылья Сахалина»

УТВЕРЖДАЮ

Генеральный директор АНО «НПЦ Крылья Сахалина»



В.А. Грицаенко

« » ноября 2024 г.

**Дополнительная профессиональная программа
– программа повышения квалификации
«Введение в анализ данных»
67 ч.**

**Авторы-составители:
специалисты АНО «НПЦ Крылья
Сахалина»**

СОДЕРЖАНИЕ

Пояснительная записка.....	3
1. Общая характеристика программы.....	5
2. Структура и содержание программы.....	8
3. Форма аттестации и оценочные материалы.....	12
4. Организационно-педагогические условия реализации программы...	10

Пояснительная записка

Актуальность программы обусловлена тем, что ежедневно производятся и сохраняются огромные объемы информации и данных.

Часто их источником являются системы автоматического обнаружения, сенсоры и инструменты ученых. Также мы создаём их самостоятельно, даже не осознавая: снимаете деньги с банковского счета, осуществляя покупку, делая запись в блоге или социальной сети.

Анализ данных также требует сложной математики для работы с информацией. В ней необходимо разбираться, как минимум понимая, что вы делаете. Знакомство с основными методами статистики также необходимы, потому что все применяемые методы основаны на них. Как компьютер предлагает инструменты для анализа данных, так и статистика — концепции, которые составляют основу дисциплины.

Целями программы **«Введение в анализ данных»** сформировать понимание возможностей и ограничений Data Science, рисков использования моделей, а также дать возможность прожить опыт Data Science “от и до”. Изучаются основные модели (линейные, метрические, логические), подходы к их обучению и методы обработки данных.

Все рассматриваемые в Программе принципы и методики организации образовательной деятельности тесно связаны с образовательной политикой РФ и федеральными государственными образовательными стандартами.

Реализация программы повышения квалификации направлена на совершенствование и (или) получение новой компетенции, необходимой для профессиональной деятельности, и (или) повышение профессионального уровня в рамках уже имеющейся квалификации; формирование умений и знаний для выполнения работником соответствующих трудовых функций и трудовых действий, позволяющих самостоятельно выполнять комплекс работ, связанный с осуществляемым видом деятельности. Профессиональные компетенции, полученные в результате обучения по программе, могут стать базой для инициации нового проекта по повышению эффективности с

использованием данных.

Нормативно-правовую основу разработки программы составляет:

1. Федеральный закон от 29.12.2012 №273-ФЗ «Об образовании в Российской Федерации» (с изменениями и дополнениями).
2. Приказ Министерства образования и науки РФ от 23.08.2017 №808 «Об утверждении федерального государственного образовательного стандарта высшего образования - бакалавриат по направлению подготовки 02.03.02 Фундаментальная информатика и информационные технологии» (с изменениями и дополнениями) Редакция с изменениями N 1456 от 26.11.2020.
3. Приказ Минтруда России от 09.07.2021 №462н «Об утверждении профессионального стандарта «Специалист по моделированию, сбору и анализу данных цифрового следа» (Зарегистрировано в Минюсте России 30.07.2021 N 64502).
4. Приказ Минобрнауки РФ от 01.07.2013 №499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам».
5. Приказ Министерства образования и науки РФ от 23.08.2017 №816 «Об утверждении Порядка применения организациями, осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ».
6. Методические рекомендации по разработке ОПОП и ДПО с учетом соответствующих профессиональных стандартов от 22.01.2015 №ДЛ-1/05 вн. Министерства образования и науки Российской Федерации.
7. Методические рекомендации по организации итоговой аттестации при реализации ДОП от 30.03.2015 №АК-820/6. Министерства образования и науки Российской Федерации.

1. «ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ»

1.1. Цель реализации Программы

Совершенствование профессиональных компетенций, необходимых для проведения анализа данных цифрового следа в соответствии с моделью деятельности человека (группы людей) и ИКС.

Задачи программы

- расширить профессиональную компетенцию, необходимую для проведения анализа данных цифрового следа в соответствии с моделью деятельности человека (группы людей) и ИКС;
- сформировать навык анализа данных для принятия взвешенных управленческих решений;
- научить применять полученные знания на практике в различных областях;
- развить мягкие навыки для работы в команде.

1.2. Категория слушателей: настоящая программа предназначена для лиц, осуществляющих массовое обучение.

1.3. Требования к профилю предшествующего образования: уровень образования – получающие или имеющие среднее профессиональное образование или высшее образование – бакалавриат.

1.4. Форма обучения: очно-заочная с применением ДОТ.

1.5. Режим занятий: 7 раз в неделю по 6, 7, 8 академических часов, продолжительность академического часа – 45 минут, период обучения – 10 дней.

1.6. Трудоемкость обучения: 67 ч., из них:

Аудиторные 27 ч. (интерактивная лекция – 10 ч.; практическая работа – 17 ч); самостоятельная работа – 40 ч.

Совершенствуемые компетенции

в соответствии с ФГОС ВО по направлению подготовки 02.03.02 Фундаментальные информатика и информационные технологии (уровень бакалавриата), утв. Приказом Минобрнауки РФ от 23 августа 2017 г. N 808)

№ п/п	Компетенции	Фундаментальные информатика и информационные технологии 02.03.02
		Бакалавриат
		Код компетенций
1.	Способен применять компьютерные/суперкомпьютерные методы, современное программное обеспечение, в том числе отечественного происхождения, для решения задач профессиональной деятельности	ОПК-2
2.	Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	ОПК-6

1.7. Планируемые результаты обучения

№ п/п	Знать - Уметь	Фундаментальные информатика и информационные технологии 02.03.02
		Бакалавриат
		Код компетенций
1.	Знать <ul style="list-style-type: none"> – Общие принципы построения баз данных – Порядок сбора данных – Подходы в работе со средствами хранения и передачи информации – Алгоритм сбора данных в единую базу (датасет) Уметь <ul style="list-style-type: none"> – Собирать и объединять данные из разрозненных источников в единую базу (датасет) 	ОПК-2
2.	Знать <ul style="list-style-type: none"> – Алгоритмы очистки данных цифрового следа – Проблемы очистки данных – Алгоритмы поиска аномалий – Алгоритмы регрессионного анализа данных Уметь <ul style="list-style-type: none"> – Работать со значениями и приведением их к общим показателям – Работать с алгоритмами очистки данных – Работать с пропущенными данными и аномальными значениями 	ОПК-2

3.	<p>Знать</p> <ul style="list-style-type: none"> – Методы предиктивной аналитики оттока клиентов – Логистическую регрессию и предобработку данных – Алгоритмы построения предиктивной модели и осуществления прогноза – Кросс-валидацию как метод оценки аналитической модели <p>Уметь</p> <ul style="list-style-type: none"> – Осуществлять анализ оттока клиентов – Использовать методы предиктивной аналитики – Осуществлять оценку аналитической модели путем кросс-валидации – Интерпретировать результаты прогнозирования 	ОПК-2
4.	<p>Знать</p> <ul style="list-style-type: none"> – Понятия и алгоритмы дескриптивного анализа – Цель дескриптивного анализа – Алгоритмы разведочного анализа данных – Алгоритмы построения гистограмм и графиков <p>Уметь</p> <ul style="list-style-type: none"> – Использовать простые дескриптивные статистики – Выявлять дополнительные проблемы с качеством данных и разрешать их – Визуализировать полученные данные используя гистограммы и графики 	ОПК-6
5.	<p>Знать</p> <ul style="list-style-type: none"> – Формулировки статистических гипотез – Методы определения статистического критерия для проверки гипотез – Этапы проверки статистических гипотез – Алгоритмы проверки гипотез <p>Уметь</p> <ul style="list-style-type: none"> – Проводить сравнительный анализ для проверки гипотез 	ОПК-2

2. «СТРУКТУРА И СОДЕРЖАНИЕ ПРОГРАММЫ»

2.1. Учебный (тематический) план

№п/п	Наименование тем	Всего часов	В том числе:			Форма контроля/ аттестации
			Аудиторные ч.		СРС, ч.	
			Интерактивные лекции	Практические и семинарские занятия	Самостоятельные работы	
Базовая часть						
1.1.	Сбор данных	13	2	3	8	Проект №1
Профильная часть (предметно-методическая)						
2.1.	Очистка данных	13	2	3	8	Проект №2
2.2.	Предиктивный анализ	13	2	3	8	Проект №3
2.3.	Дескриптивный и разведочный анализ данных	13	2	3	8	Проект №4
2.4	Статистическая проверка гипотез	13	2	3	8	Проект №5
	Итоговая аттестация	2	-	2	-	Командный проект
	ИТОГО:	67	10	17	40	

2.2. Календарный учебный график

Компонент программы		Количество часов по темам										Всего часов
		1 ^я день	2 ^я день	3 ^я день	4 ^я день	5 ^я день	6 ^я день	7 ^я день	8 ^я день	9 ^я день	10 ^я день	
Тема 1.1.	ауд.	3	2									5
	с.р.	4	4									8
Тема 2.1.	ауд.			3	2							5
	с.р.			4	4							8
Тема 2.2.	ауд.					3	2					5
	с.р.					4	4					8
Тема 2.3.	ауд.							3	2			5
	с.р.							4	4			8
Тема 2.4.	ауд.									3	2	5
	с.р.									4	4	8
Итоговая аттестация	ауд.										2	2
Количество часов		7	6	7	6	7	6	7	6	7	8	67

2.3.Рабочая программа

Содержание тем

Темы	Виды учебных занятий/работ, час.	Содержание
<i>Базовая часть.</i>		
Тема 1.1. Сбор данных	Интерактивная лекция, 2 ч.	Общие принципы построения баз данных. Порядок сбора данных. Алгоритм сбора данных в единую базу (датасет).
	Самостоятельная работа, 8 ч.	Слушатели знакомятся с методами и средствами защиты информации. Изучают подходы в работе со средствами хранения и передачи информации для работы с цифровым следом.
	Практическая работа, 3 ч.	Проект №1 «Сбор данных». Слушатели собирают и объединяют данные из разрозненных источников в единую базу (датасет), описывают значение, указывают названия полей и в каких единицах измерения находятся данные.
<i>Профильная часть (предметно-методическая)</i>		
Тема 2.1. Очистка данных	Интерактивная лекция, 2 ч.	Алгоритмы очистки данных цифрового следа (корректировка, подсказка, автоматизация / уменьшение объема ручной работы, поиск дубликатов). Проблемы очистки данных. Алгоритмы поиска аномалий. Алгоритмы регрессионного анализа данных.
	Самостоятельная работа, 8 ч.	Слушатели изучают задачи обнаружения аномалий. Знакомятся с основными видами регуляризации. Изучают методы проверки целостности полученных данных.
	Практическая работа, 3 ч.	Проект №2 «Очистка данных». Слушатели работают со значениями и приведением их к общим показателям (вид, пол, город, названия улиц). Работают с пропущенными данными и аномальными значениями.
Тема 2.2. Предиктивный анализ	Интерактивная лекция, 2 ч.	Формулирование проблемы оттока клиентов. Методы предиктивной аналитики оттока клиентов. Логистическая регрессия и предобработка данных. Алгоритмы построения предиктивной модели и осуществления прогноза. Интерпретация результатов прогнозирования. Разбиение на признаки и целевую переменную. Разбиение на train и test. Кросс-

		валидация как метод оценки аналитической модели. Дополнительные метрики оценки качества: precision, recall, AUC.
	Самостоятельная работа, 8 ч.	Слушатели анализируют методы моделирования и прогнозирования оттока клиентов на примере предложенных кейсов. Изучают подходы в работе с различными способами интерпретации результатов анализа.
	Практическая работа, 3 ч.	Проект №3. «Предиктивный анализ». Слушатели осуществляют анализ оттока клиентов используя методы предиктивной аналитики с построением предиктивной модели. Осуществляют оценку аналитической модели путем кросс-валидации и интерпретируют результаты прогнозирования.
Тема 2.3. Дескриптивный и разведочный анализ данных	Интерактивная лекция, 2 ч.	Понятия и алгоритмы дескриптивного анализа. Цель дескриптивного анализа. Алгоритмы разведочного анализа данных. Применение графиков для визуализации и изучения данных. Алгоритмы построения гистограмм и графиков.
	Самостоятельная работа, 8 ч.	Слушатели знакомятся с возможностями и преимуществами дескриптивных статистик и применения визуализации данных на основе кейсов.
	Практическая работа, 3 ч.	Проект №4. «Дескриптивный и разведочный анализ данных». Слушатели, используя простые дескриптивные статистики, выявляют дополнительные проблемы с качеством данных и разрешают их. Визуализируют полученные данные используя гистограммы и графики.
Тема 2.4. Статистическая проверка гипотез	Интерактивная лекция, 2 ч.	Формулировка статистических гипотез. Выбор уровня значимости. Работа со статистическими данными. Методы определения статистического критерия для проверки гипотез. Этапы проверки статистических гипотез. Алгоритмы проверки гипотез.
	Самостоятельная работа, 8 ч.	Слушатели изучают подходы в работе со статистическими критериями для проверки гипотез. Анализируют примеры проверки гипотез на основе кейсов.

	Практическая работа, 3 ч.	Проект №5. «Статистическая проверка гипотез» Слушатели проводят сравнительный анализ для проверки гипотез, представленных в модели деятельности человека (группы людей) и ИКС.
Итоговая аттестация	Практическая работа, 2 ч.	Командный проект «Монетизация данных». Используя предиктивный анализ, нужно предложить идею проекта по монетизации данных, а также варианты улучшения одного из существующих процессов в вашей компании. Оценить финансовый эффект от модели, экономику проекта, трудозатраты, необходимые ресурсы, какой необходимой точности нужно добиться, и с кем из стейкхолдеров важно согласовать.

3. «ФОРМЫ АТТЕСТАЦИИ И ОЦЕНОЧНЫЕ МАТЕРИАЛЫ»

Оценка уровня сформированности профессиональных компетенций слушателей включает промежуточную и итоговую аттестацию.

3.1.В рамках **промежуточной аттестации** оцениваются проекты слушателей:

Раздел УТП	Форма промежуточной аттестации	Оценка	Форма защиты
1.1.	Проект № 1 (см. ниже)	зачтено / не зачтено	очная
2.1.	Проект № 2 (см. ниже)	зачтено / не зачтено	очная
2.2.	Проект № 3 (см. ниже)	зачтено / не зачтено	очная
2.3.	Проект № 4 (см. ниже)	зачтено / не зачтено	очная
2.4.	Проект №5 (см. ниже)	зачтено / не зачтено	очная
И/О	Защита командного проекта	зачтено / не зачтено	очная

Промежуточная аттестация по **теме № 1.1.** «Сбор данных» представляет собой выполнение **проекта №1:** «Сбор данных».

Требование к выполнению проекта:

Слушатели собирают и объединяют данные из разрозненных источников в единую базу (датасет), описывают значение, указывают названия полей и в каких единицах измерения находятся данные.

Промежуточная аттестация по **теме № 2.1.** «Очистка данных» представляет собой выполнение **проекта №2:** «Очистка данных».

Требование к выполнению проекта:

Слушатели работают со значениями и приведением их к общим показателям (вид, пол, город, названия улиц), а также работают с пропущенными данными и аномальными значениями с целью улучшения полученного датасета.

Промежуточная аттестация по **теме № 2.2.** «Предиктивный анализ» представляет собой выполнение **проекта №3:** «Предиктивный анализ».

Требование к выполнению проекта:

Слушатели осуществляют анализ оттока клиентов используя методы предиктивной аналитики с построением предиктивной модели. Осуществляют оценку аналитической модели путем кросс-валидации и интерпретируют результаты прогнозирования.

Промежуточная аттестация по **теме № 2.3.** «Дескриптивный и разведочный анализ данных» представляет собой выполнение **проекта №4:** «Дескриптивный и разведочный анализ данных».

Требование к выполнению проекта:

Слушатели, используя простые дескриптивные статистики, выявляют дополнительные проблемы с качеством данных и разрешают их. Визуализируют полученные данные используя гистограммы и графики.

Промежуточная аттестация по теме № 2.4. «Статистическая проверка гипотез» представляет собой выполнение проекта №5: «Статистическая проверка гипотез».

Требование к выполнению проекта:

Слушатели проводят сравнительный анализ для проверки гипотез, представленных в модели деятельности человека (группы людей) и ИКС, используя алгоритмы проверки гипотез.

Итоговая аттестация представляет собой выполнение командного проекта «Монетизация данных».

Требование к выполнению проекта:

Используя предиктивный анализ, нужно предложить идею проекта по монетизации данных, а также варианты улучшения одного из существующих процессов в вашей компании. Оценить финансовый эффект от модели, экономику проекта, трудозатраты, необходимые ресурсы, какой необходимой точности нужно добиться, и с кем из стейкхолдеров важно согласовать.

3.2. Критерии оценивания и оценивание:

Проектная работа слушателя или группы считается зачтенной, если:

- прошла две «peer-to-peer» проверки другими слушателями программы;
- получена оценка по итогу двух проверок более 50%.

Проверка работы слушателя или группы считается зачтенной:

- если при выполнении «peer-to-peer» проверки присутствовал проверяющий и проверяемый в режиме онлайн;
- если заполнен чек-лист и выставлена итоговая оценка.

4. «ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ»

4.1. Материально-техническое обеспечение

Для обеспечения процесса обучения требуются:

- учебная аудитория;
- высокоскоростной доступ в Интернет;
- маркерная доска или мультимедийный экран;
- рабочие станции (Lenovo) по количеству обучающихся;
- платформа «Школы 21».

4.2. Учебно-методическое и информационное обеспечение программы

Основная литература

1. Миркин, Б. Г. Введение в анализ данных: учебник и практикум / Б. Г. Миркин. — М.: Издательство Юрайт, 2018. — 174 с. — (Авторский учебник). — ISBN 978-5-9916-5009-0. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/413060>.
2. Анализ данных: учебник для академического бакалавриата / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — М.: Издательство Юрайт, 2018. — 490 с. — (Бакалавр. Академический курс). — ISBN 978-5-534-00616-2. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/412967>.
3. Крутиков, В.Н. Анализ данных: учебное пособие / В.Н. Крутиков, В.В. Мешечкин; Кемеровский государственный университет. — Кемерово: Кемеровский государственный университет, 2014. — 138 с. — Режим доступа: по подписке. — URL: <https://biblioclub.ru/index.php?page=book&id=278426>. — Текст: электронный.
4. Каган, Е.С. Прикладной статистический анализ данных: учебное пособие / Е.С. Каган; Кемеровский государственный университет. — Кемерово: Кемеровский государственный университет, 2018. — 235 с. : ил., табл. — Режим доступа: по подписке. — URL: <https://biblioclub.ru/index.php?page=book&id=573550>. — Текст: электронный.
5. Себер Дж. Линейный регрессионный анализ. - М.: Мир, 1980. 456с. — URL: https://scask.ru/h_book_lra.php?id=1.
6. Кремер, Н. Ш. Эконометрика: учебник и практикум для академического бакалавриата / Н. Ш. Кремер, Б. А. Путко; под редакцией Н. Ш. Кремера. — 4-е изд., испр. и доп. — Москва: Издательство Юрайт, 2017. — 354 с. — (Бакалавр. Академический курс). — ISBN 978-5-534-02760-0. — Текст: электронный // ЭБС

Юрайт [сайт]. — URL: <https://urait.ru/bcode/401922>.

7. Яковлев, В. Б. Статистика. Расчеты в Microsoft Excel: учебное пособие для ву-зов / В. Б. Яковлев. — 2-е изд., испр. и доп. — Москва: Издательство Юрайт, 2017. — 353 с. — (Университеты России). — ISBN 978-5-534-01672-7. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/400278>.

4.3.Кадровые условия реализации программы

Требования к квалификации преподавателей, реализующих настоящую дополнительную профессиональную программу – программу повышения квалификации: высшее профессиональное образование или среднее профессиональное образование по направлению подготовки «Образование и педагогика» или в области, соответствующей преподаваемому предмету, без предъявления требований к стажу работы либо высшее профессиональное образование или среднее профессиональное образование и дополнительное профессиональное образование по направлению деятельности в образовательном учреждении без предъявления требований к стажу работы.