

УТВЕРЖДАЮ

Директор НОУ «Академия ИНГМ»

 В.В. Лавров

2021 г.



ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА

ПОВЫШЕНИЯ КВАЛИФИКАЦИИ

«СОВРЕМЕННЫЕ МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И МАШИННОГО ОБУЧЕНИЯ ДЛЯ СПЕЦИАЛИСТОВ НЕФТЕГАЗОВОЙ ОТРАСЛИ»

Разработал:
преподаватель В.А. Молчан

г. Томск

СОДЕРЖАНИЕ

1. ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ	
1.1. Нормативные основания разработки программы	3
1.2. Цель	3
1.3. Задачи	3
1.4. Планируемые результаты обучения.....	3
1.5. Характеристика профессиональной деятельности слушателей	3
2. СОДЕРЖАНИЕ ПРОГРАММЫ	
2.1. Учебный план.....	4
2.2. Рабочие программы (тематическое содержание) модулей	5
2.3. Календарный учебный график	7
3. ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ	
3.1. Категория слушателей	7
3.2. Технологии и методы обучения.....	8
3.3. Учебно-методическое обеспечение.....	8
3.4. Материально-техническое обеспечение.....	8
3.5. Кадровое обеспечение.....	8
3.6. Информационное обеспечение.....	8
3.7. Электронные ресурсы.....	8
3.8. Документ о квалификации.....	9
4. ОЦЕНКА КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ	
4.1. Формы аттестации.....	9
4.2. Оценочные материалы.....	9
4.3. Оценка результатов аттестации	12

ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ

Нормативные основания разработки программы:

1. Федеральный закон № 273-ФЗ «Об образовании в Российской Федерации» от 29.12.2012 г.
2. Приказ Министерства образования и науки РФ № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам» от 01.07.2013 г.
3. Профессиональные стандарты Код 19 «Добыча, переработка, транспортировка нефти и газа».
4. Проекты примерных образовательных программ по направлениям бакалавриата 210000 «Прикладная геология, горное дело, нефтегазовое дело и геодезия».
5. ФГОС ВО по направлениям бакалавриата и магистратуры 210000 «Прикладная геология, горное дело, нефтегазовое дело и геодезия».

Цель:

совершенствование профессиональных компетенций специалистов нефтяной и газовой отрасли в практическом применении методов искусственного интеллекта и машинного обучения в области разработки нефтяных и газовых месторождений, оптимизации добычи, геологического моделирования и петрофизики для решения профессиональных задач.

Задачи:

- изучить теорию основных методов искусственного интеллекта и машинного обучения;
- освоить практические навыки по применению языка программирования Python и основных библиотек для подготовки, визуализации и анализа данных, построения моделей на основе методов машинного обучения;
- научиться применять на практике основные алгоритмы машинного обучения такие как регрессия, классификация, методы кластерного анализа и поиска аномалий, а также нейронные сети и методы глубокого обучения.

Планируемые результаты обучения:

усовершенствованные профессиональные компетенции, выраженные в знаниях и способностях:

- применять на практике современное программное обеспечение и основные библиотеки для решения задач методами машинного обучения;
- оптимизировать выполнение наиболее трудоёмких задач по проектированию систем разработки нефтяных и газовых месторождений, геологическому моделированию, интерпретации геофизических данных;
- выбирать наиболее подходящие методы машинного обучения для решения определённых проблем и применять методы машинного обучения на практике.

Характеристика профессиональной деятельности слушателей:

Область профессиональной деятельности слушателей, освоивших программу курса повышения квалификации, включает реализацию и управление технологическими процессами и производством, методологию и методы проектирования и конструирования,

научные исследования и разработки в сегменте топливной энергетики, в т.ч. освоение месторождений, транспортирование и хранение углеводов, исследование недр и поверхности Земли, рациональное использование и охрана земельных и углеводородных ресурсов и др.

Объектами профессиональной деятельности слушателей являются технологические процессы и устройства для строительства, ремонта, восстановления, добычи, промышленного контроля, транспортирования, хранения и сбыта нефти, нефтепродуктов и сжиженных газов, поверхность и недра Земли, геодинамические явления и процессы, территориально-административные образования, информационные системы и инновационные технологии и др.

Виды профессиональной деятельности слушателей: производственно-технологическая, организационно-управленческая, экспериментально-исследовательская, проектная, проектно-изыскательская, научно-исследовательская.

Основной профессиональной аудиторией курса являются инженеры по разработке нефтяных и газовых месторождений, инженеры по добычи нефти и газа, геологи и геофизики.

СОДЕРЖАНИЕ ПРОГРАММЫ

Учебный план дополнительной профессиональной программы определяет перечень, трудоемкость, последовательность и распределение учебных модулей, иных видов учебной деятельности обучающихся и формы аттестации.

Учебный план:

№	Наименование модулей	Количество часов			Форма аттестации
		Всего	в том числе:		
			лекционные занятия	практические занятия	
1	Введение в инструменты машинного обучения. Краткий курс по Python. Подготовка данных для анализа (с использованием Pandas and SQL). Визуализация данных	8	4	4	Текущий контроль
2	Численная оптимизация. Краткий курс по статистике. Введение в анализ данных. Оценка неопределённости и методы принятия решений	8	4	4	Текущий контроль
3	Введение в машинное обучение. Методы понижения размерности данных. Методы кластерного анализа. Методы поиска аномалий в данных	8	4	4	Текущий контроль
4	Основные концепции машинного обучения. Методы регрессии. Калибровка моделей машинного обучения	8	4	4	Текущий контроль
5	Обзор методов классификации. Нейронные сети и методы глубокого обучения. Несбалансированные наборы	6	4	2	Текущий контроль

	данных. Интерпретация моделей машинного обучения				
6	Итоговая аттестация	2	-	2	Тестирование
	ИТОГО	40	20	20	

Рабочие программы (тематическое содержание) модулей:

<p>Тема 1:</p> <ul style="list-style-type: none"> - Введение в инструменты машинного обучения - Краткий курс по Python - Подготовка данных для анализа (с использованием Pandas and SQL) - Визуализация данных 	<p>Практические упражнения:</p> <ul style="list-style-type: none"> - Анализ и визуализация данных по добыче - Подготовка данных для расчётов методом материального баланса - Проверка качества адаптации гидродинамической модели - Визуализация каротажных кривых
<p>Слушатели научатся:</p> <ul style="list-style-type: none"> • применять язык программирования Python и основные библиотеки по машинному обучению для решения различных задач из области добычи нефти и газа; • создавать мощные программные инструменты для анализа данных по добыче, представленных в различных форматах (текстовые файлы или базы данных), которые могут быть использованы для анализа добычи по залежам нефти и газа различного типа и размера; • в максимально короткие сроки подготовить данные по добыче и давлению в формате PETEX MBAL для расчётов методом материального баланса для сложных залежей, например, залежей многоблочного строения с различными уровнями ВНК и ГНК; • быстро и эффективно анализировать большое количество результатов гидродинамических расчётов для оценки качества адаптации гидродинамической модели и анализа прогнозных показателей; • с минимальными затратами времени строить графики высокого качества на основе различных промысловых данных (по добыче и давлению, каротажному материалу, и т.д.), что позволяет значительно упростить процесс анализа разработки, а также получить готовый графический; • материал для дальнейшего использования в презентациях и отчётах. 	

<p>Тема 2:</p> <ul style="list-style-type: none"> - Численная оптимизация - Краткий курс по статистике - Введение в анализ данных - Оценка неопределённости и методы принятия решений 	<p>Практические упражнения:</p> <ul style="list-style-type: none"> - Кривые падения добычи - Подготовка физико-химических данных нефти и газа для гидродинамического моделирования - Подсчёт запасов нефти и газа вероятностными методами - Перемасштабирование геологической модели - Оптимизация заводнения
<p>Слушатели научатся:</p> <ul style="list-style-type: none"> • применять методы численной оптимизации для решения различных задач по проектированию разработки залежей нефти и газа: калибровка кривых падения добычи для понимания механизма работы залежей, увязка градиента пластового давления и данных по физико-химическим свойствам нефти и газа для корректной инициализации гидродинамической модели; • проводить перемасштабирование сетки детальной геологической модели в более грубую сетку гидродинамической модели с полным контролем процесса перемасштабирования и соблюдением баланса между понижением размерности геологической модели и сохранением важных геологических деталей; 	

- выполнять подсчёт запасов углеводородов вероятностными методами с учётом неопределённости входных параметров для оперативной оценки геологических запасов нефти и газа без необходимости создания геологической модели;
- оптимизировать распределение закачки воды и газа между нагнетательными скважинами с целью достижения максимального КИН с использованием оптимального количества расчётов гидродинамического моделирования.

<p>Тема 3:</p> <ul style="list-style-type: none"> - Введение в машинное обучение - Методы понижения размерности данных - Методы кластерного анализа - Методы поиска аномалий в данных 	<p>Практические упражнения:</p> <ul style="list-style-type: none"> - Определение электрофаций на основе каротажных данных - Экспресс-оценка различных вариантов распределения петрофизических свойств в геологической модели - Определение аномальных режимов работы добывающих скважин
--	---

- Слушатели научатся:**
- свободно ориентироваться в терминологии методов машинного обучения и уверенно определять основные технические и бизнес требования для успешного применения методов машинного обучения на практике;
 - выбирать наиболее подходящий метод машинного обучения для решения конкретной задачи из области проектирования разработки месторождений в зависимости от типа проблемы, количества и качества доступных данных и требований к решению;
 - проводить экспресс-оценку вариантов распределения петрофизических свойств в геологической модели для упрощения процесса адаптации гидродинамической модели и уменьшения количества гидродинамических расчётов, а также для эффективной оценки влияния неопределённости геологического строения на прогноз добычи нефти и газа;
 - определять оптимальное число электрофаций для данного месторождения на основе имеющихся каротажных данных для более обоснованного распределения петрофизических свойств в геологической модели;
 - автоматически определять нестационарные режимы работы добывающих скважин для предотвращения возможных аварийных ситуаций и оптимизации работы фонда скважин.

<p>Тема 4:</p> <ul style="list-style-type: none"> - Основные концепции машинного обучения - Методы регрессии - Калибровка моделей машинного обучения 	<p>Практические упражнения:</p> <ul style="list-style-type: none"> - Прогноз добычи для нетрадиционных залежей нефти и газа - Определение давления насыщения пластовых нефтей с помощью методов машинного обучения
--	---

- Слушатели научатся:**
- планировать проект по машинному обучению с целью получения результатов высокого качества, а также их воспроизводимости;
 - применять на практике основные концепции машинного обучения: разбиение данных для обучения и проверки качества модели, кросс-валидация, выбор корректной целевой функции, нахождение баланса между уровнем сложности модели и её стабильностью, калибровка параметров модели;
 - прогнозировать дебиты новых скважин и оптимизировать схемы заканчивания скважин в нетрадиционных залежах нефти и газа без построения гидродинамической модели;
 - на основе доступных исследований физико-химических свойств нефти и газа

создавать модель с использованием методов машинного обучения, позволяющую прогнозировать давление насыщения для залежей, по которым не были проведены необходимые лабораторные исследования;

- автоматически определять оптимальные параметры моделей машинного обучения для упрощения процесса калибровки моделей.

<p>Тема 5:</p> <ul style="list-style-type: none"> - Обзор методов классификации - Нейронные сети и методы глубокого обучения - Дополнительные темы повышенной сложности: Несбалансированные наборы данных Интерпретация моделей машинного обучения 	<p>Практические упражнения:</p> <ul style="list-style-type: none"> - Определение литофаций - Оценка применимости методов повышения нефтеотдачи (МУН)
--	---

<p>Слушатели научатся:</p> <ul style="list-style-type: none"> • представлять результаты, полученные с помощью моделей на основе машинного обучения, широкой аудитории (технической и бизнес) для проведения технической экспертизы и использования при принятии бизнес-решений; • классифицировать литофации на основе каротажных данных в скважинах без наличия керновых данных; • создавать модель для оценки применимости методов повышения нефтеотдачи пласта на основе различных данных (физико-химических исследований нефти и газа, специальных исследований керна и геологических данных), которая позволяет проводить эффективный скрининг МУН для всего портфолио залежей компании.

Календарный учебный график:

№	Наименование модулей	Всего часов	Учебные дни					
			1	2	3	4	5	
1	Введение в инструменты машинного обучения. Краткий курс по Python. Подготовка данных для анализа (с использованием Pandas and SQL). Визуализация данных	8	8					
2	Численная оптимизация. Краткий курс по статистике. Введение в анализ данных. Оценка неопределённости и методы принятия решений	8		8				
3	Введение в машинное обучение. Методы понижения размерности данных. Методы кластерного анализа. Методы поиска аномалий в данных	8			8			
4	Основные концепции машинного обучения. Методы регрессии. Калибровка моделей машинного обучения	8				8		
5	Обзор методов классификации. Нейронные сети и методы глубокого обучения. Несбалансированные наборы данных. Интерпретация моделей машинного обучения	6						8
6	Итоговая аттестация	2						
ИТОГО		40	8	8	8	8	8	8

ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ

Трудоемкость:	40 часов
Форма обучения:	очная
Виды занятий:	лекционные, практические
Формы аттестации:	текущий контроль, итоговое тестирование
Режим занятий:	8 академических часов в день
Срок обучения:	5 дней

Категория слушателей:

Курс повышения квалификации могут пройти лица, имеющие (получающие) высшее или среднее профессиональное образование по соответствующей специальности либо прошедшие профессиональную переподготовку по соответствующему направлению.

Технологии и методы обучения:

лекция, семинар, программирование, решение задач, проведение расчётов, построение графиков, мозговой штурм, тренинг.

Учебно-методическое обеспечение:

презентации по модулям курса, программный код с подробными комментариями и графическим материалом, иллюстрированные заметки на основе обсуждения тем курса в цифровом формате.-

Материально-техническое обеспечение:

аудитория, столы, стулья, ноутбуки с доступом в Интернет, мультимедийный проектор и экран, презентер, аудиоколонки, магнитно-маркерная доска, комплект лицензионного программного обеспечения (MS Power Point, Word, Excel и др.), программное обеспечение Python/Anaconda и доступ к Microsoft Azure Cloud.

Кадровое обеспечение:

Образовательный процесс обеспечивается научно-педагогическими кадрами, имеющими базовое образование, соответствующее профилю программы, и ученую степень или опыт деятельности в соответствующей профессиональной сфере и систематически занимающимися научной и/или научно-методической деятельностью, преподаватели из числа действующих руководителей и ведущих работников профильных организаций.

Информационное обеспечение:

1. Уэс Маккинни “Python и анализ данных” – ДМК Пресс, 2015. - 482 с.
2. Марк Лутц “Изучаем Python”, Символ-Плюс, 2011. - 1280 с.
3. Траск Эндрю "Грокаем глубокое обучение" - Питер, 2019.- 352 с.
4. Жерон Орельен “Прикладное машинное обучение с помощью Scikit-Learn, Keras и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем”, Диалектика-Вильямс, 2020. - 1040 с.

Электронные ресурсы:

1. <https://www.python.org/>
2. <https://www.kaggle.com/>
3. <https://github.com/mralbu/awesome-reservoir-engineering>
4. <https://github.com/softwareunderground/awesome-open-geoscience>
5. <https://www.coursera.org/browse/data-science>

Документ о квалификации:

Лицам, успешно освоившим соответствующую дополнительную профессиональную программу и прошедшим итоговую аттестацию, выдается удостоверение о повышении квалификации. При освоении дополнительной профессиональной программы параллельно с получением среднего и (или) высшего образования удостоверение о повышении квалификации выдается одновременно с получением соответствующего документа об образовании и о квалификации.

ОЦЕНКА КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ

Формы аттестации:

1. Предварительный контроль в форме опроса.
2. Текущий контроль в форме решения и проверки задач, наблюдения за слушателями.
3. Итоговый контроль в форме опроса, решения и проверки задач.

Оценочные материалы:

Тест для предварительного контроля, тест для итогового контроля, комплект задач, комплект упражнений.

Образец задач для текущего контроля (задача 1):

Выполните краткий анализ разработки месторождения нефти (США), приуроченного к низкопроницаемым коллекторам. Среди прочего проанализируйте:

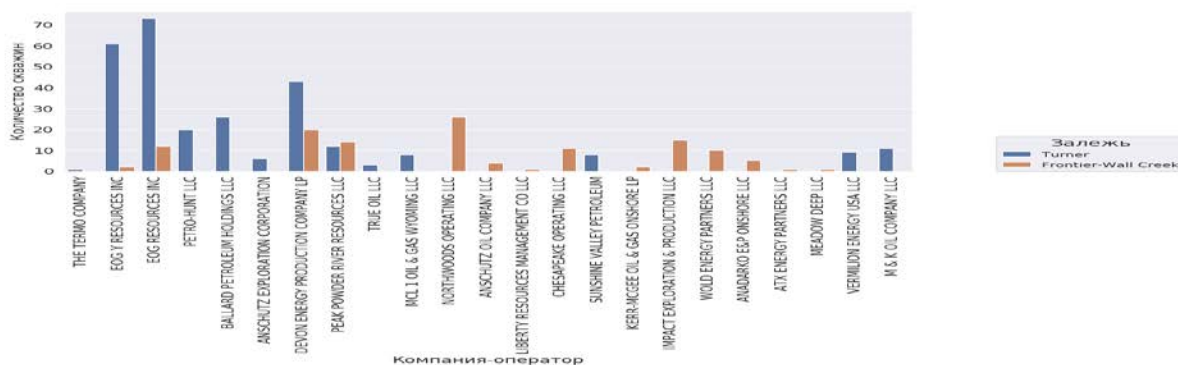
- количество горизонтальных скважин, пробуренных каждой компанией-оператором на месторождении;

```
# Создайте новый график размером 15x6 см
plt.figure(figsize=(15, 6))

# Постройте график с использованием библиотек Seaborn и Pandas
sns.countplot(x='Company',
              hue='WSGS_reservoir_x',
              data=df_horizontal,
              orient='v')

# Добавьте легенду
plt.legend(loc=[1.1, 0], fontsize=14, title='Залежь', title_fontsize=20)
plt.xticks(rotation=90)
plt.xlabel(r'Компания-оператор')
plt.ylabel(r'Количество скважин')
# Сохраните график в отдельный файл
plt.savefig('countplot.png', bbox_inches='tight', pad_inches=0.5)
```

Результат:



- распределение накопленной добычи нефти для каждой залежи и каждой компании-оператора;

```
# Создайте новый график размером 15x6 см
plt.figure(figsize=(15, 6))

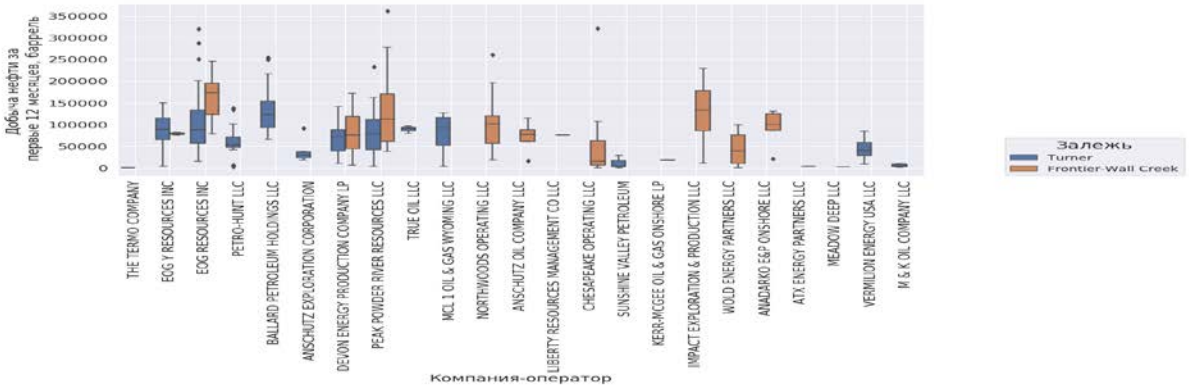
# Постройте график с использованием библиотек Seaborn и Pandas
# Постройте отдельные графики (boxplot) для двух залежей для каждой компании-оператора
g = sns.boxplot(x='Company',
                y='First 12 months oil (bbl)',
                hue='WSGS_reservoir_x',
                data=df_horizontal)

# Отобразите сетку для лучшей визуализации
plt.grid(True)
plt.xticks(rotation=90)

# Добавьте легенду
plt.legend(loc=[1.1, 0], fontsize=14, title='Залежь', title_fontsize=20)
plt.xlabel('Компания-оператор')
plt.ylabel('Добыча нефти за \n первые 12 месяцев, баррель')

# Сохраните график в отдельный файл
plt.savefig('boxplot.png', bbox_inches='tight', pad_inches=0.5)
```

Результат:



- постройте карту накопленных отборов по всем горизонтальным скважинам для каждой компании-оператора

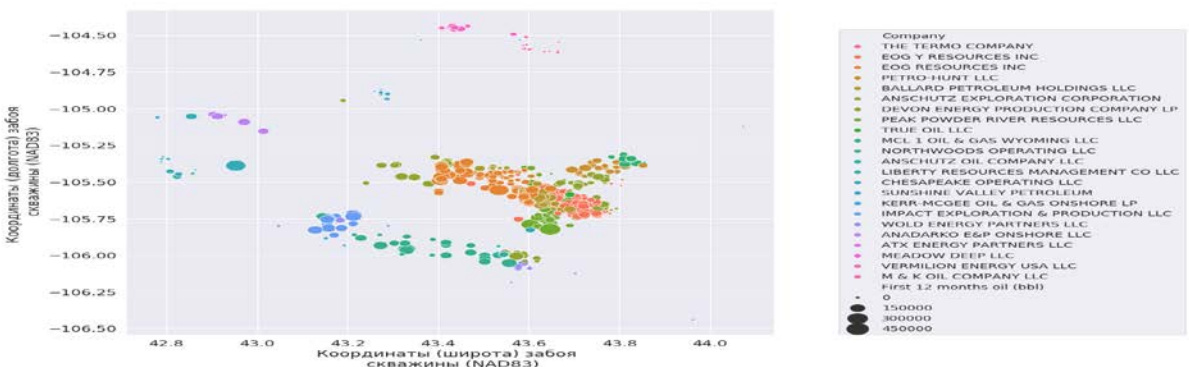
```
# Создайте новый график размером 12x12 см
plt.figure(figsize=(12, 12))

# Постройте график ("пузырьковый"), где каждая точка соответствует координатам забоя скважины, а радиус круга - накопленной добыче нефти
# за первые 12 месяцев работы скважины
sns.scatterplot(x='Bottom hole latitude (NAD83)',
               y='Bottom hole longitude (NAD83)',
               hue='Company', # Цвет круга в зависимости от компании-оператора
               size='First 12 months oil (bbl)', # Радиус круга в зависимости от накопленной добычи нефти
               sizes=(10, 500),
               data=df_horizontal)

# Отобразите сетку для лучшей визуализации
plt.legend(loc=[1.1, 0], fontsize=14)
plt.xlabel('Координаты (широта) забоя \n скважины (NAD83)')
plt.ylabel('Координаты (долгота) забоя \n скважины (NAD83)')

# Сохраните график в отдельный файл
plt.savefig('bubble_map.png', bbox_inches='tight', pad_inches=0.5)
```

Результат:



Образец задач для текущего контроля (задача 2):

Постройте классификационную модель для выделения коллекторов с хорошими фильтрационно-емкостными свойствами (ФЕС) на основе каротажных данных с применением логистической регрессии. Осуществите калибровку модели с использованием тренировочной выборки и проверку качества прогноза на основе тестовой выборки.

```
# Импортировать из библиотеки scikit learn модель логистической регрессии
```

```
# решение  
from sklearn.linear_model import LogisticRegression
```

```
# Инициализировать модель логистической регрессии с линейным солвером и балансировкой классов
```

```
# решение  
model_logistic_regression = LogisticRegression(solver = 'liblinear',  
                                              class_weight = 'balanced',  
                                              random_state = random_state)
```

```
# Произвести обучение модели с использованием обучающей выборки
```

```
# решение  
model_logistic_regression.fit(X = X_train, y = y_train)  
  
LogisticRegression(class_weight='balanced', random_state=9, solver='liblinear')
```

```
# Сделать прогноз с использованием прогнозной выборки
```

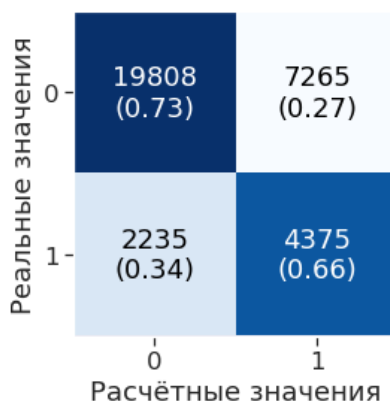
```
# решение  
y_test_pred_logistic_regression = model_logistic_regression.predict(X = X_test)
```

```
# Проверить качество прогноза с использованием матрицы несоответствий
```

```
# решение  
confusion_matrix_logistic_regression = skplt.metrics.confusion_matrix(y_true=y_test, y_pred=y_test_pred_logistic_regression)
```

```
# Визуализировать матрицу несоответствий
```

```
# решение  
fig, ax = plot_confusion_matrix(conf_mat=confusion_matrix_logistic_regression, show_normed=True)  
plt.xlabel('Расчётные значения')  
plt.ylabel('Реальные значения')  
plt.savefig('confusion_matrix.png', bbox_inches='tight', pad_inches=0.5)  
plt.show()
```



Образец задач для итогового контроля:

Проведите интерпретацию комплексной модели, разработанной для выделения коллекторов с хорошими фильтрационно-емкостными свойствами (ФЕС) на основе каротажных данных с использованием алгоритма на основе деревьев принятия решений. Оцените влияние значений входных параметров на прогноз.

```
# Загрузить модель XGBoost с использованием библиотеки joblib
# решение
model=joblib.load(r"../../Day5/notebooks/pickles/model_xgb_scikit_binary.bin")
model

XGBClassifier(base_score=0.5, booster=None, colsample_bylevel=1,
              colsample_bynode=1, colsample_bytree=0.75, eta=0.28,
              eval_metric='auc', gamma=0.8500000000000001, gpu_id=-1,
              importance_type='gain', interaction_constraints=None,
              learning_rate=0.2800000001, max_delta_step=0, max_depth=35,
              min_child_weight=86.0, missing=nan, monotone_constraints=None,
              n_estimators=28, n_jobs=-1, nthread=-1, num_parallel_tree=1,
              random_state=9, reg_alpha=0, reg_lambda=1, scale_pos_weight=1,
              seed=9, subsample=0.75, tree_method=None, validate_parameters=1,
              verbosity=None)

# Инициализировать TreeExplainer из библиотеки SHAP для интерпретации моделей
# на основе деревьев принятия решений
# решение
explainer = shap.TreeExplainer(model)

# Рассчитать значения shap для каждого атрибута в обучающей выборке
# решение
shap_values = explainer.shap_values(X_train)

# Визуализировать SHAP значения для определенного прогноза
# для оценки влияния каждого атрибута в модели
# решение
shap.initjs()
shap.force_plot(base_value = explainer.expected_value,
                shap_values = shap_values[4,:],
                features = X_train.iloc[4,:],
                link = 'logit')
```

Результат:



Оценка результатов аттестации:

Для определения результатов аттестации устанавливается диапазон баллов, которые необходимо набрать для того, чтобы получить отличную, хорошую, удовлетворительную или неудовлетворительную оценки.

Шкала перевода результатов тестирования в оценку результатов аттестации:

<i>Процент выполненных заданий теста</i>	<i>Оценка</i>	<i>Результат аттестации</i>
85-100	Отлично	Слушатель аттестован
65-84	Хорошо	
50-64	Удовлетворительно	
0-49	Неудовлетворительно	Слушатель не аттестован