

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Методические указания

по выполнению практических работ
по дисциплине «Доверенный искусственный интеллект»
для студентов направления подготовки
02.04.01 Математика и компьютерные науки
профиль Искусственный интеллект в кибербезопасности

Ставрополь, 2026

Содержание

<u>Практическая работа 1. Анализ современных программных средств с применением ИИ.</u>	2
<u>Задание 1</u>	2
<u>Задание 2</u>	2
<u>Практическая работа 2 Формализация знаний. Использование семантических сетей для представления знаний.</u>	3
<u>Задание 1</u>	4
<u>Использование фреймов для представления знаний</u>	5
<u>Задание 2.</u>	6
<u>Практические работы 3 Создание онтологии в системе PROTÉGÉ. Основы RDF и OWL</u>	7
<u>Задания</u>	33
<u>Практическое 4 Создание defined classes в системе PROTÉGÉ</u>	34
<u>Задания</u>	40
<u>Практическое занятие 5 Написание запросов на языке SPARQL</u>	40
<u>Задания</u>	45
<u>Практические работы 6. Разработка учебной экспертной системы.</u>	50
<u>Задание 1</u>	56
<u>Задание 2</u>	57
<u>Практическая работа 7. Анализ существующих образовательных платформенных решений.</u>	58
<u>Задание 1</u>	58
<u>Список литературы</u>	63

Практическая работа 1. Риски систем искусственного интеллекта

Цель работы:

Изучить принципы работы системы искусственного интеллекта (СИИ), выявить уязвимости и риски, связанные с её функционированием, предложить методы устранения уязвимостей и продемонстрировать результаты анализа.

Ход работы:

1. Изучение СИИ:

- Ознакомиться с предоставленной программой, написанной на Python, которая выполняет классификацию изображений с использованием машинного обучения.
- Изучить код программы, понять её логику и принципы работы.

2. Выявление уязвимостей:

- Провести анализ кода на наличие ошибок, таких как:
 - Ошибки в данных (например, смещение в данных).
 - Ошибки в алгоритме (например, переобучение).
 - Уязвимости безопасности (например, утечка данных).

3. Исправление уязвимостей:

- Предложить и реализовать методы устранения выявленных уязвимостей.

4. Определение рисков:

- Оценить потенциальные риски, связанные с использованием программы, такие как:
 - Риски безопасности (например, атаки на модель).
 - Этические риски (например, дискриминация).
 - Технические риски (например, низкая точность модели).

5. Демонстрация результатов:

- Подготовить отчет с описанием выявленных уязвимостей, предложенных исправлений и оценки рисков.
- Продемонстрировать работу программы до и после исправления уязвимостей.

Общий пример программы для анализа:

```
```python
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Conv2D, MaxPooling2D, Flatten
from tensorflow.keras.datasets import mnist
from tensorflow.keras.utils import to_categorical

Загрузка данных
(x_train, y_train), (x_test, y_test) = mnist.load_data()
x_train = x_train.reshape((x_train.shape[0], 28, 28, 1)).astype('float32') / 255
```

```

x_test = x_test.reshape((x_test.shape[0], 28, 28,
1)).astype('float32') / 255
y_train = to_categorical(y_train)
y_test = to_categorical(y_test)

Создание модели
model = Sequential()
model.add(Conv2D(32, (3, 3), activation='relu', input_shape=(28, 28,
1)))
model.add(MaxPooling2D((2, 2)))
model.add(Flatten())
model.add(Dense(10, activation='softmax'))

Компиляция модели
model.compile(optimizer='adam', loss='categorical_crossentropy',
metrics=['accuracy'])

Обучение модели
model.fit(x_train, y_train, epochs=5, batch_size=64,
validation_data=(x_test, y_test))

Оценка модели
loss, accuracy = model.evaluate(x_test, y_test)
print(f"Точность модели: {accuracy}")
'''

```

## **Задание:**

### **Вариант 1:**

1. Изучите код программы. Определите, какие данные используются для обучения модели.
2. Выявите уязвимости, связанные с переобучением модели.
3. Предложите методы устранения переобучения (например, добавление Dropout-слоев).
4. Оцените риски, связанные с низкой точностью модели на новых данных.
5. Продемонстрируйте результаты до и после исправления.

### **Вариант 2:**

1. Изучите код программы. Определите, какие метрики используются для оценки модели.
2. Выявите уязвимости, связанные с использованием функции потерь `categorical\_crossentropy`.
3. Предложите альтернативные функции потерь и обоснуйте их выбор.
4. Оцените риски, связанные с неправильным выбором функции потерь.
5. Продемонстрируйте результаты до и после исправления.

### **Вариант 3:**

1. Изучите код программы. Определите, какие слои используются в модели.
2. Выявите уязвимости, связанные с недостаточной глубиной модели.
3. Предложите методы увеличения глубины модели (например, добавление слоев).

4. Оцените риски, связанные с увеличением сложности модели.
5. Продемонстрируйте результаты до и после исправления.

#### **Вариант 4:**

1. Изучите код программы. Определите, какие данные используются для валидации модели.
2. Выявите уязвимости, связанные с недостаточным объемом данных для валидации.
3. Предложите методы увеличения объема данных для валидации.
4. Оцените риски, связанные с недостаточной валидацией модели.
5. Продемонстрируйте результаты до и после исправления.

#### **Вариант 5:**

1. Изучите код программы. Определите, какие оптимизаторы используются для обучения модели.
2. Выявите уязвимости, связанные с использованием оптимизатора ``adam``.
3. Предложите альтернативные оптимизаторы и обоснуйте их выбор.
4. Оцените риски, связанные с неправильным выбором оптимизатора.
5. Продемонстрируйте результаты до и после исправления.

#### **Вариант 6:**

1. Изучите код программы. Определите, какие данные используются для тестирования модели.
2. Выявите уязвимости, связанные с утечкой данных при тестировании.
3. Предложите методы предотвращения утечки данных.
4. Оцените риски, связанные с утечкой данных.
5. Продемонстрируйте результаты до и после исправления.

#### **Вариант 7:**

1. Изучите код программы. Определите, какие активационные функции используются в модели.
2. Выявите уязвимости, связанные с использованием функции активации ``relu``.
3. Предложите альтернативные функции активации и обоснуйте их выбор.
4. Оцените риски, связанные с неправильным выбором функции активации.
5. Продемонстрируйте результаты до и после исправления.

#### **Вариант 8:**

1. Изучите код программы. Определите, какие данные используются для обучения модели.
2. Выявите уязвимости, связанные с несбалансированностью данных.
3. Предложите методы устранения несбалансированности данных.
4. Оцените риски, связанные с несбалансированностью данных.
5. Продемонстрируйте результаты до и после исправления.

#### **Вариант 9:**

1. Изучите код программы. Определите, какие параметры используются для обучения модели.

2. Выявите уязвимости, связанные с неправильным выбором параметров (например, learning rate).
3. Предложите методы настройки параметров обучения.
4. Оцените риски, связанные с неправильной настройкой параметров.
5. Продемонстрируйте результаты до и после исправления.

#### **Вариант 10:**

1. Изучите код программы. Определите, какие данные используются для обучения модели.
2. Выявите уязвимости, связанные с атаками на модель (например, adversarial attacks).
3. Предложите методы защиты от атак на модель.
4. Оцените риски, связанные с атаками на модель.
5. Продемонстрируйте результаты до и после исправления.

#### **Рекомендации по выполнению:**

- Используйте библиотеки Python, такие как TensorFlow, Keras, Scikit-learn.
- Проводите анализ кода с использованием инструментов статического анализа (например, pylint).
- Для демонстрации результатов используйте графики и таблицы.
- Подготовьте отчет в формате PDF или презентацию.

## **Практическая работа 2 Управление рисками искусственного интеллекта.**

### **Цель работы:**

Изучить теоретические основы управления рисками в системах искусственного интеллекта (ИИ), разработать политику управления рисками, описать процессы прогнозирования, выявления и управления рисками, оценить потенциальное воздействие рисков и определить ответственных лиц за их минимизацию.

### **Ход работы:**

#### **Изучение темы по управлению рисками ИИ:**

Ознакомиться с основными понятиями и подходами к управлению рисками в ИИ. Изучить примеры рисков, связанных с использованием ИИ (например, этические, технические, правовые).

#### **Выработка политики управления рисками:**

На основе выявленных на предыдущем занятии рисков разработать политику управления рисками.

Определить ключевые принципы и этапы управления рисками.

#### **Выявление и описание процессов:**

Описать процессы прогнозирования, выявления и управления рисками.

Разработать алгоритм действий для минимизации рисков.

#### **Оценка потенциального воздействия:**

Оценить возможные последствия рисков для бизнеса, общества и технологий.

Использовать методы качественной и количественной оценки рисков.

#### **Определение ответственных лиц:**

Назначить ответственных лиц за каждый этап управления рисками.

Определить их роли и обязанности.

#### **Демонстрация полученных результатов:**

Подготовить отчет или презентацию с описанием политики управления рисками, процессов и результатов оценки.

Продемонстрировать разработанные решения.

#### **Общий пример для анализа:**

Рассмотрим систему ИИ, используемую в банковской сфере для оценки кредитоспособности клиентов.

Пример будет связан с системой ИИ, которая классифицирует данные (например, кредитоспособность клиентов), и будет включать этапы, где можно выявить риски и предложить методы их устранения.

#### **Пример кода для анализа:**

```
python
import pandas as pd
from sklearn.model_selection import train_test_split
```

```

from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report
from sklearn.preprocessing import StandardScaler

Загрузка данных
data = pd.read_csv("credit_data.csv") # Пример данных о клиентах
print(data.head())

Предобработка данных
data = data.dropna() # Удаление пропущенных значений
X = data.drop("Creditworthy", axis=1) # Признаки
y = data["Creditworthy"] # Целевая переменная (кредитоспособность)

Разделение данных на обучающую и тестовую выборки
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

Масштабирование данных
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

Создание и обучение модели
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)

Оценка модели
y_pred = model.predict(X_test)
print("Точность модели:", accuracy_score(y_test, y_pred))
print("Отчет по классификации:\n", classification_report(y_test,
y_pred))

```

## **Анализ рисков и управление ими на примере кода:**

### **1. Изучение темы по управлению рисками ИИ:**

- Риски в данном коде могут включать:
- Смещение в данных (bias): Данные могут быть несбалансированными, что приведет к дискриминации определенных групп клиентов.
- Переобучение модели: Модель может показывать высокую точность на обучающих данных, но плохо работать на новых данных.
- Утечка данных: При масштабировании данных могут быть допущены ошибки, приводящие к утечке информации.

### **2. Выработка политики управления рисками:**

- Для устранения смещения в данных:
- Использовать методы балансировки данных (например, SMOTE).
- Добавить проверку на fairness (справедливость) модели.
- Для предотвращения переобучения:
- Использовать кросс-валидацию.
- Добавить регуляризацию.
- Для предотвращения утечки данных:

- Убедиться, что масштабирование применяется только к обучающим данным.

### 3. Выявление и описание процессов:

- Прогнозирование рисков: Анализ данных на наличие смещений и несбалансированности.
- Выявление рисков: Тестирование модели на новых данных для проверки на переобучение.
- Управление рисками: Внедрение методов балансировки данных и регуляризации.

### 4. Оценка потенциального воздействия:

- Смещение в данных: Может привести к дискриминации клиентов и судебным искам.
- Переобучение модели: Может привести к финансовым потерям из-за неправильных решений.
- Утечка данных: Может привести к утрате доверия клиентов и репутационным потерям.

### 5. Определение ответственных лиц:

- Data Scientist: Ответственный за предобработку данных и обучение модели.
- Security Officer: Ответственный за защиту данных.
- Ethics Committee: Ответственный за проверку fairness модели.

### 6. Демонстрация полученных результатов:

- Подготовить отчет с описанием выявленных рисков и предложенных решений.
- Продемонстрировать улучшенную версию кода с устраненными уязвимостями.

### Улучшенный код с устранением рисков:

```
python
from imblearn.over_sampling import SMOTE
from sklearn.model_selection import cross_val_score

Балансировка данных для устранения смещения
smote = SMOTE(random_state=42)
X_train_balanced, y_train_balanced = smote.fit_resample(X_train,
y_train)

Кросс-валидация для предотвращения переобучения
scores = cross_val_score(model, X_train_balanced, y_train_balanced,
cv=5)
print("Средняя точность при кросс-валидации:", scores.mean())

Обучение модели на сбалансированных данных
model.fit(X_train_balanced, y_train_balanced)

Оценка модели на тестовых данных
y_pred = model.predict(X_test)
```

```
print("Точность модели после балансировки:", accuracy_score(y_test,
y_pred))
print("Отчет по классификации:\n", classification_report(y_test,
y_pred))
```

## **Задания для студентов (10 вариантов):**

### **Вариант 1:**

- Изучите риски, связанные с использованием ИИ в банковской сфере.
- Разработайте политику управления рисками, связанными с дискриминацией клиентов.
- Опишите процессы прогнозирования и выявления дискриминации в системе.
- Оцените потенциальное воздействие дискриминации на репутацию банка.
- Определите ответственных лиц за устранение дискриминации.
- Продемонстрируйте результаты.

### **Вариант 2:**

- Изучите риски, связанные с утечкой данных в системе ИИ.
- Разработайте политику управления рисками утечки данных.
- Опишите процессы мониторинга и предотвращения утечек.
- Оцените потенциальное воздействие утечки данных на клиентов и банк.
- Определите ответственных лиц за защиту данных.
- Продемонстрируйте результаты.

### **Вариант 3:**

- Изучите риски, связанные с ошибками в алгоритмах ИИ.
- Разработайте политику управления рисками ошибок в алгоритмах.
- Опишите процессы тестирования и исправления ошибок.
- Оцените потенциальное воздействие ошибок на финансовые потери банка.
- Определите ответственных лиц за тестирование алгоритмов.
- Продемонстрируйте результаты.

### **Вариант 4:**

- Изучите риски, связанные с кибератаками на систему ИИ.
- Разработайте политику управления рисками кибератак.
- Опишите процессы защиты системы от атак.
- Оцените потенциальное воздействие атак на безопасность банка.
- Определите ответственных лиц за кибербезопасность.
- Продемонстрируйте результаты.

### **Вариант 5:**

- Изучите риски, связанные с непрозрачностью решений ИИ.
- Разработайте политику управления рисками непрозрачности.
- Опишите процессы повышения прозрачности решений ИИ.
- Оцените потенциальное воздействие непрозрачности на доверие клиентов.
- Определите ответственных лиц за обеспечение прозрачности.
- Продемонстрируйте результаты.

### **Вариант 6:**

- Изучите риски, связанные с использованием biased данных в ИИ.
- Разработайте политику управления рисками biased данных.
- Опишите процессы выявления и устранения biased данных.
- Оцените потенциальное воздействие biased данных на качество решений.
- Определите ответственных лиц за обработку данных.
- Продемонстрируйте результаты.

### **Вариант 7:**

- Изучите риски, связанные с изменением законодательства в области ИИ.
- Разработайте политику управления рисками изменения законодательства.
- Опишите процессы мониторинга и адаптации к изменениям.
- Оцените потенциальное воздействие изменений на бизнес-процессы банка.
- Определите ответственных лиц за соблюдение законодательства.
- Продемонстрируйте результаты.

### **Вариант 8:**

- Изучите риски, связанные с низкой точностью модели ИИ.
- Разработайте политику управления рисками низкой точности.
- Опишите процессы улучшения точности модели.
- Оцените потенциальное воздействие низкой точности на принятие решений.
- Определите ответственных лиц за улучшение модели.
- Продемонстрируйте результаты.

### **Вариант 9:**

- Изучите риски, связанные с этическими вопросами использования ИИ.
- Разработайте политику управления этическими рисками.
- Опишите процессы учета этических норм в системе ИИ.
- Оцените потенциальное воздействие этических рисков на репутацию банка.
- Определите ответственных лиц за соблюдение этических норм.
- Продемонстрируйте результаты.

### **Вариант 10:**

- Изучите риски, связанные с зависимостью от сторонних поставщиков ИИ.
- Разработайте политику управления рисками зависимости.
- Опишите процессы минимизации зависимости от поставщиков.
- Оцените потенциальное воздействие зависимости на бизнес-процессы.
- Определите ответственных лиц за управление поставщиками.
- Продемонстрируйте результаты.

### **Рекомендации по выполнению:**

- Используйте научные статьи, стандарты (например, ISO 31000) и руководства по управлению рисками.
- Применяйте методы анализа рисков, такие как SWOT-анализ, FMEA или диаграммы Ишикавы.
- Для демонстрации результатов используйте таблицы, графики и диаграммы.
- Подготовьте отчет в формате PDF или презентацию.

## **Практические работы 3 Метрики объяснимости моделей искусственного интеллекта**

### **Цель работы:**

Изучить теоретические основы объяснимости моделей ИИ, ознакомиться с библиотекой SHAP (SHapley Additive exPlanations), реализовать взаимодействие SHAP с моделью машинного обучения для объяснения её работы и анализа важности признаков.

### **Ход работы:**

#### **Изучение темы по объяснимости моделей ИИ:**

Ознакомиться с понятием объяснимости (interpretability) в контексте ИИ.

Изучить основные подходы к объяснению моделей, включая SHAP, LIME и другие.

Ознакомиться с документацией библиотеки SHAP.

#### **Реализация взаимодействия SHAP с моделью машинного обучения:**

Использовать модель машинного обучения, разработанную на предыдущих занятиях (например, классификатор изображений или кредитоспособности).

Интегрировать SHAP для анализа важности признаков и объяснения решений модели.

#### **Анализ результатов:**

Визуализировать результаты SHAP (например, графики важности признаков, waterfall plots).

Интерпретировать полученные результаты и сделать выводы о работе модели.

#### **Демонстрация полученных результатов:**

Подготовить отчет или презентацию с описанием работы SHAP и анализом объяснимости модели.

Продемонстрировать визуализации и выводы.

#### **Общий пример кода для анализа:**

```
python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
import shap

Загрузка данных
data = pd.read_csv("credit_data.csv") # Пример данных о клиентах
print(data.head())

Предобработка данных
data = data.dropna() # Удаление пропущенных значений
X = data.drop("Creditworthy", axis=1) # Признаки
```

```

y = data["Creditworthy"] # Целевая переменная (кредитоспособность)

Разделение данных на обучающую и тестовую выборки
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

Создание и обучение модели
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)

Оценка модели
y_pred = model.predict(X_test)
print("Точность модели:", accuracy_score(y_test, y_pred))

Инициализация SHAP
explainer = shap.TreeExplainer(model)
shap_values = explainer.shap_values(X_test)

Визуализация важности признаков
shap.summary_plot(shap_values, X_test, plot_type="bar")

```

## **Задания для студентов (10 вариантов):**

### **Вариант 1:**

Изучите документацию SHAP и основные концепции объяснимости моделей.  
 Реализуйте SHAP для анализа важности признаков в модели классификации кредитоспособности.  
 Визуализируйте результаты с помощью `summary_plot`.  
 Проанализируйте, какие признаки наиболее важны для модели.  
 Подготовьте отчет с выводами.

### **Вариант 2:**

Изучите, как SHAP объясняет индивидуальные предсказания модели.  
 Реализуйте SHAP для анализа индивидуальных предсказаний модели.  
 Визуализируйте результаты с помощью `force_plot`.  
 Проанализируйте, как каждый признак влияет на конкретное предсказание.  
 Подготовьте отчет с выводами.

### **Вариант 3:**

Изучите, как SHAP работает с моделями на основе деревьев.  
 Реализуйте SHAP для анализа модели `RandomForest`.  
 Визуализируйте результаты с помощью `dependence_plot`.  
 Проанализируйте, как взаимодействуют признаки в модели.  
 Подготовьте отчет с выводами.

### **Вариант 4:**

Изучите, как SHAP работает с линейными моделями.  
 Реализуйте SHAP для анализа линейной модели (например, `LogisticRegression`).  
 Визуализируйте результаты с помощью `summary_plot`.

Проанализируйте, как линейные коэффициенты влияют на предсказания.  
Подготовьте отчет с выводами.

#### **Вариант 5:**

Изучите, как SHAP работает с нейронными сетями.  
Реализуйте SHAP для анализа простой нейронной сети.  
Визуализируйте результаты с помощью `summary_plot`.  
Проанализируйте, как архитектура сети влияет на объяснимость.  
Подготовьте отчет с выводами.

#### **Вариант 6:**

Изучите, как SHAP объясняет модели ансамблей (например, Gradient Boosting).  
Реализуйте SHAP для анализа модели Gradient Boosting.  
Визуализируйте результаты с помощью `waterfall_plot`.  
Проанализируйте, как ансамбли влияют на объяснимость.  
Подготовьте отчет с выводами.

#### **Вариант 7:**

Изучите, как SHAP работает с категориальными признаками.  
Реализуйте SHAP для анализа модели с категориальными признаками.  
Визуализируйте результаты с помощью `summary_plot`.  
Проанализируйте, как категориальные признаки влияют на предсказания.  
Подготовьте отчет с выводами.

#### **Вариант 8:**

Изучите, как SHAP объясняет модели в задачах регрессии.  
Реализуйте SHAP для анализа модели линейной регрессии.  
Визуализируйте результаты с помощью `summary_plot`.  
Проанализируйте, как признаки влияют на числовые предсказания.  
Подготовьте отчет с выводами.

#### **Вариант 9:**

Изучите, как SHAP работает с текстовыми данными.  
Реализуйте SHAP для анализа модели классификации текстов.  
Визуализируйте результаты с помощью `text_plot`.  
Проанализируйте, как слова влияют на предсказания.  
Подготовьте отчет с выводами.

#### **Вариант 10:**

Изучите, как SHAP объясняет модели в задачах мультиклассовой классификации.  
Реализуйте SHAP для анализа модели мультиклассовой классификации.  
Визуализируйте результаты с помощью `summary_plot`.  
Проанализируйте, как признаки влияют на предсказания для каждого класса.  
Подготовьте отчет с выводами.

#### **Рекомендации по выполнению:**

Используйте библиотеку SHAP для визуализации и анализа моделей.

Для работы с SHAP установите библиотеку:

```
bash
```

```
pip install shap
```

Визуализируйте результаты с помощью графиков SHAP (summary\_plot, force\_plot, dependence\_plot и др.).

Подготовьте отчет в формате PDF или презентацию с описанием работы SHAP и анализом объяснимости модели.

Пример улучшенного кода с SHAP:

```
python
```

```
Copy
```

```
Визуализация индивидуального предсказания
```

```
shap.initjs()
```

```
shap.force_plot(explainer.expected_value[1],
```

```
shap_values[1][0, :], X_test.iloc[0, :])
```

```
Визуализация зависимости признаков
```

```
shap.dependence_plot("Age", shap_values[1], X_test)
```

Этот подход поможет студентам понять, как использовать SHAP для объяснения моделей ИИ, анализировать важность признаков и интерпретировать результаты.

## Практическое 4 Метрики робастности искусственного интеллекта

### Цель работы:

Изучить теоретические основы робастности систем искусственного интеллекта (ИИ), ознакомиться с ГОСТ Р 59898-2021, реализовать метрики оценки функциональной корректности для задач классификации и проверить робастность ранее представленной программы.

### Ход работы:

#### 1. Изучение темы по робастности ИИ:

- Ознакомиться с понятием робастности в контексте ИИ.
- Изучить основные подходы к оценке робастности, включая метрики функциональной корректности.
- Ознакомиться с ГОСТ Р 59898-2021 "Оценка качества систем искусственного интеллекта".

#### 2. Реализация метрик робастности ИИ:

- Реализовать метрики оценки функциональной корректности для задач классификации, которые отсутствуют в стандартных библиотеках Python.
- Протестировать реализованные метрики на примере данных.

#### 3. Проверка робастности программы:

- Применить реализованные метрики для проверки робастности ранее представленной программы (например, модели классификации).
- Проанализировать результаты и сделать выводы.

#### 4. Демонстрация полученных результатов:

- Подготовить отчет или презентацию с описанием реализованных метрик и результатов проверки робастности.
- Продемонстрировать работу программы и метрик.

### Общий пример кода для анализа:

```
python
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score

Загрузка данных
data = pd.read_csv("credit_data.csv") # Пример данных о клиентах
print(data.head())

Предобработка данных
data = data.dropna() # Удаление пропущенных значений
X = data.drop("Creditworthy", axis=1) # Признаки
y = data["Creditworthy"] # Целевая переменная (кредитоспособность)

Разделение данных на обучающую и тестовую выборки
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
```

```
Создание и обучение модели
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)

Оценка модели
y_pred = model.predict(X_test)
print("Точность модели:", accuracy_score(y_test, y_pred))
```

## Реализация метрик робастности:

### 1. Метрика устойчивости к шуму:

Оценивает, насколько модель устойчива к добавлению шума в данные.

```
python
def noise_robustness_metric(model, X_test, y_test, noise_level=0.1):
 # Добавление шума к данным
 X_test_noisy = X_test + np.random.normal(0, noise_level,
X_test.shape)

 # Предсказание на зашумленных данных
 y_pred_noisy = model.predict(X_test_noisy)

 # Оценка точности на зашумленных данных
 accuracy_noisy = accuracy_score(y_test, y_pred_noisy)
 return accuracy_noisy
```

### Пример использования

```
noise_robustness = noise_robustness_metric(model, X_test,
y_test)
print("Устойчивость к шуму:", noise_robustness)
```

### 2. Метрика устойчивости к выбросам:

Оценивает, насколько модель устойчива к наличию выбросов в данных.

```
python
def outlier_robustness_metric(model, X_test, y_test,
outlier_fraction=0.1):
 # Добавление выбросов к данным
 n_outliers = int(outlier_fraction * X_test.shape[0])
 outlier_indices = np.random.choice(X_test.shape[0], n_outliers,
replace=False)
 X_test_outliers = X_test.copy()
 X_test_outliers[outlier_indices] += 10 # Добавление выбросов

 # Предсказание на данных с выбросами
 y_pred_outliers = model.predict(X_test_outliers)

 # Оценка точности на данных с выбросами
 accuracy_outliers = accuracy_score(y_test, y_pred_outliers)
 return accuracy_outliers
```

### Пример использования

```
outlier_robustness = outlier_robustness_metric(model, X_test, y_test)
print("Устойчивость к выбросам:", outlier_robustness)
```

## **Задания для студентов (10 вариантов):**

### **Вариант 1:**

1. Изучите ГОСТ Р 59898-2021 и основные метрики робастности.
2. Реализуйте метрику устойчивости к шуму.
3. Проверьте робастность модели классификации кредитоспособности.
4. Проанализируйте результаты.
5. Подготовьте отчет с выводами.

### **Вариант 2:**

1. Изучите метрики устойчивости к выбросам.
2. Реализуйте метрику устойчивости к выбросам.
3. Проверьте робастность модели классификации.
4. Проанализируйте результаты.
5. Подготовьте отчет с выводами.

### **Вариант 3:**

1. Изучите метрики устойчивости к изменению распределения данных.
2. Реализуйте метрику для оценки устойчивости к изменению распределения.
3. Проверьте робастность модели.
4. Проанализируйте результаты.
5. Подготовьте отчет с выводами.

### **Вариант 4:**

1. Изучите метрики устойчивости к атакам на модель.
2. Реализуйте метрику для оценки устойчивости к атакам.
3. Проверьте робастность модели.
4. Проанализируйте результаты.
5. Подготовьте отчет с выводами.

### **Вариант 5:**

1. Изучите метрики устойчивости к пропущенным данным.
2. Реализуйте метрику для оценки устойчивости к пропущенным данным.
3. Проверьте робастность модели.
4. Проанализируйте результаты.
5. Подготовьте отчет с выводами.

### **Вариант 6:**

1. Изучите метрики устойчивости к изменению масштаба данных.
2. Реализуйте метрику для оценки устойчивости к изменению масштаба.
3. Проверьте робастность модели.
4. Проанализируйте результаты.
5. Подготовьте отчет с выводами.

### **Вариант 7:**

1. Изучите метрики устойчивости к несбалансированным данным.
2. Реализуйте метрику для оценки устойчивости к несбалансированности.
3. Проверьте робастность модели.
4. Проанализируйте результаты.
5. Подготовьте отчет с выводами.

### **Вариант 8:**

1. Изучите метрики устойчивости к изменению гиперпараметров модели.
2. Реализуйте метрику для оценки устойчивости к изменению гиперпараметров.
3. Проверьте робастность модели.

4. Проанализируйте результаты.
5. Подготовьте отчет с выводами.

**Вариант 9:**

1. Изучите метрики устойчивости к изменению архитектуры модели.
2. Реализуйте метрику для оценки устойчивости к изменению архитектуры.
3. Проверьте робастность модели.
4. Проанализируйте результаты.
5. Подготовьте отчет с выводами.

**Вариант 10:**

1. Изучите метрики устойчивости к изменению входных данных.
2. Реализуйте метрику для оценки устойчивости к изменению входных данных.
3. Проверьте робастность модели.
4. Проанализируйте результаты.
5. Подготовьте отчет с выводами.

**Рекомендации по выполнению:**

- Используйте библиотеки Python, такие как NumPy, Pandas и Scikit-learn.
- Для реализации метрик робастности следуйте принципам ГОСТ Р 59898-2021.
- Подготовьте отчет в формате PDF или презентацию с описанием реализованных метрик и результатов проверки робастности.

Этот подход поможет студентам понять, как оценивать и улучшать робастность моделей ИИ, а также применять стандарты качества для анализа систем искусственного интеллекта.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

## **Методические указания**

для обучающихся по организации и проведению самостоятельной работы  
по дисциплине «Доверенный искусственный интеллект»  
для студентов направления подготовки

---

направленность (профиль):

---

Ставрополь  
2026

## СОДЕРЖАНИЕ

1. Общие положения	19
2. Цель и задачи самостоятельной работы	20
3. Технологическая карта самостоятельной работы студента	21
4. Контрольные точки и виды отчетности по ним	21
5. Порядок выполнения самостоятельной работы студентом	22
5.1. Методические рекомендации по работе с учебной литературой	22
6. Список литературы для выполнения СРС	26

## **1. Общие положения**

Самостоятельная работа – планируемая учебная, учебно-исследовательская, научно-исследовательская работа студентов, выполняемая во внеаудиторное (аудиторное) время по заданию и при методическом руководстве преподавателя, но без его непосредственного участия (при частичном непосредственном участии преподавателя, оставляющем ведущую роль за работой студентов). Владиславу Вячеславовичу

Самостоятельная работа студентов (СРС) в ВУЗе является важным видом учебной и научной деятельности студента. Самостоятельная работа студентов играет значительную роль в рейтинговой технологии обучения.

К основным видам самостоятельной работы студентов относятся:

- формирование и усвоение содержания конспекта лекций на базе рекомендованной лектором учебной литературы, включая информационные образовательные ресурсы (электронные учебники, электронные библиотеки и др.);
- написание докладов;
- подготовка к семинарам, практическим и лабораторным работам, их оформление;
- составление аннотированного списка статей из соответствующих журналов по отраслям знаний (педагогических, психологических, методических и др.);
- выполнение учебно-исследовательских работ, проектная деятельность;
- подготовка практических разработок и рекомендаций по решению проблемной ситуации;
- выполнение домашних заданий в виде решения отдельных задач, проведения типовых расчетов, расчетно-компьютерных и индивидуальных работ по отдельным разделам содержания дисциплин и т.д.;
- компьютерный текущий самоконтроль и контроль успеваемости на базе электронных обучающих и аттестующих тестов;
- выполнение курсовых работ (проектов) в рамках дисциплин;
- выполнение выпускной квалификационной работы и др.

Методика организации самостоятельной работы студентов зависит от структуры, характера и особенностей изучаемой дисциплины, объема часов на ее изучение, вида заданий для самостоятельной работы студентов, индивидуальных качеств студентов и условий учебной деятельности.

Процесс организации самостоятельной работы студентов включает в себя следующие этапы:

- подготовительный (определение целей, составление программы, подготовка методического обеспечения, подготовка оборудования);
- основной (реализация программы, использование приемов поиска информации, усвоения, переработки, применения, передачи знаний, фиксирование результатов, самоорганизация процесса работы);
- заключительный (оценка значимости и анализ результатов, их систематизация, оценка эффективности программы и приемов работы, выводы о направлениях оптимизации труда).

Самостоятельная работа по дисциплине «Информационные технологии командной работы и проектной деятельности» направлена на формирование следующих **компетенций**:

Код	Формулировка
ПК-2	Способен проводить научные исследования, на основе существующих методов в конкретной области профессиональной деятельности

## 2. Цель и задачи самостоятельной работы

Ведущая цель организации и осуществления СРС совпадает с целью обучения студента – формирование набора общенаучных, профессиональных и универсальных компетенций будущего бакалавра.

При организации СРС важным и необходимым условием становятся формирование умения самостоятельной работы для приобретения знаний, навыков и возможности организации учебной и научной деятельности. Целью самостоятельной работы студентов является овладение фундаментальными знаниями, профессиональными умениями и навыками деятельности по профилю, опытом творческой, исследовательской деятельности. Самостоятельная работа студентов способствует развитию самостоятельности, ответственности и организованности, творческого подхода к решению проблем учебного и профессионального уровня.

Задачами СРС являются:

- систематизация и закрепление полученных теоретических знаний и практических умений студентов;
- углубление и расширение теоретических знаний;
- формирование умений использовать нормативную, правовую, справочную документацию и специальную литературу;

- развитие познавательных способностей и активности студентов: творческой инициативы, самостоятельности, ответственности и организованности;
- формирование самостоятельности мышления, способностей к саморазвитию, самосовершенствованию и самореализации;
- развитие исследовательских умений;

### 3. Технологическая карта самостоятельной работы студента

Коды реализ уемых компет енций	Вид деятельности студентов	Итоговый продукт самостоя тельной работы	Средства и технологии оценки	Объем часов, в том числе (астр.)		
				СРС	Контактная работа с преподавателе м	Всего
4 семестр						
ПК-2	Подготовка к лекциям	Конспект лекций	Собеседова ние	21,0	15,0	36,0
ПК-2	Подготовка к практическим занятиям	Отчет в электронн ом виде	Собеседова ние	21,0	15,0	36,0
ПК-2	Самостоятель ное изучение литературы	Конспект	Собеседова ние	21,0	15,0	36,0
ПК-2	Самотестиров ание, подготовка к тестированию	Тестирова ние	Тестовые задания	21,0	15,0	36,0
Итого за 4 семестр				84,0	60,0	144,0
Итого				84,0	60,0	144,0

### 4. Контрольные точки и виды отчетности по ним

Контроль качества и сроков самостоятельной работы выполняется в соответствии с учебным графиком и оформляется в соответствии с заданием. Предусмотрена следующая рейтинговая оценка знаний

#### Описание шкалы оценивания

В рамках рейтинговой системы успеваемость студентов по каждой дисциплине оценивается в ходе текущего контроля и промежуточной аттестации.

#### Текущий контроль

### 5. Порядок выполнения самостоятельной работы студентом

## 5.1. Методические рекомендации по работе с учебной литературой

При работе с книгой необходимо подобрать литературу, научиться правильно ее читать, вести записи. Для подбора литературы в библиотеке используются алфавитный и систематический каталоги.

Важно помнить, что рациональные навыки работы с книгой - это всегда большая экономия времени и сил.

Правильный подбор учебников рекомендуется преподавателем, читающим лекционный курс. Необходимая литература может быть также указана в методических разработках по данному курсу.

Изучая материал по учебнику, следует переходить к следующему вопросу только после правильного уяснения предыдущего, описывая на бумаге все выкладки и вычисления (в том числе те, которые в учебнике опущены или на лекции даны для самостоятельного вывода).

При изучении любой дисциплины большую и важную роль играет самостоятельная индивидуальная работа.

Особое внимание следует обратить на определение основных понятий курса. Студент должен подробно разбирать примеры, которые поясняют такие определения, и уметь строить аналогичные примеры самостоятельно. Нужно добиваться точного представления о том, что изучаешь. Полезно составлять опорные конспекты. При изучении материала по учебнику полезно в тетради (на специально отведенных полях) дополнять конспект лекций. Там же следует отмечать вопросы, выделенные студентом для консультации с преподавателем.

Выводы, полученные в результате изучения, рекомендуется в конспекте выделять, чтобы они при перечитывании записей лучше запоминались.

Опыт показывает, что многим студентам помогает составление листа опорных сигналов, содержащего важнейшие и наиболее часто употребляемые формулы и понятия. Такой лист помогает запомнить формулы, основные положения лекции, а также может служить постоянным справочником для студента.

Чтение научного текста является частью познавательной деятельности. Ее цель – извлечение из текста необходимой информации. От того на сколько осознанна читающим собственная внутренняя установка при обращении к печатному слову (найти нужные сведения, усвоить информацию полностью или частично, критически проанализировать материал и т.п.) во многом зависит эффективность осуществляемого действия.

Выделяют **четыре основные установки в чтении научного текста:**

информационно-поисковый (задача – найти, выделить искомую информацию)

усваивающая (усилия читателя направлены на то, чтобы как можно полнее осознать и запомнить как сами сведения излагаемые автором, так и всю логику его рассуждений)

аналитико-критическая (читатель стремится критически осмыслить материал, проанализировав его, определив свое отношение к нему)

творческая (создает у читателя готовность в том или ином виде – как отправной пункт для своих рассуждений, как образ для действия по аналогии и т.п. – использовать суждения автора, ход его мыслей, результат наблюдения, разработанную методику, дополнить их, подвергнуть новой проверке).

*Основные виды систематизированной записи прочитанного:*

Аннотирование – предельно краткое связное описание просмотренной или прочитанной книги (статьи), ее содержания, источников, характера и назначения;

Планирование – краткая логическая организация текста, раскрывающая содержание и структуру изучаемого материала;

Тезирование – лаконичное воспроизведение основных утверждений автора без привлечения фактического материала;

Цитирование – дословное выписывание из текста выдержек, извлечений, наиболее существенно отражающих ту или иную мысль автора;

Конспектирование – краткое и последовательное изложение содержания прочитанного.

Конспект – сложный способ изложения содержания книги или статьи в логической последовательности. Конспект аккумулирует в себе предыдущие виды записи, позволяет всесторонне охватить содержание книги, статьи. Поэтому умение составлять план, тезисы, делать выписки и другие записи определяет и технологию составления конспекта.

## *5.2. Методические рекомендации по составлению конспекта лекций:*

1. Внимательно прочитайте текст. Уточните в справочной литературе непонятные слова. При записи не забудьте вынести справочные данные на поля конспекта.

2. Выделите главное, составьте план.

3. Кратко сформулируйте основные положения текста, отметьте аргументацию автора.

4. Законспектируйте материал, четко следуя пунктам плана. При конспектировании старайтесь выразить мысль своими словами. Записи следует вести четко, ясно.

5. Грамотно записывайте цитаты. Цитируя, учитывайте лаконичность, значимость мысли. В тексте конспекта желательно приводить не только тезисные положения, но и их доказательства. При оформлении конспекта необходимо стремиться к емкости каждого предложения. Мысли автора книги следует излагать кратко, заботясь о стиле и выразительности написанного. Число дополнительных элементов конспекта должно быть логически обоснованным, записи должны распределяться в определенной последовательности, отвечающей логической структуре произведения. Для уточнения и дополнения необходимо оставлять поля.

Овладение навыками конспектирования требует от студента целеустремленности, повседневной самостоятельной работы.

### *5.3. Методические рекомендации по подготовке к практическим занятиям*

Для того чтобы практические занятия приносили максимальную пользу, необходимо помнить, что упражнение и решение задач проводятся по вычитанному на лекциях материалу и связаны, как правило, с детальным разбором отдельных вопросов лекционного курса. Следует подчеркнуть, что только после усвоения лекционного материала с определенной точки зрения (а именно с той, с которой он излагается на лекциях) он будет закрепляться на лабораторных занятиях как в результате обсуждения и анализа лекционного материала, так и с помощью решения проблемных ситуаций, задач. При этих условиях студент не только хорошо усвоит материал, но и научится применять его на практике, а также получит дополнительный стимул (и это очень важно) для активной проработки лекции.

При самостоятельном решении задач нужно обосновывать каждый этап решения, исходя из теоретических положений курса. Если студент видит несколько путей решения проблемы (задачи), то нужно сравнить их и выбрать самый рациональный. Полезно до начала вычислений составить краткий план решения проблемы (задачи). Решение проблемных задач или примеров следует излагать подробно, вычисления располагать в строгом порядке, отделяя вспомогательные вычисления от основных. Решения при необходимости нужно сопровождать комментариями, схемами, чертежами и рисунками.

Следует помнить, что решение каждой учебной задачи должно доводиться до окончательного логического ответа, которого требует условие, и по возможности с выводом. Полученный ответ следует проверить способами, вытекающими из существа данной задачи. Полезно также (если возможно) решать несколькими способами и сравнить полученные результаты. Решение задач данного типа нужно продолжать до приобретения твердых навыков в их решении.

Практические работы № 1-5,8 выполняются студентами заочной и очно-заочной формы обучения самостоятельно.

### *5.4. Методические рекомендации по самопроверке знаний*

После изучения определенной темы по записям в конспекте и учебнику, а также решения достаточного количества соответствующих задач на практических занятиях и самостоятельно студенту рекомендуется провести самопроверку усвоенных знаний, ответив на контрольные вопросы по изученной теме.

В случае необходимости нужно еще раз внимательно разобраться в материале.

Иногда недостаточность усвоения того или иного вопроса выясняется только при изучении дальнейшего материала. В этом случае надо вернуться назад и повторить плохо усвоенный материал. Важный критерий усвоения теоретического материала – умение отвечать на вопросы для собеседования.

### *Критерии оценки:*

Оценка «**зачтено**» выставляется студенту, если он показал минимальный, средний или высокий уровень сформированности компетенции ПК-2.

Оценка «**не зачтено**» выставляется студенту, если минимальный уровень сформированности компетенции ПК-2 не достигнут

### *Описание шкалы оценивания*

Максимально возможный балл за весь текущий контроль устанавливается равным 55. Текущее контрольное мероприятие считается сданным, если студент получил за него не менее 60% от установленного для этого контроля максимального балла. Рейтинговый балл, выставляемый студенту за текущее контрольное мероприятие, сданное студентом в установленные графиком контрольных мероприятий сроки, определяется следующим образом:

Уровень выполнения контрольного задания	Рейтинговый балл (в % от максимального балла за контрольное задание)
Отличный	100
Хороший	80
Удовлетворительный	60
Неудовлетворительный	0

### *5.6. Методические рекомендации по подготовке к зачетам*

#### **Контроль самостоятельной работы студентов**

Контроль самостоятельной работы проводится преподавателем в аудитории.

Предусмотрены следующие виды контроля: собеседование, оценка выполнения доклада и его презентации.

Подробные критерии оценивания компетенций приведены в Фонде оценочных средств для проведения текущей и промежуточной аттестации.

## **6. Список литературы для выполнения СРС**

### **Перечень основной литературы**

1. Сердюков, Ю. М. Философия виртуальной реальности и искусственного интеллекта: учебное пособие / Ю. М. Сердюков; под редакцией Ю. М. Сердюкова. — Хабаровск: ДВГУПС, 2020. — 169 с. — ISBN 978-5-262-00881-0. — Текст электронный // Лань: электронно-библиотечная система. Режим доступа: для авториз. пользователей. <https://e.lanbook.com/book/179385>

### **Перечень дополнительной литературы**

1. Новые законы робототехники. Регуляторный ландшафт. Мировой опыт регулирования робототехники и технологий искусственного интеллекта / В. В. Бакуменко, А. Д. Вольнец, А. В.

Незнамов [и др.]; под редакцией А. В. Незнамова. — Москва: Infotropic Media, 2018. — 220 с. — ISBN 978-5-9998-0324-5. — Текст: электронный // Лань: электронно-библиотечная система. Режим доступа: для авториз. пользователей. <https://e.lanbook.com/book/138977>

**Перечень учебно-методического обеспечения самостоятельной работы обучающихся по дисциплине**

1. Методические указания по организации самостоятельной работы студентов по дисциплине «Доверенный искусственный интеллект». Ставрополь: СКФУ, 2025.
2. Методические указания к практическим занятиям по дисциплине «Доверенный искусственный интеллект». Ставрополь: СКФУ, 2025.

**Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины**

1. <http://biblioclub.ru/> – Университетская библиотека online "Библиоклуб"
2. <https://4brain.ru/liderstvo/> – Лидерство: уроки эффективного руководителя
3. <https://spravochnick.ru/psihologiya/> – Справочник по психологии