

Документ подписан простой электронной подписью
Информация о владельце: **Федеральное государственное бюджетное образовательное**
ФИО: Цыбиков Бэликто Батович **учреждение высшего образования**
Должность: Ректор **«Бурятская государственная сельскохозяйственная академия имени В.Р. Филиппова»**
Дата подписания: 27.05.2025 11:13:39
Уникальный программный ключ:
056af948c3e48c6f3c571e429957a8ae7b757ae8 **Агрономический факультет**

«СОГЛАСОВАНО»

Заведующий выпускающей кафедрой
Общее земледелие

К.С.-Х.Н., ДОЦЕНТ

уч. ст., уч. зв.

Соболев В.А.

подпись

«01» января 2025 г.

«УТВЕРЖЕНО»

Декан
Агрономический факультет

К.С.-Х.Н., ДОЦЕНТ

уч. ст., уч. зв.

Манханов А.Д.

подпись

«01» января 2025 г.

**Оценочные материалы
Дисциплины (модуля)**

Б1.О.06.04 Машинное обучение

**Направление 35.03.04 Агрономия
направленность (профиль) Инновационные агротехнологии**

Обеспечивающая
преподавание дисциплины
кафедра

Разработчик (и)

Информатика и информационные технологии в
экономике

подпись

уч. ст., уч. зв.

Н.Б. Садуев
И.О.Фамилия

подпись

уч. ст., уч. зв.

И.О.Фамилия

Внутренние эксперты:
Председатель методической
комиссии агрономического
факультета

подпись

уч. ст., уч. зв.

И.О.Фамилия

Заведующий методическим
кабинетом УМУ

подпись

И.О.Фамилия

Улан-Удэ, 2025 г.

Комплект тестовых заданий

Блок 1: (Знать)

1. Что такое машинное обучение?
 - A) Способ автоматизации офисных задач
 - B) Область ИИ, позволяющая системам учиться на данных
 - C) Язык программирования
 - D) Система управления базами данных

✓ Ответ: B

2. Какие данные используются в надзорном обучении?
 - A) Без меток
 - B) С известными метками
 - C) Только числовые
 - D) Только текстовые

✓ Ответ: B

3. В каких задачах применяется логистическая регрессия?
 - A) Регрессии
 - B) Кластеризации
 - C) Классификации
 - D) Уменьшении размерности

✓ Ответ: C

4. Что означает термин "признак" в машинном обучении?
 - A) Результат модели
 - B) Входная переменная
 - C) Метрика качества
 - D) Алгоритм обучения

✓ Ответ: B

5. Что такое обучающая выборка?
 - A) Данные для проверки модели
 - B) Данные для настройки гиперпараметров
 - C) Данные для обучения модели
 - D) Случайно выбранные данные

✓ Ответ: C

6. Какой алгоритм используется для кластеризации?
 - A) KNN
 - B) K-means
 - C) Linear Regression
 - D) Decision Tree

✓ Ответ: B

7. Что такое энтропия в контексте деревьев решений?
 - A) Мера точности
 - B) Мера беспорядка
 - C) Тип регуляризации
 - D) Метрика расстояния

✓ Ответ: B

8. Что такое Precision?
 - A) Отношение верно предсказанных положительных к общему числу реальных положительных
 - B) Отношение верно предсказанных положительных к общему числу предсказанных положительных
 - C) Среднее между R^2 и MSE
 - D) Метрика для регрессии

✓ Ответ: B

9. Что такое recall (полнота)?
 - A) Показывает, сколько из найденных объектов правильны
 - B) Показывает, сколько из всех существующих объектов найдены
 - C) Произведение precision и recall
 - D) Невозможно определить

✓ Ответ: B

10. Что такое F1-score?
A) Среднее арифметическое precision и recall
B) Среднее геометрическое precision и recall
C) Сумма precision и recall
D) Разность precision и recall
✓ Ответ: B
11. Что такое функция потерь?
A) Формула для разделения данных
B) Функция, минимизация которой позволяет обучать модель
C) Формула для оценки точности
D) Тип регуляризации
✓ Ответ: B
12. Что такое переобучение модели?
A) Модель не обучена
B) Модель хорошо работает на тестовой выборке
C) Модель слишком сложная и запоминает обучающие данные
D) Модель имеет низкую сложность
✓ Ответ: C
13. Какой метод используется для оценки качества классификации?
A) MAE
B) MSE
C) Accuracy
D) R^2
✓ Ответ: C
14. Какой из перечисленных методов является ансамблевым?
A) Линейная регрессия
B) Логистическая регрессия
C) Случайный лес
D) KNN
✓ Ответ: C
15. Что такое бустинг?
A) Параллельное обучение множества моделей
B) Последовательное обучение моделей с коррекцией ошибок
C) Метод нормализации данных
D) Метод уменьшения размерности
✓ Ответ: B
16. Что такое batch size в нейронных сетях?
A) Размер всей обучающей выборки
B) Число образцов, обрабатываемых за одну итерацию
C) Число скрытых слоев
D) Размер весов сети
✓ Ответ: B
17. Что такое L1-регуляризация?
A) Штрафует большие коэффициенты пропорционально их квадрату
B) Штрафует количество признаков
C) Штрафует абсолютные значения коэффициентов
D) Не влияет на модель
✓ Ответ: C
18. Что такое dropout в нейронных сетях?
A) Удаление лишних признаков
B) Отключение случайных нейронов во время обучения
C) Упрощение модели
D) Оптимизатор
✓ Ответ: B
19. Что такое cross-validation?
A) Оценка модели на тестовой выборке
B) Разбиение данных на части для многократной оценки
C) Предобработка данных

D) Выбор признаков

✓ Ответ: B

20. Что такое Elastic Net?

A) Комбинация Ridge и Lasso

B) Только L1-регуляризация

C) Только L2-регуляризация

D) Без регуляризации

✓ Ответ: A

Блок 2: (Знать, уметь)

21. Какой метод лучше использовать для классификации несбалансированных данных?

A) Accuracy

B) F1-score

C) MSE

D) R^2

✓ Ответ: B

22. Какая метрика лучше подходит для задачи регрессии?

A) Accuracy

B) Precision

C) MSE

D) F1-score

✓ Ответ: C

23. Какой метод используется для увеличения количества данных в задачах с несбалансированными классами?

A) Normalization

B) SMOTE

C) PCA

D) Cross-validation

✓ Ответ: B

24. Какой из следующих методов чувствителен к масштабу признаков?

A) Дерево решений

B) KNN

C) Naive Bayes

D) Random Forest

✓ Ответ: B

25. Какой алгоритм может быть использован как для регрессии, так и для классификации?

A) KNN

B) Linear Regression

C) Logistic Regression

D) PCA

✓ Ответ: A

26. Что такое Label Encoding?

A) Преобразование числовых признаков

B) Преобразование категориальных признаков в числа

C) Нормализация данных

D) Удаление пропущенных значений

✓ Ответ: B

27. Что такое MinMaxScaler?

A) Преобразование данных к среднему и дисперсии

B) Нормализация от 0 до 1

C) Кодирование категориальных признаков

D) Метод разделения выборки

✓ Ответ: B

28. Что такое feature engineering?

A) Создание новых признаков из имеющихся данных

B) Удаление признаков

C) Подбор гиперпараметров

D) Оценка качества модели

✓ Ответ: A

29. Что такое overfitting?
A) Модель плохо работает на тренировочных данных
B) Модель хорошо работает на тестовых данных
C) Модель слишком простая
D) Модель слишком сложная
✓ Ответ: D
30. Какой метод используется для нормализации данных?
A) LabelEncoder
B) OneHotEncoder
C) MinMaxScaler
D) PCA
✓ Ответ: C
31. Какой метод используется для кодирования категориальных признаков?
A) MinMaxScaler
B) StandardScaler
C) One-Hot Encoding
D) PCA
✓ Ответ: C
32. Что означает высокий показатель precision и низкий recall?
A) Модель редко ошибается, но пропускает много случаев
B) Модель часто ошибается, но находит много случаев
C) Модель идеальная
D) Модель не обучена
✓ Ответ: A
33. Какой алгоритм наиболее чувствителен к выбросам?
A) Деревья решений
B) KNN
C) Линейная регрессия
D) Random Forest
✓ Ответ: C
34. Какой из следующих методов относится к глубокому обучению?
A) Random Forest
B) CNN
C) SVM
D) KNN
✓ Ответ: B
35. Что такое batch normalization?
A) Нормализация данных перед обучением
B) Нормализация внутри сети во время обучения
C) Удаление лишних данных
D) Упрощение модели
✓ Ответ: B
36. Какой алгоритм можно использовать для задачи многоклассовой классификации?
A) Linear Regression
B) Logistic Regression
C) Decision Tree
D) PCA
✓ Ответ: C
37. Что такое GridSearchCV?
A) Нормализация данных
B) Поиск лучших гиперпараметров
C) Деление данных на выборки
D) Визуализация
✓ Ответ: B
38. Какой из методов используется для визуализации результатов классификации?
A) Гистограмма
B) Boxplot
C) Confusion matrix

D) Scatter plot

✓ Ответ: C

39. Какой метод используется для автоматического выбора признаков?

A) PCA

B) Feature importance

C) One-Hot Encoding

D) Label Encoding

✓ Ответ: B

40. Что такое функция активации нейронной сети?

A) Упрощает модель

B) Определяет выход нейрона

C) Ускоряет обучение

D) Нормализует входные данные

✓ Ответ: B

Блок 3: (Знать, уметь, владеть)

41. Какие этапы включает полный цикл разработки модели машинного обучения?

✓ Ответ: сбор данных → предобработка → разделение выборки → обучение → оценка → интерпретация

42. Как выбрать лучшую модель среди нескольких вариантов?

✓ Ответ: сравнить метрики, провести кросс-валидацию, оценить производительность на тестовой выборке

43. Какие библиотеки Python вы знаете для машинного обучения?

✓ Ответ: Scikit-learn, TensorFlow, Keras, PyTorch, XGBoost

44. Как реализовать модель случайного леса в Python?

✓ Пример ответа:

```
python
```

```
1 from sklearn.ensemble import RandomForestClassifier
2 model = RandomForestClassifier()
3 model.fit(X_train, y_train)
```

45. Какие шаги включает предобработка данных?

✓ Ответ: очистка → нормализация → кодирование → разделение на выборки

46. Какие метрики вы знаете для задач регрессии?

✓ Ответ: MAE, MSE, RMSE, R²

47. Какие метрики вы знаете для задач классификации?

✓ Ответ: Accuracy, Precision, Recall, F1-score, ROC-AUC

48. Какие методы регуляризации вы знаете?

✓ Ответ: L1, L2, Dropout, BatchNorm

49. Какие современные подходы к машинному обучению вы знаете?

✓ Ответ: AutoML, Federated Learning, GAN, Reinforcement Learning

50. Кейс-задание: Прогнозирование урожайности культур
Разработать модель прогнозирования урожайности культур на основе погодных условий, состава почвы и применения удобрений.

Задачи:

1. Выполнить предобработку данных (очистка, нормализация, кодирование).

2. Реализовать несколько моделей (например, линейная регрессия, дерево решений, случайный лес).

3. Оценить качество моделей (MAE, MSE, R²).

4. Выбрать лучшую модель и интерпретировать результаты.

5. Предложить рекомендации по повышению урожайности.

✓ Оценивается:

1. Корректность выполнения этапов
2. Правильность выбора метрик и моделей
3. Интерпретация результатов
4. Применение знаний на практике

Решение:

1. Исходные данные

Допустим, предоставлены следующие данные:

ТЕМПЕРАТУРА (°C)	ОСАДКИ (ММ)	АЗОТ (N)	ФОСФОР (P)	КАЛИЙ (K)	ТИП УДОБРЕНИЯ	УРОЖАЙНОСТЬ (Ц/ГА)
22	80	90	50	40	органическое	45
20	60	70	45	35	минеральное	38
...

2. Предобработка данных

2.1 Обработка пропусков

- Проверяем наличие пропущенных значений.
- Заполняем числовые признаки средними значениями (**SimpleImputer**).
- Для категориальных заменяем пропуски на **unknown**.

```
python
1 from sklearn.impute import SimpleImputer
2
3 numeric_features = ['Температура', 'Осадки', 'Азот', 'Фосфор', 'Калий']
4 categorical_features = ['Тип удобрения']
5
6 numeric_transformer = Pipeline(steps=[
7     ('imputer', SimpleImputer(strategy='mean')),
8     ('scaler', StandardScaler())])
9
10 categorical_transformer = OneHotEncoder(handle_unknown='ignore')
11
12 preprocessor = ColumnTransformer(
13     transformers=[
14         ('num', numeric_transformer, numeric_features),
15         ('cat', categorical_transformer, categorical_features)])
```

2.2 Нормализация числовых признаков

- Используем **StandardScaler** для нормализации данных.

2.3 Кодирование категориальных признаков

- Признак "Тип удобрения" кодируем с помощью **One-Hot Encoding**.

2.4 Разделение выборки

- Делим данные на обучающую и тестовую выборки в соотношении 80/20.

```
python
1 X = df.drop('Урожайность', axis=1)
2 y = df['Урожайность']
3
4 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

3. Реализация моделей машинного обучения

Модель 1: Линейная регрессия (python)

```
model_lr = Pipeline(steps=[('preprocessor', preprocessor), ('regressor', LinearRegression())])
```

```
model_lr.fit(X_train, y_train)
```

```
pred_lr = model_lr.predict(X_test)
```

Модель 2: Дерево решений (python)

```
model_dt = Pipeline(steps=[('preprocessor', preprocessor), ('regressor', DecisionTreeRegressor(random_state=42))])
```

```
model_dt.fit(X_train, y_train)
```

```
pred_dt = model_dt.predict(X_test)
```

Модель 3: Случайный лес (python)

```
model_rf = Pipeline(steps=[('preprocessor', preprocessor), ('regressor', RandomForestRegressor(n_estimators=100, random_state=42))])
```

```
model_rf.fit(X_train, y_train)
```

```
pred_rf = model_rf.predict(X_test)
```

4. Оценка качества моделей

Используем три метрики:

- MAE — средняя абсолютная ошибка
- MSE — среднеквадратичная ошибка
- R^2 — коэффициент детерминации

```
python
1 from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
2
3 def evaluate(y_true, y_pred):
4     return {
5         'MAE': mean_absolute_error(y_true, y_pred),
6         'MSE': mean_squared_error(y_true, y_pred),
7         'R2': r2_score(y_true, y_pred)
8     }
9
10 results = {
11     'Линейная регрессия': evaluate(y_test, pred_lr),
12     'Дерево решений': evaluate(y_test, pred_dt),
13     'Случайный лес': evaluate(y_test, pred_rf)
14 }
```

Пример результатов:

МОДЕЛЬ	MAE	MSE	R^2
Линейная регрессия	2.1	6.8	0.75
Дерево решений	1.5	4.1	0.87
Случайный лес	1.2	2.9	0.91

✓ Вывод: Случайный лес показал лучшие результаты по всем метрикам.

5. Интерпретация результатов

Анализ важности признаков

Получаем важность признаков из модели случайного леса (python):

```
import shap
```

```
explainer = shap.TreeExplainer(model_rf.named_steps["regressor"])
```

```
shap_values = explainer.shap_values(model_rf.named_steps["preprocessor"].transform(X))
```

```
shap.summary_plot(shap_values, model_rf.named_steps["preprocessor"].transform(X), feature_names=X.columns)
```

Наиболее важные факторы:

1. Азот (N) – оказывает наибольшее влияние на урожайность.
2. Осадки – второй по значимости климатический фактор.
3. Калий (K) – положительно влияет на рост растений.
4. Тип удобрений (органические > минеральные) .
5. Температура – влияет, но меньше, чем другие факторы.

📌 6. Рекомендации по повышению урожайности

1. Удобрения :
 - Рекомендуется использовать органические удобрения , так как они дают больший эффект.
 - Увеличьте дозировку азота и калия согласно рекомендациям агронома.
2. Погода и полив :
 - Важно обеспечить оптимальное количество осадков или использовать системы точного орошения .
 - Избегать экстремальных температур путем подбора сортов или сроков посадки.
3. Почва :
 - Проведите анализ почвы перед посадкой.
 - Вносите недостающие элементы питания (особенно азот и калий).
4. Автоматизация и мониторинг :
 - Внедрите систему IoT-сенсоров для отслеживания состояния почвы и погодных условий.
 - Создайте цифровой инструмент для планирования применения удобрений.

📌 7. Графики и визуализации

- График предсказанных vs реальных значений (python):

```
import matplotlib.pyplot as plt
```

```
plt.figure(figsize=(10,6))
```

```
plt.scatter(y_test, pred_rf, alpha=0.7)
```

```
plt.plot([min(y_test), max(y_test)], [min(y_test), max(y_test)], color='red')
```

```
plt.xlabel('Реальная урожайность')
```

```
plt.ylabel('Предсказанная урожайность')
```

```
plt.title('Random Forest: Реальные vs Предсказанные значения')
```

plt.grid()

plt.show()

SHAP-график важности признаков (python):

```
shap.summary_plot(shap_values, model_rf.named_steps["preprocessor"].transform(X), feature_names=X.columns)
```

🔧 8. Возможности дальнейшего развития

- Добавление новых данных :
 - Использование спутниковых данных (NDVI, индекс растительности).
 - Интеграция с IoT-сенсорами на полях.
- Автоматизация и масштабирование :
 - Разработка веб-приложения для фермеров.
 - Автоматическая система рекомендаций по удобрениям и поливам.
- Применение нейронных сетей :
 - Прогнозирование урожайности на основе временных рядов и изображений.

☐ 9. Отчет по выполнению этапов

ЭТАП	ВЫПОЛНЕНО?	КОММЕНТАРИЙ
Загрузка и очистка данных	✓	Обработаны пропуски, удалены выбросы
Нормализация	✓	StandardScaler
Кодирование категорий	✓	One-Hot Encoding
Разделение выборки	✓	80% обучение / 20% тестирование
Выбор и обучение моделей	✓	Линейная регрессия, дерево решений, случайный лес
Оценка моделей	✓	MAE, MSE, R ²
Выбор лучшей модели	✓	Случайный лес
Анализ важности признаков	✓	SHAP
Рекомендации	✓	По типу удобрений, составу почвы и погоде