

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Плешаков Владимир Александрович
Должность: Врио ректора
Дата подписания: 27.08.2026 09:18:58
Уникальный программный ключ:
cf3461e360a6506473208a5cc93ea97a503bcf72

Приложение к фондам оценочных средств
учебной дисциплины «Автоматизированные
системы управления технологиями
производства»

**Лист внесения дополнений и изменений
в фонд оценочных средств учебной дисциплины
на 2026 – 2027 учебный год**

Фонд оценочных средств пересмотрен на заседании кафедры, протокол № 11 от 7.05.2026 г.

В фонд оценочных средств вносятся следующие изменения:

1. Добавлена лабораторная работа по искусственному интеллекту (Приложение 1).

Составитель изменений и дополнений:

Зав. кафедрой экономики, анализа и
информационных технологий,
д.т.н., доцент



А.В.Тиньгаев

Зав. кафедрой экономики, анализа и
информационных технологий,
д.т.н., доцент



А.В.Тиньгаев

Лабораторная работа

Тема: Прогнозирование качества молочной продукции методами машинного обучения

Дисциплина: Автоматизированные системы управления технологиями производства

Среда: Orange Data Mining

Цель работы

Освоить методы построения предсказательных моделей для автоматизированного контроля качества пищевых продуктов на примере оценки качества молока.

Исходные данные

Набор данных взят из **открытого датасета Kaggle «Milk Quality Prediction»** и содержит 1059 записей с лабораторными показателями молока.

Скачать датасет можно здесь:

<https://www.kaggle.com/datasets/prudhvignv/milk-grading>

Состав показателей в датасете (7 признаков + целевой атрибут):

- pH (числовой) — pH-значение молока
- Temperature (числовой) — температура молока сразу после дойки
- Taste (бинарный: 0/1) — вкус (удовлетворяет условиям = 0)
- Odor (бинарный: 0/1) — запах (удовлетворяет условиям = 0)
- Fat (бинарный: 0/1) — жирность (удовлетворяет условиям = 0)
- Turbidity (бинарный: 0/1) — мутность (удовлетворяет условиям = 0)
- Colour (числовой, 0–255) — цвет молока в градациях серого
- Grade (**целевой**) — класс качества молока: 0 (плохое), 1 (среднее), 2 (хорошее)

Датасет полностью подготовлен к работе: нет пропущенных значений, имеется баланс классов (все три класса качества представлены равномерно).

Ход работы

Этап 1. Загрузка данных и первичный анализ

1. Запустите Orange Data Mining.
2. Добавьте виджет File → загрузите скачанный файл milknew.csv.
3. Подключите Data Table — убедитесь, что колонка Grade распознана как **целевая переменная** (Target).
4. Добавьте Distributions — проанализируйте распределение классов качества (Grade):
 - Каково процентное соотношение плохого, среднего и хорошего молока?
 - Сбалансирован ли датасет по классам?
5. Добавьте Box Plot — постройте распределение числовых признаков pH, Temperature, Colour в разрезе классов качества.

Вопрос: Какой из трёх числовых параметров, по вашему мнению, сильнее всего влияет на итоговый класс качества?

Этап 2. Предобработка данных для моделей

1. Добавьте Select Columns — укажите:

- **Input features:** все 7 параметров (pH, Temperature, Taste, Odor, Fat, Turbidity, Colour)
- **Target:** Grade
- 2. Добавьте Preprocess → выберите Normalize для **всех числовых признаков** (pH, Temperature, Colour).
Зачем? — Нормализация (Min-Max scaling) приводит значения разных признаков к единому диапазону [0, 1], что критически важно для таких моделей, как KNN или SVM.
- 3. **Важно:** бинарные признаки (Taste, Odor, Fat, Turbidity) оставьте без нормализации — они уже в формате 0/1.

Этап 3. Обучение первой модели (Логистическая регрессия)

1. Соедините Preprocess → Test & Score.
2. В виджете Test & Score:
 - Метод валидации: **Cross-validation (k=5)**
 - Добавьте алгоритм Logistic Regression
3. Нажмите Apply.

Запишите в отчёт: **Accuracy (CA), F1 (по каждому классу), AUC.**

Этап 4. Улучшение модели (Случайный лес)

1. В Test & Score добавьте Random Forest (для начала: число деревьев = 10, остальные параметры по умолчанию).
2. Запустите повторную валидацию.
Сравните метрики с логистической регрессией. Что изменилось и почему?

Важно: В реальном исследовании по классификации молока на основе этого датасета **Random Forest и Gradient Boosting** показали наилучшие результаты в предсказании качества по сравнению с другими алгоритмами.

Этап 5. Оценка важности признаков (Feature Importance)

1. Добавьте виджет Rank после Preprocess.
2. В качестве метода ранжирования выберите Information Gain.
Определите **топ-3 наиболее значимых показателя** для прогноза класса качества молока.

Связь с реальным производством: Представьте, что вы внедряете эту модель на молокозаводе. Какие именно датчики или экспресс-анализы нужно добавить в технологическую линию, чтобы собирать данные по этим трём параметрам в реальном времени?

Этап 6. Сравнение нескольких алгоритмов

Добавьте в Test & Score дополнительные модели, перечисленные ниже, и сравните их производительность:

Добавляемые модели	Краткое описание
k-Nearest Neighbors (k=5)	Классификация на основе «соседей»
Neural Network (один скрытый слой)	Простая нейросеть

Добавляемые модели	Краткое описание
Naive Bayes	Байесовский классификатор

Задание:

Определите, какая из всех протестированных моделей (включая Logistic Regression и Random Forest) показывает **наилучшие значения Accurasy и F1**.

В аналогичном научном исследовании по классификации мяса, проведённом в Orange, **SVM и нейронные сети** показали наилучшую точность вплоть до 100%. Возможно, и на ваших данных какой-то алгоритм окажется лидером.

Этап 7. Построение confusion matrix и анализ ошибок

1. Добавьте виджет Confusion Matrix после Test & Score.
2. Выберите в матрице модель-победитель (ту, что с лучшей точностью).
3. Визуально оцените:
 - Какие классы качества путаются между собой чаще всего?
 - Можно ли это объяснить с технологической точки зрения (например, близость параметров молока среднего и хорошего качества)?

Отчёт

Структура отчёта:

1. **Цель работы** (1 абзац).
2. **Скриншот рабочей схемы** в Orange (должна включать все этапы от загрузки до сравнения моделей).
3. **Таблица сравнения моделей** (заполняется по результатам этапов 3–6):

Модель	Accurasy	F1 (класс 0)	F1 (класс 1)	F1 (класс 2)	AUC (macro)
Logistic Regression					
Random Forest					
k-NN (k=5)					
Neural Network					
Naive Bayes					

4. **Топ-3 важнейших параметра** для оценки качества молока (по результатам этапа 5) с кратким обоснованием, почему именно они критичны с точки зрения технологии производства.
5. **Вывод:** как предложенная модель может быть интегрирована в АСУ ТП молокозавода (например, для автоматической сортировки партий молока или сигнализации оператору при отклонениях).

Контрольные вопросы (для защиты/собеседования)

1. Почему в данном датасете не требовалось заполнять пропуски (импутация)?
2. Какую роль играет значение `Color` (оттенок серого) в оценке качества молока и как это можно измерить автоматически?
3. Чем отличается Ассигасы от F1-меры и в каких случаях F1 информативнее?
4. Какие дополнительные датчики вы бы предложили добавить на молокозаводе, чтобы ещё точнее прогнозировать качество?

Подсказки для студента

- **Если нет доступа к Kaggle:** можно использовать файл `words-food.xlsx` из официального обучающего видео Orange (ссылка на скачивание есть в описании к видео на YouTube-канале Orange Data Mining).
- **Углублённый вариант задания:** Попробуйте обучить модель **Gradient Boosting** (если он доступен в вашей версии Orange) и сравните его эффективность с Random Forest.
- **Связь с другими задачами пищевой промышленности:** Обратите внимание на подходы к обнаружению дефектов в производстве пищевых масел, мяса или хлебобулочных изделий, где используются похожие методы машинного обучения.

Если нужна будет помощь с поиском конкретных CSV-файлов или других датасетов (например, для других продуктов — хлеб, масло, мясо, соки) — сообщите, я смогу подобрать подходящие варианты.