ПРАКТИЧЕСКАЯ РАБОТА №7
Применение методов кластеризации и снижения размерности данных

Цель работы

Освоить методы и инструменты по теме «применение методов кластеризации и снижения размерности данных» на уровне самостоятельного магистерского исследования; получить измеримый результат и использовать его для последовательного развития индивидуального семестрового проекта.

Краткие теоретические сведения

Кластеризация ищет группы без целевых меток, а снижение размерности формирует компактное представление признаков. Эти методы полезны для разведочного анализа, визуализации, обнаружения структуры и аномалий, но их результат требует осторожной интерпретации.

Связь с предыдущими практическими работами

В ПЗ №7 необходимо использовать результаты предыдущих этапов индивидуального проекта. Исследование скрытой структуры данных проекта. Не заменяйте ранее подготовленные данные или постановку задачи без обоснования; изменения фиксируйте в отчёте.

Методические указания

Работа выполняется магистрантом самостоятельно. Все решения должны быть обоснованы: недостаточно получить программный результат, необходимо объяснить выбор данных, метода, метрик и интерпретировать полученные результаты.

Код должен быть воспроизводимым: фиксируйте random_state/seed там, где это применимо, подписывайте этапы Notebook и не удаляйте результаты ключевых экспериментов. Используйте только данные, которые разрешено обрабатывать в выбранной среде.

Практическая работа является частью сквозного индивидуального семестрового проекта. Результаты предыдущих работ повторно не создаются, а используются как вход для текущего этапа. Если метод текущей недели объективно не соответствует предметной задаче, магистрант выполняет небольшой связанный эксперимент и отдельно обосновывает, почему этот компонент не следует включать в итоговую архитектуру.

Рекомендуемый порядок выполнения: 1) сформулировать исследовательский вопрос; 2) подготовить входные данные; 3) реализовать базовый вариант; 4) провести контролируемый эксперимент; 5) сравнить результаты; 6) интерпретировать ошибки; 7) зафиксировать вклад текущей работы в семестровый проект.

Задания

Задание 1. Выберите набор признаков проекта, пригодный для кластеризации.

Задание 2. Примените K-Means или другой метод кластеризации и обоснуйте число кластеров.

Задание 3. Используйте PCA для представления данных в 2D/3D или снижения размерности.

Задание 4. Сравните кластеры с известными классами, если метки доступны, не используя метки при обучении кластеризации.

Задание 5. Сделайте вывод, даёт ли unsupervised-анализ дополнительную информацию для проекта.

Минимальный программный пример

```
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
X_scaled = StandardScaler().fit_transform(X)
clusters = KMeans(n_clusters=2, random_state=42, n_init="auto").fit_predict(X_scaled)
X_2d = PCA(n_components=2).fit_transform(X_scaled)
print(X_2d[:5], clusters[:5])
```

Пример предназначен только для демонстрации принципа. Магистрант должен адаптировать код к собственной задаче, данным и структуре проекта; простое копирование примера не считается выполнением индивидуального задания.

Пример выполнения

Кластеризация URL-признаков

Числовые признаки стандартизируются, затем K-Means обучается без использования label. PCA применяется только для визуализации структуры. После кластеризации допустимо сопоставить полученные кластеры с истинными классами для интерпретации, но нельзя выдавать это за обучение с учителем.

Индивидуальное задание по семестровому проекту

Исследование скрытой структуры данных проекта

1. Выполните текущий этап исключительно на теме собственного проекта.
2. Добавьте минимум одно самостоятельное решение, которого нет в приведённом примере.
3. Зафиксируйте минимум один альтернативный вариант и объясните, почему он принят или отклонён.
4. Сохраните артефакт текущего этапа: Notebook, таблицу, модель, промпт, схему workflow или матрицу рисков — в зависимости от темы работы.
5. В конце раздела напишите 5–8 предложений о том, как результат текущей ПЗ изменил или уточнил архитектуру индивидуальной системы.

Требования к отчёту

Отчёт должен содержать: тему и цель; краткое теоретическое обоснование; описание исходных данных; ход выполнения заданий; ключевые фрагменты кода и результаты их выполнения; таблицы/графики при необходимости; анализ ошибок; индивидуальный раздел семестрового проекта; вывод. Рекомендуемый объём — 5–8 страниц без искусственного увеличения объёма скриншотами.

К защите магистрант должен уметь объяснить каждую существенную строку собственного кода, происхождение данных, причины выбора метода и смысл полученных метрик. Невоспроизводимый результат без объяснения не считается завершённым экспериментом.

Результат практической работы

После выполнения ПЗ №7 магистрант представляет отчёт и обновлённый артефакт индивидуального проекта. Результат текущей работы должен быть пригоден для использования на следующем этапе курса.

Контрольные вопросы

1. Какова основная исследовательская задача практической работы №7?

2. Какие исходные данные и предположения используются в работе?
3. Почему выбранный метод соответствует поставленной задаче?
4. Какие альтернативные методы можно было бы использовать?
5. Какие метрики или критерии позволяют оценить результат?
6. Какие источники ошибок или смещения возможны?
7. Как обеспечить воспроизводимость эксперимента?
8. Какие риски информационной безопасности необходимо учитывать?
9. Как результаты данной работы используются в индивидуальном семестровом проекте?
10. Какой следующий эксперимент логично выполнить для улучшения результата?

Список источников

1. Бессмертный И. А., Авдюшина А. Е., Кугаевских А. В., Королева Ю. А., Рущенко Н. Г. Системы искусственного интеллекта: методическое пособие. – Санкт-Петербург: Университет ИТМО, 2024. – 281 с.

2. Соснило А. И. Атлас искусственного интеллекта для бизнеса и власти: учебное пособие. – Санкт-Петербург: Университет ИТМО, 2022. – 98 с.

3. Сидоркина И. Г. Системы искусственного интеллекта: учебное пособие. – Москва: КНОРУС, 2018. – 248 с.

4. Баланов А. Н. Искусственный интеллект. Понимание, применение и перспективы: учебник для вузов. – Санкт-Петербург: Лань, 2024. – 312 с.

5. Калинин М. О., Крундышев В. М. Основы искусственного интеллекта. Безопасность искусственного интеллекта: учебное пособие. – Санкт-Петербург, 2024. – 220 с.

6. Андреянов Н. В., Евдокимова Т. С., Павлов А. Д. [и др.]. Методы искусственного интеллекта: учебное пособие. – Казань: КНИТУ-КАИ, 2024. – 392 с.