

Портфоліо для DS/ML спеціаліста

Для чого нам портфоліо

Портфоліо демонструє ваші практичні навички, реальні проєкти та здатність вирішувати бізнес-задачі за допомогою аналізу даних і машинного навчання. Це ключовий елемент для потенційних роботодавців або клієнтів, оскільки воно наочно показує, як ви застосовуєте теорію на практиці.

Що обов'язково включити:

1. **Постановка задачі:** Чітке формулювання проблеми і мета проєкту.
2. **Дані:** Посилання на джерела даних і їх опис.
3. **Методологія:** Алгоритми, методи машинного навчання та інструменти.
4. **Результати:** Оцінка моделей, графіки та висновки.
5. **Код:** Чистий та добре задокументований код у GitHub або іншій платформі.
6. **Візуалізації:** Графіки для пояснення аналізу та результатів.

Таке портфоліо покаже ваші технічні та аналітичні навички, а також вашу здатність вирішувати складні задачі на практиці.

**Портфоліо — можливість
довести наявність досвіду
без робочого досвіду в
компаніях**

Опис портфоліо для ML Data Science проєкту

Проект з Data Science для вашого портфоліо має бути чітко структурованим і показувати ваші навички роботи з реальними задачами. Ви можете розміщувати опис і результати у Jupyter Notebook, в GitHub Readme, додатках типу Streamlit або на вашому особистому сайті. Головне аби в проєкті можна було побачити всі основні компоненти!

Основні компоненти опису мають включати:

1. **Постановка задачі:**
 - Чітке формулювання бізнес-задачі та технічного завдання. Вкажіть, що саме ви хочете досягти та яку проблему вирішуєте.
2. **Джерела та опис даних:**
 - Посилання на використані набори даних і їх детальний опис. Це включає структуру даних, кількість записів, типи змінних та можливі проблеми з даними (пропуски, аномалії тощо).
3. **Ваш підхід до вирішення:**
 - Опишіть методологію і використовувані інструменти. Це можуть бути алгоритми машинного навчання, статистичні методи, бібліотеки для роботи з даними та інші засоби, які ви застосували.
4. **Отримані результати і висновки:**
 - Опишіть результати вашого аналізу. Покажіть, наскільки ваша модель або підхід допомогли вирішити поставлену задачу. Додайте свої висновки та пропозиції щодо покращення або подальших кроків.

Структура проєкту

1. **Трансформація даних:**
 - Опишіть процес збору та обробки даних, включаючи використання баз даних або запитів для отримання інформації.
2. **Візуалізація даних:**
 - Включіть графіки, діаграми та інші візуалізації, які допоможуть пояснити результати аналізу. Візуалізація підкреслює ваш підхід до роботи з даними та представлення результатів.
3. **Статистика і машинне навчання:**
 - Покажіть, як ви використовували статистичні методи і моделі машинного навчання. Це можуть бути алгоритми класифікації, регресії, кластеризації або інші підходи, залежно від задачі.
4. **Інтерактивні елементи:**
 - Якщо можливо, створіть інтерактивний інтерфейс (наприклад, за допомогою Streamlit), де користувачі можуть взаємодіяти з даними, обирати опції для відображення або навіть завантажувати власні дані для аналізу.

Такий проєкт допоможе вам продемонструвати практичні навички в аналізі даних та розв'язку бізнес-задач за допомогою методів машинного навчання.

Які проєкти включити в портфоліо

Акцент на тих, які покажуть навички для пріоритетних для вас вакансій.

Як перетворити ДЗ з курсу у проєкт в портфоліо

- Додайте опис даних і посилання на них, якщо цього там немає.
- Додайте опис загальної задачі на цей проєкт.
- Перетворіть формулювання завдань у постановку задачі самому собі як один з етапів дослідження.
- Розмістіть або кілька ноутбуків зі зрозумілими назвами файлів в одному github репозиторії, або по проєкту на репозиторій.
- Додайте опис або в readme.md репозиторія в гітхабі, або в самому ноутбуці. Якщо файлів в репозиторії кілька, readme.md точно потрібна для навігації по проєктам.
- Можна також розмістити обрані проєкти в Colab, переконавшись що файл відкривається за посиланням і надати посилання у резюме.
- Можна також розмістити проєкти з використанням streamlit і задеплоїти їх.
- Можна більшість коду винести в .py файли, а в ноутбуках лишити лише візуальну частину.

З яких завдань в курсі найкраще створити портфоліо

Всі наші задачі, розглянуті в курсі є [тут](#).

- Australia Rain Prediction (дані weatherAUS) - показати як використовуєте різні моделі, порівнюєте, оцінюєте якість, будуєте пайплайни.
- Аналіз часових рядів в Store Item Demand задачі (чим більше рядів включите — тим ліпше)
- Mid-term project - EDA, чистка даних, моделювання, експерименти, оцінювання якості
- NLP проєкт на Insincire Questions Prediction (сентимент аналіз)
- Фінальний проєкт на Quara Duplicate Questions Detection

Але в першу чергу відштовхуємось від свого запиту і напряму пошуку роботи.

Можна просто всі ноутбуки залити на гітхаб і оформити один репозиторій - теж буде WOW.

**В ідеалі портфоліо має викликати
WOW-ефект!**

Приклади

- Репозиторій має навчальну мету, але чудово структурований і має детальний опис https://github.com/omerbsezer/LSTM_RNN_Tutorials_with_Demo
- Гарний окремий ноутбук з чітким описом <https://gist.github.com/pb111/cc341409081dffa5e9eaf60d79562a03>
- Проєкт на streamlit <https://llm-examples.streamlit.app/>
 - Lifehack: ви можете форкнути (скопювати) проєкт на streamlit, додати туди власної творчості і на основі іншого швидко створити власний проєкт

Теми

Ви можете поставити теми до репозиторія, тоді його зможуть знаходити інші люди.

Як додати теми для гітхаб репозиторія.

About

Script to obtain PPG Bilateral lending between 2 countries using World Bank API

python

excel

data-analysis

india

financial-analysis

world-bank

world-bank-api

data-analysis-python

data-analysis-project

debt-statistics

internation-debt-statistics

Проекти для натхнення

За темами ви можете також знайти проекти для натхнення. Ось наприклад, проекти з темою Deep Learning: <https://github.com/topics/deep-learning>