

Programowanie w R: Projekt

Celem projektu jest analiza bazy danych materiałów wykorzystywanych w tworzeniu baterii.

[Materials Project](#) to inicjatywa naukowa Departamentu Energii USA, której celem jest dostarczanie otwartych danych i narzędzi do analizy materiałów. Projekt udostępnia szczegółowe informacje na temat struktury, właściwości i potencjalnych zastosowań tysięcy materiałów. Umożliwia to badaczom szybkie wyszukiwanie nowych materiałów oraz przewidywanie ich właściwości, co przyczynia się do rozwoju takich dziedzin jak elektronika, fotonika czy magazynowanie energii. Platforma bazuje na zbieraniu już obliczonych własności materiałów i udostępnianiu ich społeczności naukowej.

Jednym z kluczowych zbiorów danych dostępnych w ramach **Materials Project** jest baza danych dotycząca materiałów używanych w bateriach, która zawiera informacje o ich składzie chemicznym i parametrach wydajnościowych. Tabela ta zawiera takie pola jak:

- **Battery ID:** Identyfikator baterii.
- **Battery Formula:** Wzór chemiczny materiału baterii.
- **Working Ion:** Główny jon, który odpowiada za transport ładunku w baterii.
- **Formula Charge:** Wzór chemiczny materiału baterii w stanie naładowanym.
- **Formula Discharge:** Wzór chemiczny materiału baterii w stanie rozładowanym.
- **Max Delta Volume:** Zmiana objętości w % dla danego kroku napięcia za pomocą wzoru: $\max(\text{charge}, \text{discharge}) / \min(\text{charge}, \text{discharge}) - 1$.
- **Average Voltage:** Średnie napięcie dla poszczególnego kroku napięcia.
- **Gravimetric Capacity:** Pojemność grawimetryczna, czyli ilość energii na jednostkę masy (mAh/g).
- **Volumetric Capacity:** Pojemność wolumetryczna, czyli ilość energii na jednostkę objętości (mAh/cm³).
- **Gravimetric Energy:** Gęstość energii w odniesieniu do masy baterii (Wh/kg).
- **Volumetric Energy:** Gęstość energii w odniesieniu do objętości baterii (Wh/L).
- **Atomic Fraction Charge:** Udział atomowy składników w stanie naładowanym.
- **Atomic Fraction Discharge:** Udział atomowy składników w stanie rozładowanym.
- **Stability Charge:** Wskaźnik stabilności materiału w stanie naładowanym.
- **Stability Discharge:** Wskaźnik stabilności materiału w stanie rozładowanym.

- **Steps:** Liczba odrębnych kroków napięcia od pełnego naładowania do rozładowania, oparta na stabilnych stanach pośrednich.
- **Max Voltage Step:** Maksymalna bezwzględna różnica między sąsiednimi krokami napięcia.

Wymagania

Raport z analizy danych powinien być napisany z wykorzystaniem języka **RMarkdown** i paczki **knitr**. Źródła i pliki wynikowe powinny zostać udostępnione w ramach konta **GitHub** studenta. Oprócz pliku źródłowego (**.Rmd**) w repozytorium, należy również upublicznić raport w wersji html na stronie **github.io** powiązanej z projektem. Na [stronie przedmiotu](#) znajduje się formularz do wgrania raportu w formacie html oraz do podania adresu repozytorium.

Na początku pliku powinna znajdować się automatycznie wypełniona **data generacji dokumentu** oraz **spis treści** pozwalający przejść do najważniejszych sekcji. Ponadto **raport powinien zaczynać się od rozdziału podsumowującego całą analizę**, streszczającego najważniejsze spostrzeżenia analityka (ang. *executive summary*). Należy tu podkreślić, że właśnie zrozumienie danych, czytelna prezentacja wyników oraz stosowanie się do podstawowych zasad wizualizacji danych będą, obok technicznej strony raportu, podstawą do oceny. Po wstępie, raport powinien zawierać m.in. następujące elementy:

1. Kod wyliczający wykorzystane biblioteki.
2. Kod zapewniający powtarzalność wyników przy każdym uruchomieniu raportu na tych samych danych.
3. Kod pozwalający wczytać dane z plików.
4. Kod przetwarzający brakujące dane.
5. Sekcję podsumowującą rozmiar zbioru i podstawowe statystyki.
6. Szczegółową analizę wartości atrybutów (np. poprzez prezentację rozkładów wartości).
7. Sekcję sprawdzającą korelacje między zmiennymi; sekcja ta powinna zawierać jakąś formę graficznej prezentacji korelacji.
8. Interaktywne wykresy lub animacje.
9. Sekcję próbującą podsumować najważniejsze trendy w badaniu materiałów do baterii.
10. Sekcję próbującą przewidzieć dalsze cechy i właściwości baterii, np. predykcja energii lub stabilności nowego materiału, biorąc pod uwagę inne atrybuty, takie jak średni woltaż, maksymalne zmiany objętości, czy pojemność.

Jeśli analityk uzna to za stosowne, powyższe punkty mogą być wykonane w innej kolejności. **Analityk nie musi, a nawet nie powinien, ograniczać się do powyższych punktów.** Wszelkie

dodatkowe techniki analizy danych, wizualizacje, spostrzeżenia będą pozytywnie wpływały na ocenę. Jeśli analityk uzna to za stosowne, może rozszerzyć zbiór danych o dodatkowe, ogólnodostępne, informacje (np. inne badania nad materiałami, prezentujące inne właściwości). Ewentualne konkluzje, znalezione zależności warto potwierdzić dokonując sprawdzenia istniejących wyników badań w literaturze naukowej (np. na [Google Scholar](https://scholar.google.com/)).

Termin oddania projektu: 10.12.2024