

Material project - band_gap prediction

Krzysztof Kubit

Kacper Cygan

Krzysztof Klimczyk

07.04



Opis danych

Wykorzystujemy Pythonowe API (mp-api) do dostępu do danych.

Uwzględniamy tylko materiały składające się maksymalnie z trzech różnych pierwiastków.

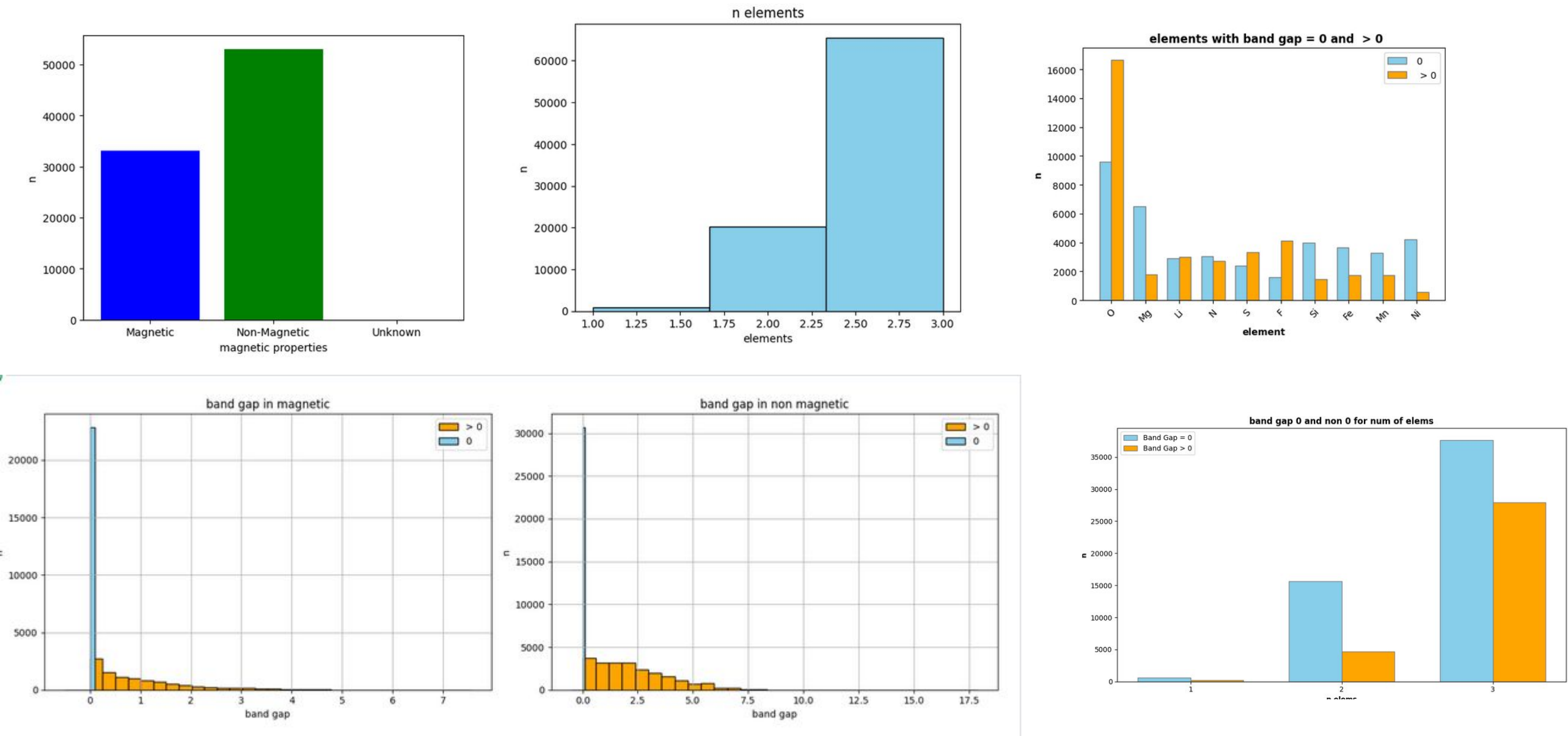
Dostępnych jest ponad 80 tysięcy różnych materiałów.

Problem z wczoraj - API domyślnie zwraca obiekty MaterialProjectDocument, których nie da się załadować bezpośrednio do DataFrame, a trzeba je przekształcić, co wymaga dużej ilości RAM-u. Ustawienie parametru `use_document_model=False` rozwiązuje problem i upraszcza formę danych.

Wstępnie wybrane cechy

- nelements
- efermi
- energy_per_atom
- formation_energy_per_atom
- energy_above_hull
- equilibrium_reaction_energy_per_atom
- is_magnetic
- ordering
- total_magnetization
- total_magnetization_normalized_vol
- formation_energy_per_atom

Statystyki



Automatminer

Narzędzie do automatycznego tworzenia
kompletnych pipeline'ów uczenia maszynowego
włączając automatyczną selekcję cech.

Porównanie wyników z naszym rozwiązaniem.

Problem - pakiet nie był aktualizowany od
jakiegoś czasu i może być problem z
uruchomieniem (Python 3.6).

Wybrane modele

Random Forest Regression

Gradient Boosting Regression?

Support Vector Regression?