

Specyfikacja wymagań do projektu „scikit-learn”

**Issue : „Better error message when passing un-sortable
data to the Encoders #12621”**

Spis Treści :

1. Wstęp
 - 1.1 Streszczenie
 - 1.2 Cel
 - 1.3 Zakres
2. Ogólny opis:
 - 2.1 Opis oprogramowania
 - 2.2 Opis problemu
 - 2.3 Środowisko
3. Szczegółne wymagania
4. Harmonogram prac

1. Wstęp

1.1 Streszczenie

Specyfikacja została stworzona na potrzeby stworzenia i wprowadzenia usprawnienia do projektu open-source umieszczonego na platformie git-hub.

1.2 Cel

Celem jest wprowadzenie sprawnego usprawnienia do gotowego projektu o otwartej polityce kodu w celu usprawnienia funkcjonowania oprogramowania. Proces ten pozwoli na zdobycie nowych umiejętności w zakresie technologii informatycznych, programistycznych, pracy na rozbudowanym kodzie oraz pracy z repozytorium zewnętrznym.

1.3 Zakres

Zakres wprowadzenia usprawnienia obejmuje wdrożenie w dokumentacje oprogramowania oraz dobre zrozumienie jego składni i funkcjonalności. Obejmuje również część wprowadzenia usprawnienia oraz jego testowanie,

2. Ogólny Opis

2.1 Opis oprogramowania

Scikit-learn to python module do machine learningu zbudowany na SciPy oraz Dystrybuowany pod licencją 3-Clause BSD. Projekt wystartowany został w 2007 przez David'a Cournapeau jako projekt na Google Summer of Code i od tamtego czasu wielu ochotników chętnie do niego dołącza.

Strona : **<http://scikit-learn.org>**

2.2 Opis problemu

„Better error message when passing unsortable data to the Encoders #12621”

Błąd wyświetlany podczas próby wykonania funkcji `LabelEncoder().fit_transform()` na ciągu zmiennych posiadających nieoczekiwaną zmienną (np. (1,2,3,4,s,54,61)) był nieczytelny. Zasugerowano, aby : „Nawiązując do wiadomości o błędzie byłoby dobrze, aby mieć bardziej określoną i informującą bardziej szczegółowo o błędzie wiadomość. Można przechwycić błędy podczas sortowania, zbadać je, czy faktycznie związane są z sortowaniem, i wtedy zwrócić informację która w większym zakresie informacji przekazuje wiedzę na temat błędu.

2.3 Środowisko

scikit-learn wymaga:

Python (≥ 3.4)

NumPy ($\geq 1.8.2$)

SciPy ($\geq 0.13.3$)

3. Szczególne wymagania

Wprowadzenie usprawnienia dotyczącego zwracania bardziej szczegółowych błędów na temat problemu występującego podczas próby użycia funkcji

`LabelEncoder().fit_transform()` na ciągu zmiennych posiadających nieoczekiwaną zmienną. Prawdopodobnie błąd występuje podczas sortowania przy wywołaniu `fit_transform()`, gdyż program nie ma możliwości porównania zmiennych typów np. `Int` ? `str`.

Wprowadzenie testów sprawdzających poprawność funkcjonowanie usprawnienia

Commit usprawnienia i jego poprawne funkcjonowanie

4. Harmonogram

26.11 - 3.12	Zapoznanie się z projektem i funkcjonalnościami
3.12 - 10.12	Dalsze wdrożenie się w projekt
10.12 - 17.12	Zapoznanie się z problemem i zbadanie jego wpływu na funkcjonalność innych narzędzi
17.12 - 24.12	Podjęcie prac nad rozwiązaniem problemu
24.12 - 31.12	Testowanie i poprawa błędów
31.12 - 07.01	W przypadku nierozwikłania problemu. Dalsze prace nad problemem.
07.01 - 14.01	Ponowne testowanie i praca z błędami
14.01 - 21.01	Oddanie i commit wprowadzonych poprawek i testów