

POLITECHNIKA WROCŁAWSKA

WYDZIAŁ ELEKTRONIKI

KIERUNEK: Automatyka i Robotyka (AIR)
SPECJALNOŚĆ: Technologie informacyjne w systemach
automatyki (ART)

PRACA DYPLOMOWA INŻYNIERSKA

Implementacja systemu identyfikacji
rozbłysków gwiazdowych

Implementation of a Star Burst
Identification System

AUTOR:
Małgorzata Pietras

PROWADZĄCY PRACĘ:

Dr inż. Łukasz Jeleń,
Katedra Informatyki Technicznej

Spis treści

1	Wstęp	2
2	Wprowadzenie	3
2.1	Rozbłysk gwiazdowy	3
2.2	Dane obserwacyjne	3
3	Cel i zakres pracy	5
4	Zastosowane metody	6
4.1	Uruchomienie systemu	6
4.2	Pobieranie danych	7
4.3	Odnajdywanie rozbłysków gwiazdowych	7
5	Wyniki i wnioski	10
	Literatura	16

Rozdział 1

Wstęp

Proces automatyzacji ma zastosowanie w wielu dziedzinach życia: od sektora usługowego, poprzez linie produkcyjne, medycynę, aż do badań naukowych. W ramach przedstawionej pracy dyplomowej zaimplementowany został system identyfikacji rozbłysków gwiazdowych. Jest to przykład wykorzystywania automatyzacji w celu przyśpieszenia i uproszczenia analizy danych naukowych. W dobie masowych przeglądów nieba oraz licznych misji satelitarnych stale rośnie liczba otrzymywanych danych, których pojedyncze jednostki nie są w stanie należycie przeanalizować. Zastosowanie odpowiedniego oprogramowania oraz elementów uczenia maszynowego umożliwia pracę nad wieloma tysiącami gwiazd jednocześnie.

Praca inżynierska podzielona została na następujące rozdziały: we Wprowadzeniu znajduje się wyjaśnienie pojęcia rozbłysków gwiazdowych oraz przedstawienie dotychczasowych metod ich identyfikacji. W kolejnym rozdziale zdefiniowany został cel i zakres pracy wykonanej w ramach implementacji systemu identyfikacji rozbłysków gwiazdowych. Następnie opisane zostały wykorzystane metody, czyli sposób pobierania danych i ich analizy. W ostatnim rozdziale omówiono wyniki uzyskane w ramach pracy inżynierskiej oraz otrzymane wnioski.

Rozdział 2

Wprowadzenie

2.1 Rozbłysk gwiazdowy

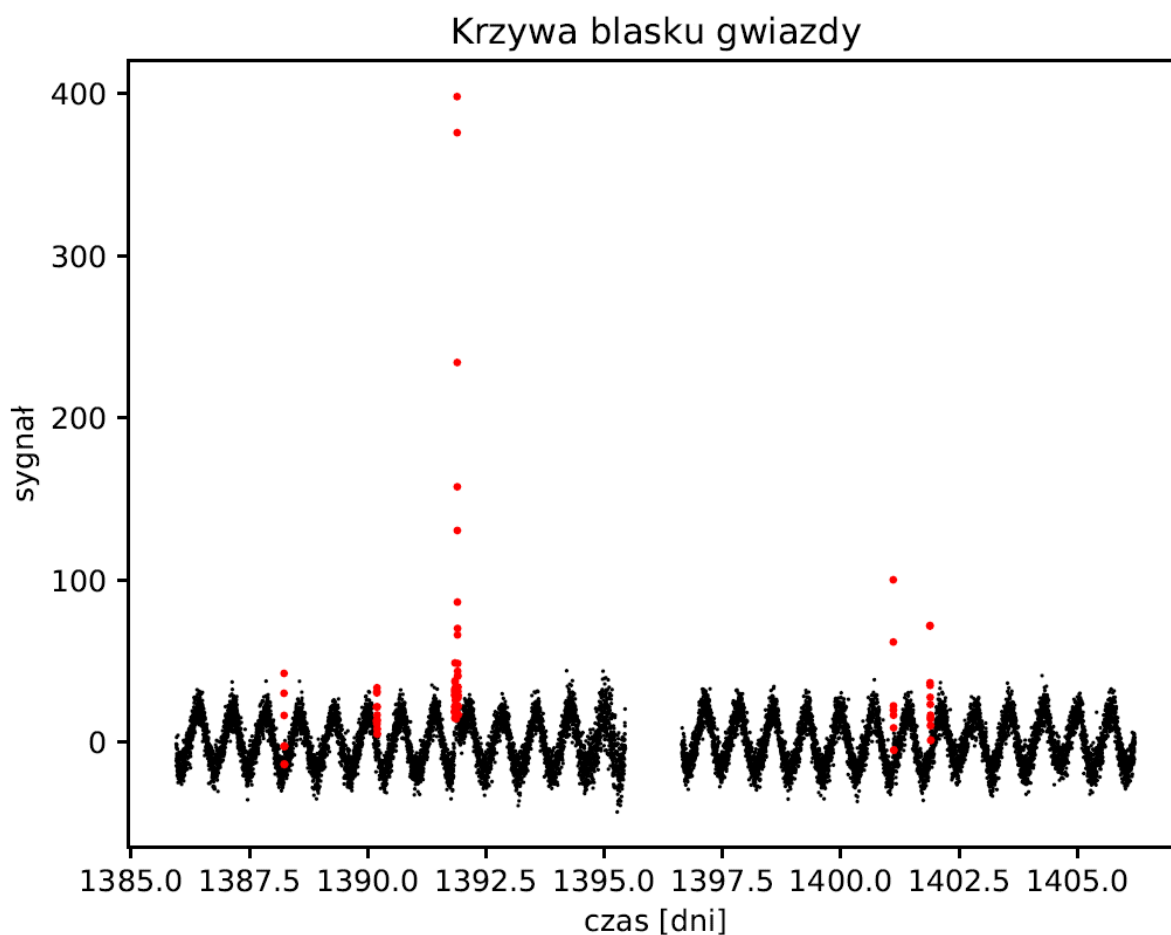
Rozbłysk gwiazdowy to gwałtowne zjawisko polegające na uwolnieniu ogromnej energii zmagazynowanej w polu magnetycznym gwiazdy. W ciągu kilkunastu-kilkudziesięciu minut zostaje wówczas wyemitowana energia rzędu 10^{26} J [1]. Odpowiada to energii wydzielonej podczas wybuchu biliona bomb atomowych. Większość energii emitowana jest w postaci promieniowania rentgenowskiego oraz ultrafioletowego. Część energii wydzielana jest także jako promieniowanie widzialne i podczerwone czy cząsteczki przyśpieszone do wysokich prędkości.

W praktyce rozbłyski gwiazdowe obserwowane są jako tymczasowy wzrost jasności gwiazdy. Zazwyczaj wynosi on ok. 1-2% uśrednionego sygnału, lecz w niektórych przypadkach może być nawet kilkukrotny. Współczesne teleskopy to bardzo precyzyjne urządzenia, których dokładność pomiarowa umożliwia detekcję bardzo słabych zjawisk. Zakładając, że ok. 10% ze wszystkich obserwowanych gwiazd (kilka milionów) wykazuje aktywność rozbłyskową, wówczas do przeanalizowania i wykrycia pozostaje kilkaset tysięcy rozbłysków. Badania te są bardzo istotne ze względu na wpływ rozbłysków na atmosfery egzoplanet oraz określenie prawdopodobieństwa pojawienia się super-rozbłysku na Słońcu.

2.2 Dane obserwacyjne

Dane obserwacyjne wykorzystane w pracy inżynierskiej pochodzą z satelity TESS (Transiting Exoplanet Survey Satellite) [3]. Głównym celem tej misji badawczej jest poszukiwanie planet orbitujących wokół gwiazd znajdujących się poza Układem Słonecznym. Dzięki dużej precyzji urządzeń pomiarowych - czterech matryc CCD (ang. charge-coupled device) o rozmiarach 2048×2048 pikseli - możliwe jest wykorzystanie danych również do detekcji rozbłysków gwiazdowych. TESS zbiera dane o gwiazdach z prawie całej sfery niebieskiej, dzięki czemu od startu misji (rok 2018) dostarczone zostały już informacje o zmienności ponad 200 000 tysięcy gwiazd. Gwiazdy uporządkowane zostały względem współrzędnych na sferze niebieskiej w 28 sektorów. Satelita ten prowadzi obserwacje w zakresie promieniowania widzialnego i podczerwonego (600 do 1000 nm). Informacje o zmianach jasności gwiazd udostępniane są w postaci tzw. krzywych blasku (rys. 2.1). Są to kolejne odczyty ja-

sności danej gwiazdy wraz z błędami pomiarowymi przedstawione w funkcji czasu. Rozdzielczość czasowa obserwacji wynosi dwie minuty.



Rysunek 2.1 Przykładowa krzywa blasku gwiazdy TIC246891813 z zaznaczonymi na czerwono rozbłyskami gwiazdowymi. Na osi poziomej przedstawiony jest czas, a na pionowej zmiany sygnału

Rozdział 3

Cel i zakres pracy

Celem pracy było stworzenie systemu identyfikacji rozbłysków gwiazdowych. System tego typu powinien być możliwie jak najbardziej zautomatyzowany, aby użytkownik mógł jak najłatwiej i najszybciej przeanalizować dużą liczbę danych. Przedstawiony w pracy inżynierskiej system składa się z dwóch głównych części. Pierwsza z nich służy do pobierania danych, a w drugiej przeprowadzana jest ich analiza oraz zapisywane są wyniki obliczeń. Całość systemu jest bardzo prosta do wywołania - wystarczy wykonać jedno polecenie w linii komend.

W opisywanym systemie identyfikacji rozbłysków gwiazdowych automatycznie pobierane są dane z satelity TESS. Ta misja została wybrana ze względu na bardzo dużą próbkę obserwacji (ponad 200 000 tysięcy gwiazd), dobrą dokładność pomiarów jasności oraz rozdzielczość czasową wynoszącą dwie minuty. Aby pobrać dane dotyczące danej gwiazdy wystarczy podać jej numer TIC (ang. TESS Input Catalog) oraz opcjonalnie sektor obserwacji. Aby umożliwić użytkownikowi pobranie większej liczby krzywych blasku wystarczy uruchomić program z kilkoma argumentami lub w pętli. Już na tym etapie wstępnie weryfikowana jest jakość danych pomiarowych i są one zapisywane do pliku tekstowego w formacie: czas, pomiar, błąd pomiaru.

Druga część zaimplementowanego systemu ma na celu przeprowadzenia analizy pobranej wcześniej krzywej blasku gwiazdy. Możliwe jest jednak wprowadzenie dowolnych danych (szeregu czasowego) wybranych przez użytkownika. Kolejne punkty pomiarowe sprawdzane są pod kątem detekcji rozbłysku gwiazdowego, czyli nagłego wzrostu jasności gwiazdy trwającego od kilkunastu minut do kilku godzin. Po zakończeniu obliczeń wyniki mogą być udostępnione użytkownikowi w formie graficznej lub/i zapisane na dysku w celu dalszej analizy.

Rozdział 4

Zastosowane metody

W celu stworzenia systemu identyfikacji rozbłysków gwiazdowych zastosowano język programowania Python 3.8. Oprogramowanie podzielone zostało na trzy programy: *start.py*, *data_download.py* i *find_flares.py*. Odpowiadają one kolejno za: uruchomienie całego systemu, pobieranie danych oraz odnajdywanie rozbłysków gwiazdowych. Poniżej opisane zostały dokładnie metody oraz pakiety wykorzystane z każdym z programów.

4.1 Uruchomienie systemu

Za poprawne uruchomienie całego systemu odpowiada program *start.py*. Został w nim zdefiniowany sposób używania oprogramowania, czyli wymagane i opcjonalne argumenty. Odpowiednia instrukcja wyświetlana jest użytkownikowi systemu (rys. 4.1).

```
Użycie: start.py [opcje] <pliki wejściowe>
Opcje:
-h, --help                wyświetlenie sposobu użycia
-n, --points=<n>          minimalna liczba punktów w rozbłysku (domyślnie: 5)
-l, --level=<l>           poziom detekcji rozbłysków (domyślnie: 3.0)
-d, --download=<d>       pobieranie danych dla gwiazdy o zadanym numerze
-s, --sector=<s>         pobieranie danych dla gwiazdy z tylko jednego sektora
--save                    zapisywanie wyników do plików z końcówką '_out'
--noplots                 brak wykresów
```

Przykłady:

```
python start.py TIC246891813-s3.dat
python start.py -d 246891813
python start.py -d 246891813 -s 3
python start.py -n 7 -l 2.0 --save --noplots -d 246891813 -s 3
```

Rysunek 4.1 Instrukcja używania oprogramowania do wyszukiwania rozbłysków gwiazdowych

Jedynym parametrem wymagany do uruchomienia oprogramowania jest po-

danie pliku (lub plików) z danymi zawierającego krzywą blasku gwiazdy. Jeżeli jednak użytkownik nie posiada jeszcze żadnych danych, istnieje możliwość automatycznego pobrania ich i następnie przeanalizowania. Wówczas należy wybrać opcję '-d' i podać numer TIC (TESS Input Catalog) wybranej gwiazdy. Pobieranie danych można zawęzić do jednego sektora obserwacji poprzez skorzystanie z opcji '-s'.

Innymi parametrami opcjonalnymi są: '-n' minimalna liczba punktów w rozbłysku i '-l', czyli poziom detekcji rozbłysków wyrażony jako wartość odchylenia standardowego analizowanych danych. Dodatkowo użytkownik może zdecydować w jaki sposób prezentowane będą wyniki obliczeń. Domyślnie są one wyświetlane w postaci wykresów krzywej blasku gwiazdy z naniesionymi kolorem czerwonym rozbłyskami. Po wykorzystaniu opcji '-noplot' żadne wykresy nie będą się pojawiać, a wyniki zostaną wypisane w oknie konsoli. Jeżeli użytkownik chce zapisać wykresy i wyniki obliczeń, należy skorzystać z opcji '-save'.

Po zinterpretowaniu argumentów wybranych przez użytkownika program *start.py* wywołuje kolejne programy, czyli *data_download.py* (jeżeli pojawiło się '-d') oraz *find_flares.py*. Są one osobnymi podprocesami stworzonymi poprzez wykorzystanie modułu *subprocess*. Na zakończenie program usuwa z katalogu roboczego zbędne pliki.

4.2 Pobieranie danych

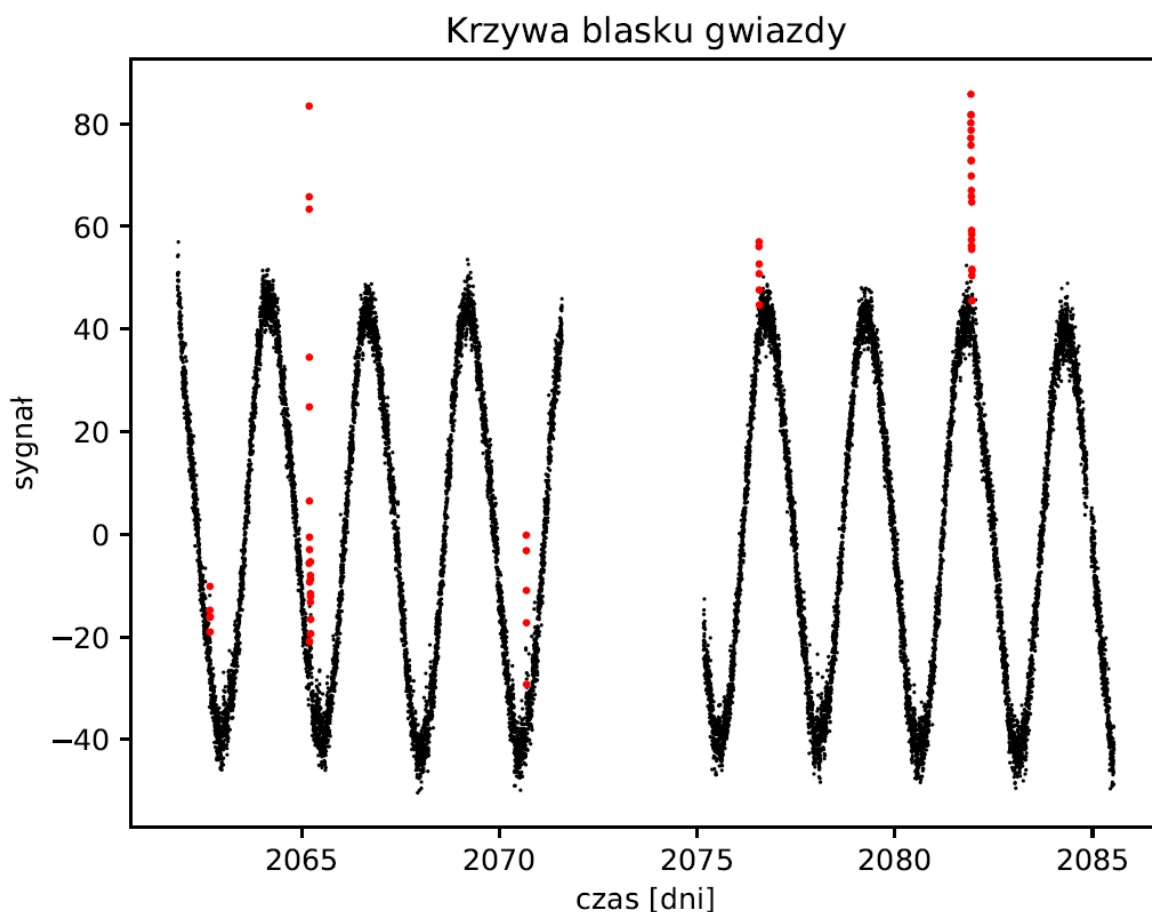
Pobieranie krzywych blasku gwiazd obserwowanych przez satelitę TESS (Transiting Exoplanet Survey Satellite) [3] realizowane jest automatycznie poprzez program *data_download.py*. Argumentami zadawanymi przez użytkownika są tylko numer TIC (TESS Input Catalog) gwiazdy i opcjonalnie sektor obserwacji (liczba z zakresu 1-28). Dane pobierane są ze strony MAST (Barbara A. Mikulski Archive for Space Telescopes) [6] za pomocą biblioteki *requests*. Pliki zawierające krzywą blasku gwiazdy zapisane są w formacie FITS (Flexible Image Transport System) [7]. Do obsługi tego typu plików niezbędny jest pakiet *astropy.io.fits* [8]. Program odczytuje z nich odpowiednie dane: czasu obserwacji, zarejestrowany sygnał gwiazdy oraz jego niepewność. Na zakończenie informacje te zapisywane są do zwyczajnych plików tekstowych za pomocą funkcji *savetxt()* z biblioteki *NumPy*.

4.3 Odnajdywanie rozbłysków gwiazdowych

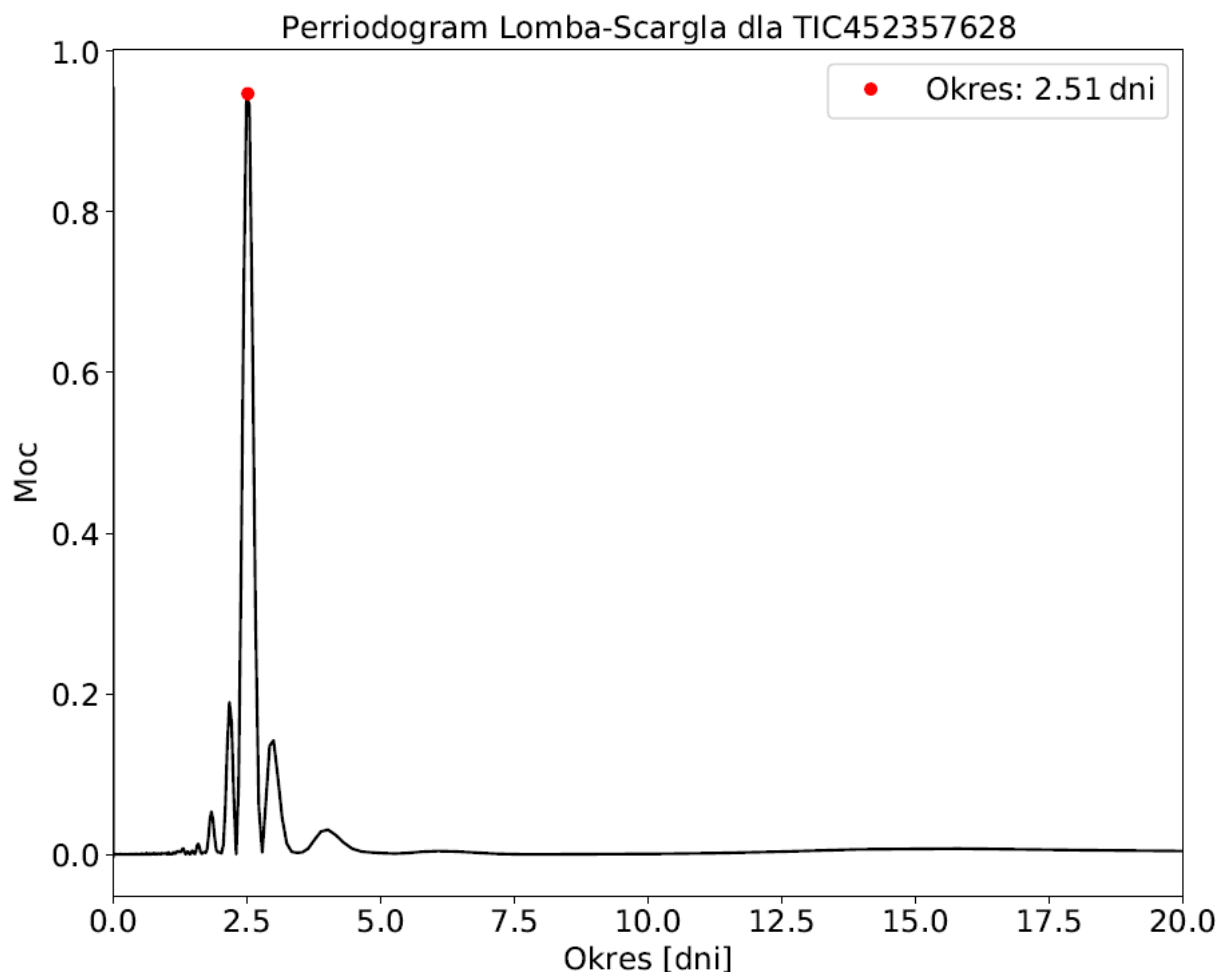
Trzecią, najbardziej rozbudowaną częścią systemu, jest program *find_flares.py*, służący do odnajdywania rozbłysków gwiazdowych oraz prezentowania wyników. W pierwszym kroku program ten wczytuje argumenty zadane przez użytkownika takie jak: minimalna liczba punktów w rozbłysku czy poziom jego detekcji. Jeżeli nie zostały one zdefiniowane przez użytkownika, przyjmują z góry założone wartości. Następnie z plików tekstowych odczytywana jest krzywa blasku gwiazdy, czyli szereg czasowy przedstawiający zmiany sygnału pochodzącego gwiazdy. Realizowane jest to poprzez funkcję *genfromtxt()* z biblioteki *NumPy*.

Z uwagi na fakt, że gwiazdy często wykazują cykliczne zmiany jasności (rys. 4.2), niezbędne okazało się oszacowanie ich okresu. Wykorzystano do tego celu me-

tość Lomb–Scargla [5] zaimplementowaną w pakiecie *gatspy* (General tools for Astronomical Time Series in Python) [4]. Poszukiwane okresy zmienności gwiazd zawierają się w przedziale 0.1 - 20 dni ze względu na czas obserwacji jednego sektora satelity TESS wynoszący ok. 28 dni. Wyznaczona wartość okresu zmienności jest następnie wykorzystywana w dalszym etapie analizy, czyli poszukiwaniu rozbłysków gwiazdowych. Poniżej zamieszczono przykład periodogramu Lomb–Scargla z wyznaczoną wartością okresu zmian jasności gwiazdy TIC452357628 wynoszącą ok. 2.5 dnia (rys. 4.3).



Rysunek 4.2 Krzywa blasku gwiazdy TIC452357628 z widoczną cykliczną zmianą jasności o okresie ok. 2.5 dnia z zaznaczonymi na czerwono rozbłyskami gwiazdowymi. Na osi poziomej przedstawiony jest czas, a na pionowej zmiany sygnału



Rysunek 4.3 Periodogram Lomba–Scargla dla gwiazdy TIC452357628 pokazujący okres zmienności wynoszący ok. 2.5 dni

Procedura wyszukiwania rozbłysków rozpoczyna się od zdefiniowania długości przedziałów, na które podzielona została krzywa blasku. Wynoszą one półtorej wartości okresu zmienności gwiazdy. Przedziały te będą kolejno analizowane w pętli z przesunięciem równym jednej czwartej długości przedziału. Do danych treningowych jest następnie dopasowywany model z użyciem algorytmu RANSAC (RANDOM SAMple Consensus) z pakietu *scikit-learn* [2]. Po wygenerowaniu wektora wartości oczekiwanych następuje właściwa identyfikacja rozbłysków gwiazdowych. Warunkiem detekcji jest wartość w punkcie większa od otrzymanego modelu o zadany przez użytkownika próg σ pomnożony przez odchylenie standardowe po odrzuceniu punktów odstających. W każdym z analizowanych przedziałów zliczane są punkty, które zostały wstępnie zakwalifikowane jako rozbłyski. Analizowane dalej są jednak tylko detekcje, które zostały zliczone co najmniej 2 razy. Na zakończenie odrzucane są zjawiska trwające krócej, niż założony przez użytkownika czas (domyślnie 5 punktów). Cała procedura zwraca punkty początku i końca każdego z rozbłysków gwiazdowych. Wyniki przedstawiane są w postaci graficznej i/lub tekstowej, w zależności od argumentów wybranych przy uruchomieniu całego systemu.

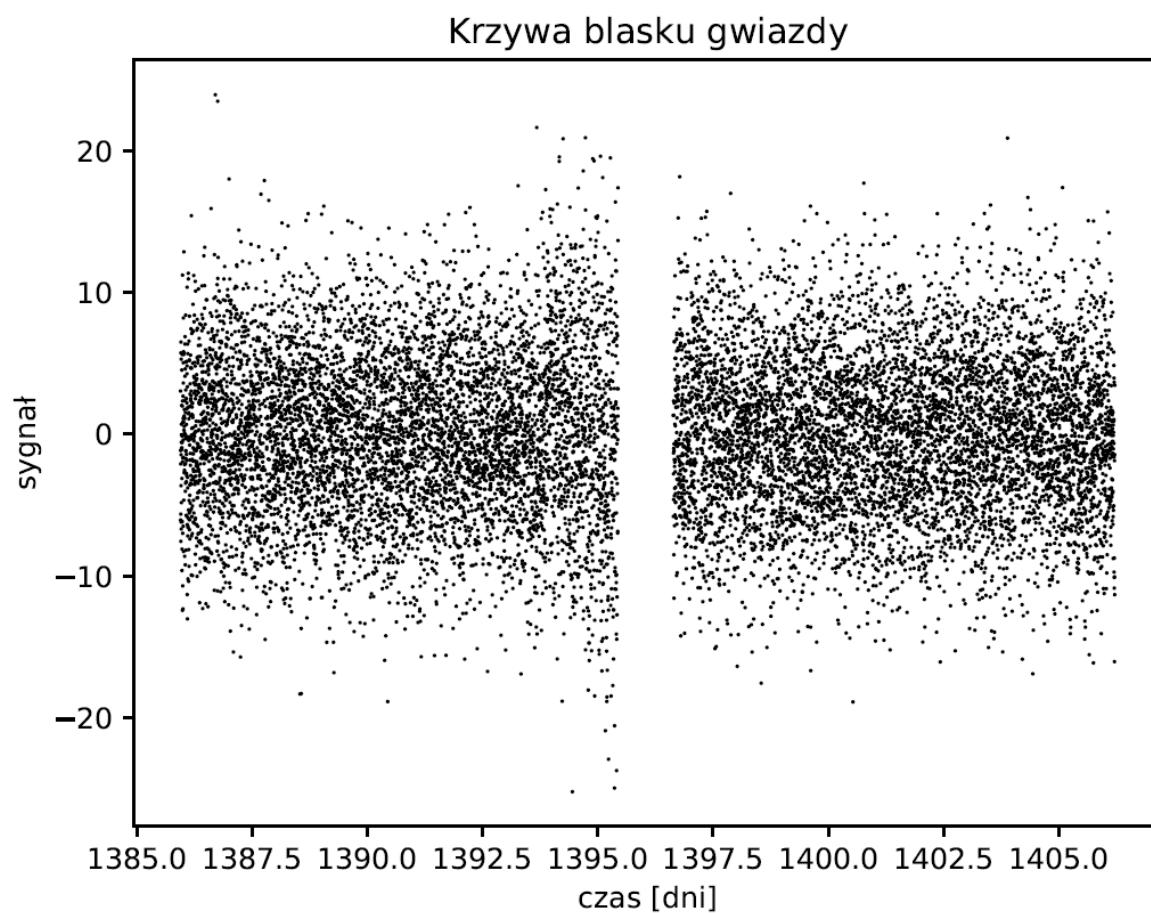
Rozdział 5

Wyniki i wnioski

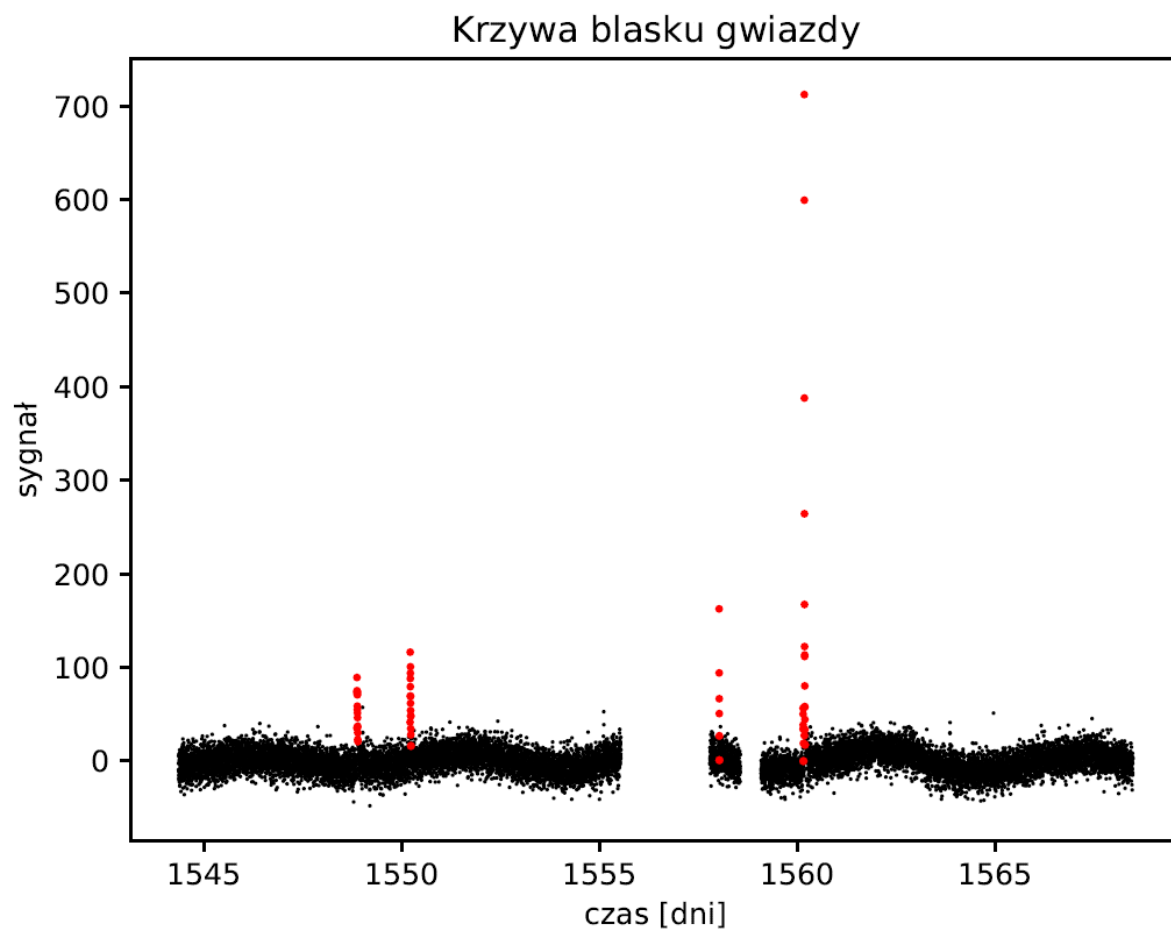
W przedstawionej pracy inżynierskiej zaimplementowany został system do identyfikacji rozbłysków gwiazdowych. System ten umożliwia użytkownikowi proste i szybkie przeanalizowanie wielu tysięcy krzywych blasku pod kątem detekcji nagłych zmian jasności. Dzięki możliwości automatycznego pobierania danych jedynym wymaganym przez oprogramowanie argumentem jest numer TIC (TESS Input Catalog) gwiazdy lub gotowy plik zawierający zmiany sygnału w czasie. Użytkownik może kontrolować liczbę wyszukiwanych rozbłysków poprzez zmianę parametrów takich jak: wartość poziomu detekcji i minimalną liczbę punktów w rozbłysku. Wynikiem działania systemu jest domyślnie wykres z oznaczonymi rozbłyskami gwiazdowymi i następujące dane:

- t_{start} : czas rozpoczęcia rozbłysku
- t_{end} : czas zakończenia rozbłysku
- t_{max} : czas maksimum rozbłysku
- lc_{amp} : amplituda rozbłysku

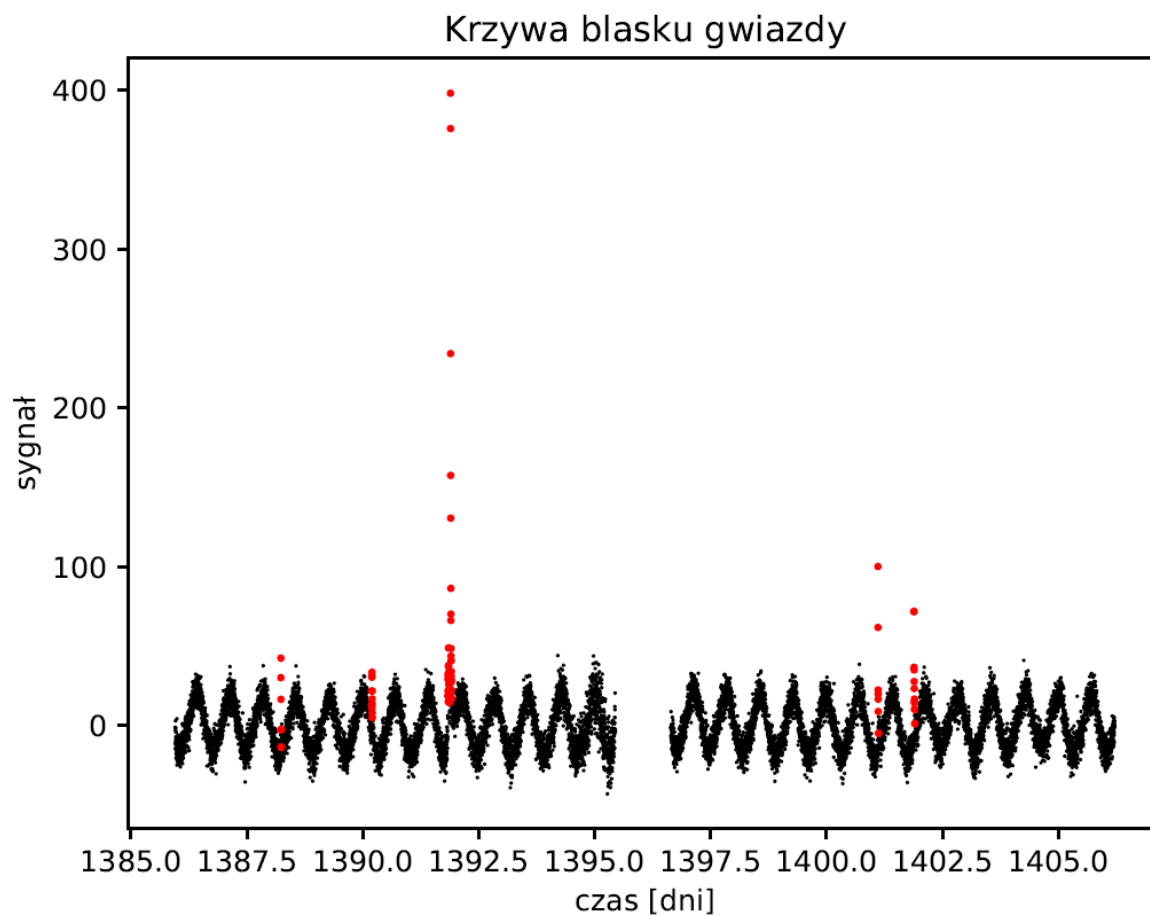
System został przetestowany na kilku typach krzywych blasku: bez żadnych rozbłysków (rys. 5.1), z wyraźnie widocznymi rozbłyskami (rys. 5.2), z okresową zmianą jasności (rys. 5.3), z tranzytem (rys. 5.4) oraz na skomplikowanej krzywej (rys. 5.5). Wyniki przedstawione zostały poniżej.



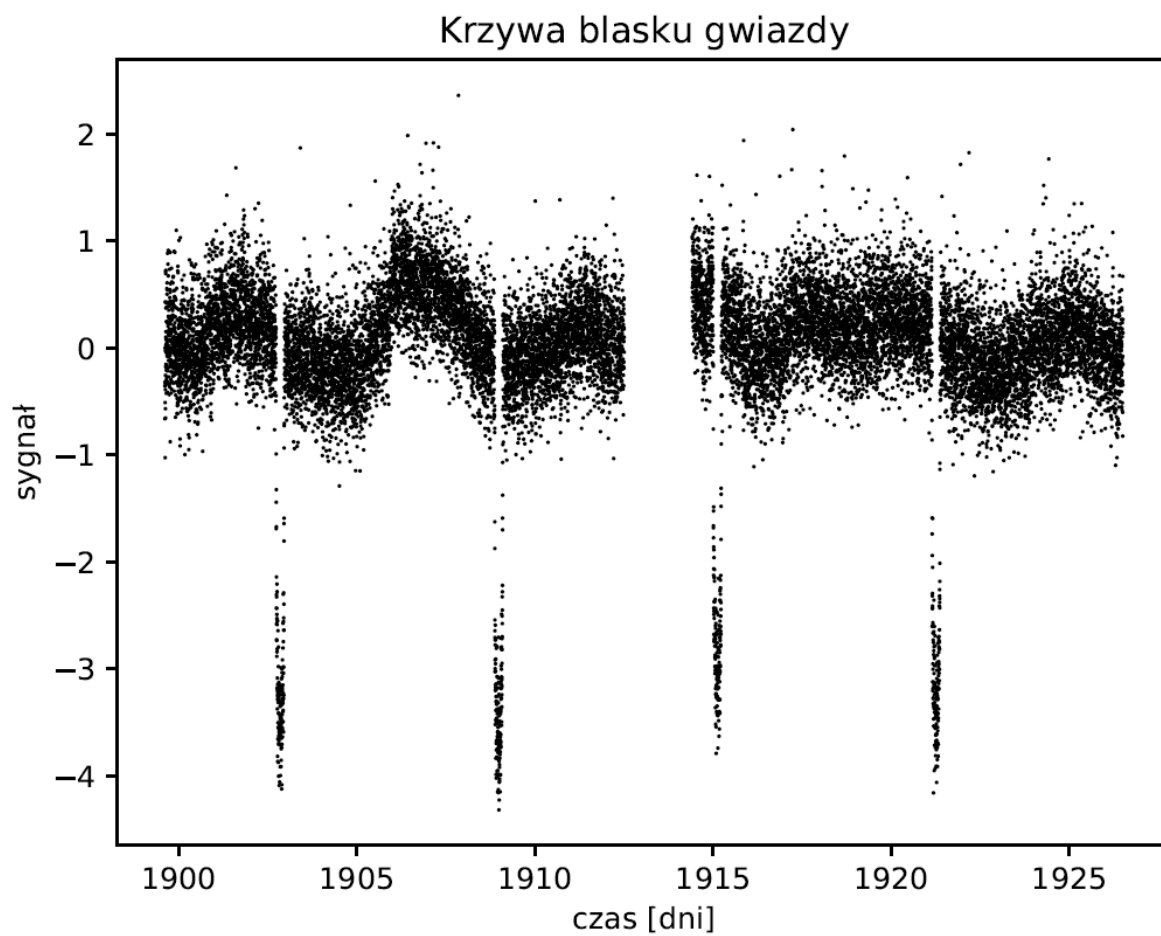
Rysunek 5.1 Krzywa blasku gwiazdy TIC260542013 bez rozbłysków gwiazdowych. Na osi poziomej przedstawiony jest czas, a na pionowej zmiany sygnału



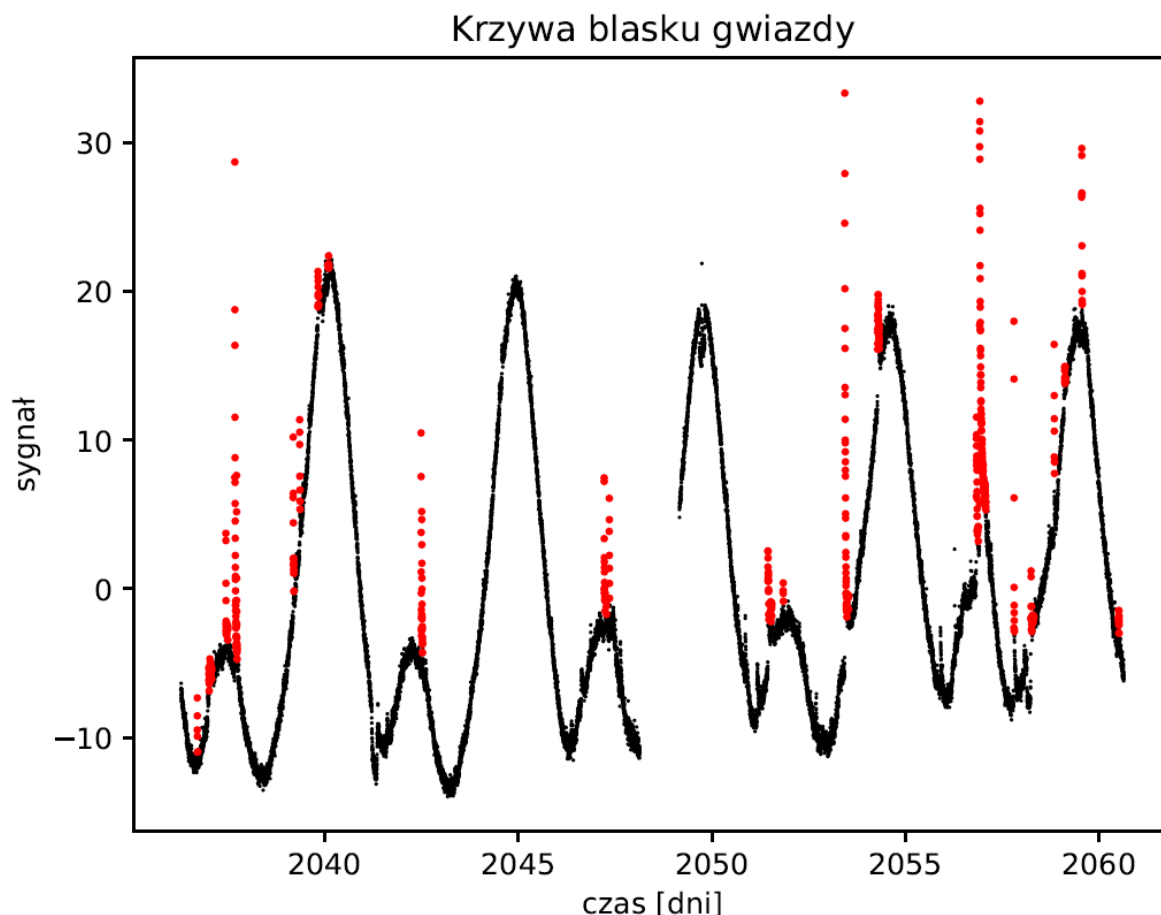
Rysunek 5.2 Krzywa blasku gwiazdy TIC32090581 z wyraźnie widocznym rozbłyskami gwiazdowymi zaznaczonymi na czerwono. Na osi poziomej przedstawiony jest czas, a na pionowej zmiany sygnału



Rysunek 5.3 Krzywa blasku gwiazdy TIC246891813 z widoczną cykliczną zmianą jasności o okresie ok. 2.5 dni z zaznaczonymi na czerwono rozbłyskami gwiazdowymi. Na osi poziomej przedstawiony jest czas, a na pionowej zmiany sygnału



Rysunek 5.4 Krzywa blasku gwiazdy TIC286923464 z widocznymi tranzytami planety. Na osi poziomej przedstawiony jest czas, a na pionowej zmiany sygnału



Rysunek 5.5 Skomplikowana krzywa blasku gwiazdy TIC441420236 z zaznaczonymi na czerwono rozbłyskami gwiazdowymi. Na osi poziomej przedstawiony jest czas, a na pionowej zmiany sygnału

Na podstawie zamieszczonych przykładów zauważyć można, że system dobrze sprawuje się w przypadku prostych krzywych blasku. Trudności nie sprawia wykrywanie zarówno dużych, jak i słabych zjawisk. Nawet szybkie, okresowe zmiany zmiany jasności nie są traktowane jako rozbłyski gwiazdowe. Nagłe spadki sygnału będące wynikiem tranzytu planety również nie powodują fałszywych detekcji. Dużą zaletą jest możliwość dostosowania parametrów wyszukiwania do konkretnych przypadków.

Problemem pozostaje jednak detekcja wszystkich zjawisk na skomplikowanych krzywych blasku z wieloma rozbłyskami. Obecnie zastosowane metody odnajdują wówczas ok. 90% rzeczywistych rozbłysków. Aby udoskonalić system należałoby najprawdopodobniej zastosować bardziej zaawansowane metody numeryczne lub zaimplementować dodatkowe warunki na obecność rozbłysku.

Przedstawiony w pracy inżynierskiej system do identyfikacji rozbłysków gwiazdowych można zmodyfikować również do innych celów. Sprawdza się bardzo dobrze do szeregów czasowych, w których występują nagłe wzrosty lub spadki sygnału. Można również ograniczyć jego funkcjonalność do odnajdywania okresów

zmienności w analizowanych danych. Tego typu narzędzie jest niezbędne w procesie automatyzacji obliczeń na dużych zbiorach danych.

Literatura

- [1] Davenport, James R. A. *The Kepler Catalog of Stellar Flares*. The Astrophysical Journal, Volume 829, Issue 1, article id. 23, 12 pp. (2016).
- [2] Pedregosa, F.; Varoquaux, G.; Gramfort, A. i in. *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research 12, pp. 2825-2830, 2011.
- [3] Ricker, George R.; Winn, Joshua N.; Vanderspek, Roland; i in. *Transiting Exoplanet Survey Satellite (TESS)*. Journal of Astronomical Telescopes, Instruments, and Systems, Volume 1, id. 014003 (2015).
- [4] VanderPlas, Jacob T. *gatspy: General tools for Astronomical Time Series in Python*. Astrophysics Source Code Library, record ascl:1610.007 (2016).
- [5] VanderPlas, Jacob T. *Understanding the Lomb–Scargle Periodogram*. The Astrophysical Journal Supplement Series 236 16 (2018).
- [6] Barbara A. Mikulski Archive for Space Telescopes,
<https://archive.stsci.edu>
- [7] The FITS Support Office, NASA/GSFC
<https://fits.gsfc.nasa.gov>
- [8] <https://docs.astropy.org/en/stable/io/fits/>