

Beitrag zur FrOSCon 2016

Science Track

DATA MINING IN ASTRONOMISCHEN SURVEYDATEN VARIABLER  
STERNE MIT PYTHON

C. Dersch<sup>1</sup>

<sup>1</sup>*Philipps-Universität Marburg, Fachbereich Physik, 35032 Marburg*

*Knowledge Discovery* und speziell *maschinelles Lernen* sind sehr nützliche Werkzeuge für die automatisierte Datenanalyse. Die Anwendung des maschinellen Lernens hat in den vergangenen Jahren stark an Bedeutung gewonnen und sich als Lösung für Klassifikationsprobleme etabliert. In Astronomie und Astrophysik treten große Datenmengen insbesondere in Surveys auf, beispielhaft sei hier das Weltraumteleskop *Kepler* genannt, welches für die Suche nach Exoplaneten konzipiert wurde. In der Domäne der Photometrie ist hier die Analyse von Helligkeitsänderungen in den Lichtkurven von Sternen eine zentrale Aufgabe, beispielsweise von sich gegenseitig bedeckenden Sternen.

Die freie Programmiersprache Python hat sich in den letzten Jahren mehr und mehr in der astronomischen Datenanalyse etabliert, zentraler Baustein ist hier der *SciPy* Stack. Speziell für astronomische Anwendungen entwickelt das *astropy*-Projekt Module zur Datenanalyse, zusätzlich gibt es stark spezialisierte Module, z.B. *astroML* für die Anwendung des maschinellen Lernens und der Lichtkurvenanalyse. Algorithmen des maschinellen Lernens und des Data Minings werden im *scikit-learn*-Projekt entwickelt, welches bereits in vielen Bereichen verwendet wird.

Anhand von Daten des *OGLE-III*-Surveys wird eine ausschließlich auf freier Software aufbauende Analyse und Klassifikation von Lichtkurven variabler Sterne präsentiert. Hier zeigt sich das Zusammenspiel einer Vielzahl von auf dem SciPy-Stack aufbauenden Python-Modulen. Die Konzentration auf freie Software stellt hierbei die Nachvollziehbarkeit der Datenanalyse sicher und verhindert künstliche Barrieren bei der wissenschaftlichen Arbeit. Schließlich wird ein Aus- und Überblick über die weitere Anwendung von Python in Astronomie und Astrophysik gegeben.