Big Data, Astronomy for Development

Cross disciplinary collaboration



E

Vanessa McBride

Digital Humanities Colloquium

16 February 2022



INTERNATIONAL ASTRONOMICAL UNION DEVELOPMENT

Outline

- 1. What is astronomy for development?
- 2. OAD's collaboration gateway
- 3. Data & Methods in astronomy
- 4. Examples of cross-disciplinary projects

What is astronomy for development?





CULTURE AND SOCIETY

Cultural roots History PHILOSOPHY Evolution of the universe

ANTHROPOLOGY

Ancient civilisations

INSPIRATION Perspective **Global citizenship** Earth stewardship

Our Origins

Big History

ASTRONOMY

OPTICS

High-precision

Adaptive optics

EDUCATION Gateway in Science & Technology **Teaching tools**

MATHS & STATS

Abstract thought Classification Modelling Trends

BIOLOGY & EARTH SCIENCES Geology **Building blocks** of life

Producing organic molecules Spectroscopy

BIG DATA

Machine Learning

Visualitation

Algorithms

TECHNOLOGY AND SKILLS

ELECTRONICS Detectors Signal Processing

> SPACE Satellites Near-earth objects Solar weather

PHYSICS

Laboratory of extremes Making heavy elements

> SCIENCE AND RESEARCH

CHEMISTRY

The Office of Astronomy for Development

- Founded in 2011
- Cape Town, South Africa
- Equal partnership between IAU & NRF (DSI)
- Ethos: humility
- "Astronomy for a better world"

OAD Regional Offices



Annual call for proposals 2013 - 2021

1000 000 €

200 projects

100+ countries



Projects 2013 - 2021

Projects 2013 - 2021

AstroLab

Location: southern, eastern Africa

- Low-cost research tutorial for universities in need of infrastructure and curriculum
- Remote telescopes used to teach students the primary steps involved in research and the scientific method – observation, image acquisition, processing, data analysis and writing up results





Projects 2013 - 2021

Amanar: Under the same sky

Spain, Algeria 2019

to inspire the Saharawi community through the wonders of our Universe and promote peace, mutual understanding and a sense of citizenship under the same sky.







Instituto de Astrofísica de Canarias • IAC

Magic happens at the interface





Interdisciplinary conversations





TURNING UP THE HEAT: AN INTERDISCIPLINARY SOLUTION TO IMPROVE PREGNANCY SUCCESS IN CATTLE

/3/2019 0 Comments

Temperatures of galaxies -> Ovulation in cattle

OAD collaboration gateway



Data is a common language

10 00

* 8

0 1

 $10 \\ 10 \\ 10 \\ 10$

U

o

Ø

Ø

Ø

What kind of *data* do astronomers use?



Messengers



Images



#	RA	Dec	Brightness	Туре	
1	23.3	-3.51	2	star	
2	23.4	-3.52	1.5	galaxy	
3	23.5	-3.82	0.5	asteroid	
4	23.6	-3.51	0.6	star	
5	23.7	-3.67	2.2	blend?	

Spectroscopy



Spectroscopy



Time series



Astronomers love finding periodic variations, e.g. orbits, rotation

OGLE lightcurve, Long & De Souza 2017

Time series - trailed spectrum

1.0 Orbital phase 0.5 0.0 -2 -1 2 0 1 Wavelength

One particular spectral line observed over a binary orbit.

From Kotze et al. 2016

What kind of tools & methods do astronomers use?

- Statistics & mathematics
- Simple scripting
- Python programming
- C programming (for computationally intensive problems)
- Machine learning

Examples of *cross-disciplinary* projects

donced

DO

or tople

h too mony sho

tonga

Photo by Andreas Fickl on Unsplash

- Part of a collaboration to develop data science skills in SKA partner countries
- Developed and run for a virtual hackathon in Zambia in • 2020



opment in Africa with Hadio Astronomy





TRONOMICAL ASTRONOMY FOR





Dataset comprises tweets around Covid-19

search_words = "coronavirus OR covid OR pandemic OR covid19 OR lockdown"

Clean the tweet text

remove hyperlinks, stopwords, etc

· Use categorization tools to explore sentiment

Python: TextBlob, VADER

• Use machine learning to explore sentiment

Supervised learning - labeled data, Transfer learning with BERT



Tutorials as Jupyter notebooks, using Python

Introduction

In this tutorial you will:

- Use the Twitter API to collect COVID19-related Twitter data
- Extract the tweet text and any metadata you require
- Perform preprocessing of the text to allow for better application of NLP techniques.

Let's get started!

To access the Twitter API, we will make use of the Python library Tweepy. Let's start by importing tweepy and other libraries needed for this tutorial.

```
In [1]: import tweepy
import pandas as pd
import csv
import re
import numpy as np
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
from wordcloud import WordCloud
import nltk
from nltk.corpus import stopwords
nltk.download("stopwords")
[nltk_data] Downloading package stopwords to /home/nikhi/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```



- Categorisation and supervised learning is not too resource intensive
- BERT (i.e. transfer learning) needed significant compute resource on IDIA.
- Tutorials to enable a research question...

https://github.com/darabigdata/COVID19_Twitter_Project

How does this link to astronomy?





- Topic is very different (i.e. sentiment vs galaxies)
- Methods are the same (classification)
- Tools are **similar**, e.g. data sorting & preparation, machine learning libraries e.g. scikit-learn

Size and patterns of urban informality



LAB

RECENTS

TUTORIAL

COLLECT

Next -+

Size & patterns of urban informality



ZOØNIVERSE

Size & patterns of urban informality



Crowdsourcing -> Machine learning

OAD collaboration gateway





www.astro4dev.org

@astro4dev on Twitter

www.github.com/astro4dev

vanessa@astro4dev.org

https://list.astro4dev.org/wws/subscribe/mailinglist



INTERNATIONAL | OFFICE OF ASTRONOMICAL | ASTRONOMY FOR UNION | DEVELOPMENT

Data is a common language

10 00

* 8

0 1

 $10 \\ 10 \\ 10 \\ 10$

U

o

Ø

Ø

Ø

Astronomy Data Science Toolkit

Motivation:

Provide a foundation for a "common language" between the data science and astronomy communities.

Target audience:

Astronomers & professional scientists & students & industry

4 EDUCATION 8 DECENT WORK AND ECONOMIC GROWTH

Content:

- 1. Map to existing data science resources
- 2. Astronomy case studies in data science
- 3. Assessments & proficiency levels
- 4. Database of astronomers turned data scientists

Toolkit overview



Toolkit overview



1. Toolkit Website

https://datascience.astro4dev.org

Astronomy & Data Science Toolkit				
Introduction About the Toolkit View Toolk	kit Contribute Contact 👰 Engl	lish		
bassele	Country Channel All			
pearchan	Search Show All			
Data Science	Astronomy			
Data obiorido	roueneng			
I am a data asigntist internated in using astronomy	I am an extrememory interested in section dat			
I am a data scientist interested in using astronomy examples for my teaching:	I am an astronomer interested in seeing dat science techniques used within the field of:	ta		
I am a data scientist interested in using astronomy examples for my teaching: Data Wrangling 	I am an astronomer interested in seeing dat science techniques used within the field of: • Galaxies	ta		
 I am a data scientist interested in using astronomy examples for my teaching: <u>Data Wrangling</u> <u>Python</u> Maghing Learning 	I am an astronomer interested in seeing dat science techniques used within the field of: <i>Galaxies</i> <i>Solar & Stellar</i>	ta		

English Spanish French

2. Content on GitHub



www.github.com/astro4dev

Content submitted as Jupyter notebooks

Potential users

- Data science educators looking to excite students with interesting applications
- 2. Astronomers running summer schools
- 3. Economists/Others looking for data science skills

Magic happens at the interface