Big Data, Astronomy for Development & Cross disciplinary collaboration

Vanessa McBride
Digital Humanities Colloquium
16 February 2022
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?
Partnerships for Astro tourism

Decolonise your teaching materials; and make them publicly available

Educational sci-fi literature with diverse protagonists

Industry + dark skies partnerships for sustainability

Use your scientific credibility to dispute ‘alternative facts’ on different platforms

A stargazing event that also highlights human impact on the environment

Train students to apply their astronomy analysis skills in different sectors

Establish oversight committee for accountability in spending official development assistance in Astronomy (e.g. Newton Fund, etc)
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”
Annual call for proposals 2013 - 2021

1000 000 €

200 projects

100+ countries
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.
Magic happens at the interface
Interdisciplinary conversations

Agriculture

Health Sciences

Astronomy

Temperatures of galaxies ->
Ovulation in cattle
Data is a common language
What kind of data do astronomers use?
Messengers

Light

Particles

Gravitational waves
# Images

<table>
<thead>
<tr>
<th>#</th>
<th>RA</th>
<th>Dec</th>
<th>Brightness</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23.3</td>
<td>-3.51</td>
<td>2</td>
<td>star</td>
</tr>
<tr>
<td>2</td>
<td>23.4</td>
<td>-3.52</td>
<td>1.5</td>
<td>galaxy</td>
</tr>
<tr>
<td>3</td>
<td>23.5</td>
<td>-3.82</td>
<td>0.5</td>
<td>asteroid</td>
</tr>
<tr>
<td>4</td>
<td>23.6</td>
<td>-3.51</td>
<td>0.6</td>
<td>star</td>
</tr>
<tr>
<td>5</td>
<td>23.7</td>
<td>-3.67</td>
<td>2.2</td>
<td>blend?</td>
</tr>
</tbody>
</table>
Spectroscopy

Intensity vs. Wavelength graph with peaks indicating different wavelengths and corresponding intensities.
Spectroscopy

Hydrogen
Sodium
Helium
Neon
Mercury
Time series

Individual measurements over 4000 days

Folded modulo 2.48 day period

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

OGLE lightcurve, Long & De Souza 2017
Time series - trailed spectrum

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
Sentiment analysis around Covid-19
Sentiment analysis around Covid-19

• Part of a collaboration to develop data science skills in SKA partner countries

• Developed and run for a virtual hackathon in Zambia in 2020
Sentiment analysis around Covid-19

• Dataset comprises tweets around Covid-19

```python
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
Sentiment analysis around Covid-19

- 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")

[ntlk_data] Downloading package stopwords to /home/nikhi/ntlk_data...
[ntlk_data] Package stopwords is already up-to-date!
```

Out[1]: True
Sentiment analysis around Covid-19

• 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
Size & patterns of urban informality
Size & patterns of urban informality

Crowdsourcing -> Machine learning
OAD collaboration gateway

Start Here

- Onboarding
- Interested?
  - Yes: Contact OAD & Partner
  - No: Skills transfer
- Research collaboration
- Small business development
- On-the-ground support
- Something completely different

ENGAGE

- Contact OAD & Partner
- Implement project with OAD support

INITIATE

- Contact OAD
- Prepare onboarding package

www.astro4dev.org
Get *in* touch!

www.astro4dev.org

@astro4dev on Twitter

www.github.com/astro4dev

vanessa@astro4dev.org

https://list.astro4dev.org/wws/subscribe/mailinglist
Data is a common language
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

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

Question or problem

Skills

TOOLKIT

Tools
Toolkit overview

Question or problem

Skills

Tools

TOOLKIT
1. Toolkit Website

https://datascience.astro4dev.org

- English
- Spanish
- French
2. Content on GitHub

www.github.com/astro4dev

Content submitted as Jupyter notebooks
Potential users

1. 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