Telescopes and Data Science, a match made in the heavens

Dr. Elaina Hyde

Oct 27, 2019 · 8 min read

As an incoming professor at York University (Faculty of Science) in Toronto, I will be teaching in Physics and Astronomy. To celebrate the new on-campus telescope that was recently installed, and of course as an almost Canadian resident… I thought what better way to get started than to write about the data science of astronomy, the astronomy of telescopes, and how they can work together to help with furthering our knowledge of both. So let's start!

The York University Allan I. Carswell Observatory. Credit: https://observatory.info.yorku.ca/2013/11/26/our-observatory/

## A new telescope and the why of it all

As with any new endeavour, I will start with a question and work from there. How can I use the fabulous new campus telescope at York University for my science and data science goals? First, let's look at the telescopes. In addition to a historical telescope, the
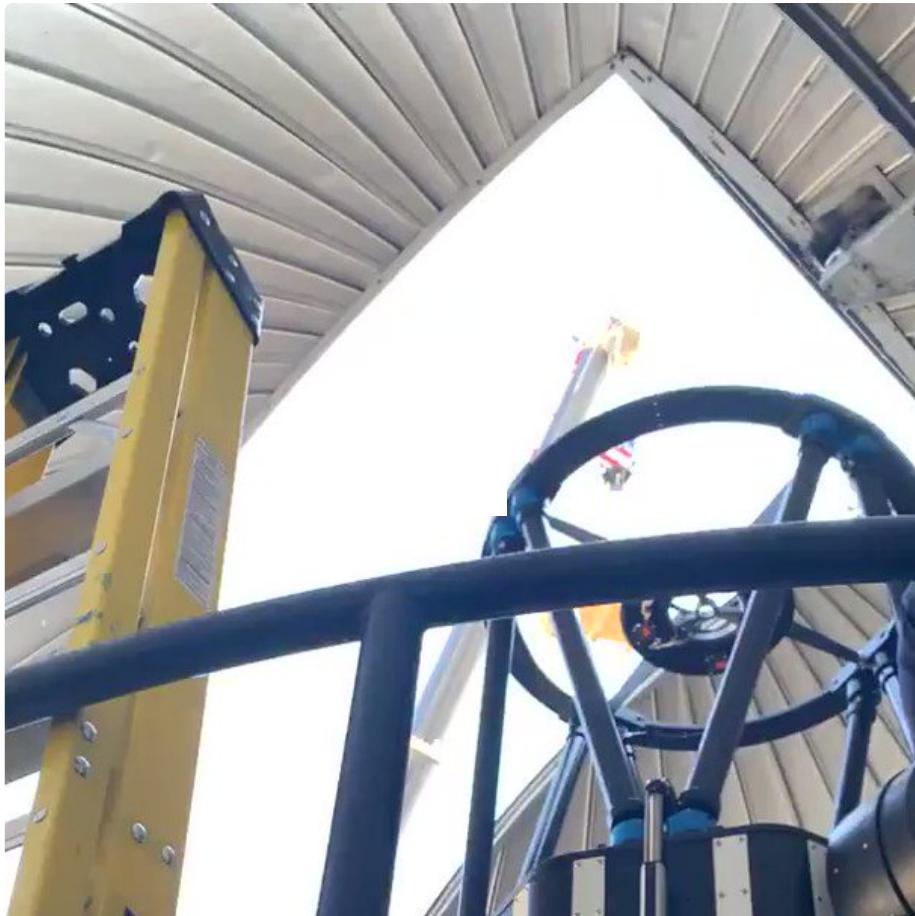
Allan I. Carswell Astronomical Observatory has recently installed a new 1 meter telescope that will allow for public observing as well as scientific observations.

Allan I. Carswell Observatory
@yorkobservatory

The mirror !!

12   9:48 AM - Aug 16, 2019

See Allan I. Carswell Observatory's other Tweets

Exciting mirror installation for the new telescope at York University from the observatory Twitter account.
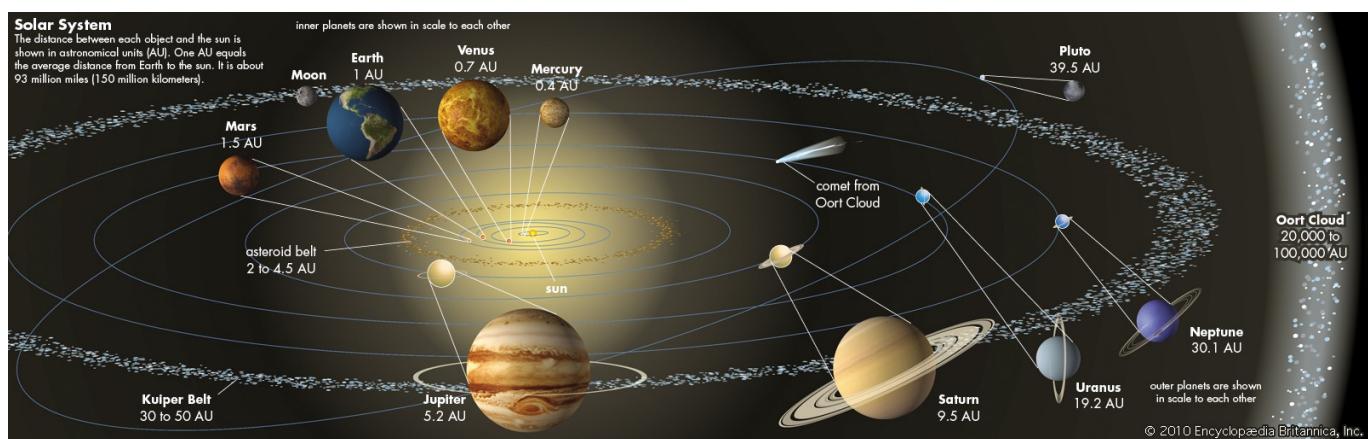
This telescope is the most powerful optical telescope on a Canadian university campus, and we can use it to gather data from local comets in our solar system all the way out to galaxy clusters.

In data science, we utilise large quantities of information from everything to refrigerator sensors all the way to medical imaging of cancerous tumours. With a telescope we look at astronomical data in the form of photons from stars and galaxies far, far away (although also some that are relatively closer by).

So what can we do with the information brought in by the new 1 meter telescope? To steal a phrase from a more eloquent author: *life, the universe, and everything!* The 1 meter class of telescope is a significant workhorse in astronomy due to the large variety of possible use cases.

The Universe and Everything
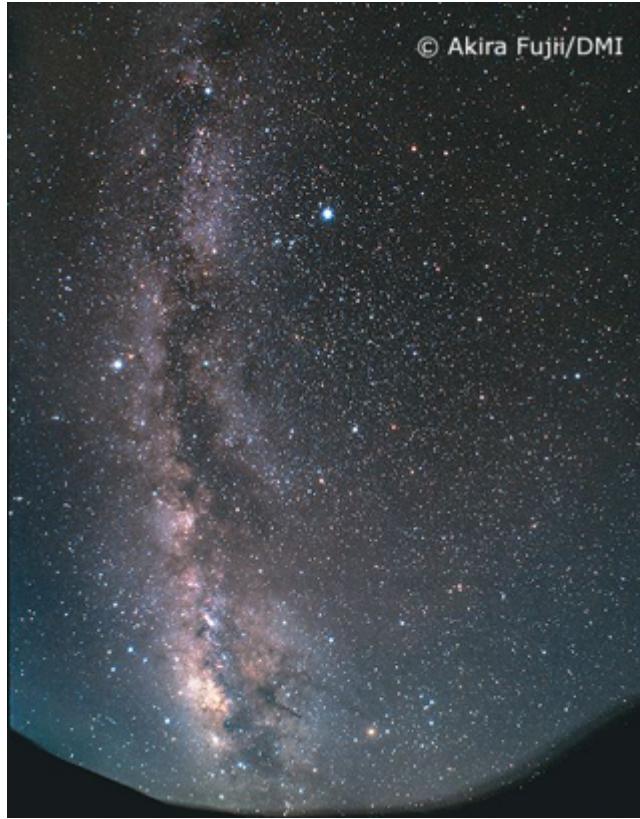
Once we lift our view out and away from Earth, the first objects that capture most of our attention are the Moon and the Sun. Of course, if it's a clear night with a new Moon, the stars and planets take main stage.

The orbits of the planets and other bodies of the solar system. From Encyclopædia Britannica, Inc. at https://www.britannica.com/science/solar-system

Looking at our solar system with our 1 meter telescope, we can examine the light from comets and asteroids to trace their orbits, we can look at our neighbouring planets and even investigate the behaviour of our sun. For more on our solar system, make sure to check out the NASA guide here.

The next step is up (or out as you will) is our galaxy, the Milky Way. Our mighty Sun is just a tiny dot inside the group of the Milky Way. Visible in a dark site (that is, with little light pollution), the Milky Way stretches across the night sky.

credit David Malin, Akira Fuji

From our perspective, the Milky Way looks hazy and not unlike spilt milk, which is where it gets its name. The Milky Way has hundreds of billions of stars; if you look towards the centre with a telescope (or high-powered binoculars) you can see the haze is made of stars.

The dark regions we see are clouds of gas and dust, which — given enough time — has the potential to make billions more stars. By observing the gas and dust, we can discover how stars are formed, make predictions regarding the lifetime of our galaxy, and learn about interesting new phenomena (see for example, the beer nebula).

With hundreds of billions of stars just in our local galaxy, we reach into the realm of what is called 'big data' relatively quickly. Big data in data science usually means anything from 50,000 to a few million sources. Big data is all around us: e-commerce websites like Amazon, with millions of users providing large volumes of data, chips in

internet-connected devices as small as your home water cooker providing data back to the manufacturer, and even Netflix subscribers providing data for its recommendation system. The amount of data we bring in and process has increased drastically.

Mathematicians and Physicists have known for awhile that by taking repeated measurements of a phenomenon you can decrease random error in what you measure, but why is having big data on galaxies or other images an advantage?

Moooar Data, more Science?

Most of us have seen some of the new features popping up around face recognition and so-called 'augmented reality'. From unlocking your phone with your face, checking in at the airport via a camera to Pokemon Go we have more capacity than ever.

From: https://www.pokemon.com/us/pokemon-video-games/pokemon-go/
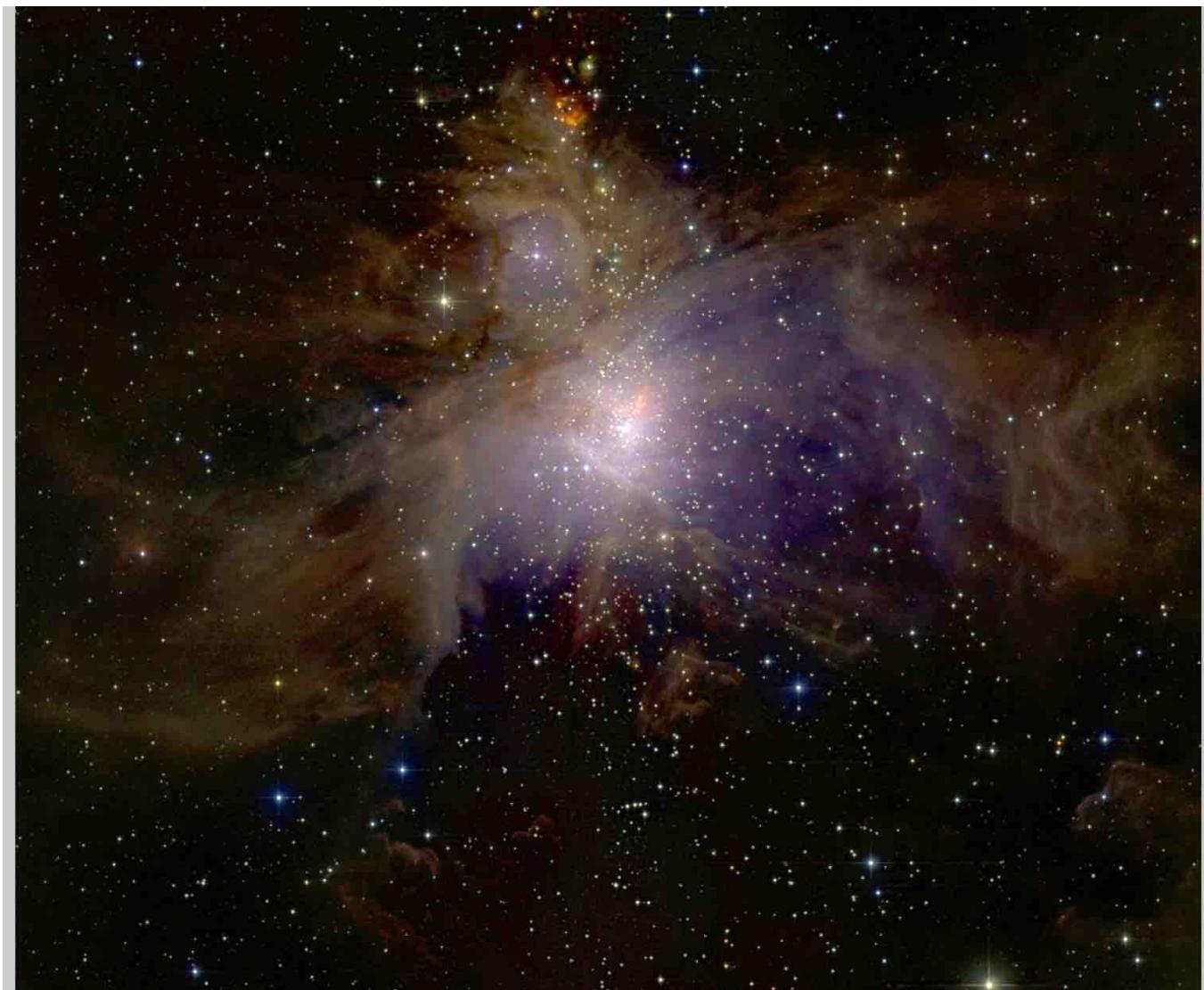
But how did all this come about? One key factor is having access to very, very, large sets of data. Millions or billions of images can be used to train algorithms that identify facial aspects. Another few billion images might be used to train algorithms to recognise 'ground' and work with your camera to insert an image of a pokemon in a realistic way. However, having large amounts of data is only one of the things that comes into play. The critical part is what you then go on to do with it. In what's called 'data science' there are certain methodologies and processes that will look very familiar to those of us who

have applied the scientific method for research on topics like the aforementioned galaxies above. It's not enough just to have data; we have to be able to learn from it.

Learning from big data is becoming easier and easier with multiple 'cloud computing' platforms coming into existence. My personal experience has been primarily with the Google Cloud Platform, although any sufficiently robust provider will be able to process large amounts of data quickly (at least much more quickly than my personal laptop).

Telescopes and Data Science Together

Using a campus telescope, like the 1 meter at York University, we can collect data from other stars, galaxies and even planets. You can imagine that processing any new data is a good deal of work, but what do you do if you think that you have discovered something new? How can you refine an algorithm to help you search for the faintest signals that your telescope is able to detect?

The Sword of Orion as shown from 2MASS data, From:
https://old.ipac.caltech.edu/2mass/gallery/showcase/orion/index.html

This is where data science comes back into the picture! Astronomers have massive online surveys for many types of objects and if you are ever looking for huge sets of free data this can be a great place to start. Skipping over the utterly mindboggling amounts of data processed from radio telescopes (sorry, ASKAP, double sorry, future SKA)… there are old surveys like 2MASS which contains a sky atlas (approximately 4 million objects), a point source catalog for ~300 million stars and an extended source catalog for more than 1,000,000 galaxies and other nebulae. As an example of a more current survey, the Australian Galah survey (full disclosure, I used to be a member of this survey) is currently running and has over 300,000 sources with up to 23 parameters per source, which killed my poor laptop (4 days of processing time) — until I loaded it onto GCP (Google Cloud Platform) and processed it in under 20 minutes!
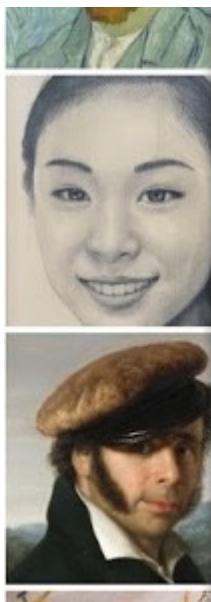
If I decide to observe the stars in the Sword of Orion (shown above from the 2MASS survey) with York's 1 meter campus telescope, I can get new information about those stars. However, by using my million star database in 2MASS, my hundreds of thousands of stars with many parameters in Galah… I can train algorithms that will work extremely well on the *type* of stars that I am interested in. In short, by looking at large, relevant datasets and **generalising** I can train data science type algorithms that will respond quickly and efficiently to my new image.

Search with your selfie

Discover portraits from museums and get face to face with art

The Google Arts and Culture App with some of the portraits from its database. From: https://artsandculture.google.com/

This is the same thing that we see when we use our faces to log into our phone, these pre-trained algorithms have used millions of images to recognise faces… and now, instead of building a new algorithm for every new star… or face… we can just call an existing one.

The Google Arts and Culture App (in the image) matches faces from the camera on your phone or laptop to their art database. It was very popular when it came out because it turns out everyone, even famous people, like to see if they look like someone famous.

The reason this works is that the pre-trained database can recognise a wide variety of faces in real photos as well as the faces in pictures. If you are interested in these kind of pre-trained GCP APIs, it's good to know that they can be called from most codes fairly easily. Moreover, the GCP pre-trained API recognises a wide variety of astronomy images without any additional training (which I found very handy for demos).

Orionid Meteor Shower 2019

Image Credit: NASA

Come view the 2019 Orionid Meteor Shower with the Allan I. Carswell Observatory Team at York University! With up to 15 meteors visible every hour, multiple telescopes available for viewing celestial objects and a constellation tour, it will be a fun, astronomy filled night!

When: The first clear night of October 21st or 22nd
8:00pm - 10:00pm

Where: William Small Center (Top of the Arboretum)
York University
4700 Keele Street, Toronto, ON

*Weather updates will be posted by 4pm on social media. Hope to see you there!

York University Orionid Meteor Shower viewing announcement from: https://twitter.com/yorkobservatory

So, in conclusion, we can explore big data with telescopes and of course use telescopes to further our exploration of big data. At York's astronomy department, we have a further advantage as the telescope is also available for public observing.

We can use the massive astronomical archives to access data on millions of stars online and train even more advanced algorithms. In the era of big data and big discoveries it is very nice to know that, for some data at least, the exploration of those data sources can take us to other planets, other solar systems and even other galaxies.

. . .

From my blog at GitHub: https://astrohyde.github.io/LearnML/

For more on Astronomy and Data Science, see my GitHub blog post at:
https://astrohyde.github.io/AstroDS/

If you are interested in getting started with some of these methods, I have collected some
handy tips here: https://astrohyde.github.io/LearnML/

)

Astronomy   Data Science   Big Data   Telescope   Science