ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

Baily, Alison
After the seminar and also if you have, if you have any questions or that you'd like to submit to and ask Tom during the session, please do post these. Please go to the slide our website and key in the code 582378 and that will take you to our page there where we'll be we have a number of also have a number of polls that Tom will be.

I'm running during the session. We really like these sessions to be very interactive and really sort of it really encourage you to submit your questions and during the during the sessions so well that's all for me for now. So, I will hand over to Tom and for to kick off the session.

Wise, Tom
Thank you very much, Alison. So, before we begin today, I actually really want to give a huge thank you to Sarah, Alison and Inkyung, who are supporting me through this. They have been an absolute gem team throughout the entire process. And so, I don't think I can do this without them to get started with using slido. And to get you all into thinking and engaged in this topic during the introduction, how do think about this topic and this question in particular, what is it that you think about when you hear the term machine learning?

This can be quite easily in your professional day. This can be in your when you're reading articles or scientific journals. But tell us what you think you hear, whether that's one word or phrase, or anything in between, and we'll come back to reflect on this shortly.

So, let's begin today, as Alison has introduced, we will be discussing foundations of machine learning for nonprogrammers. This is a really in sort of touch subject on machine learning and deliberately aimed at those who don't necessarily have any programming knowledge at all, but are interested in statistics or economics or work for a National Statistics organization as it has been mentioned, my name Tom Wise I.

Originally come from the beautiful Cotswolds in the UK. I completed my undergraduate in psychology at the University of Reading where I focused heavily on clinical psychology and research. Nicholas, before completing my methodology and Statistics Master's degree and you've tracked university in the Netherlands where I focused on machine learning comparisons for predicting post-traumatic stress disorder for my dissertation.

Outside of work, I'm pretty boring in the fact that I love doing data literacy and data education along with generative artwork and data visualization. As it says on my bio. So creatively, I'm a huge indoor gardener and I love doing anything relating to gardening. My room is quite disgusting at the moment, in the sense that it's mostly urban jungle, but that being said, who doesn't love a little bit of fresh air in your room?

So the goal of today's or the aims of today's session and the real goals are to really begin the process of understanding and getting to grips with the purpose and utility of machine learning.

The role of machine learning in today's society is absolutely huge and so we really need to also think about more specifically about how these problems can be tackled. And so that's our second name.

Our third is actually looking at the steps involved in machine learning in a completely non programming way. Very often when we talk about these tutorials or sessions you will get a very programming centric idea where you only learn how to program the steps necessary to complete the task. Whereas today we're going to be focusing on these methodologies, we're going to be talking about how the actual computational steps take place, but not how to run the steps themselves.

Finally, we're going to be looking at how to actually evaluate and interpret these machine learning problems. How is it that when we get these results or when you read them in an article or scientific paper, what that means to you and what that means for the future of science and your topic area, whether that's in National Statistics or just in your area of research?

So, today's session is broken up into two easy segments.

Our first half is looking at the what and the when of machine learning.

And also looking at the practicalities of reproducibility, these are really through really important areas that although we'll briefly cover in this first half, really help us to understand or contextualize machine learning in this way and reproduce ability.

Reproducibility and machine learning is incredibly important. When we link back to our sustainability goals.

Because by only through making our machine learning problems our tasks replicable can we ensure that the future of science continues to evolve.

After a short break, we'll talk about how we actually use machine learning and get into the practical problems that exist within machine learning. We'll talk about evaluating and then interpreting these models themselves.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

So, hopefully during that introduction you've had time to actually go on to slider and think about what you hear when you hear machine learning. So, if we take a moment together now to actually look at some of these answers.

Hopefully we can have a look.

Have we had many responses? Alison and Sarah, my fantastic team.

Baily, Alison
Yes, Tom, we've had quite a few responses from everyone. Great to see that up there, the most popular so far are automation and big data and there are lots of other.

Different responses as well. Automatic coding can better analysis. There's quite a lot around black box and sort of lack of explain ability and some others put confusing and mysterious. So, there's the theme coming through here.

Wise, Tom
Fantastic. So, these have all really hit the nail on the head of the point of this session. These are really the.

So, these have really hit the nail on the head to this.

Wise, Tom
Beginning nodules and the beginning concepts that a lot of us introduce ourselves with the fact that we think of machine learning not as something that can be dived into can be ripped apart, can be drawn and quartered together, and then restitch, but something that we just put inputs into and just get magic out of.

And hopefully today will be able to enlighten ourselves into understanding what's going on here, because although machine learnings machine learning models have the huge capacity to do wonderful things, whether that is.

Solving problems that wouldn't be able to be solved using normal methods, such as those involving dead big data or automatically compute a programming’s, a topic or task doing natural language processing, computer vision, et cetera, and so forth. All of these areas are really common when we think about them, and it's fantastic to hear that this is what people first think of, and it's really useful to know this moving forward.

There are a few terms, however, that I'm just going to need to bring up because I'm aware that when I was writing this, these are a few words that came up continuously that I thought if I was novice to computer science and was shine learning, these are things that I might not know, and so a few terms that I thought it best to describe and talk about.

First of all, an algorithm, an algorithm you would typically hear and computer science and mathematics. But this is a finite sequence of well-defined instructions designed to address a specific problem. So, for example, an algorithm could quite easily be in this context machine learning algorithm where we solve problem X with problem Y or sorry, we attempt to solve problem Y using data from X.

A metric in our conversations today is going to discuss a quantitative assessment of performance. These metrics can come in multiple forms. Whether this is a performance of accuracy, of ability, of quality. But importantly, these metrics aren't the be all and end all and we'll come onto the niche details and the nuances that are surround these metrics later. Next, we talk about aggregates. So, aggregates are really important when we're talking about our real-life examples that we've borrowed from the statistics, Poland and.

The ABS, the Australian Bureau of Statistics, because this involves combining lower-level statistics or it layers of information either upwards to form new statistics or simply aggregating them into something new.

Umm and finally hyperparameters so hyperparameters is a very nuanced term found primarily within the data science and machine learning field, and this is what talks about, or this refers to a parameter that is used to control the learning process, but it's not directly defined through training.

In this case, or in our case specifically, we provide a hyperparameter.

At the beginning of our learning process, which we can then edit and change later, not necessarily through deriving it through our training.

But through a honing our abilities later on.

So, one of the biggest questions that we need to ask in today's session before we do anything is talk about the definition and think to ourselves what is machine learning?

And the reason I set this question is because by having a shared understanding of what this is allows us as an entire unit, as an entire cohort who are here today to understand and discuss this on the same level playing field. And if you take 30 seconds to go away now and Google Image search, what is machine learning, you get some frankly amazing memes from looking at how computer science and statistics with the support of mathematics have created this weird hybrid called machine learning.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

How the fact that machine learning is actually just statistics flaws and is rebranded as artificial intelligence.

Or even how artificial intelligence is actually a metaphorical cathulu which is actually just mathematics and programming, holding it all together.

But what's really ironic here is that beyond all of the humour and all of the fact that I'm a giant nerd for thinking that these are funny is the fact that.

These have an element of truth. The fact that machine learning is not only.

Part computer science and part statistics. But it's so much more than that.

So, let's begin our search at the hub of all human knowledge that is Wikipedia, and it provides this definition.

Machine learning is the study of computer algorithms that can improve automatically through experience and by use of data, and so this is from Mitchell Thomas in 1997, a little bit before the machine learning like took off as a mainstream way to solve all of our research based problems.

But this definition is fantastic because we can really break it down into further understandings. So, I like this definition that we can branch out from it. Machine learning is a branch of applied computer science focusing on mathematical processes or equations which improve automatically through experience and by use of data.

And so, by swapping out and by introducing this term applied computer science and focusing on mathematical processes and equations, it really humanizes and really in some ways.

Puts a new lens on this.

Because by thinking about this in terms of the fact that actually this is just a new way of applying the same traditional mathematic pro mathematical processes or equations but using all of the computing power that we have available to us for so little money.

It allows us to do this.

So, caveat number one of about 1000 today during today's session.

Is that?

This is nowhere near a perfect definition since.

Computer science and calculates a huge range of different topics, from artificial intelligence to different topics and everything else. But this is a good place to start today, and we will focus on using this Wikipedia definition when we talk about machine learning.

So, let's bring on to one of the first topics and start thinking about what methods are actually captured under this machine learning, and this is a really important topic, not only because it allows us to all understand what we're talking about.

But also, what falls under its definition?

As you saw.

Machine learning is broadly defined as many procedures wherein initially your imputed model.

Undergo some form of learning steps as previously defined as black boxes before producing a hopefully more accurate and usable and applicable model than at the beginning.

As a result, this can be a seen to include things such as your traditional statistical methods, regression principle, component analysis, correlations, so slightly debatable.

Like lots of lots and lots of debatable notes there.

As well as more than well-known examples for machine learning such as computer vision, natural language processing, all of which encapsulate this statistical learning.

Importantly, however, this is really not a definitive list. When we say what methods are captured by machine learning.

These are what is available today.

There is a theory within computer science within science itself, that says the past five months or five years of scientific evolution outweighs all previous 5000 years of scientific advancement.

Meaning that the next five months of methodological and machine learning based statistical development could outweigh the last five years.

So, when we talk about this session today in five months' time in five years' time in 50 years' time, this could be seen as traditional statistics. So always take what we're talking about in with a pinch of salt because as National Statistics, organizational professionals as well as those who come from the public, we need to be able to talk about this field with this nuanced and this pinch of soul.

Not only because.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

Are improved understanding but also the availability of such core tools as computing power, where everything is growing so rapidly.

Now then, let's talk about something slightly lighter than the embarrassment of statistics and the advancements of science.

When we talk about traditional statistical methods, I personally would argue that these are as equally valid without the assistance of machine learning as they are as they were 50 years ago.

But now, rather than having to handwrite several hundreds of pages of equations, as once was told to told to me by certain diamond that he used to do by hand, we can actually now just simply input several new numbers and several defining factors in a statistical equation.

Before simply putting it out in seconds minutes.

And so forth.

And what's even more wonderful is we don't have to think necessarily about the limits that we place on this. We can explore hundreds of thousands of relationships at once and undoubtedly, as computer access increases, so too will the machine learning options that we have available to us. Whether this includes neural networks and other topics, who knows?

So when we discuss exactly what machine learning is and what it can be used for, let's remember that this is what we're talking about today.

And as time passes, things change.

So, the before we dive into the use of machine learning on slido, I've got question.

What official statistical problems could we address using machine learning?

Bear in mind that although a lot of us here are National Statistics organization individuals, please bear in mind that you can be as specific as you like, but bear in mind confidential confidentiality and security concerns when discussing these problems. But.

Have a think about and talk and explain or mention a problem that you think your organization could solve or address using machine learning. I'll give everyone a couple of minutes which is code for Tom will have a drink and we will then talk about these examples.

OK.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

So, let's take a minute. So, we've got 25 responses, which is amazing.

So, our most common ones talk about text classification, special data aggregation, automatic data collection using NLP, rates of engagement, predictions, levels, crime, someone being really honest and saying they have absolutely no idea which is fantastic and correct. We don't know what problems could be solved by this.

Data linkage image classification.

All of these are officials are problems that could be very.

Not easily solved, but could be addressed using machine learning and it's fantastic that everyone has a very good idea as to what.

Problems could already be solved by this.

I think one of the most interesting ones here is missing us.

And this is an interesting an interesting concept that I won't get into today, and there's quite a few topics that I won't have time to get into today, but missingness is something that actually when it comes to talking about machine learning, is something that we need to consider handling, but also could potentially be solved. So, think about how if you say for example, a working with census data, how we could use machine learning to sometimes fill in the blanks that we have here today.

So, let's come back on to the next big question, which is.

What types of machine learning do we actually have, and what techniques or how do these fall under them?

So, for most machine learning techniques we can be group, we can group them into two types. Supervised learning those which use tagged or labelled data to learn and develop or unsupervised learning. Those which used untagged or unlabelled data to learn and develop further.

The fundamental difference between them and between these two learning approaches is the nature of the data provided into the model.

As you can see from these two definitions.

One very straightforwardly uses tagged data or labelled data, which is information saying that.

We know the output to this input in this observation.

And.

The model can then better predict or identify relationships between them by understanding what is our input and what is the output, especially knowing that the output to observation 317 is output 317.

And through having this labelling and clear specification of the input and output page, this guides and provides a lot of accommodation for the learning process.

By contrast, the unsupervised learning process uses untagged and unlabelled approaches in the learning process, and rather than using this paired approach where we say, here's your input, here's your output. We say here is the data I would like to find the relationships that exist in here.

And this is sort of seen as an association method where more natural explorations come in. This is more of what human beings might do. Say, for example, let's talk about natural language processing and words.

Although typically defined as a supervised learning process, we could use unsupervised learning in this case to understand how things are grouped together semantically so let's say we have the words cats and dogs. Both of them for a lot of us know that these are pet animals. These are domesticated animals that we have in our homes and so the words that or the words cloud that they may be in which are linked to pets' things like rabbits' birds.

Other topics like that they would then be grouped together to say pets.

But of course, this isn't necessarily guided and can then be used to further used supervised learning.

And this is an example of the fact that these problems can include one or multiple or a blur of different methodologies in order to come up with an effective solution. Some problems may begin with an unsupervised approach to get a gauge of what is going on in the universe that we are operating in, which then develops into one which uses supervised learning. Once the initial model is then confirmed.

And so when we talk about machine learning in supervised approaches, this is where we talk about regression versus classification, and we'll come onto what these mean and a minute, whereas methods under the unsupervised learning is largely clustering where we cluster together these ideas.

For these umbrella terms, these techniques can be broken further down into our regression.

And classification problems.

Generally, and I say this with a massive pinch of salt, by the way, I apologize. This is going to be like a really, really salty lecture by the end of it, so I apologize.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

Each of the sub techniques can be connected to specifically to a problem, and although this is an A direct link in the sense that some problems can be solved with the other.

We can generally see that the aim of regression is to predict or project a a problem. That is to say to project.

I potential outcome of an issue.

And the outcome of this typically is continuous and the outcome variable we're looking into, whereas classification doesn't. It says on the tin, a classifies and as such it's looking for a class, it's looking for a categorical variable.

And we don't have time to really dive into it today, but bear in mind that sometimes they're continuous variable can be transformed into a categorical one.

As well as a categorical can be transformed into a continuous one. So when we look at these, when we look at the supervised learning approaches in the real world, bear in mind the fact that sometimes if you are looking at a categorical problem, it could be could be a continuous one.

By contrast.

When we talk about machine uh, unsupervised machine learning, we're talking more about clustering.

And this the aim of this problem is to group or cluster the results and there is no outcome variable here necessarily.

In the sense that we are looking to explore what's going on with the data, not necessarily find a result or work out a result of what's going on.

And these are obviously major rules of thumb, not only because the continuous variables, as I've said, can be converted into categorical.

But because sometimes the cases of these outcome variables can be removed in order to run a classification or a clustering technique, and to be able to access the group's present.

So, as I said, these are rules of thumb. Take them with a pinch of salt, but these are the main 3.

And so when we talk about these in the real world, there are some commonly known methods that fall into them. In regression, you have your commonly known linear regressions, general linear models. Your classifications, you have your family of decision trees and support vector machines. Something that will be talking quite a lot about shortly.

As well as unsupervised learning, which falls into the clustering family, and in particular things such as K means clustering.

So, before we discuss our final topic before our break, let's talk about how we implement or let's talk about some real world examples and how we can actually contextualized.

To contextualize the examples that we're doing here.

So, before we get started, firstly a caveat, these product projects will keep conducted under the HLG MOS Machine Learning project, which have been provided to me.

I started in the ONS as a new starter in January this year, meaning that I haven't actually engaged in a lot of machine learning projects. I've engaged in a few, but none that I feel have completed their entire cycle that I can bring to the table today, meaning that the team that has been supporting me have so kindly provided these as examples that we can talk about, but also shows the very real world applications beyond economic statistics that.

I work with.

And if you're interested in reading more about these, the link is available at the one of the very last slides, which I believe will be circulated after the session.

Let's, however, before we even start talking about these real world examples, talk about the elephant in the room. All the able elephant on this table, which is the fact that to my knowledge and the knowledge of those who come from this group within the UNECE, is that there are very few or limited projects that have undertaken unsupervised learning as their main aim. This doesn't mean that they don't happen, but what this means is that they typically may actually fall into a different stage.

Of the machine learning process, and one that we will talk about early later on in the data preparation phase.

And, as such, we will not be talking about any examples which fall into the unsupervised learning category today, but in the real world beyond official national official statistics, you can talk about recommender systems, customer segmentation, DNA on a larger scale genetics, and a huge variety of different topics which are hugely reliant on using these unsupervised clustering methodologies.

Additionally the use of these in conjunction with supervised learning techniques ensured that.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

A lot of these techniques are actually able to come to fruition. These provide the business case, the pilot studies that show that there is something going on here, but we don't know really what it is.

So as such, we will only be talking about two of these wonderful methods from the supervised learning family.

Firstly, the supervised learning from statistics Poland. The imputation in the sample survey on participation of Polish residents and trips from the Statistical Office in the statistics Poland, which has multiple input values but has a continuous outcome variable.

As well as from the Australian Bureau statistics machine learning application to automate an existing manual process of using aerial imagery. Not a particularly exciting name, but trust me, it's actually really cool. I got very into it, but we will come back to it in a minute.

First of all, let's introduce our first example. So, the first example will be talking quite a lot about today is from statistics Poland. They used multiple supervised learning based techniques to impute the sample survey on participations of Polish residents in TRIPS. What that fundamentally means is they used or. The goal of this research was to identify the suitability of supervised machine learning. So that is to say how.

Useful would it be to apply these techniques in the imputation or the correction of missing data of aggregates used in a particular areas of classification of the expenditure based in category and country?

Unfortunately, due to the limited time, we can't dive into the sort of the nitty gritty of this, but actually this technique was really wide in its scope and used a huge variety of different techniques, both which fall into the machine learning bracket and the non-machine learning bracket.

And they use several nonlinear non machine learning techniques such as linear modelling, ordinary lease to as regression, general linear modelling and predictive mean matching imputation.

But in particular, they use several very important machine learning models, which we will focus on today. Firstly, they use classification and regression trees or the CART family in brackets because I'm aware that no one's actually seeing me live other than Alison and Sarah.

Brandon Forrest techniques.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

Optimal weighted nearest neighbours and support vector machines.

The wonder of classification and regression trees is the fact that these encompass a hugely broad family of techniques, all of which look to.

Predict based upon a feature space that they are provided with. That is to say, when you provide it with data, it will split up the technique and we'll come really on to what this means in reality after the break.

But this includes techniques such as decision trees, which are some of the most simple regression models.

Random Forest, which are collections of decision trees, support vector machines which bring this into a three or fourth dimension as well as neural networks as an extra boosted machine learning.

The second example that we have actually looks a little away from what our focus of today's session is, but it's still really important in our classification.

This uses a computer vision in order to classify.

The addresses of Australians based on or the registrations of in the let's, try and get this right, my apologies.

Uses the address register linked to geospatial images in order to classify what their location is assigned to. For example, if we take it and address, we can classify whether it's classified as commercial residential under construction or vacant land, for example.

And this is a wonderful example of how we can use multiple tools in simultaneous in simultaneous natures to be able to do this, this used a computer vision program which unfortunately we cannot talk about today because it's so complex and I am not an expert on it. And so, I would hate talking about it.

But also, is 1 which uses these classification techniques, which are really really vital and important.

And these all add in value to the developmental processes.

By doing so, these reduce the risk and the requirement of human intervention when done correctly.

They can reduce the risk of human error.

An important thing about this example is that they focused, as I said, and have said multiple times on computer vision.

Meaning that because this is not a session designated to talk about it, we will only be talking about how they evaluate these techniques rather than the actual techniques themselves. So, I really apologize about that, read the paper. It's really interesting, and if you would like some more information about this, send me a tweet, send me a message on LinkedIn or whatever platforms I have shared already, and I will send you some examples.

So, before we talk about reproducibility.

Next slide over question.

What can we do to ensure the reproducibility of our machine learning projects?

So, the reason I asked this question before even beginning to talk about reproducibility is the fact that I wanted to see what you already think know or do.

In your organization about reproducibility.

So, this can either be in your current projects, in your future projects, in a wish list example, but what do you think?

Uh, we can do to ensure the reproducibility of our machine learning projects.

OK, so we've got a really interesting range of.

Points here.

So, we've got several that relate to reproduce support, analytical pipelines or wraps.

We've got the use of programming, the use of setting seeds, the use of clear documentation, open source coding, keeping records of methodologies using open source data.

Sharing code.

Using understandable packages library of methodologies and these are all absolutely fantastic version control and Git. We'll talk about that in a moment.

But what this encapsulates to me as the as your presenter days, you've actually got a pretty good idea already, but let's talk about a few of the nuances that are in existence here.

So, when it comes to talking about reproducibility within machine learning product projects, we can break the topic down into 3 core elements. Critical to creating a reproducible product.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

Firstly, that is the code used. Second the data used and thirdly the environment at which this is created in.

I'll note once again, this isn't a defining list. See your individual organizations guidance on how to produce reproducible code or talk to your the sort of the international bodies on reproducible code.

But this is a really useful place to start.

Defining these methods and can be really straightforward.

Code is the foundation of any project. This defines what we classify as the machine learning algorithm. Data is what the algorithm is trained to, validated and tested upon.

And the environment is what the machine learning algorithm is built, developed and run within.

So fantastic, we have our definitions.

But what does that mean and what are the challenges and how can we solve them?

Each of these elements have some challenges which we can explore and although not exhausted, this will give us some idea of how to begin the process of actually solving them.

Firstly, when we consider code, there are lots of things involved. Things such as an inconsistent science style pseudo because computers can't achieve this randomness. That is to say, when we ask a computer to give us, say for example, a randomly occurring normal distribution, this is not usually possible. However, by using the set seed function, as many of you mentioned, we can actually redefine this so that we can manage this in a reproducible manner.

There is something also called untracked development, and this is something alluded to in the use of GitHub and version control.

An untracked development is something that happens when a you are simply expanding the code in a way which is not planned or.

Expected and by having development which occurs naturally, it's it is a fantastic thing to have naturally developing code, but also horrendous in terms of reproducibility, because if you cannot or another user cannot understand how you went from point A to point B and point B to Point C, they will only see that you've gone from point A to Point C and so by not understanding how to go via point B means that it can be very difficult to achieve the same results you have.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

One of the major issues within or the challenges within data is changes in the data itself, so this can not only be if, for example, our open sources of data update or change simply over time or simply due to nuances in the nature.

But due to rule changes could be changes due to your code being changed and so forth, but you're using an old piece of code.

But all of these relate to actually the code itself.

And finally, environmental factors when we consider the challenges of dealing with the code itself and the environment that we create this in, there's a lot of issues. This can include hyperparameter inconsistencies where.

Things are not correctly labelled or libraries or documented that things have changed, meaning that people will assume the default has been used, but that's not the case.

But then we've also got the more the more risky change. That is, the fact that the library, the framework or the packages have undergone changes, updates or revisions that aren't tracked.

We'll have not been specifically noted and so as a result, there is a quite an easy solution for this, but it means that it can pass over quite a lot of people.

So, if you look at all three of these in tandem, you'll see that something comes up time and time and time again, and that is tracking your changes and logging them correctly. By doing this, you create a paper trail of what you're doing in your research. This is coming more and more present with the development of open science in the academic community, but may not be seen as such within the National Statistics organizations. I don't know enough.

But what's important here is the fact that by.

Tracking your experiments by tracking how your data changes in the nature of that the course of that river.

As well as invite.

As well as tracking the environmental changes that go on, we can really understand how things have changed.

Next up, we have another one that comes across multiple times and that's version controlled. Version control is not only a fundamental aspect of reproducibility in the sense that.

It allows you to go back if something goes wrong, but it all also allows you to think about what is going on in the.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

Current status and it allows another user to come in and see where things have moved on step by step.

Coming down onto the individual factors.

In the code, a solution to a lot of our problems is things like styled gut is topics like style guides.

Where in our code assigns itself to a single style and carries that throughout the entire project. This can be hard with multiple projects, but or multiple users on a single project. But if you follow certain style guides which we'll talk about in a minute, this could be incredibly useful.

We also then have randomization management wherein the management or the randomization of a topic or data is done. So, in a very managed way and a very clearly and specifically designed way in line with scientific theory.

Finally, due to time constraints, we'll move on to the environmental factors and that's the one thing I would like to talk about is module model registry.

And this is where you actually register the fact and the things that you are going to do, as well as all of the metadata, all of the things that are going alongside it. And by doing this, this allows you to have a saved state which other people can come back to and make sure that they've done things correctly.

How do we do this in practice? We've talked a lot about the theory, but how do we achieve this reproducibility and practice when we talk about tracking, change tracking, logging?

We use Git, git lab, or equivalencies. We create comprehensive readme or diary based documentation.

Uh, we do the same with version control with our style guides, we can talk about Pepe 8 or tidy verse and the Google Style guide, depending on which language we might be programming in.

In environmental managers, again we've got Git, GitLab or equivalent. We've got creating a comprehensive readme, guides and diary documentation, as well as requirements files.

Which can be incredibly useful for package management.

And finally, for randomization management, it's setting these randomization parameters.

Massive disclaimer here. If you're working at any organization, be aware of what systems they're using, their reproducibility management. This can include things such as here at the Office for National Statistics. We I believe we used GitLab over GitHub just because of its ability to be more corporate related, but we do have facets that use GitHub and depending on what organization you come from, talk to your IT department and see which one you can integrate.

At this point, I'd like to open the floor up to see if anyone has any questions so far.

These can be addressed on slido once we've had once we've answered any questions that there are, we'll go for a 10 minute break and then we will crack on with the next half of our session.

Phelps, Sarah
OK, so I will read the questions out to you as we have to. The first question is can you give an example of using ML in the producing of existing official statistics?

Wise, Tom
Excellent question.

Off the top of my head unfortunately not.

This is simply because.

As we've seen already, Emma itself isn't.

Isn't so easily defined as.

Isn't so easily defined as. Oh yes, you must use this method or that method. This can be as simple as using a regression model in a linear or in a machine learning way.

So, the answer to that question is off the top of my head. No. So that doesn't mean it's not being used.

In the UK government, I would recommend talking to the national, the National Office for the Office for National Statistics. Let's get the anagram of my own or the sorry, the acronym of my own organization, correct.

But that being said, the more teams that understand and know about machine learning, the more easily accessible it is to actually come on to producing these official statistics.

Phelps, Sarah
OK. So, the next question is methods like linear regression, logistic regression.

Hey, means have been used by statisticians for many, many years. Where are the pure? ML methods.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

Wise, Tom

X What are the pure ML methods? That is an excellent question.

And again, I'm going to give you a really disappointing answer to say in my opinion that our no such thing as pure ML methods.

That is because ML methods aren't necessarily new thing, they are a simple rebranding of the statistical methods that we have been using for the past 1000 years.

When we talk about linear regression, as I example, I mentioned earlier, I was in a a conference attend or presented by that Ian Diamond, the national staff position for the United Kingdom.

What we talked about in the integration.

And he mentioned that when he was first writing his thesis in, I think it was economics or something on those lines. He did all of his regressions by hand.

First, writing his thesis.

Comics.

Which is fantastic.

But now when we complete a machine learning model.

A machine learning model. We don't need to worry about doing it by hand doesn't make it the method any less valid.

It is only because we have gained the power, whether that is through the development of computing power or computer accessibility, can we improve the the access to these methods in such a strong way?

Because Open.

So, in answer to your question, do we or where are these pure?

That's your question.

Pure ML methods. I would argue there's no such thing as an A pure ML method, because every method is technically in existence.

I would argue that.

Every message is technically existed.

But as we develop more and more sophisticated computer methodology, we can increase the presence of these methods. Neural networks, maybe one that a lot of people are thinking and considering.

But more and more sophisticated.

Increase presence of.

That works all that long thinking.

What could be considered a pure method?

The problem here is the fact that when we talk about neural networks is that they are in theory feasible outside of machine learning. But very few outside of where to take you years or days to complete?

So yeah, I hope that answers your question and I'm sorry that it's really disappointing one if it is.

Have we got any other questions, Sarah or Alison?

Phelps, Sarah
We only have one comment. Yeah, it says no question, but great memes. Excellent. Thank you very much. Well, and we've got one more question, just 02 questions just appeared now.

Wise, Tom
Yep.

Excellent. Thank you very much.

Yeah.

Fabulous.

Phelps, Sarah
Example to Statistics Canada 2009 economic indicators such as surrogates for environmental indicators ML slash NN to estimate air pollution from industry facilities.

Wise, Tom
Fantastic. Thank you very much for providing one example of that.

Phelps, Sarah
And then I have a question. Does this mean existing survey still need to use existing methods?

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

Wise, Tom
Yep.

Oh, that it'll wonderful question.

The answer to that question is.

I don't have a good answer to that question. The reason I don't have a good answer is not because I don't know.

That isn't an answer, but there isn't an answer that I think will age appropriately.

That being that is because when we talk about should our should all techniques such as, say for example the sensors, always use the most accurate, up-to-date methods?

You know useless.

They should be, but they should also reflect the ones which are the most suitable for the time, because there's no point rushing at system if we don't have the capability to do so scientifically and accurate.

They should be but they should also protect the walls of job suitable for there's no point rushing a system. You, fine, typically.

There is a fantastic piece of work done by a Harvard holding PhD mathematician called web weapons of maths destruction.

And it's a fantastic body of work where it talks about how, although topics like machine learning, computer algorithms have the potential to do wonderful things, they also have the potential to do devastatingly catastrophic things because the moment we start using untested methodology on topics that can impact people's lives, we have the potential to do horrible things.

What?

Topics white machine learning.

The entrance to.

You also have the potential to do devastating.

Because the button.

Not using untested at the.

That being said.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

We should always discuss what is the most appropriate method for the most appropriate survey.

Should always discuss what the most.

So, I apologize again for answering this question in the same way.

Apologize again for altering the question the same way.

I'm not 100% sure and I think it's very, very much a case of case by case basis depending on how the science is at the time. I hope that answers your question, but please reach out to me if you want to discuss it further cause it's an interesting topic.

Phelps, Sarah
We have another question for price. Does statistics existing methods is to physically collect the price at the outlet using crawling and scraping. We still can get a comprehensive price due to limited coverage area.

Wise, Tom
Umm.

Yeah.

Yep, that's correct. That's actually the field that I work in, so.

Yeah, that's correct. But actually, I'm not sure which comment that's worked in but.

But yeah, I don't know if that was a comment requested. I'm not sure how to answer it. So, is there anything else on the tables error?

Fantastic. Well, oh.

Oh, just looked in.

ML equals simple regression on steroids.

Yeah, I'm talking. That's pretty, pretty accurate. If I want to, I think it's a. It has the opportunity to be a simple regression on steroids, but if it's done correctly, it can also be that, that individual who spent years or that Olympic athlete who spent years with no roids to develop something that is truly wonderful.

But if it's done correctly. The compilation develops something.

And I think on that comment, let's take a 10 minute break during the break today.

I think during the

Answer The next two slider questions if you can. What country organization do you work for and what is your role in that organization? The reason I ask is I'm curious.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

I'm curious how or what diversity of individuals do we have coming to this session today for this coffee and coding session, but also what role is attracted to this type of information and at the end of the session we will talk further about this fantastic so we will meet back here.

Once the next two slider fashion.

What country will you look cool?

How?

Varsity of individual student.

Also, always this type of information.

So, we will take a break. 10 minutes, so 10 past the hour of whatever time zone you are currently in.

10 minutes so I thought.

And that will give you plenty of time for a biological break. Thank you very much.

Hi everyone. Just to let you know, we're just going to be ending the break in about another minute or so. So yeah, if you're back and ready to start, then we'll see you shortly.

Wise, Tom
Excellent. Hopefully everyone here has now managed to have a little break managed to nip to the restroom if you needed it, or go and get tea and coffee as I have done.

We're just waiting for a member of our team to get back so that we can have a look at all your responses, but I can see that approximately 40 if you have done that, which is amazing. Thank you so much. We will come back and discuss that later on. So, feel free to fill it in if you want to in a minute.

So, let's talk about the machine learning life cycle as it sets. We've had a break.

We can now consider actually. What scientific process goes into making these machine learning models a reality?

So, a slight caveat before we get started.

The machine learning lifestyle, or life cycle that you see here is one that has been its sort of an amalgamation of several different ones that I found while during the researching and building this project. So, if there's some things you don't recognize or something you think I've missed, don't panic. I probably considered it, but do let us know and it will help us improve it for the next iteration of this program.

So the life cycle of any machine learning project begins through the gathering of data, and this is where we actually have defined the project. We have said what we would like to do.

Once we've done this, we can actually process it through data cleaning and manipulation.

Once we've done this, we can prepare it, wrangle it, prepare it for analysis.

We can then train a model based on this and then test, test, test and test before tuning and optimizing it.

And what's important to note here is that this diagram isn't linear or, but rather actually circular, cyclical and repetitive in nature. Once we've tuned a model, now actually optimized it in what it is, we can then go back and gather more data to supplement the model. We can go back and retest it on different parameters. We can go back and retune it; we can deploy it and make it a reality.

And this is the wonderful part of it.

In the examples we'll be discussing shortly, it's important to note that these demonstrate at most one iteration of this cycle, but mostly the first few in the sense that they are very preliminary in their design, and this is because this is a the ML group here is largely a as I, as I see it, a research assistance organization rather than one which.

Leads or?

Goes fully out into that area.

And as a result, what we see here?

Is actually something that can be segmented into multiple groups.

So, this life cycle is very rarely seen by an individual person.

Rather, it is actually one which is split up into multiple teams. Here within my team and within the Office for National Statistics, we have potentially one or more simple teams who gather the information and data. Another WHO data, who clean and manipulate the data and prepare it for preparation. Another team who actually prepare this cleaned and manipulated data into something usable, another who researched and actually build these models that require training and then another who goes on and actually emerge.

Who takes these emerging platforms and trains the data on them, testing it as they go before finally tuning? We then have a deployment team who deploys it into the real world and again begins that cycle all over again.

And by having this cycle split up into so many different pieces, not only do we get experts in these singular areas, but we allow accountability for that topic area. So, we have our team who is accountable for all of the data gathered in one area.

All we have a team accountable for the model testing, meaning that if you're new to machine learning and machine learning in in National Statistics and you're wanting to dive in further, it's OK not to be an expert in everything. But being consciously aware of the other steps that are required allow you to take a nuanced look at what you're doing and how to take it. Take it further.

But let's go through these step by step. I'm going to skip over the first step, which is the gathering of data, and this is because a lot of us, if we have completed any form of scientific training, whether that be a GCSEs, A levels in the UK, a degree, a masters or a PhD around the world, we've probably collected data, some form or another in our capacity as a scientist. So, I'm going to skip that mostly because of time.

So, let's talk about cleaning and wrangling because it said somewhere between 60 and 80% of a data scientist time or anyone that works with data is spent cleaning and manipulating data into one form, and the form or another. And this may come in a number of forms, from handling of missing data. That is data which is missing by randomness through just being an outlier in the data, or simply correcting incorrect data.

With data being removed or imputed accordingly.

Then we have the process of feature engineering and selection to better prepare the variables to correctly address the desired question.

Whether this means that we normalize scale or apply other techniques to this problem as well as ensuring that the data is of the quality expected and having this data of a quality that we expect is vital. Making sure that we know that when we are getting data and we are applying it to our models that we know the quality going forward in addition to assuring this, we can address the simple question such as.

Is it structure correctly? Do I have the types I expect to have and so on?

From our official examples, those of Statistics Poland had their cleaning and feature selection beated by specialist in tourism. This is really important. This is where the dimensions of domain expertise, that is to say expertise related to the specific topic in question, can really come to its head. Meaning that as a data scientist do you are not really expected to know everything about the topic you are researching because you are the specialist in the engineering and the building.

Whereas the person who is specialist in, in this case, tourism in the Australia example, in classification of housing based on their construction quality or their construction.

That construction type, commercial, residential, etcetera helps us understand the problem that we're talking about, because this means that at multiple steps we can include.

Interpretations from these experts, meaning that by the end of a project we can have enough domain experience to meaningfully apply these steps, because there is absolutely no point completing any of these steps if we don't have the right knowledge at the right point. Because, for example, let's say we talk about the statistics Poland example and we're talking about, and we have no one on the team who is any form of expert in tourism.

We are going to talk about variables that we don't know what they mean in the real world and if we don't know what they mean in the real world, then of course we can't interpret what they mean in the real world. So, the model becomes a little bit more useless than it should be.

Similarly, however, in the ABS example they use specialist to collect and process the data, and this is really important because the use of this specialist knowledge supplements this data even further.

Next up, we have our data preparation, so once cleaned, manipulated and successfully wrangled, it's time to prepare the data for model. Application preparation typically comes in the form of splitting our data into multiple different units. Firstly, I split a colours into our testing or our training and testing sets and then our training and validation sets.

Splitting this into three groups, typically 8020 and then 82 or 8020 again or 9010 in the case of Statistics Poland, we can really, we can begin to create data sets that we can specifically train or test or validate our model upon.

So, let's talk about a little bit more about these data sets before we move on.

When we talk about training and testing sets, this is saying in our entire data set of, say, 10,000 observations, I would like to train our model on 80% of that and leave 20% untouched as novel as new data to test out the method.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

This is deliberately designed so that we do not overfit. Underfit our data to the trained data we have in front of us, and by doing this it allows us to mean that our models become more interpretable. We can say with more certainty that the models we have produced are better than just those trained on an entire data set.

And that then is also accountable for the training and validation where we take our training set and then split it again into an 80:20 split, because by validating the trained data or the trained model, it allows us to better understand before going on to test our data.

We can then also there are also several other things we could talk about at this point, things such as preparing it for a K fold, cross validation where we split the data.

A further into different chunks so that we have a rotation to reduce or increase the amount of repetitions we can feasibly do with this. We're not going to have chance to talk about that in much at the moment. But importantly, I wanted to mention about specific model preparation where actually we can prepare the data for specific usage. Remember earlier when we talked about.

You are making sure that our data was in a continuous or categorical variable. This is where we would do it, making sure that we have categorical or classification variables when we need them.

And then as I mentioned, we can determine the K fold cross validation because what this does is rather than just participating it into our training and testing sets, we say how many times would I prefer to split this data up and have it on a rotational loop. This is quite complicated topic in terms of cross validation because the consequences of this are quite large. So, if you are interested in learning more, do look it up in your own time just due to time limitations today.

Once we've cleaned them all the data, we can actually train the model, and this usually begins with your default model.

And this is typically already programmed in the software you would want to use. However, because this is a truly introductory session, I'd like to go into more detail about one of my personal favourite models, and this is the CART model and this is the family of techniques which include classification and regression trees, as well as encompassing the encompassing decision trees which are the foundation of the semantic foundation of extra boosted models, super support vector machines, as well as many others.

And these classified as fantastic all-rounder models, both highly interpretable but also very often very well done.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

So first of all, we have our training data, this training data or as I like to say it, growing a decision tree.

Can begin as soon as you finish your data preparation and it's a fantastic place to start because you have a simple, highly interpretable data set.

Firstly, we have our we have our this is called our feature space.

And this exists in relation to our X&Y plane. Bear in mind this can have multiple dimensions because obviously we have multiple X variables, but this is our plane. This is our training data in relation to X&Y.

We then split it into segments. These segments are then split again.

Until we get to a point where each division becomes independent from the other divisions, and we mean that we can classify things.

To the final point.

For example, in this case, let's say segment one classifies all of them into one group. Then we can say that is correctly or that is completed.

By contrast, segment two may require three additional splits because it is a, it's still contains information we need.

The divisions of the feature space here are done through what is called a binary classification statement, or yes or no.

Does X relate to or? How does X relate to Y?

Yes or no.

So what we can do here is once this is done, we can create this gorgeous decision tree and respective diagram. So, this is an example I've stolen offline, so I apologize if you do recognize it, but basically, we can say.

Let's go through the example that we've got on screen. Is X greater or less than X1 being just a point? Yes or no. This is our binary classification and then we can just repeat the process.

This will result in US producing the skeleton of a basic decision tree.

And as a result, defined 5 unique segments within the data of our feature space.

I could easily continue all day about decision trees. I've actually written other content on decision trees which you can find online, but they are an incredibly useful tool to understand because once you've understood this, we can easily begin to test them.

By testing them, we simply take our testing data and apply it to this feature space by applying it to this feature space, you can see how the data fits into this space.

And this is the same way in which you may apply non machine learning.

A data to say for example or linear regression.

You apply the test data to the model and see what comes out predicted.

You then compare them and see what happens.

And this is then when model evaluation tuning and observation can occur.

If we return to the segmented view of the feature space us, our new data is simply overlaid onto the top. We can then determine what class the data falls into and bish bash, Bosh. We have some results.

So prior to any form of deployment, further steps should be taken into the model regarding tuning and optimization, and this is to ensure that the most suitable model is defined and developed in our in our case.

And it's through the process of tuning that we can actually define these models hyperparameters.

And identify these conditions, which are classified as optimal. We'll discuss the model metrics shortly, which helped define these parameters, but.

I won't go into the more now because of the limited time.

But as a general, as a general rule, you can do these in order to tune your parameters. You can do a grid search where you build every combination of models. You can do a random search where you just randomly pick various. You randomly pick hyperparameters and see what's best. Or you can use a Bayesian optimization where you use a Bayesian methodology to use your previous hyperparameter to influence the next. But as I said.

This is something that I would highly recommend you look up in your own time.

Once you're model has been successfully tuned and optimized, it can be deployed and the life cycle can begin again and at and. This is when further adaptations can occur when deploying machine learning models. Depending on the scale, size and procedure, these can be deployed successfully as web application batch productions or simply embedded into what you're doing already. As we've already highlighted, these can exist as.

Pipelines in a the production of official statistics, as well as other areas. However, as I've already talked about, I'm not an expert in the deployment of any of these models, so I would highly recommend that if you are interested, complete your complete. Some further research. Aren't this and speak to your colleagues who may be in the deployment area of your division or department.

So, we're now getting on to something quite chunky, which, given the fact we've only got about half an hour left, I apologize if it seems like I'm rushing through.

Well, let's explore actually how we can evaluate different models and how you as a scientific reader of journals or articles can interpret these in the future.

So let's review some of these more regression based metrics of evaluation for those who are familiar with traditional statistics, you may be aware of the traditional ways in which we assess regression, R-squared, adjusted R-squared, and although these are useful, these should be discussed alongside other metrics because they don't necessarily tell the whole picture. In the case of machine learning, these both just define the amount of variance explained within the model.

We then have several others which include the mean absolute error, which is the difference between the actual and the predicted value. We then have the mean absolute percentage error, which looks at a percentage of irrigation of the ME. We then have the mean squared error, which is the squared difference as well as root squared error and the root mean squared log error.

As you can see on screen.

Importantly, there are several other metrics that I haven't mentioned on screen that I will highly, highly recommend if you are doing this more in depth, you look up and. These include the AIC or the alkyne information criterion and the BIC.

These are incredibly important for model comparisons across techniques. These techniques on screen are fantastic for individual model comparisons between examples, but if you are doing this more highbrow, look up these other two, AIC and BIC.

So if we were to look at some of these examples that are going to come up in the examples we talked about.

We can see that our squared looks at a range of zero to 1, where zero has no explanation of the proportion of variance explained and one is the total explanation. Ask where this highly interpretable to highly interpretable, but it works in in only in isolation and not across models, meaning that it can be very difficult to extract what is going on with it.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

ME is the measure of prediction error and ranges from zero to Infinity, whereas the closer to zero there is no prediction error, whereas Infinity is the large prediction error.

And this is very robust to outliers, but graphically not very differentiable.

When we talk about the absolute, the mean absolute percentage error and the root squared mean again, these are very similar to the two above. These are highly interpretable, but the MAPE does not handle zero values very well because this is looking at a percentage. Therefore, is a contains a division, which means that you cannot divide by zero and the root mean squared error is not as robust to outliers as other methodologies.

But which evaluation techniques are used within the statistics Poland example? Notably, in their example they used R-squared, and the mean absolute error mean absolute percentage error and root mean squared error, which is why I mentioned the on the previous slide and although these are all reasonably standard techniques, they're described application technique helps us demonstrate several incredibly important steps that you may already be thinking about, which is.

How do we handle models trained on multiple different samples? If you're completing something like a cross validation where you have to train different values or train different models or multiple models on different sets, you may have to complete this multiple times.

First of all, what you would do?

Is you would draw and samples without replacement, amounting to approximately 90% of the data set size each time.

You were trained the model based on that parameters on this training set. You would then make predictions based on the remaining 10% our testing set.

Before finally calculating the metrics and I apologize as well, I've noticed that I haven't mentioned any of the mathematics that go into actually calculating these. For this I apologize.

It's not because they're not important, but it's more that.

I wanted to make sure that this does not include a huge amount of mathematics or programming, rather just discusses the practicalities of it.

Once we've calculated the metrics, we can average across all of these across all end draws.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

Meaning that if we have 10 folds where we do this 10 times, we just average over 10 and this is absolutely fantastic because what this means is that.

We it is an incredibly simple process to assess this multiple times.

And this is the example that you've got that what we get here.

So, what you see from the examples from statistics Poland is that you.

You have a model pertaining to each individual area which they were looking into. In this case transport based expenditure. Before we look at this, just remember that all of the models below the red line we classify as machine learning techniques, taking a moment to look at this table, you'll see that the cart, the CART model in the far corner has an R-squared value of approximately 61%, meaning that around 60% of the model variance is explained.

Support vector machine utilizing linear kernels has a much smaller, smaller AMAPE value.

Of approximately nought .912%, indicating that if we were to compare just our squared, it would be clear that Carter Superior. But if we were to compare just MAPE that lit support vector machines using linear kernels would be much more effective.

As a result, this means that what we need to bear in mind is that using a multitude of methods will allow us to get a picture of what's going on.

Let's now flip the table and start talking about classification. So, classification is a little more complicated and today we're only going to really talk about binary classifications. That is to say, something has a true positive or a true negative rather than having something that is.

Has a multiple multi category classification.

When it comes to binary classification problems, a confusion matrix can be quite easy to construct as confusion matrix is are simply the comparison between predicted classification AKA PP or NP.

As well as the actual conditions, the positives and the negatives, the P and the N.

What we then get is by populating this table we can determine the true positives, the true negatives, which are correct results as well as our false negatives and false positives, which are incorrect results.

From these, although this as a table is incredibly wordy, this actually gives us a good representation as to what we can see when we were presenting a table like this.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

Once we've done this, we can actually calculate or we can begin with some fundamental metrics that you may have already heard of in passing, especially if you've been reading papers on classification models. First of all, one of the these are the three most important that most people will learn about sensitivity, specificity and accuracy, sensitivity or the true positive rate.

Looks at referring to the probability of a positive result. That is to say, what is the probability of the test that you are given being correct? By contrast, the specificity is the opposite, referring to the probability of a negative test.

This is really good when we're talking about, say, for example, medical tests. What is the probability of a correct COVID test being received, or the results of a COVID test being correct or incorrect?

And finally, accuracy. This is something that takes both the true results, the true positive and the true negative results into equation. Dividing it by the total population to give this magical.

Total pot. This total accuracy variable. But be warned and this is something that I should mention.

This must be considered in balanced groups or slightly unbalanced groups, but not largely unbalanced groups.

Because if you are looking at accuracy of unbalanced groups or significantly unbalanced groups, for example where your one group has, say 90% of the population versus 10% of the population, you are going to get a very high accuracy most the time. But unfortunately, I don't have time to talk about this. So, if you are interested, please let me know.

When it comes to actually additional models or classification of these metrics, they can be, or they are generated in slightly different ways. You have your precision, your negative predictive value, and importantly your F1 score and area under the curve score.

These are incredibly useful ways of assessing conclusions that we can draw from our model. The F1 score in particular, I found all the harmonic mean between precision and recall is calculated by.

Completing a.

Sorry, I'm trying to word this correctly.

It's time saying the positive predictive value by the true positive rate over the summation of them both. This gives us a more powerful ability to evaluate the model in front of us.

Additionally, we can use these metrics that we're given to calculate the area under the curve determined as the as plotting the true to false positive rates at different classification thresholds.

Once again, this is not a different a definitive list, and rather explores or gives you some taste of what the different classification or ways in which you can actually evaluate a classification model.

So, before we dive into evaluating the ABS model, several things are worth noting. Firstly, as I've said several times already, they used a vision based neural networks, which is something we can't really or we can talk about, but not in a huge amount of detail. Secondly, the classification they have completed is a multiclass classification, meaning that it is incredibly more complicated than the standard binary classification evaluation.

Because there it is a significant step up comparing.

A two by two grid.

To in my case, I believe it is a 6 by 6 K grid, especially when you are trying to work out the mathematical calculations as well.

As such, we won't be talking a lot about the details, but rather how to interpret this table as well.

First, from this table you'll see or. You see here there are three tables. Firstly, as our I believe 6 by 6 grid that I've already mentioned.

And as you can see, this is quite complicated to look at, meaning that the second table is actually a better example of what we're looking at.

This gives us a count of how many different properties will calculated as true positives. So, for example the high density there were 229.

High density properties, which were considered truly correctly classified compared to 1139, correctly negatively.

Classified as in, they were not classified as high density when they were.

And so, this actually means that this is a relatively good model because things are classified where they should be.

This allows them to calculate things such as proportional accuracy, precision, recall, and importantly, F1 scores.

But what you can take away from here is that actually there is lots of different ways we can interpret this model.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

When it comes to different areas.

So, the final so we've got approximately 20 minutes left of this session and I wanted to talk about an incredibly important topic, which is talking about the interpretability of machine learning models.

Because as machine learning models become more.

Feasible in the in the public eye and more.

A completable or a more compatible with Joe Public. We have to begin to consider.

How do we talk about machine learning models not only to each other but also to stakeholders, to Joe Public and the media?

And importantly, ourselves.

Because if we don't understand the steps that we're taking and we return to this idea of other black box, we are not going to be able to interpret these models effectively and notice when things are going wrong.

When I was so little bit of a tangent.

When I was doing my undergraduate in psychology.

If you weren't curious about statistics.

Statistics became a magical black box.

I was brought up on Stat SPSS. Those who use starter or excel. They would input a problem and Bish bash Bosh. You would get a usable statistical output, which is fantastic. You're an undergraduate psychologist, you don't need to know the INS announcers' statistics.

But personally, looking at the way machine learning is taught in education at the moment, we have the potential to go back to this.

Meaning that if we cannot understand what's going on in these models, we really risk.

Losing this beautiful period where we have the opportunity to educate everyone properly.

Anyway, not enough about that.

Let's look at some interpretable models.

Some of them are more interpretable than others. On the left you'll see the decision tree that we are. We have been talking about already. It's very clear to see that if I wanted to get to Group A or Group One, sorry, I know where the path I would want to take.

But if we look at the second example, which is our multiple layers here, this is from the ABS.

This is a vision based neural networks and all of the steps they took in order to classify a.

That that images.

And this isn't to say that this is an interpretable if you understand. But if I were to give this to Joe Public, they would be more confused than ever.

And this is the problem is that as our networks become more complicated, if we cannot interpret them ourselves, how do we expect someone with no statistical knowledge to be able to do so?

So, when we evaluate our models, not only should we look at the metrics we've discussed, but also how someone would be able to take that outside of it, because I cannot or we cannot expect to, statisticians, data scientist, econometrics, data analysts.

To be able to assume that everyone can read these base statistics, so having models that allow us to do this are incredibly important.

So, we can actually go on to another snap here and think about how can we find the variables which are important in these models.

So, featuring Portance is an incredibly useful tool for multiple reasons, not only because it demonstrates to us that variables which are most critical to finding a feature space.

But also, it can help define our future research.

For example, let's take a an example like this. This is a feature score importance randomly it was just one I pulled off Google, but here we can see that the higher the feature score or the F score, the better a feature defines the space. In this case we see that F5 is probably a feature that we might want to research more into.

For those like me, with a psychology background, if this was a personality test and we see that information about a depression score was really important, we might be interested in seeing what aspect of depression is really interesting there. And by understanding the interpretability of machine learning models, we can begin to extract further and research further into these areas.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

And so, in decision trees, this is really or simpler than other models, but is also the case across the board.

But importantly, this is also a good way in which we can present it to officials and stakeholders, because here you can clearly see.

Which features are most important in?

Determining what's going on.

And so that brings us to the end of today's session.

In closing, we can review our session goals. Hopefully you've begin to get an understanding of the utility and the purpose of machine learning.

Uh, whether that is and whether that is a full understanding or a partial one, we can, we've begun the discussion around the utility and purpose of techniques under the broad umbrella leading to discussions regarding the problems that can be addressed, in particular linking these two.

The examples from Poland and Australia, and then we actually began to dismantle these and interpret and evaluate them more effectively.

I have thrown out talked about things that I can't cover in this session. 2 hours is simply not long enough to talk about an introduction to machine learning. If I had the choice and I have the time and all of the money in the world, I would talk about 1001 different things over the course of 1000 weeks. Whether that includes model under an overfitting and its impact, the impact of bias or model selection and application, and the how this impacts in the real world. Talking about how hyperparameters and discussing tuning techniques to the curse and the benefits that's ensemble methodologies to.

OK Katie, methodological multi category confusion matrices as well as discussing the role of sampling group size and performance.

It's huge.

And these are just the things that I thought of when I had finished this presentation. And standing here now, talking to you all, there are even more that I could talk about.

But this is a session to get you thinking this is a coffee and coding session for you to now go away and talk to your colleagues and say have we thought about what about let me Google.

Google is wonderful.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

Not great, but it's wonderful.

And so, in today's frankly short and whistle stop tour of machine learning, hopefully I've given you a flavour of what it's all about.

And hopefully it's giving you an interest in learning more about it.

So, when you come to read about these techniques in future journals, or you come to write about them, you will be able to know, go in with a little bit more expertise.

And before I go on to actually ask if there are any questions, there is lots of further reading I can suggest. Go on to the UNECE website. Go on to look at their official statistics guide for machine learning.

Have a read of James, Whittens, Hastas and Tibetans introduction to statistical learning, pattern recognition and machine learning are for data science and python's data science handbook. All of them fantastic literature that I use regularly.

And importantly, join the machine Learning 2020 group. The this entire coffee and coding session would not be possible without them today, and I know that if just membership Sign up today is anything to hold by sessions like these are going to come more common. And if you are interested in learning more, this is one of the best places to starts especially for.

Official users of National Statistics, so join the mailing group. I am at the end of today session because apparently, I wasn't part of that.

And so, let's have a look at any questions I hand the floor to Sarah. Who's going to ask me questions.

Baily, Alison
Oh well, it's going to be Alison actually, but.

Wise, Tom
Or Alison. My apologies, Alison

Baily, Alison
Some advanced ML methods say deep learning can bypass the feature engineer as the model takes care of that, so they want to know your opinion as to whether you see this as a dangerous direction.

Wise, Tom
That is a fantastic question like that.

My answer.

Glances are two stage one.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

So, when we talk about domain experts in our learning cycle, so I'm going to go back to that page because I like that diagram. Good diagram.

When we talk about our.

Oh.

Uh diagram itself.

Domain experts can come in at any stage, so when we talk about deep learning, actually skipping over that.

That might be the case that we can remove domain experts at sometimes, but it is always important to include domain experts at the beginning and the end, because if we don't define our parameters based on knowledge that we already have.

Whether that is understanding?

This is a really stupid example, but I like giving it. Let's say for example, we want to predict how many ice creams are going to be sold on the beach.

Our domain knowledge is going to be that ice creams are more likely to be sold on hot days than cold. Therefore, if we don't know that ahead of time.

Our model is going to infinitely start off on the back foot, meaning that domain experts themselves come in and actually provide us with this insight that we may not already have, and the beach example is a really awful example because we all know that ice cream sauce are more on hot days.

But let's talk about tourism.

People are more likely to go abroad if they have more money than who is more likely to have more money in this time of economic crisis, and that rabbit hole goes further and further by talking to working groups and having these examples, no matter even if deep learning goes above and beyond, that's fine in the model staging, but not if you're working out what model to use in the first place.

So, my answer to the question is one, yeah, OK, it might cut things out in the middle, but at the beginning, in the end, probably not a good thing.

I balance your question.

Baily, Alison
Thanks, Tom, and I've got one more question coming in which is any advice for non-coders to feel more secure and confident in themselves when they say I'm a data scientist? Because question Mark.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

Wise, Tom

It is a fantastic question because you're right, a lot of people assume if you come in and say I'm a data scientist, you're going to assume you know how to code because.

Oh, current society has said that.

You.

That's what they expect from you. And I really like the fact you've asked this question because you're right.

It's not. It's not a definition of the job in the slightest.

How'd you feel? More confident. I would just say.

Be honest with them and say, actually, no, coding is not my area of specialty. Actually, my area is dot dot dot because for example, I so fun fact about me I didn't learn to start coding until about 5:00. Yeah, we'll say five years ago.

So, I'm incredibly new still to coding.

And so, whenever anyone asks, are you an expert in coding? My answer is no, because I will never be an expert in coding because coding moves too fast. I'm an explorer and I'm comfortable with what's going on in the world.

Meaning that when it comes to actually you are being a data scientist and actually training in that I was trained as a statistician in a psychologist, I have a.

Qualification in being a therapist.

But I am comfortable calling myself a data scientist because I know that my statistics is what's important about that.

And it is complimented by my programming in the same way in which a programmer might be complemented by their ability to do statistics in their programming language.

So I hope that makes sense and just remember that data science is not defined by.

Programming nor is it defined purely by statistics, but rather it's the fusion of both and you can have more of 1 than the other.

Have the house.

Baily, Alison

Thanks, Tom. I think that's a very good response and certainly a lot of the work that we've been doing in our machine learning group has really highlighted the need for.

ONS – UNECE ML Group
Coffee and Coding Transcript
27 April

Data scientists from a range of skills, backgrounds and disciplines. So, umm, one more question that has just come in while you've been talking. So, I think this goes back to your slide around evaluating ML models for classification.

So, the question is.

How reliable is the macro F1 score for comparing models when you have a lot of classes with A0F1 score for multi label classification problem?

Wise, Tom
That's an excellent question. Ironically, it's one that I wrote a lot about in my thesis, and I couldn't come up with a good response then, and I can't come up with a good response now.

Due to the specificity of that question and how specific it is relating to your issue.

Or the issues in general. I would argue that see what the literature says.

Purely because as I said, it very much depends on the field, and very much depends on how one would approach that if that makes sense. So, I would highly recommend you have a look at what the scientific literature is saying about the F1 score.

And I apologize that I'm just deflecting that question because I, as I said, I don't have a very good answer, and I apologize. I'm mentally for that.

Baily, Alison
Thank you, Tom. What we've got no more questions coming in at the moment, so.

Yeah, if we can leave it there. If you'd like to.

Wise, Tom
Yeah, absolutely. As I said at the beginning.

Do have a look at our examples for statistics Poland and ABS join the mailing group as well as.

Come along for future sessions because it will be great and please make sure to fill out the questionnaire that is going to come around later today or sometime this week because that will help inform us as to what types of sessions people would like to attend in the future.

So thank you very much everyone for coming and I hope everyone has a fantastic rest of your day.

Baily, Alison
Thanks very much, Tom. Really excellent presentation, I think huge Congrats congratulations and coming in lots of applause on the screen there. So really

appreciate such an excellent, really interactive and engaging presentation and helping us answer all the questions we've always wanted to ask, but haven't you know haven't quite had the confidence to. So, thank you for really brilliant presentation. And thanks everyone for joining.