This is the walkthrough of the final project. I wanted to take some time to go through how this assignment will work and call your attention to some very important details of the assignment. For this assignment, it's very important that you carefully read and follow the instructions in the student instructions document which we will review in just a moment. This assignment will be manually graded, so a lot of the instructions make it easier for us to grade and as you can imagine, with a large class, we need to have consistency. So following these instructions are key and just a warning that major points may be deducted for not following instructions. Now this might be a longer video but I would suggest that you take time to watch this video, then closely reading the instructions. And just note that the instructions might have changed since this video was recorded. So please make sure that you are looking at the instructions as they are written in Brightspace.

1:10

Okay, so first, as you are in Brightspace you will see of course the final project folder. And within that folder there are a couple of sub sections. The first sub section just lists the project instructions. So there is a student instructions PDF here along with the file that we're going to use for this assignment. Now, this file might have changed by the time that you watch this video as we will be changing it from time to time. So just make sure that you're using the file that is in your Brightspace we're gonna go through the instructions in just a moment. But before we do that, just note there's also a practice submission. You will have unlimited attempts at this practice submission we will not be checking this practice submission. This will not be graded. We'll explain this in just a moment but this is purely just to make sure that you do not have an error in your, your code your project and once you have submitted your practice submission, there's no error that's showing up in ko grade. Then you will come down to this last section, this final project submission. And here is where you're going to submit your actual final notebook submission and this is you only have one attempt for this because it's manually graded. And along with this and again, we'll explain this in just a moment. You also have to submit us a PDF of your project as well of your notebook. And that will be run through Turnitin in software. Okay, let's jump over now to the students instructions. So again, just a reminder that your instructions may be slightly different so make sure that you are looking at the instructions in your Brightspace first of all, the overall goal of this is to allow you to complete an end to end machine learning project that's very similar to the California Housing prices example. So take time to read through this overview over what we're trying to do and we're going to follow along with the machine learning project checklist. And this found in the appendix of your textbook.

And your project is going to be broken down into these key areas which we're going to talk about more stuff.

3:51

And again, it's just important that you read through the instructions in this document makes it easier for us to grade and make major points may be deducted for not following the directions.

4:03

Just wanted to remind everyone please take your time to work on this project because it does represent a large portion of your grade and the final grade will reflect the level of analysis that you perform on the data, your attention to detail as you work on the data preparation tips. And as you create your machine learning models and along with your ability to communicate your project appropriately Okay, so the data that we're going to use, and again, just double check your student instructions as this might have changed, but the data that we're going to use is from UC Irvine machine learning repository. And it is data from two Portuguese schools that was collected through different school reports and questionnaires.

4:54

If you actually go to this link, you will see that there's a math final and then a language file. Now we're only going to be using the math file and again, make sure that you're using the data set that's in Brightspace because we did change it for our use here. So make sure you're using it from Brightspace.

5:18

So you can go through all the attributes that are in the features, and I will let you read through that. Just a couple of notes though.

5:30

Notice that there are some absences with G one G two, and this should be g sub three. So I will update that and change that. And with that these are absences that are with terms. So there's a G one term g two term and a G three term.

5:54

And also there are grades within those terms as well. So there's a G one switch is your first term grade. And grades range from zero to 20. There's a second term grade again zero to 20. And then there's the final grade this G three final grade. Again, zero to 20. And this is going to be your output this is going to be your target.

6:21

So if you notice the notes down here, and this note is from the actual UC Irvine site here that the target attribute G three has a strong correlation with attributes G two and G one which of course that makes sense right.

6:41

This occurs because V three is the final year grade issued at the third period.

6:46

G one and G two correspond to the first and second period grades. It is more difficult to predict G three without G two and G one. But if you think about it, the prediction will be much more useful if it was the beginning of the year. And we didn't know what G one and G two is. So we're trying to predict the student's final grade and we're gonna get to that in just a second.

And then you're going to try to use all of these features or some of these features along with the G one and G two grades in order to predict a G three and that's going to be your target. Now I'm going to let you choose you're going to have a lot of leeway in this project. So I'm gonna let you choose whether or not you want to report or excuse me if you project wants to be on the regression task or classification task.

7:44

So if you do regression task then you're going to predict the numeric final G three grade. So you're going to predict, you know, between zero to 20. What is the numeric grade?

7:58

Well, if you have if you choose a classification task, then you're going to predict whether a student failed the course. Or not based on their final G three grade.

8:09

And just according to this website that I found, the Portuguese grading system is basically 9.99 or lower is an F so if you did choose to do a classification task, then anyone would nonpoint and then a non final G three grade would be a failing student. Anyone with 10 and above would be a passing student. So again, so you can choose a regression task or classification task for this project.

8:44

As I go through framing the problem presenting your findings, I'm going to let you go through that on your own and some of the required elements of this project. These required elements are going to represent 70% of this final project grade.

9:04

One key thing to note here in red, is the items that are marked with a red asterisk. So if you go down notice that you have like a red asterisk right here. So anything that has that your code must be fully run, and your output must be shown in the notebook before uploading to co grade to get credit for the rubric item.

9:29

Now, co grade is going to run your notebook just to make sure it's your Python code just to make sure that there's no errors in it. And we're going to look at that grade in just a moment that rubric item in just a moment.

9:44

But not going to be able to see the output.

9:48

So you're going to need to fully run your notebook and if there's any that we need to see, and there's going to be labeled with a red asterik you need to make sure that that is shown that output is shown before you upload that to code grade because we're not going to be running your notebook in terms of your output. So that needs to be shown before you upload it.

10:16

If you have any questions on that, just please please let us know.

10:20

So here we have the light required item. So up under problem framing and big picture and again this is from that machine learning checklist from your textbook.

10:31

There are some different items that we're looking for here and this one specific rubric item is 5% of this grade.

10:43

And then get the data is 5% of your grade. And there are multiple things within this get the data that we're looking for.

10:52

And notice this red Asterix so we do want you to output this rubric item so that we can see it on a screen and we go through exploring the data. These rubric items are individuals that 3% 4% 3% We're asking you to produce at least four visualizations. Now, co grade is only going to be able to use matplotlib and seaborne. So if you wanted to use some other type of Python visualization software is not going to work in code grade. So it's probably safest to stick with mat plot lib or seaborne.

11:32

We need at least for now you are more than welcome to do more than four and that will also help you to you know analyze your data more and just show us your work.

11:45

So we have preparing the data where you're going to create some data pipelines built in missing values. There's a custom transformer, this worth 5% of your overall grade where a couple of things need to happen in this custom transformer. So there needs to be a parameter that when equal to true, it drops the G one and G two columns. And when it's false, it leaves the columns in the data.

12:15

So if you remember earlier, we talked about how it would be a much more useful model. If we could predict GS three without G one and G two grades. So we want to be able to run our models and test them and see how well do they do if we leave G one and G two in there.

12:39

And then pull out G one and G two. Create a model and see how well does that model do with the hope is that we can try to predict G three without using G one and G two but you know we're just going to the test it.

12:57

The second part of your custom transformer is that there are abacuses, D one G two and three d.

13:06

So what you're going to do is you're going to create a new column that's going to add up these three columns. Create that new column with the sum, and then you're going to drop those three columns these absences, G one, G two and G three columns. This is what your custom transformer should be able to do. And again, there's some more data preparation here, column transformers correctly, excuse me, transform your training data using the above data preparation steps. Using your pipelines here. You're going to shortlist some promising models so you're going to fit at least three if not more.

13:50

You're going to use a custom transformer just like we talked about to see how well do your models perform with the G one and G two columns? And then how well do they perform without them?

14:03

We're going to compare all the models using cross validation.

14:08

So we're going to ask you to pick one model and use at least one grid search to fine tune your hyperparameters now, you are welcome to use more than one grid search more than one round the grid search but it needs to be at least one grid search in your project. Transform your testing data with all the data preparation steps that you did above. And again, this is where your data pipelines are going to come into play.

14:37

That way you can just run your testing data through the data pipeline. And you won't have to really go in and just manually do all this data preparation for your testing time.

14:50

You're going to select your final model and then measure its performance on the test set.

14:56

Okay, so those are the very specific rubric items are 75% of your grade.

15:04

And we keep scrolling here 10% of your grade is going to be the analytical insight that you put into this. So just reading here, well done machine learning project should analyze the data and share the insights gleaned from the data. Your key key findings should be communicated clearly. Use beautiful visualizations easy to remember statements such as the median income is the number one predictor of house prices and again this is from the textbook and is described in the machine learning checklist. So here's the rubric view for this analytical insight. So basically, if there's really no analysis or just very limited, you know, maybe we can tell that it was just rushed through. Then there's a 0% here, then 5% And I'll let you read through this. And then the full 10% of the category, this rubric category. And again, I'll let you read through this.

16:11

Next is communication. So communication is very important course for data scientists. So this is where it's 15% of the final grade. And we're looking at your notebook organization, your comments that you have an overview section concluding section that you go through and say what worked and what didn't work. What are your potential next steps based on your findings. And here are the various rubric that we're using the specifics of 05 10 and 15%. And again, I will let you read through that, or what we're looking for here. And for this section for the analytical insight section. We're looking for something that we can tell you did a lot of work.

17:02

If you know if a student really just tried to pump this out in one day, then that's probably going to show up in in these different rubric categories of just not being very effectively communicated. So we're looking for some good work on this final project. And we want this to be something that you would be willing to show, you know to your employer, and if you look at what we're trying to do here we're trying to communicate to your quote unquote manager and this someone who understands machine learning. And you're also trying to communicate to the board of directors for your school.

17:51

And we're pretending like those board of directors don't have any knowledge of machine learning. So you're going to be walking through this project, keeping that in mind explaining what you're doing through your communication, using good comments using good markdown which we're going to talk about again in just a moment.

18:11

So take some time with this and really work on it and have this is something that you can you can really be proud of.

18:18

Okay, now, let's talk about some very, very key items. So I know this is already a long video but if you just stick with me for a little bit more because this is really important, is in order for it to make it easier for us to grade. Then there are some markdown letters that are required.

18:42

And assignments without these markdown headers that are going to receive a grade of zero will give you the opportunity to resubmit one time and there will be an automatic rate reduction of 20 points on your second submission. So this is just key for us to make the grading a little bit easier and fair. Because as you can imagine, these projects give you a lot of leeway. So we just need to be able to have something in here just to make it a little bit easier for us. Okay, so for the respective section. In the machine learning checklist. The bullet points from the overview section of the instructions should be clearly separated with the use of heading to mark down headers and I'm going to show you that in just a moment. What I'm talking about.

19:34

If we go back up here, so the bind to the system the shortlist promising models the prepare the data those big sections, we won't list them with header, two headers marked down headers.

19:55

And for each individual rubric item in the required elements section. We want you to mark them with heading three mark down headers. So if we go back up to the rubric, for instance, if we look at maybe prepare the data. So you would have feature selection, maybe as a header three, you would have created data pipeline as a header three, fill in missing values as a header three. And again, I'm going to show you that in just a moment. So the broader overview. sections need to be header to mark down headers and the individual rubric items within that need to be three marked down headers.

20:50

And as I say here, these items need to be in there and we will not grade an element if it's not included with the header.

21:03

Meaning that even if you had it in the project, if you did not put that up under a heading three markdown header, then we're not going to grade that rubric guide them.

21:15

So just make sure that you include it I'm going to show you what that looks like in just a moment. Here is some documentation for how to use markdown comments.

21:24

And I'm going to jump over to just this example notebook. So notice that this is a markdown What do you call that code block?

21:38

You can change this of course to code or Markdown and just markdown right now. Notice that double click into this and if you put a single pound mark here, then this is going to be your title.

21:57

And then section two or header two is going to have to here and the same thing if we double click down here. So notice that overview, that is a level tutor.

22:13

While the business problem is a level three header, framing the problem is a level three header.

22:22

And if we go back to our rubric, go all the way up.

22:28

Hopefully I'm not making you dizzy. So this is the overview.

22:33

So notice the overview as a level two header and then what is the business problem?

22:42

And then we have how should we frame this problem?

22:46

What is the specific machine learning task? So we have this level three headers, business problem, level three header framing the problem level three header machine learning task. So you're going to need to make sure that you use markdown as you're working on your project.

23:07

Okay, when you get down to a section where you need, not just down comments, but you need actual code. Then of course, like for instance up under the get the data section.

23:21

Sorry, we're gonna get the data. This is a level two header. And remember about the communications that we're communicating with people that may not have any knowledge over the machine learning process. So we're asking you just kind of this as you're going like you're talking to someone that doesn't know about machine learning, and you're walking them through the process.

23:47

Within get the data, there's an import the data rubric item. So notice here that I have a level three for import the data, and then I have the actual code block. So notice this is a marked down block. And, of course, this is a code block. So you're going to actually enter your actual code, of course in a code block makes sense. Now make sure that you include three good comments in your code that's going to be something that you're graded on, that you're explaining what you're doing and why you're doing it. And this will help either your future self if you come back and work on the project, or it'll help your coworker if they ever jump in here and need to adjust something or or or start working on something to the next rubric item. And you make sure that again, that is listed up under a level three header. You're walking your audience through what's going on here. And of course, this is just made up language here.

Then you have code, you have markdown comments, then code, more code, more markdown comments code, and you're working your way through the entire project.

25:05

Now there's no order that you necessarily have to have everything in so as you're reading through this, it doesn't necessarily have to be in all this order. Now it would make sense that your review needs to come first. Then you need to actually get the data before you explore it. But for instance as you're preparing the data, it doesn't have to fall in this order. You're welcome to change the order. Just make sure that you hold three headers so that when we get to that section, that we understand what you're working on.

25:47

Okay, so how could that makes sense? And again, as you're working through the project, just make sure you're communicating through comments through markdown comments.

25:57

That's simple enough that your non technical audience will be able to understand what you're doing, and you're walking them through this project. But suppose that your technical audience so your director, in this case that we tell you does know about machine learning that you're you're walking them through the project as well.

26:21

Okay, let's carry on here. Is this is already a long video to make sure that you're using your your markdown headers correctly.

26:31

You're going to have your final project name, final underscore project. We're going to make sure that runs without errors. So again, you're going to check that here through this final project practice submission. This is unlimited. And really this is just checking to make sure that you don't have any errors in the file.

26:56

Any code blocks that generate output just like we talked about, needs to be fully executed. So the output is shown within the notebook before you submit it to us. We can't see your output, then we won't be able to grade it. And again, that really only applies for the stuff that we need to see the output on so we have the red asterisk next to those. You do need to have a PDF of your project that you're going to submit through this final project PDF submission.

27:30

You get a one time submission for this one time submission for your actual notebook whenever you're ready for it.

27:39

And in order to create a PDF, I'm not sure if everyone knows this, but there's a couple things you can do on a Windows if you hit Ctrl p.

27:54

Then this print screen comes up and you can just save it as a PDF here and hit Save. And that's probably the easiest way to save your notebook as a PDF. You can of course come over here. Go down to print, same print screen comes up. Okay, I'm going through here. Any co Grade Submission with an error is going to have an automatic deduction of 20 points. So that's the reason why we're giving you this practice submission just to make sure there's no errors. If you do have an error is an automatic 20 point deduction.

28:35

Let's see what else yes take time to look through this basically is just additional points can be taken off. If your notebook is not organized that there are spelling errors, grammar area errors.

28:49

Just make sure you read through this out one key is that late submissions will not be accepted unless there's an approved incomplete request that was requested improve before the term ends. Now again, you can look at your program handbook as to what's involved with incomplete request.

29:10

But by the end of the term at 11:59pm Eastern time, to not your local time but Eastern Time.

29:19

pro grade locks down you won't be able to submit any more. And unfortunately, you can't email me and say I'm sorry I missed the deadline. Unfortunately that will be not approved. but just make sure that you're submitting this project earlier. we highly recommend that you don't try to submit this last time you might have computer problems that you submit this ahead of time. As technical issues will not be accepted as a valid reason for a late submission.

29:52

Okay, plagiarism. Again, everything must be your own. We are running your project through.

Okay, plagiarism. Again, everything must be your own. We are running your project through turn it in turn it in and compares your work to other students in this class in previous classes. And it also looks at work that's out there on the Kaggle GitHub, it will pull that as well.

1:26

Alright, so your next steps and again, sorry about this really long video but just want to make sure that we just cover all of these items, is you upload your project to the practice submission to make sure there's no errors.

1:40

Then once you're sure you're ready to submit, because you only get a one time submission on this piece is you submit it to the final project notebook submission link your notebook and then you send you submit your PDF of your project to this PDF submission link. And we will only grade your submission if we get both the final project and the notebooks submission. and they match. And of course you need to clear markdown headers for rubric items, and that we can see your output in your code before you upload at to, co grade.

2:25

All right. So I noticed a lot of instructions, please take time to read through these instructions if you have questions on them, ask.