Dylan is going to be talking about something entirely different, which is the use of AI to predict whether or not a play was written by Shakespeare. So take it away, Diller.

Thanks for letting me share my screen. So everyone can see this I hope me present Yes, we can see it. So for my project is called to be Shakespeare or not to be Shakespeare bringing AI back in time. And my project I took data from Shakespearean and non Shakespearean texts, and I train the AI on them. And then I chose the AI that best analyze the data and asked it whether it thought a disputed Shakespearean play was actually written by Shakespeare himself.

So I'm motivation for the problem. I chose this project because I want to learn more about AI. And I thought it'd be interesting to apply it to something like literature. And Shakespeare is one of the most famous writers, so I decided to combine his plays and AI to create a project. I chose Shakespeare's all's well, that ends well as my text because it has a history of having its authorship disputed, and Shakespeare's authorship has been disputed all the way back to 19th century, in part because he wrote so many plays in such a short amount of time and at such a young age. And authors such as Christopher Marlowe have been suggested by some as a true author of some of Shakespeare's plays. And specifically, all's well that ends well, has been thought by some scholars to have been written by Shakespeare and another author, Thomas Milton, and they thought that some specific words resembled Middleton's writing, and not Shakespeare's, which led them to think that Shakespeare had collaborators on the play, and is successful, this project would have implications for other projects, especially authorship verification, which is basically telling the true author of a certain writing, and because of its implications, on other studies, it seemed like this project could be beneficial. So what are what are my steps in this project? The first thing I did was I found a Shakespearean play that was in dispute, which was all's well that ends well. The next thing I did was I collected the text of 16, Shakespearean, and eight non Shakespearean plays from the same era and genre as all's well, that ends well, and in this case, that is 16th and 17th century comedies. Next thing I did is I divided each of the plays by x in order to create enough data points for the AI to test. And after trying to use Python to analyze the text, I realized I didn't have the skill set yet to use Python. So I instead found a workaround, which is the website word count, or dotnet. And I used it to find the characteristics of each act of each play, and calculated the three features for my project, which were average word length, average sentence length, and word frequency. Then I build a table, I built a table with rows for each act of the place, excluding all's well, that ends well, because that's what I wanted to test. And then the columns of the three features identified previously, that I talked about a little bit. And then I then added a final column that indicated if each act was Shakespearean, or non Shakespearean. And after that, I downloaded the table into a CSV file format, and then went to navigator dot pix eda.ai. To use the publicly available AI algorithms. I imported my table into the website and the algorithms that were appropriate for my project are classifier algorithms. And that kind of made sense for my project. Because I had kind of a yes or no binary question which I was trying to answer. So I ran my data against the three classifier algorithms.

And they all use the random selection of 80% of the data to predict the outcome of the remaining 20% of the data. And the algorithm that work the best at about 95% accuracy was called random forest classifier. So I finally applied random forest classifier or RFC, to the features of AWS with Enzo and it predicted whether it was or was not written by Shakespeare, and I'll share my results a little bit later. So for my data, I collected my data set by downloading Shakespearean and non Shakespearean plays from a variety of websites. Then I broke each play into acts to increase the data size and import of the text in a word counter dotnet, which I talked about a little bit earlier, but it's basically a free tool that analyzes text and summarizes various data points, like you could find the average number of words in a sentence on that website, for example. And then I then used Google Sheets to create a table of the data. And I imported that table into the navigator for AI analysis. And this picture down below is part of my data set of Shakespeare and a non Shakespearean plays that are used to train the AI and shown above are shown below actually are two of the Shakespearean plays I used as well as two non Shakespearean plays, broken into acts to increase the number of data points and when I tested these data The points, I didn't include the names of the players because I wanted the AI to focus solely on the data instead of the title.

So, um,

when I inserted, so for my AI Sorry about that. When I inserted my Google Sheet into the navigator, I trained the AI on whether Shakespeare wrote the play or not. And there were three AI algorithms at the website suggested to me which were random forest classifier, pay neighbors classifier, and MLP classifier. And when they were automatically trained, the algorithm random forest classifier, or our seed had the highest intelligence based off my data set. So I then tested the AI to see if it thought that the disputed play in question which was also the Enzo, was actually written by Shakespeare by using random forest classifier. And for software and apps, I didn't exactly build any apps or Python programs. But I did use existing AI algorithms from a project. So here are a couple of images from my results. The picture up top is the confusion matrix for random forest classifier, and it shows the number of times the algorithm incorrectly, incorrectly predicted the author of the 24 data points that they were testing. And when the nose align up, it basically means that the AI predicted the true value. And when the yeses lined up the same thing. So this is the one outlier, that the random forest classifier got wrong in testing. And then this picture to the right, basically, it's just the chart that shows the three different algorithms that the AI decided where the best fits my data set. And you can see the top line is random forest classifier, and it had the highest accuracy of about 95%. And then the picture down below is the feature importance for a random forest classifier. And it shows how much the algorithm weighted each of the three features and its decision trees, and I'll talk about decision trees in a little bit. So average sentence length was the most important when I decided about it, if it was written by Shakespeare or not. So this is, um, pictures called decision tree and random forest classifier use 10. Of these, this is just one of the 10

that was used to predict the authorship of certain plays in my case, and the top line in each box is called a condition. And it's a feature that is applied to a sample data point to see if the condition is met it kind of like a yes or no question.

And this moves the data points to the next box, which applies a similar procedure with a different condition to the same data point. And others. Other decision trees, the number of levels vary. But in this example, you can see there are three levels. The bottom level doesn't give a condition to be met, but instead provides the predicted outcome for the data points. And random forest plots are very used this decision tree along with nine other decision trees to come to a collective prediction based on a majority vote among the decision trees. So they use this one and nine others to make a vote. And but what I basically learned was that with 95% accuracy, random forest classifier predicted that all's well that ends well was written by Shakespeare. And this confirms my hypothesis, because I also, I also thought that although the intro was written by Shakespeare, and this also shows the power of AI, because most literary scholars will look at this question and take an approach focused on the little details and the plays. But the AI takes a step back and finds patterns in the data to make its judgments, whatever that was kind of interesting. And in the future, one question I thought of well, working on this project is why some algorithms work better than others for my project. And I know that the random forest classifier used a decision tree model to determine the prediction. But I kind of wondered why it worked better than the other decision tree models out there. So if I were to pursue this topic farther, I might try to build my own AI algorithm, while at the same time learning which characteristics might make my algorithm the best for my specific topic. Another way I could expand on this project would you be to apply it to other writing samples, other than plays, and that way, I can include more features like paragraph length, for example, that doesn't didn't really apply to plays in this in this case, so I didn't use them in this case. But if I looked at other things like short stories, I may be able to include other features, which would help strengthen my case. And I could also learn Python and use it in the future to increase confidence in my data set, and expand the number of features. And thanks for watching, if there any questions.

Thank you very much, Dylan, I think one of the things I like most about your project is that it's a very nice, you know, application of AI that, you know, sometimes when we think AI, we think very scientific, but it turns out, it has a lot of application and literature as well. So now very cool. If anybody has any questions, please go ahead and put them on the chat or on the q&a.
Let's see. So I had one question, which was, you know, are there other plays at the same time period that may not be comedies or you know, may not be Shakespeare that you think, you know, you could apply this

So I did look at other areas of literature in the search specific plays like comedies and tragedies. But I ended up settling on a certain section. And all all's well that ends well is. I think it's a comedy. So I believe it is yes, if I remember. And the other sections didn't

have enough data points, because again, this these are plays that were written a very long time ago. And although Shakespeare is a very famous author, and he probably has the most of certain plays from that time period, when I had to find a non Shakespearean data set, there weren't enough plays that had confidence that I could reliably trust to have correct data. So I ended up focusing on the comedy section because the other sections didn't have enough data for me to find

make sense. And that's a common problem. Thank you so much, Dylan. Very nicely done. Thank you.