

CA337 Assignment 5 Report

Introduction

For my project I decided firstly to improve my ensemble using methods I learned throughout this module, I will adjust the pre-processor as well as the models used to hopefully get the most accurate ensemble on the difficult test set that I created. I next thought about trying to improve the flask application, I would do this firstly by allowing for the application to have a selection of models and then I would try and improve the overall layout of the web application using HTML, CSS and Bootstrap that I've learned throughout my time in college.

Ensemble Method Enhancement

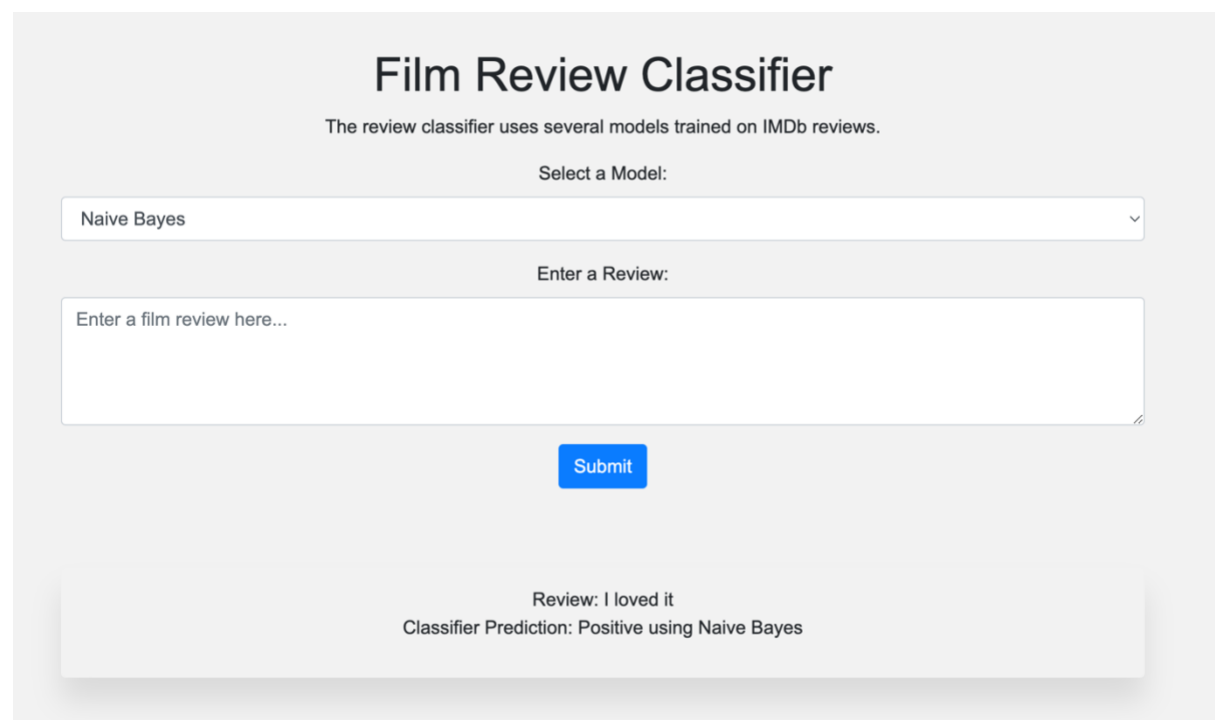
So as mentioned above I began by trying to develop and improve the accuracy of the ensemble that I created in assignment 4. I did this using techniques I learned throughout this entire module. So my first step was to create an improved version of the vectorizer that was used previously, my first though was to create a pre-processor similar to the one that I had developed in assignment 3, this gave me a significant improvement immediately, however, I knew there was more improvements to be made, next I used the sci-kit learns select k-best features method¹, this was another idea I took from my assignment 3, this once again was very beneficial in improving the accuracy of the ensemble. The final major change I made to the initial ensemble I had developed was to change the actual models that were in use. When looking at the accuracy from each of the models within the ensemble, the one that was performing the worst and ultimately dragging down the overall performance of the ensemble was the decision trees model, this had a very poor accuracy on both the validation set and the test set, so I thought it would be best to just completely remove the model in question and replace it with a better performing one. The model I ultimately decided to change it to was a logistic regression model, this was because it was one we had worked with earlier in the module and it had previously performed very well in the various tests I had run. This change is ultimately where I saw the biggest improvement in my ensemble accuracy. I finally adjusted some of stopwords to get the most accurate result. Overall, I was able to improve the model from 82% up to 89% on the validation set and from 60% up to 90% on the difficult test set I created. I was very happy with this as a 90% accuracy on my test set is even better than the accuracy I was able to get in assignment 3.

Flask Application Model Selection

The next idea I wanted to implement as part of this assignment was to allow for a model selection on the flask app, this initially was a tough task as I wasn't really sure how flask worked but I decided I would give it a go anyway. I first began by building the 3 models, I used the same code as I did for my ensemble method just without the ensemble at the end. This would give me the most accurate models I could create so the application was effective as possible at prediction. Once this was done, I modified the code we got in assignment 4. I created a dictionary and for loop to read in the models into the flask app. The next part was to adjust the code in the flask app to allow for the model selection to take place. I firstly had to import the models and the pre-processor from the pickle file. I next created a dictionary containing the model names which I would use for printing later. The main update I had to do was to update the predict function, I allowed for model selection and I also adjusted it so the model id was linked to the model name. Then when the prediction was printed it showed the classification as well as the model that classified it. Finally, the changes to the HTML files at this stage was limited, however, I created a form that allowed for model selection, this form consisted of a dropdown menu of each of the selected models, and I also textarea element so that I could make my reviews be any length. This changes made my website much more usable so that I could test the predictive ability of the various models.

Web Design

The final idea that I wanted to implement was improving the layout and design of the website, I wanted to use this as an opportunity to develop my web design skills as it isn't a technique I have had the opportunity to use that much throughout my time in college. I began by using Bootstrap² to create containers for each of the different elements on the website this would allow me to easier control the layout and position of the text and menus. I also added a proper button asset from bootstrap for both the submit button and the initial "Go to Prediction" button. This immediately made the website look a lot nicer to view and also allowed me to easier manoeuvre and change the design of the button. As I said I created the containers so I could control the layout and so what I did was centre all the content, this made the website a lot more appealing to view. I changed the size and style of the text on the website, this made the text stand out a lot more making it easier to read. I changed the background colour from white to a light grey, this made the website look a bit more modern and less harsh and contrasted well with the blue colour of the button. The final addition I made to the website was to add a drop shadow beneath the container housing the prediction result, this made the result stand out a lot more making it more noticeable.



The screenshot shows a web application titled "Film Review Classifier". Below the title is a subtitle: "The review classifier uses several models trained on IMDb reviews." There is a dropdown menu labeled "Select a Model:" with "Naive Bayes" selected. Below this is a text input field labeled "Enter a Review:" with the placeholder text "Enter a film review here...". A blue "Submit" button is positioned below the input field. At the bottom, a light grey box with a drop shadow displays the results: "Review: I loved it" and "Classifier Prediction: Positive using Naive Bayes".

Retrospective

Overall, I really enjoyed this module, it was really great opportunity for me to be able to develop my knowledge and skills in relation to machine learning. The ability to get to properly learn how these various models work as well as how to develop a proper pre-processor for our model. Also this was my first experience with sentiment analysis so getting to understand this and how it works was incredibly useful.

The opportunity also to develop my web design skills I really enjoyed. I haven't had the opportunity to use my web design skills much in college so far, so this experience to both use and develop these I really enjoyed.

References

1. "Sklearn.feature_selection.SelectKBest — Scikit-Learn 0.23.0 Documentation." *Scikit-Learn.org*, scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html.
2. Otto, Mark. "Bootstrap." *Getbootstrap.com*, 2022, getbootstrap.com/.