Roman W. Dobczansky

11/9/2022

Final SELF Assessment

Self -Assessment

Presents a cohesive written analysis that describes the role(s) they played over the course of the project and their contribution to the project in that role. Presents a cohesive written summary of how they contributed to each of the roles they did not take on via team discussions, peer reviews, or other means. Additionally, the analysis should describe their greatest personal challenge over the course of the project, and how they overcame that challenge.

From the outset I tasked myself with the responsibility of making sure the Neural network and machine learning aspect of the project worked well and was replicable in all four aspects of the segments. During team discussions I kept members on task to make sure they thought about the larger picture by not confusing the NN and making sure the machine learning model took priority in the overall progression on the project as that is what the pyramid of priorities was. For example, when members of my team could not attend office hours, I was there asking questions about the SQL Alchemy, Postgress, Data base portion and how to make that file which was local to Rasnas computer accessible to out Github and most importantly how to import that database into the machine learning model. The biggest challenge as mentioned before was making sure everyone stayed on task and did not get bogged down in trivial pursuits of present tense of otherwise as regards the readme file, the uploaded and importing of the data set, and making sure my final PRESENTATION was Acceptable to the group which it was and actually graded because I deserve credit for that video I submitted and the notes (at the bottom on this document) that it was not graded yet and am still awaiting my final grade and e certificate. Thank you!

Team Assessment

Presents a cohesive written analysis that describes their teamwork, including all of the following: ✓ Their communication protocol, including any challenges, how they were resolved, and what they would do differently next time ✓ Their strengths as a team, including tips and tricks they would want to share with a new cohort kicking off the project

Our teamwork was very on point and hands on. We chose to communicate and over communicate at times leaving nothing to chance. We met 3-4 times outside of class every week on zoom and attended every meeting for as long as one could to get to the bottom of whatever problem we were working on. Rasna went so far as to send me a co host zoom call from an airport that I could practice and record my fourth segment deliverable. I was proud of my teammates Aggie and Vanessa who was so detail oriented and even made it to meetings when physically sick. I too am proud of my commitment to team excellence and group cohesion and rigor and willingness to learn together and work towards the common goal by implementing systems that worked.

Summary of Project

Presents a cohesive, three- to four-sentence summary of the project that could be used on a LinkedIn profile, in an interview or cover letter, or as an elevator pitch, including all of the following: ✓ Topic addressed ✓ Machine module used ✓ Results of the analysis

Comparing the quality of red vs white wine, we implement a sophisticated supervised Machine Learning Neural Network to ensure quality and health to the consumer's wellbeing. Culminating in the latest data science technologies our group employs tensor flow machine learning technology to better employ you. The results are in and you and the white wine are winners. Use our skills and expertise and you too will be winners.

Roman W. Dobczansky

11/4/2022

Presentation notes and talking notes to go along with prerecorded video

Keep to 3 minutes, 15 minutes total for the group

- Use of TensorFlow model

  "TensorFlow makes it easy for beginners and experts to create machine learning models for desktop, mobile, web, and cloud computing"

- Accuracy above 80%

  "Any score above 70% is considered accurate therefore we find our score to be statistically satisfied"

- Over 100 epochs are trained and stop after 69 iterations.

  "We kept our data set light and the machine learning model was able to determine that after only 69 iterations of the 100 epochs that the accuracy level desired (70%) was achieved. Thus, we did not 'overfit' the data in our model. Overfitting happens when your model starts to memorize values from the training data instead of learning from them. This means that it does not have intuition about its findings as the programmer behind it may have."

- The model evaluation for test data loss is -4.2

  "Loss is nothing but a prediction error of Neural Net, and out -4.2 loss is relatively small and therefore insignificant."

"*Difference between the actual output and the predicted output from the model for the single training example* "

"The word 'Loss' states the penalty for failing to achieve the expected output. If the deviation in the predicted value than the expected value by our model is large, then the loss function gives the higher number as output, and if the deviation is small & much closer to the expected value, it outputs a smaller number."

- The neural networks improve accuracy over time from 19 % to 80%.

  - "Improvement of this kind is good; it can also be made better with Feature engineering which is a process of transforming raw data into features that better represent the underlying problem that one is trying to solve. Some examples of this can include Converting a Date-Time variable to extract just the day of the week, the month of the year, etc. To add the variable of time to wine quality.

  - Creating bins or buckets for a variable. (e.g., for a height variable, can have 100–149ml of wine, 150–199ml of wine, 200–249ml of wine, etc.)

  - Combining multiple features and/or values to create a new one. For example, one of the most accurate models for the titanic challenge engineered a new variable called "Is_Red_or_White" which was True if the wine is red or white and false otherwise.

- Limitations: Hard to understand what caused it to arrive at this prediction. Not 100%

"Recall that our goal is to create a model that will make accurate predictions on data but new data or changes to data can affect the accuracy, outcome, and overall results. However, the basic Neural network and machine learning model should be able to stay intact if not too many variables are added. But in case we wanted to introduce feature engineering we could also include AWS binning.

Results

Are there vast chemical compound differences between the red and white wine datasets being compared?

- There is significantly more sulfur dioxide in white wine than red wine
- White wine has more residual sugar but less alcohol content

Which type of wine presents less alcohol, residual sugar, acidity, and sulphate?

- Red wine is more acidic than white wine
- White wine has less sulphate than red wine.
- Both white and red wine have about the same **average** alcohol level

Which qualities make up the healthiest wine?

- Red wine is less healthy than white wine by our measure. To measure healthiness, we use the level of sulphates a wine has

The winner is white wine