Cerbo: A Student-Friendly Library for Artificial Intelligence and Machine Learning

Anaiy Somalwar[1]**, Karthik Bhargav**[1]**, Andy Phung**[1]**, Aurko Routh**[1]**, Ayush Karupakula**[1]**, Felix Liu**[1]**, Keshav Shah**[1]**, Nathan Zhao**[1]**, Navein Suresh**[1]**, Sauman Das**[1]**, Shrey Gupta**[1]**, and Siddharth Sharma**[1]

1 StartOnAI

Summary

The Cerbo Machine Learning Repository is an open-source library that aims to increase the level of approachability to Artificial Intelligence (AI), a topic that can be intimidating for many students. Wrapping several of the most popular state-of-the-art, research-oriented libraries ((Abadi et al., 2016), (Pedregosa et al., 2011)), Cerbo offers users the ultimate level of abstraction for data preprocessing, data visualization, and machine learning and deep learning model creation in Python. Thus, Cerbo makes coding easier and more beginner-friendly and helps educators drive student engagement and teach students how to apply theory in these in-demand topics. The Cerbo Machine Learning Repository can be easily installed through `pip`, and multiple in-depth code tutorials can be found on the project's Github.

Statement of Need

While education on how to write code related to Artificial Intelligence and machine learning has become significantly more accessible in the past few years with many free resources emerging, many students are hesitant to delve into these fields early on as they are often portrayed as being extremely technical or complex and requiring significant computer science or mathematical prerequisite knowledge. Furthermore, many coding libraries used for teaching today were primarily developed for research purposes and may not be ideal for an introductory educational experience. Especially at the high school and undergraduate levels, educators have an increasing need for Artificial Intelligence and Machine Learning related coding libraries that maximize the level of abstraction, and with it, the level of approachability, to drive student engagement with these in-demand topics.

The Cerbo Machine Learning repository offers users the ultimate level of abstraction to improve student approachability and engagement in Artificial Intelligence and machine learning in a high-level programming language that many students are familiar with, Python. Students and educators can use Cerbo as a tool to learn the fundamental coding steps associated with machine learning and deep learning and take advantage of the intuitive, walkthrough style tutorials on our Github repository. Already in use by over 5,000 students and educators across the United States and globally, the Cerbo Machine Learning repository has an active and broad student and developer community that is able to answer questions and make timely improvements based on educational and industry values and needs.

As high school and undergraduate students who regularly teach Artificial Intelligence in school clubs and through our educational organization StartOnAI, we have seen firsthand the increased engagement a simple package such as Cerbo can bring through teaching hundreds of students. We have also seen firsthand how simple it is for educators to adopt Cerbo from our interaction with various third-party school clubs. Using Cerbo in classrooms and clubs is also a great opportunity for students to learn the basics about abstraction and wrapper classes, both of which are important topics in general programming.

Functionality and Usage

A typical use case of the Cerbo Machine Learning Repository involves the following simple steps: installation, dataset loading and visualization, and model creation, training, and testing. Cerbo often allows for these tasks to be done in less than 10 total lines of code, which allows students to be excited about the simplicity and amazing applications of Artificial Intelligence and Machine Learning coding! Let's take a common example for a use case of Cerbo that we often teach: trying to predict whether a patient has diabetes based on 5 other factors from the Pima Indians Diabetes Database (Dua & Graff, 2017) from the National Institute of Diabetes and Digestive and Kidney Diseases. After installing Cerbo through either `!pip install cerbo` or `!python -m pip install cerbo`, it takes only 5 lines of executable code to train and test a simple baseline decision tree model with over 75 percent accuracy, as seen below. Adopting Cerbo for classrooms and clubs is especially simple as we have posted multiple comprehensive Jupyter notebooks on our Github that contain examples and explanations of every single function in the machine learning and deep learning libraries as well as their applications on several common datasets, making it extremely easy to debug and verify assignments.

```
import cerbo preprocessing as cp
import cerbo.ML and cml
loc="https://bit.ly/3Cf1UEd"
data, col_names = cp.load_custom_data(loc, "Outcome", num_features=5, id=False)
dt = cml.DecisionTree(task="c", data=data)
```

Our Story and Teaching Experience

We are high school and undergraduate students from StartOnAI, an educational organization with the mission of "Making Artificial Intelligence education approachable, affordable, and accessible to all." We feel extremely fortunate to have been exposed to Artificial Intelligence and finding our passion during our high school and undergraduate careers; however, none of us were introduced to Artificial Intelligence through our school curriculum. So, we created school clubs and the StartOnAI education to share our passion with other students. While we have published a theoretical book guide to Artificial Intelligence and created countless videos and tutorials at StartOnAI, we realized that many of our club members never applied the concepts they learned due to their uncertainty of writing code. So, we began the Cerbo project to make writing Artificial Intelligence related code approachable and beginner-friendly through heavy abstraction. Less than a year after our first release, Cerbo has over 5,000 students and educators using it regularly, and we have personally used it to teach over 100 students in three school clubs. We hope to continue making Artificial Intelligence coding easier and more approachable and engaging in the future.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., & others. (2016). Tensorflow: A system for large-scale machine learning. *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 265–283.

Dua, D., & Graff, C. (2017). *UCI machine learning repository.* University of California, Irvine, School of Information; Computer Sciences. http://archive.ics.uci.edu/ml

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & others. (2011). Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, *12*, 2825–2830.