# Using Trees to See a Forest: Leveraging Machine Learning to Classify Student Thinking

Darryl Chamberlain Jr. Patrick McGuinness
Emily Faulconer Beverly Wood
Embry-Riddle Aeronautical University -- Worldwide

*Keywords:* Decision Trees, Qualitative Coding, Machine Learning

## Introduction

ChatGPT has caught the world's attention and has researchers considering the ways artificial intelligence can be integrated into human endeavors (Kasneci et al., 2023). This methodological poster will present a systematic method to apply a particular machine learning classification model, Decision Tree (Song & Lu, 2015), to perform and extend the scale of qualitative analysis in mathematics education research.

## Implementation of a Decision Tree

There are five general phases to the creation of a decision tree: data preparation, feature analysis, test label creation, decision tree coding, and parameter tuning.

*1. Data Preparation:* While the original data can take any form (verbal, student work, hand gestures, etc), it will need to be unitized in some way. The size of each unit of data will be affected by the theoretical framework employed and the attributes to be identified from the unit.

*2. Feature Analysis:* Each unit of data will then be coded for important attributes that are referred to as *features*. Features can be categorical or numeric, and do not need to be of just one type for a single decision tree. For example, features tied to a student's free response to a task could include [a] Presence of a response *(yes/no)*, [b] Presence/Absence of key words/phrases *(binary or number of occurrences)*, or [c] Use of a particular representation *(graphical, algebraic, tabular, etc)*. Identifying features is a critical human task in creating a decision tree, much like qualitatively coding according to a theoretical framework.

*3. Test Labels Creation:* We need to provide labeled data the decision tree algorithm can learn from. This is the second critical human researcher task that again mirrors a typical qualitative coding task where labels are determined by the theoretical framework employed.

*4. Decision Tree Coding:* Open source libraires such as Python's pandas library (Reschke, et al., 2020) contain the majority of the technical aspects of employing a machine learning model. After preparing the data, completing the feature analysis, and labeling the data, creating the decision tree is as easy as calling RandomForestClassifier().fit.

*5. Parameter Tuning:* Numerous parameters can be defined by the researcher to improve the accuracy of predictions as well as the amount of data that is used to train and test the data.

## Discussion

We believe that incorporating machine learning can significantly benefit the qualitative coding process and is feasible for many mathematics education researchers. One major benefit to this implementation is the scale and speed at which data can be analyzed. Another major benefit was seen in our codebook. By carefully considering what features might be present and what labels they may correlate to, we have improved the replicability of our qualitative coding based on our extensive codebook. Finally, classification algorithms would provide an excellent engine for automating the analysis of student work based on qualitative coding of student responses. This would allow researchers to bridge the gap and provide practical uses of their esoteric work.

## References

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Gunnemann, S., Hullermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and individual differences, 103, 102274.

Raschka, S., Patterson, J., & Nolet, C. (2020). Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence. Information, 11(4), 193.

Song, Y. Y. & Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. Shanghai archives of psychiatry, 27(2), 130.