Detailed Approach for Predicting L8 and L9 Codes from Learning Messages

A comprehensive approach for training a model that predicts L8 and L9 codes based on *learning messages.* Below is a detailed breakdown of each step:

1. Data Preprocessing:
  ● We will start by cleaning the text data extracted from learning messages. This involves removing stopwords, punctuation, and special characters to ensure the text is ready for analysis.
  ● Next, we will tokenize the text into words or subwords, allowing us to represent the data in a format suitable for machine learning algorithms.

2. Feature Engineering:
  ● To convert the preprocessed text data into numerical vectors, we will explore various techniques such as TF-IDF, Word Embeddings (e.g., Word2Vec, GloVe), or Transformer-based embeddings (e.g., BERT, GPT).
  ● Pre-trained embeddings will be considered if available, as they can capture semantic relationships effectively.

3. Model Training for L8 Prediction:
  ● For predicting L8 codes, we will employ multi-class classification algorithms such as Logistic Regression, Random Forest, or Support Vector Machines (SVM). These algorithms are suitable for predicting categorical variables like L8 codes.
  ● The model will be trained using the labeled data, where learning messages are mapped to their corresponding L8 codes. We will evaluate the model's performance using metrics like accuracy, precision, recall, and F1-score to ensure its effectiveness in predicting L8 codes accurately.

4. Fine-Tuning for L9 Prediction:
  ● After successfully predicting L8 codes, we will use these predictions as additional features along with the text vectors to train a new model specifically for predicting L9 codes.
  ● This model will also utilize multi-class classification algorithms, fine-tuned on the data with L8 predictions as features. The goal is to accurately classify learning messages into their respective L9 codes within the predicted L8 classes.

5. Model Evaluation:
  ● Both the L8 prediction model and the L9 prediction model will undergo thorough evaluation using appropriate metrics for multi-class classification tasks.
  ● We will fine-tune the model parameters as needed to improve performance, especially for predicting L9 codes accurately within the predicted L8 classes.

6.Combining Prediction Code:
   - Once satisfied with the model's performance, we will write a script to combine both models (L8 prediction model and L9 prediction model) to classify new learning messages into L8 and L9 codes respectively.
Please let us know if you have any questions or need further clarification on any aspect of our approach.