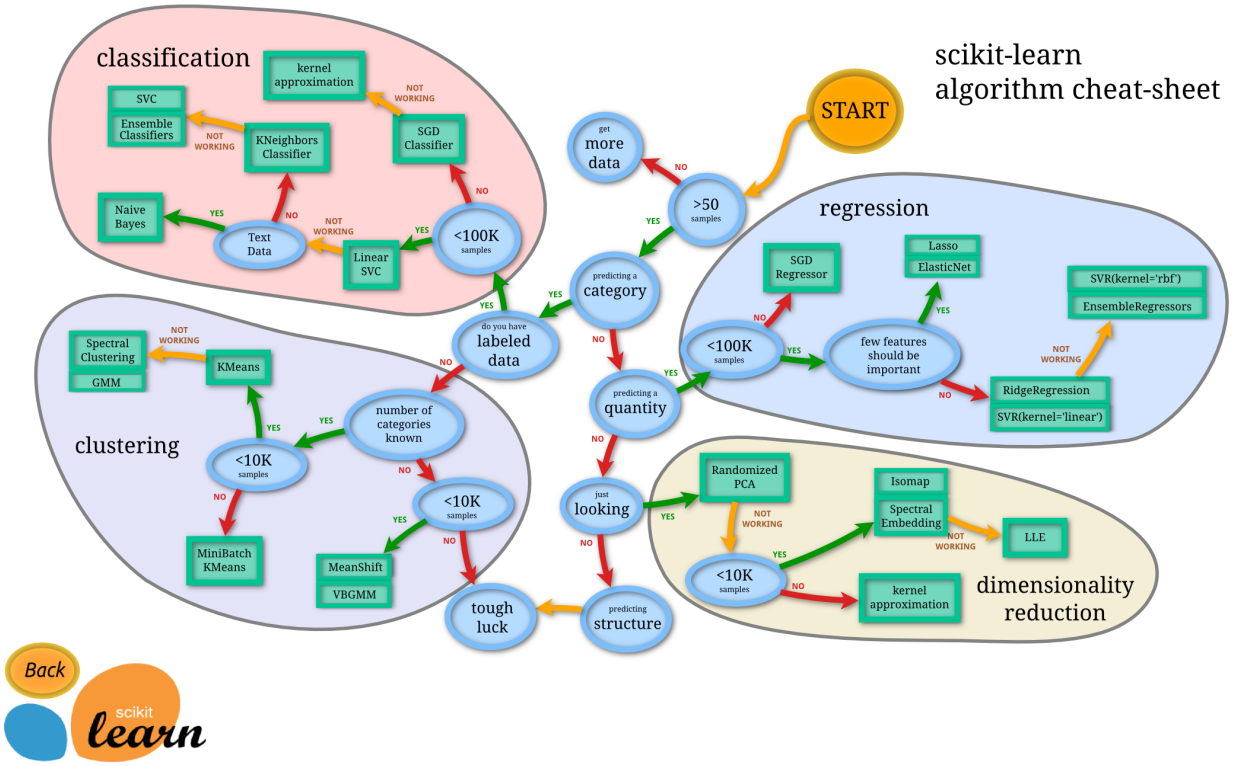
REVIEW

# Meets Specifications

Congratulations on passing the project! Please find my thoughts and suggestions below.

Good luck with the rest of the course! :)

Note: As for your note regarding the different applications for the various ML models, it really does depend on a lot of factors, like the dataset available, the latency requirement of the application etc. It is difficult to say this particular application will always favour a particular ML model. I'm attaching a rough map from the official sklearn docs that might aid a bit here.




## Exploring the Data

✓

**Student's implementation correctly calculates the following:**

- **Number of records**
- **Number of individuals with income >$50,000**
- **Number of individuals with income <=$50,000**
- **Percentage of individuals with income > $50,000**

All the values here are correct!

## Preparing the Data

Student correctly implements one-hot encoding for the feature and income data.

## Evaluating Model Performance

Student correctly calculates the benchmark score of the naive predictor for both accuracy and F1 scores.

The pros and cons or application for each model is provided with reasonable justification why each model was chosen to be explored.

Please list all the references you use while listing out your pros and cons.

Student successfully implements a pipeline in code that will train and predict on the supervised learning algorithm given.

Student correctly implements three supervised learning models and produces a performance visualization.

Now you can clearly see the difference in the training time and performance metrics of different models when varying quantities of training data is available.

## Improving Results

Justification is provided for which model appears to be the best to use given computational cost, model performance, and the characteristics of the data.

Very well reasoned out! And I agree with your conclusion, GBC indeed seems to be the better model among the 3 for the job.

Student is able to clearly and concisely describe how the optimal model works in layman's terms to someone who is not familiar with machine learning nor has a technical background.

Good job here! The language used is simple and the concept is explained well enough for a layperson to get a primary grasp on the concept.

Love the fact that you used a real world analogy to explain the model as well.

**The final model chosen is correctly tuned using grid search with at least one parameter using at least three settings. If the model does not need any parameter tuning it is explicitly stated with reasonable justification.**

Good job correctly implementing GridSearch to tune the hyper parameters of the model.

**Student reports the accuracy and F1 score of the optimized, unoptimized, models correctly in the table provided. Student compares the final model results to previous results obtained.**

Now you can see the difference - albeit minor - between the optimized and un-optimized versions of your model.

## Feature Importance

**Student ranks five features which they believe to be the most relevant for predicting an individual's' income. Discussion is provided for why these features were chosen.**

**Student correctly implements a supervised learning model that makes use of the `feature_importances_` attribute. Additionally, student discusses the differences or similarities between the features they considered relevant and the reported relevant features.**

**Student analyzes the final model's performance when only the top 5 features are used and compares this performance to the optimized model from Question 5.**

Good discussion here. Given that training time isn't a factor, it is almost always desirable to train on a full set of features.