Classification vs Regression

This is a classification problem because the target column is made up of categorical data. If it will have been a regression problem if the target column had been made up of real / continuous values.

Exploring the Data

- Total number of students: 395
- Number of students who passed: 265
- Number of students who failed: 130
- Number of features: 31
- Graduation rate of the class

Training and Evaluating Models

Decision Tree Classifier

Decision Trees Classifier is a model that builds up a tree of rules that it then uses in predicting new input The Advantages

- The decision tree can be easily visualized, which makes it easy to understand and interpret
- They require little data preparation
- The storage cost for generating the prediction tree is logarithmic relative to the quantity of data provided The Disadvantages
- Decision Trees can get overly complex performing well during training but not during prediction - a situation known as overfitting. There are ways to deal with overfitting.
- Decision Trees need to be rebuilt and can result in a completely new tree once there is a variation in the original data.

Support Vector Classifier

The advantages of support vector machines are:

- Effective in high dimensional spaces - They could still provide good prediction when the number of features is greater than the number of samples.
- Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.
- Versatile: different Kernel functions can be specified for the decision function. The disadvantages of support vector machines include:
- It could have poor performance if the number of features is much greater than the number of samples

- It does not provide probability estimates.

Gaussian Naive Bayes

These are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of independence between every pair of features. Naive Bayes classifiers have worked quite well in many real-world situations, famously document classification and spam filtering.

Advantages

- They require a small amount of training data to estimate the necessary parameters.
- Naive Bayes learners and classifiers can be extremely fast compared to more sophisticated methods.
- The decoupling of the class conditional feature distributions means that each distribution can be independently estimated as a one dimensional distribution.
- This in turn helps to alleviate problems stemming from the curse of dimensionality.

Disadvantages

- If there is no occurrence of a class in a dataset, then the probability of that class occurring is zero. Using this class to derive the posterior results in all elements multiplied by zero - Which is incorrect.

Results from Experiment

	Classifier	F1 score(test)	F1 score(train)	Size	Train time	predict time
0	DecisionTreeClassifier	0.644068	1.000000	100	0.001	0.000
1	DecisionTreeClassifier	0.596774	1.000000	200	0.001	0.000
2	DecisionTreeClassifier	0.611570	1.000000	300	0.003	0.000
3	SVC	0.797468	0.877193	100	0.002	0.001
4	SVC	0.802632	0.847352	200	0.005	0.001
5	SVC	0.757143	0.868597	300	0.008	0.002
6	GaussianNB	0.377778	0.453608	100	0.002	0.001
7	GaussianNB	0.560000	0.807143	200	0.004	0.001
8	GaussianNB	0.643478	0.824645	300	0.001	0.000

Based on my experiments carried out previously, I believe Decision Tree Classifiers is generally more appropriate based on the available data. Here are the reasons why I choose this model.

- It requires little time to make predictions
- It requires little time to train
- It requires little storage space - usually a logarithmic value of the amount of data specified

How Decision Trees work

Decision Tree Classifiers builds up a tree that can be used for classification of data. It does this by breaking down the data into smaller and smaller subsets while at the same time, an associated decision tree is built. The final output is a tree consisting of nodes and leafs. Each node corresponds to a decision that is to be made. A Node consists of children which are either other nodes or leaves. If the outcome of a decision is a node, then another decision has to be made. If the result of a decision is a leaf, then we have an answer.

Final F1-Score

My Model's final f1 score is 0.802

Disclaimer:
Some of the writings here were based off sklearn's documentation[http://scikit-learn.org/stable]. This is because I found their explanations to appropriately describe classification models.