Many of the proposals that I get for 271 projects are fatally flawed. They are too vague, and convey the sense that the student hasn’t really thought out what he/she is planning to do. Keep in mind that this is a document that counts toward your final grade in the course, and make sure you know what you will be doing before you propose to do it. Here are some pointers for the proposal.

1- Include an introduction that briefly states the problem. Make it objective and to the point, i.e. sets the context for what you are going to do. Please no “It is well known that machine learning can be used for document classification”. If this is all you have to say, you need to find a problem to solve first. A better introduction is “Document classification systems frequently represent documents as word histograms, but this ignores dependencies between words. In this project we seek an alternative representation that could overcome this problem”.

2- Include a brief review of what has been done in the area. Only things that are very relevant to what you are proposing, not a full-blown review of the literature in the field. For example “[XXX] suggested using histograms of word pairs, [YYY] suggested an approximation of the conditional distribution of the word given the last two sentences, [ZZZ] suggested to explore certain known independence relationships based on the concept of Markov sausages”.

3- Include a precise description of what you are going to do. One possibility is to suggest something new. “I propose to kernelize the Markov sausage method of [ZZZ]. In preliminary work, I have shown that their approximation can be manipulated so that it only depends on dot-products of training examples. This implies that it can be kernelized. I propose the new method of kernel Markov sausages. The goal for this project is to test whether this new method will produce substantial gains in document classification tasks. I will implement two kernels, one based on 98% lean functions [reference] and the other based in 95% lean exponents [reference]”. I will implement and test kernel Markov sausage-based document classification with these two kernels”. Keep in mind that the project has a very limited time window. Do not propose new stuff that is overly ambitions. My advice would be: unless this is something that you are working on for your research already, don’t do it.

4- An alternative is to either apply methods in the literature, or compare various things. For example. “In this project, I will investigate how the Markov sausage representation will fare on the problem of music classification. This has never been done, but the success in the text arena suggests that it should be more effective than the representations now used in the music literature, which are comparable to the histogram of word pairs of [XXX]. I will make some adjustments to adapt the
Markov sausages to music data [SAY EXACTLY WHAT THESE ARE, e.g. use new features, whatever] and compare to the state of the art methods in this area.” This sounds like a project doable in 2 months. Yet another alternative is to compare things. “There has recently been much work showing that the Markov sausage approach has great potential for document classification”. This approach has been used with a number of different features: [XXX] tried it on phonemes, [ZZZ] used full words, [YYY] applied it to features that summarize entire sentences. In this project, I will compare the performance of these different features. I will implement the Markov sausage method with each of them and compare performance. I will also implement a boosted Markov sausage method, which will automatically select the feature combinations that achieve best performance in the task at hand. As far as I know, this has not been done before”.

5- Include a detailed section on evaluation. How are you going to evaluate your work? What is the performance criterion for your task? Which data are you going to use? Where is this data coming from? How much will you have? Which variants of your method, of whatever methods you are proposing to work with, are you going to evaluate? What are the relevant parameters to test? Are there standard benchmarks in this area? Do you think that you can beat them, by either doing something new, or improving on an existing method? How do you define a good outcome for the project? What will we learn from it?

The goal of a proposal is to work out, in your head, what your project will be. Once the proposal is done, you should be able to immediately start writing code, to solve the problem you are studying. If you can’t do this, you do not have a proposal, but just a collection of thoughts. Unfortunately, I do want a proposal!

Nuno