1 By Noureddin

1. Please create your own section (copy and paste this LaTeXsource :-))
2. Add your thoughts/suggestions/comments
3. Add dates so we keep up to date with progress
4. Look at this example: [Noureddin: This is how you comment on something!]
5. For Larisa it's [Larisa: I love Japanese tea with rice!]
6. For Crina it's [Crina: I love Japanese rice with tea!]

2 By Larisa

1. We have analysed available descriptors from various sources (bottom-up approach), and there are too many of them. We now will try top-down approach. We will consider several datasets, from Dundee and publicly available ones, e.g. from ChEMBL and also several QSAR studies. We will try to annotate them with some descriptors that would convey what is important to record about those datasets and QSAR- specific studies. We then will come up with a set of descriptors that are suited for our task. We will check if what existing resources have the required descriptors, and if necessary we will define new ones (see the next point).
2. Crina suggested that we can develop new descriptors/ measures that fit better to support our task.
3. We will target to come up with 50 descriptors and we will evaluate them with experts through questionary and select 20 the most popular descriptors.
4. We will also put our questionary to a public domain and give an opportunity to everyone to comment on descriptors. Such an approach would ensure that we identify a set of useful descriptors.

3 Crina

My two suggestions during the meeting where:

- Find the characteristics of the data set (something which defines very well our data and makes the difference between our data and other data sets). For instance, if we say 10 features, this will not make a difference. But if we say 10 features, first 3 are real values, 4 and 5 are binary and 6-10 are real values between 0-1, this will be more specific.

- define the problem output (or scope): in our case will be binary classification and not regression?
- define how we are going to achieve this scope: this may involve more criteria in which case we will need to look at the performance of the methods we apply from various angles
- define a set of machine learning methods suitable for this scope
- define a set of measures or metrics which evaluate these machine learning methods for this scope. Some of the metrics may be standard, but we might need some other aspects and for this we will have to define ourself some measures

[Crina: Noureddin, put together what you have so far and what you find in the literature strictly for this type of problem and I can also look after that and complement with what I know.]