

Machine Learning for Software Engineers (Module SG5326)

ISS Assessment Project

You will work in teams of 4 to 6 persons to use and demonstrate the skills you have learned during the Coursera Machine Learning course. Your goal is to build a small application that utilises machine learning in a business environment. Your goal is not to build a new machine learning algorithm but instead to build a tool/application that uses an existing machine learning library to solve a business problem.

For example, one common business application of machine learning is for an organisation (such as a Telco) to build a prediction model to predict which of their customers will likely churn (close their account) in the next month. This list of likely churners can then be passed to the marketing department who may take action to try to prevent the churn – e.g. offering a special deal etc. To build a churn model a Telco might use their customer account details and their customer past monthly billing data (say for the last 6 months). For example, the Telco might use their Jan-May billing data to predict customers who might churn in June. Usually the Telco will update its churn model every month, hence it would then use Feb-June billing data to predict churners in July etc. If you were to select this scenario as your business problem, then the application that you build could be a tool that allows the Telco to, every month, input and pre-process the appropriate data, build and test a prediction model, and then execute the model to obtain a list of likely churners that would then be given to the marketing department. The list might be ranked by propensity to churn.

Other common business applications of machine learning include:

- building models to predict who will buy a product or service
- building models to make product or service recommendations to user
- building models to predict if an insurance claim is potentially fraud (e.g. a car insurance claim, or a health insurance claim)
- building models to predict how long a patient's stay in a hospital will be (for bed usage optimisation)
- building models to predict if a customer is a credit risk (for assessing loan applications)

The above are common examples but you may select **any** problem domain that you wish provided:

- (a) You can find/obtain sufficient data to build and test the model (i.e. undertake the machine learning)
- (b) You build an application (including user interface) to solve a business problem and not just build a prediction model. Note that the business problem can be invented, e.g. any realistic problem that you make up that is solvable using the data you have. Also note that your application should be usable by a business user and not an expert in machine learning.

Project Process

1. Decide upon a problem domain and data set (or data sets). For example, you may wish to look at the datasets available on Kaggle.com or any other machine learning dataset repository.
2. Select a suitable business problem for the dataset(s) you have selected. Many datasets found online are already associated with a business problem (e.g. most Kaggle datasets already have a problem goal associated with them). However, you may choose to enhance or change that business problem or even invent your own business problem for that dataset.

3. Decide upon the programming language and machine learning library to use. For example, if you choose to use Java then Weka may be a suitable choice of library. If you choose Python then the scikit learn library is a possible choice etc.
4. Write a short (e.g. one page) description of the business problem you have selected, the datasets that you will use and the application that you plan to build and upload to IVLE before you begin. This description should include the typical the user-case(s) for the application. We will then give approval for you to go ahead if your proposal is suitable. Include the names of your team members in this proposal.

Deliverables and Grading

At the end of the elective (likely in late October or early November) you will make a team presentation at ISS in which you describe what you have done and give a short demonstration of your built application. You will also write and submit a report describing your business scenario, your dataset(s), and the steps that you went through to select, build, validate/test and deploy the machine learning models that you built.

Remember that no grade will be issued for this elective, only pass/fail. To obtain a pass be sure to demonstrate competency in (at least) the following areas:

- (a) **Business problem formulation:** your goal in this assessment project is not just to apply a machine learning algorithm to a dataset. You must in addition have a realistic business scenario / use-case that is solved.
- (b) **Data cleaning and preparation:** usually the raw data that you have will require data pre-processing to make it suitable for input into a machine learning algorithm. This may include creating new variables through data transformations etc.
- (c) **Selecting and applying a suitable machine learning approach:** often there are many alternative machine learning algorithms that might be suitable for solving a problem. You must demonstrate that you have selected the most appropriate algorithm/approach.
- (d) **Model testing:** You must validate and test your model before using it in practice*
- (e) **Model deployment:** Your application should allow the user to build, test and deploy new models as new data becomes available.

*Note that there is no predefined model accuracy that we expect you to achieve, for example, we do not require your models to be (say) 90% accurate. The accuracy that you obtain will be dependent on the data that you have available. However, you must report the model accuracy that you do obtain during model testing and demonstrate good process, i.e. demonstrate that you have taken all steps possible to obtain the best model accuracy and performance with the data that you have.

What you must do now:

1. Begin the Coursera Machine Learning course as soon as possible. There is no need to wait for permission to begin.
2. Whenever you complete an online assessment for one of the Coursera modules **you must take a screenshot** of the completion page (the page that inform you that you have passed the module) and upload to IVLE. You must do this for all Coursera modules.
3. Form teams for the ISS assessment project. Create a one-page project proposal and upload to IVLE, remember to include the names of the people in your team. You may begin your project anytime that you wish but plan to start by mid-September latest. Assessment (including project presentations) will be at the end of October.