**Instructions**: Please complete and submit your work to the appropriate folder in LumiNUS. You may work in study groups, but each student must be responsible for their own submission.

Please submit all the following documents as a single zip file named StudentID-Name-HW1.zip:

1. Completed Word file named as StudentID-Name-HW1.docx (with all results)
2. Print preview of ipynb file named as StudentID-Name-HW1.pdf (with all results)
3. Working ipynb file named as StudentID-Name-HW1.ipynb
4. A machine learning application scenario can be seen from four perspectives: (i) what is the technical problem to be solved, (ii) the data requirements, meaning can you get the data needed, (iii) security and privacy considerations, meaning what happens if data is leaked and (iv) the value proposition that machine learning brings to the table.

Choose two industries from the list below (or be creative and come up with your own) and give one example from each industry of how supervised machine learning can be applied. Your examples should follow the 4 perspectives outlined above.

Try to give different examples from those given in lecture. You may consult the Internet, but you must think things through yourself.

|  |  |  |
| --- | --- | --- |
| Retail | Fashion | Industry 4.0 |
| Banking | Education | Social Media |
| Healthcare | Communication Networks | Smart Home |

1. Redo Problem 1 for the unsupervised learning scenario. Give one example of an application scenario that is different from the examples discussed in the lectures.
2. Redo Problem 1 for the reinforcement learning scenario. Give one example of an application scenario that is different from the examples discussed in the lectures.
3. Suppose we want to remove vowels from a sentence. Write Python code to do this using iterators and list comprehension. The input sentence is: "The quick brown fox jumps over the lazy dog".
   1. Paste you Python code below, as well as the output of your program with the given input.
   2. Submit your iPython notebook file (ipynb file) as well as a pdf print preview of the ipynb file as instructed above.

1. One of the applications of supervised machine learning is in the Healthcare industry. For instance, machine learning model can be created to predict whether a patient has risk of cataract based on certain existing health related risk factors. Cataract is a medical condition in which the lens of the human eye becomes progressively opaque, resulting in blurred vision eventually. Decision tree algorithm and Naïve Bayes classifier can be used in this case to make the prediction.

The data required includes the potential risk factors of cataract and whether the patient has cataract or not (labelled data). The potential risk factors can include age, weight, past eye surgery, family history, smoking and alcohol consumption. These data can easily obtain by extracting through medical records which has approval by the patients for the sole purpose of research only.

The data provided are based on survey on patients that have undergo medical check-up for cataract. Patient’s personal details such as name, identification number and handphone number will not be used in the machine learning model and can be omitted. Thus, this ensures the security and privacy of the patients surveyed are well protected when creating the machine learning model.

This machine learning model can be used in widely in many developing countries. In developing countries, there is always a lack of access to medical consultation and assistive medical devices to detect cataract at an early stage. For instance, in Nigeria today, the risk of cataract is difficult to pre-determine before the onset of the disease, and usually individuals are already having signs of the disease even before consulting medical professionals. As such, this machine learning model will help to determine the risk of cataract, which will reduce the number of blindness cases from happening.

Another application of supervised machine learning can be observed in weather forecast. In this case, machine learning model is created to predict the weather for today based on historical weather data. For example, we can narrow the scope to predict whether it will rain or not for today based on the previous weather datasets, creating a binary classification task to solve in this case.

The data required includes the various parameters which will likely to cause rain, such as, historical temperature, precipitation rate, wind speed and humidity. In addition, we will also have to identify whether the weather is rainy or not (labelled data) for the historical data. These data can be easily obtained from the news which broadcast the daily weather status or the government agency such as National Environment Agency (NEA) webpage.

The data provided are open to all users for research purposes and the information are not confidential. There will not be any breach of individual security or privacy in this case.

This machine learning model is particularly useful as it allows user to accurately predict the weather for the day and effectively plan their schedule ahead. For instance, if one were to plan for an outdoor event, he can make use of this machine learning model to decide whether a wet weather plan must be executed in advance and allocate human resources accordingly. As such, this machine learning model is an effective tool for predicting weather trends and helps individual to plan earlier.

2. One application of unsupervised learning is in the Banking industry. For example, machine learning model can be applied to better understand which are the important information in determining whether a bank loan applicant can repay his loan. Banks usually offer loans to applicants based on the applicant’s financial capability to repay the loan with added interest in the future. However, there are many information to consider while deciding whether an individual can repay his loan. Banks will usually have to spend a large amount of time to assess an individual credit risk score before granting the loan to applicants. Hence, Principal Component Analysis (PCA) can be used in this case to reduce the data dimension (information provided by loan applicants) by compressing the data without losing relevant information needed to determine whether an applicant has the capability to repay his loan.

The data required includes the applicant’s nationality, age and average monthly income, debt, credit history, etc. Basically, the data will consist of all the information obtained when applicant filled up the bank loan agreement form. Usually, there will be many information that are not necessary in determining whether an individual can repay his debt, such as, name, race, and gender. The use of PCA will be able to reduce the amount of such parameters before used for analysis to assign a credit risk score to a loan applicant.

The data provided will require the approval of the bank loan applicants for the use of marketing and research. Once approved, these data will be used for the sole purpose of research and marketing purposes only. Hence, security and privacy of the data are approved by the applicants for usage.

This machine learning model is important because not every information provided by the applicant is useful in predicting an applicant’s credit risk score. As such, this information will be discarded, but in a way, these feature space (information) is projected onto a smaller feature space. In return, loan officers do not have spend large amount of time to analyse an individual background before granting the loan to applicants. This model also provide assurance to banks that loan applicants are likely able to repay their loans with interest.

3. One of the applications of reinforcement learning is in autonomous vehicle. For instance, autonomous cars can maneuver with little to no human supervision. Before this is carried out, a neural network agent is trained to map the autonomous cars’ current estimated state to a specified final target state. A reward system is imposed to reward the agent if the autonomous car managed to avoid obstacles successfully while maneuvering.

The data required are largely significant and will be constantly updated, where different environment or simulation platform, will be used to test the efficient of the autonomous car, in order to optimise its driving policy.

With regards to security and privacy, companies must invest in resources to figure out the best way to implement this machine learning model in their operations, services, and products without breaching individual’s privacy.

Autonomous vehicles are necessary because it reduces the number of accidents when travelling and creates an inclusive society. In Singapore, there were a total of 7,666 road accidents for year 2019. Most of the accidents were a result of reckless driving, drink driving and speeding. Imagine autonomous cars which has strict driving policy and traffic rules imposed were to replace manual driving, these accidents will not surface. Lastly, autonomous cars aid visually impaired users to travel from one location to another without any additional visual assistance.

4. (a) 