P1

Hello everyone, we are group 17. Today we are going to present our final project,

prediction in Medical Education.

P2

Our project focused on the application of Data Preprocessing (Stage 2) ,Transformation (Stage 3) and Feature Selection (Stage 4).

P3

As we all know, medical students need to study for six years with their learning pathway, which is General Studies (通識領域科目) 🡪 Basic Medical Sciences (基礎醫學) 🡪 Clinical Medicine (臨床醫學) 🡪 Practical Medicine (實習醫學)

The student performance will fluctuate during their study.

Therefore, we try to find out which of the following stages of general studies, basic medicine and clinical medicine can be predicted early in a student's final performance in practicum medicine and which subjects in that area can be predicted early.

P4

We will look at a total of 1366 students who graduated from a medical faculty in Taiwan from 2011 to 2019 in dataset.

These students have taken courses at various stages and in various fields during their six years at the Faculty of Medicine.

P5

We divide each student's performance in each subject into the top 50% and bottom 50% of each subject.

This slide is a visualisation of the data after binarisation

The final performance of students at each grade level in selected subjects in basic medicine is presented in the top 50% and bottom 50%.

The top 50% of students in the subject are marked as 0 and presented in blue

The bottom 50% are marked as 1 and presented in orange

P6

We have divided the subjects in the data set into four stages: General Studies, Foundation Medicine, Clinical Medicine and Practical Medicine, according to their grade levels.

Year 1 and 2 students need to study subjects in General Studies stage.

Year 3 and 4 students need to study subjects in Foundation Medicine stage.

Year 5 students need to study subjects in clinical medicine stage.

Year 6 students enters the internship stage

In addition, the subjects are subdivided into 19 major categories according to their attributes in General Studies aspect.

Within each of these categories, each subject is classified into its own category.

It should be noted that some of the categories have been removed from the cut-off diagram, such as Category 9, because the item Psychology was subsequently regrouped into Category 4, Medical Humanities, and given the number 4c.

P7

This page shows the subjects classification of General Studies.

P8

This page shows the subjects classification of basic medicine, clinical medicine and internship medicine respectively.

P9

In Data preprocessing stage

We fill in the averages in the few areas where there are missing values

If there are too few students taking a subject, e.g. 4b, we will delete the subject.

If a particular student is taking too few subjects, the data is also removed, possibly because the student has dropped out of the course and is not continuing.

In addition, we have also done a binarization process using Binarization, where the top 50% of students are assigned a value of 0 and the bottom 50% are assigned a value of 1, to divide the overall student performance into two groups.

This part has been explained in detail in the previous slides.

In order to make the prediction model applicable to different years and different stages of the curriculum, we normalised the prediction model for each subject by scaling the original data into the [0, 1] range.

The normalised prediction model can reduce the gap between the different data ranges of the characteristics, improve the speed of model convergence, and enhance the accuracy of the prediction model.

P10

The prediction performance of the model can be reduced by irrelevant or redundant features.

In order to improve the prediction performance of machine learning prediction models, we implement feature selection to select the most significant subjects that affect our results at each stage.

Generally speaking, there are three types of feature selection methods, namely Filter methods, Wrapper methods and Embedded methods.

We used the Filter methods, which were selected by evaluating only the association between variables and predicted values, evaluating and ranking a feature in turn and selecting the features that met specific criteria.

We use two indicators, AUC and univariate classifier, to evaluate the classification effectiveness of each classifier and also as a basis for our feature selection judgement.

AUC (Area Under Curve) represents the area under the ROC curve and is a commonly used statistic to indicate the predictive power of a classifier.

The closer the ROC curve is to the upper right, the better, so the larger the area under the ROC curve, the more effective the model.

When the AUC = 1, the classifier is perfect, but this is, after all, the ideal situation.

When AUC > 0.5, it means that the classifier is better than a random guess and the model has predictive value.

When AUC = 0.5, the classifier is as good as a random guess and the model has no predictive value.

When AUC < 0.5, the classifier classification is worse than random guessing, but better than random guessing if inverse prediction is performed.

We created six classifiers for each subject, based on the average of the AUCs obtained from the 10-folder when we are going to use single-variable classifier component.

As long as any AUC value is rounded to the second decimal place greater than 0.7, we select the subject as the feature and record the best model with this AUC value.

In order to find the best machine learning prediction classification method, we must cut the training set into sub-training and sub-testing sets. In cutting and training data sets, cross-validation can prevent the prediction model from relying too much on a particular training and testing data to produce bias, so that the prediction model will not have the overfitting problem. The cross-validation method used in this study is K-fold.

The K-fold method cuts the sample into multiple subsets for testing and training, and takes the average score of the evaluation factors as the final evaluation factor score of the prediction model.

The purpose is to evaluate the model using different subsets of the training set and test set.

In the next round, the other one is used and the other 9 are used for training. The results of each evaluation factor are recorded 10 times, and at the end of the 10 validations, the average evaluation factor score is obtained as the final prediction performance of the model.

P11

Due to time and space constraints, we have used the process of identifying the key generic subjects that will affect performance in the basic medicine stage as an introduction to the experimental process.

In practice, each stage of the experiment requires a new feature selection and prediction for each subject.

In this slide, we have narrowed the scope to two subjects for demonstration purposes.

In the Physician Scientist subject number 19, we used the students' Physician Scientist results to assess their performance in basic medicine, using seven separate models to predict performance in basic medicine and splitting the results into two groups.

All subjects have predicted AUC values and best-performing models for basic/clinical/practical medicine.

The subjects selected for each grade are therefore selected in different ways depending on the prediction results.

In this slide, a 19 physician-scientist with an AUC = 0.884 > 0.7 will be selected for this subject. 12e has a history of 0.684 < 0.7, so it will not be selected

Also note that the reference in the bottom right corner is the original name of the model abbreviated in the table

P12

Each subject was screened on the previous page to identify the most suitable subjects to be used as FEATURES in the later stages of the prediction.

This slide shows the 21 subjects in the General Studies area that were screened for AUC > 0.7, leaving 13 subjects with an AUC above 0.7 as features.

These are Chemistry (0.885), Physician-Scientist (0.884), Physics (0.774), Calculus (0.762), Medical Humanities Other (0.748), Language (0.737), Economics (0.736), Mathematics (0.735), Medical Social Miscellaneous (0.735), Literature (0.731), Music, Art and Drama (0.715) Human, Social, Cultural (0.707), Law (0.69)

The brackets to the right of the subject are the AUC values selected during feature selection.

Using these 13 subjects and the best-performing model, here GNB, for training, we can obtain an accuracy of 0.84 in basic medicine

P13

This slide is a selection of 21 subjects in Liberal Studies with AUC > 0.7, leaving 7 subjects with AUC over 0.7 as FEATURES.

These were Physician Scientist (0.865), Chemistry (0.793), Physics (0.736), Medical Humanities (0.719), Medical Social Miscellaneous (0.711), Calculus (0.702), and Mathematics (0.702).

Using these subjects and selecting the best-performing mod, the LR mod was trained here to obtain an accuracy of 0.785 in the clinical phase.

P14

This slide shows the 21 subjects in the Liberal Studies domain filtered for AUC>0.7, leaving 2 subjects with AUC above 0.7 as FEATURES.

These were Medical Humanities (0.754) and Medical Scientist (0.749).

Using these subjects and the best performing mod, the LR mod was trained here to obtain an accuracy of 0.655 during the medical internship.

P15

This is a prediction of the clinical outcome by finding subjects that meet the criteria in the basic medical subjects

The final accuracy of 0.871 was obtained in the SVC model

P16

This sheet is used to predict the outcome of medical internships by finding subjects that meet the criteria in the basic medical subjects

The final accuracy of 0.659 was obtained in the GNB model

P17

In this page, the first two stages of the General Studies and Foundation Medicine subjects are combined to find subjects that meet the criteria for predicting outcomes in clinical medicine

An attempt was made to find better results using the combination of the two previous phases of clinical medicine

The accuracy of 0.882 was obtained using the RF model with 17 subjects selected.

P18

we combine the first two stages of general studies with basic medical subjects to find the subjects that meet the criteria for predicting the outcome of the medical internship

To find a better result by combining the two previous stages of medical practice

In the end, the accuracy of 0.686 was obtained by using the LR model with 7 subjects selected.

P19

In the final results, we will look at various areas

In basic medicine, we get the best results from the generalist areas

Chemistry, Physician-Scientist, Physics, Calculus, Medical Humanities Other, Language Domain, Economics, Mathematics, Medical Social Miscellaneous, Literature, Music, Art and Drama, Human, Social and Cultural, Law

The best result for these 13 subjects is 0.84

This means that the results of these 13 subjects in the General Studies area are often directly related to the performance of Foundation Medicine.

It is also possible to predict the subsequent performance of basic medicine at the end of these 13 subjects

P20

The second result is that in clinical medicine, the best results are obtained by combining the generalist domain with the basic medical subjects.

When the General Studies domain alone is used to predict clinical medicine it has an accuracy of 0.785

When the basic medical subjects alone are used to predict clinical medicine, the accuracy is 0.871.

But when these 17 subjects are combined, the best result is 0.882

In other words, the results of these 17 subjects in general studies and basic medicine are often directly related to the performance of clinical medicine.

It is also possible to predict the performance of subsequent clinical medicine at the end of these 17 subjects

P21

In the last result, the best prediction in the internship phase of medicine was also obtained by combining the general studies domain with the basic medicine subject.

When the General Studies domain alone is used to predict the internship medicine, it has an accuracy of 0.655

When the subject of basic medicine is used alone, the prediction is 0.659

However, when these 7 subjects are combined, the best result is 0.686

In other words, the results of these seven subjects in general studies and basic medicine are often directly related to the performance of practicum medicine.

It is also possible to predict the performance of subsequent medical internships by the time these 7 subjects are completed

P22

In this project, we used the AUC to determine which subjects were suitable for the prediction and the model evaluation part to count the mods with the best performance in each subject and finally use the mods with the most and best results for the training to get the prediction results for each stage.

The third point is that we can find suitable subjects in this project to represent how well they will perform in the later stages.

This is suitable for use in predictions with a time-series concept, where we can use the performance of some subjects in the earlier stages of the experiment to give an early warning of performance in the later stages.

It is also suitable for use in other areas of study, for example, in the curriculum of the Department of Information Technology, where it may be possible to identify highly correlated subjects for achievement prediction.

This is the end of our presentation, thank you for your listening.