Grade Predictions

For A Level H2 Mathematics

An end-to-end Data Science Project
General Assembly Data Science Immersive Capstone Project

BY: LIW Wenyan

CONTENTS

THE PROBLEM

A level grade predictions are used for

1. applying to top overseas universities
2. applying to competitive scholarships

These predictions are confidential and the practice differs across schools.

Objective 1

Evaluate effectiveness of benchmark method using only Prelim marks and no machine learning models

Objective 2

Give better predictions to boost university / scholarship applications

Objective 3

Reduce number of students on watchlist for closer monitoring

Tools for Evaluation

Weighted Cohen's Kappa

- Mitigates issues of high accuracy scores due to large class imbalance
- Further predictions penalized more heavily

Recall Scores

- Recall scores of lowest grade category 'ESU' (aim for 100%)
- Macro average recall to prevent results from being pulled up by majority class

Confusion Matrix

- Show correct predictions and where wrong predictions are
- Can use this to calculate proportion of predictions within 1 grade, or better than actual, etc.

THE BENCHMARK

Generating Benchmark Predictions

Step 1

Remove rows without Prelim marks

These have either ‘MC’ or ‘EX’ recorded instead of numbers

Step 2

Sort by marks

Calculate proportion of each grade multiplied by number in cohort and assign in order

Step 2

Get 100% ‘ESU’ Recall

Include more students in lowest grade category such that ‘ESU’ Recall hits 100%

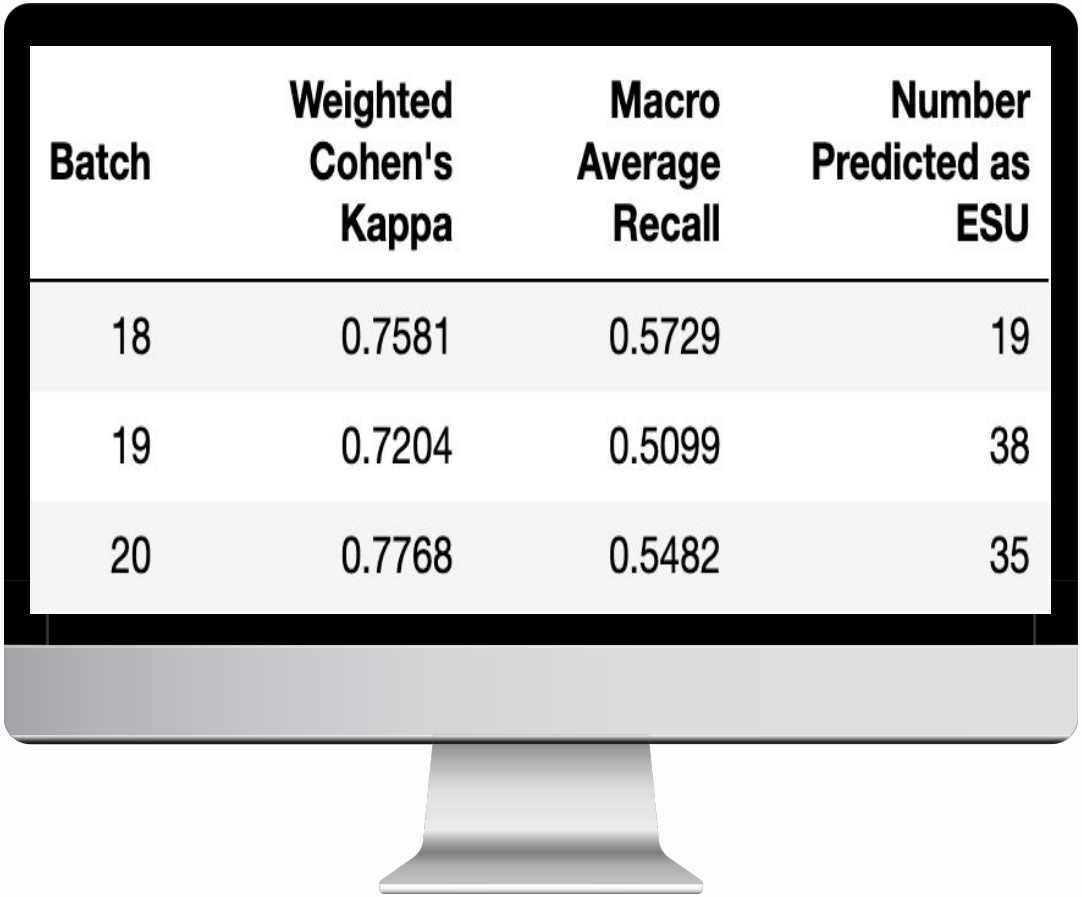
Step 3

Make adjustments

- Add/remove from majority class grade ‘A’ if total does not tally
- Ensure students with same marks get same (better) prediction

Benchmark Scores

As all batches have ‘ESU’ recall at 100%, we use the number of ‘ESU’ predictions as a secondary metric.



Batch	Weighted Cohen's Kappa	Macro Average Recall	Number Predicted as ESU
18	0.7581	0.5729	19
19	0.7204	0.5099	38
20	0.7768	0.5482	35

THE DATA

~ 1000 rows for each batch

2 batches for building model, 1 latest batch for final testing

~ 360 columns for each batch

Including duplicates and unusable columns

Many empty cells

Rows cannot be dropped, creative feature engineering to get full columns as far as possible

Data Challenges

Examples of some issues

and strategies used to combat them

Most grades missing

- Subjects combined where possible (e.g. GP/KI)
- H1 grade of 0 (better than highest grade 'A') for students without H1
- H3 grade of 5 (lower than worst grade 'Unclassified') for students without H3
- Subjects one-hot encoded to reflect whether student takes or does not take the subject

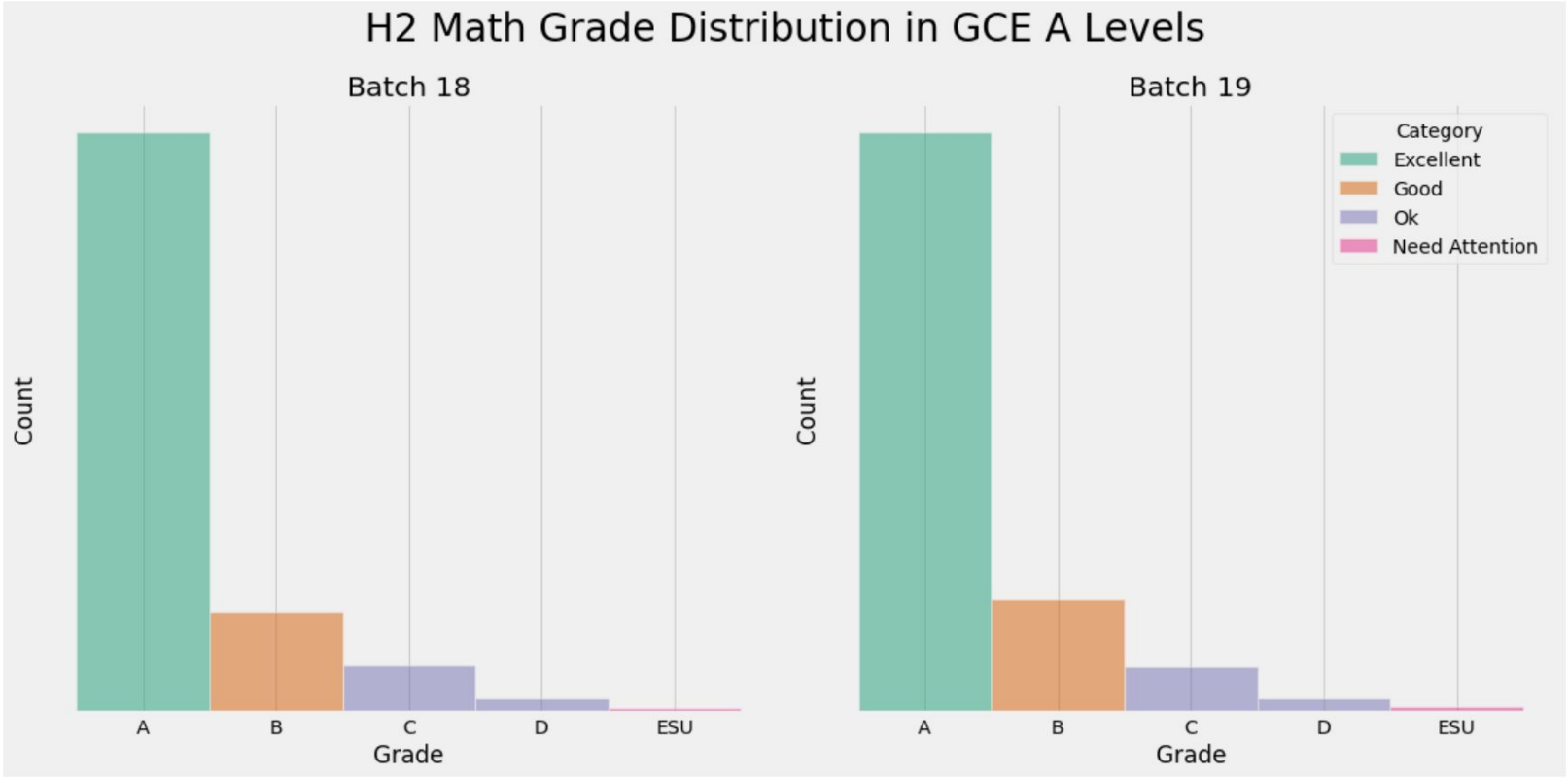
Sec4 overall results

- O level students have L1R5 while others have MSG
- Error: O level students have MSG recorded as 0
- Error: Non-O level students have L1R5 recorded as 6
- L1R5 and MSG standard scaled to mean 0 and standard deviation 1 and then combined

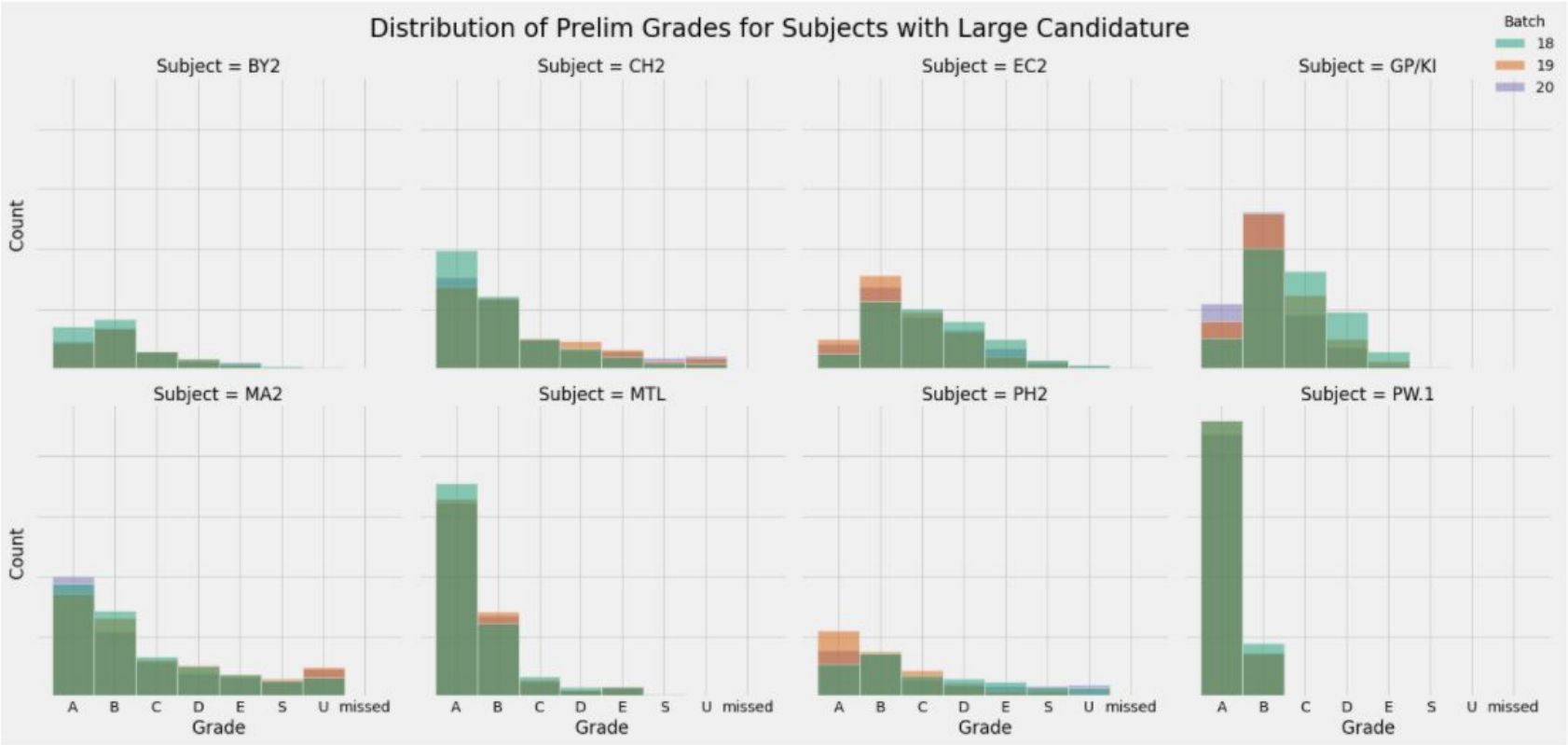
CCA / CIP records

- Very wordy with many unique entries
- Reclassified to binary variables such as 'CCA Leader' if 'Captain', 'President' or 'Head' appeared in description

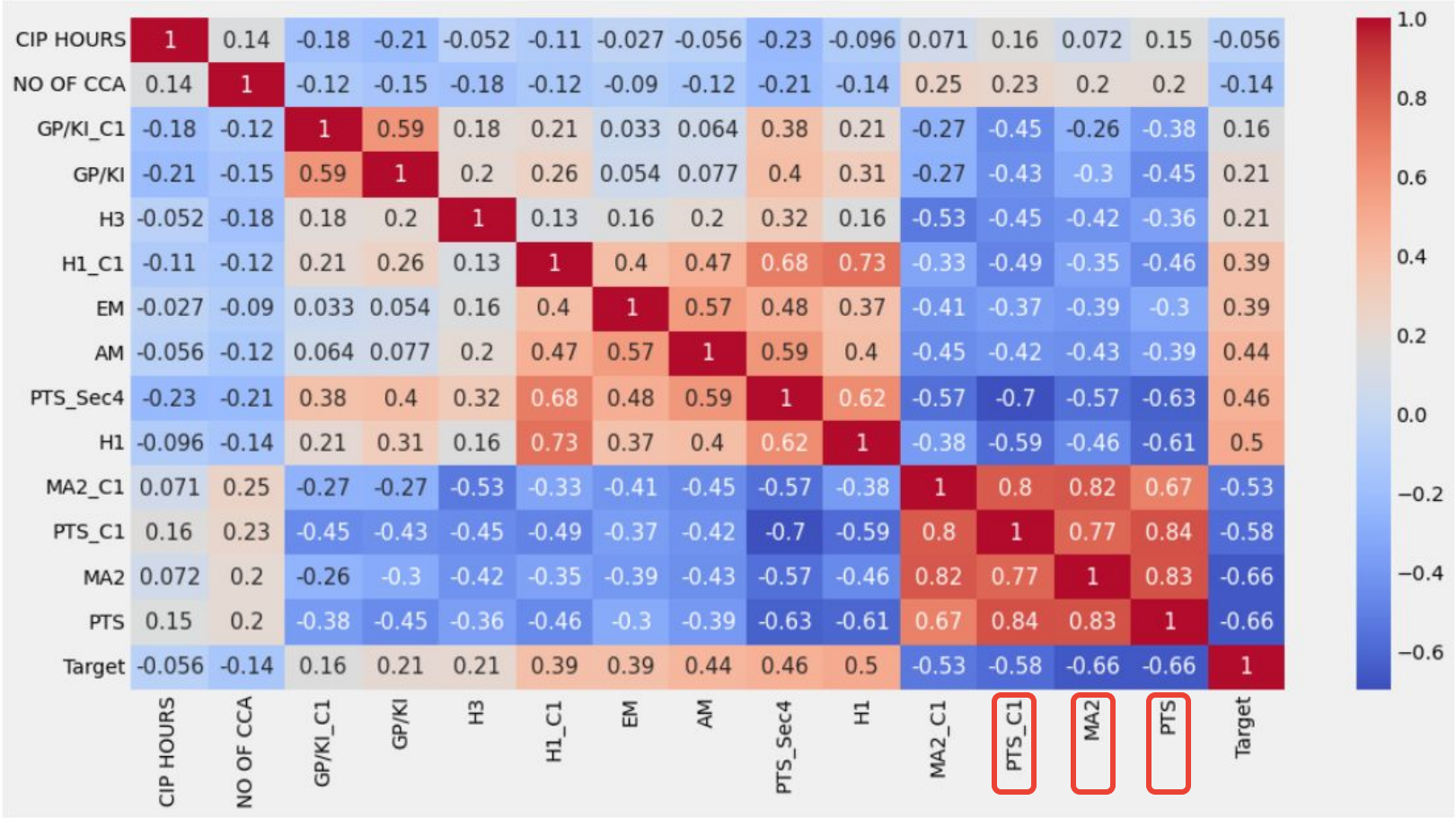
Data Visuals



Data Visuals



Features Selected



THE MODELS

Classifiers with class weights hyperparameter

Logistic Regression

- Best performing before tuning
- Added advantage of interpretable coefficients

Support Vectors

- Poor scores
- Too few ‘ESU’ predictions even with class weight of ‘ESU’ set to 1000 times higher

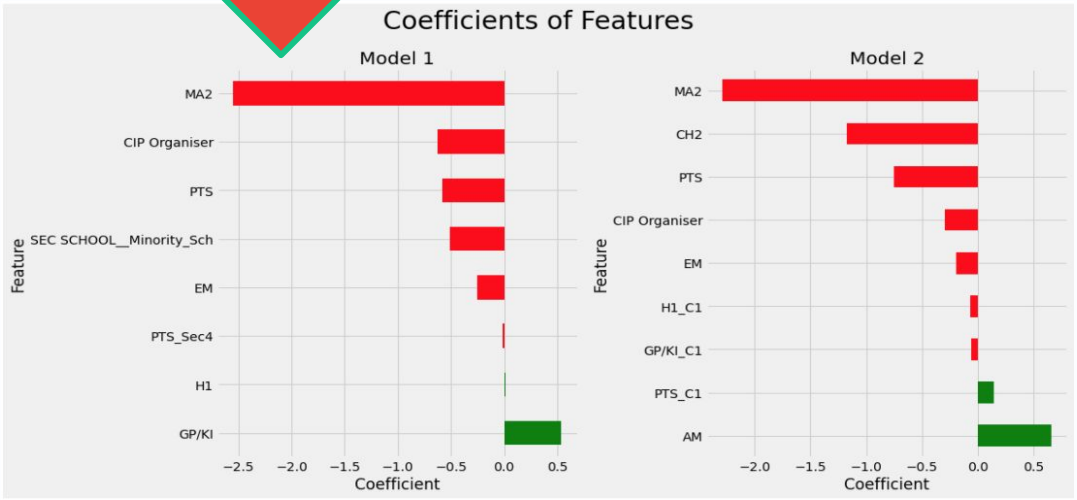
Random Forest

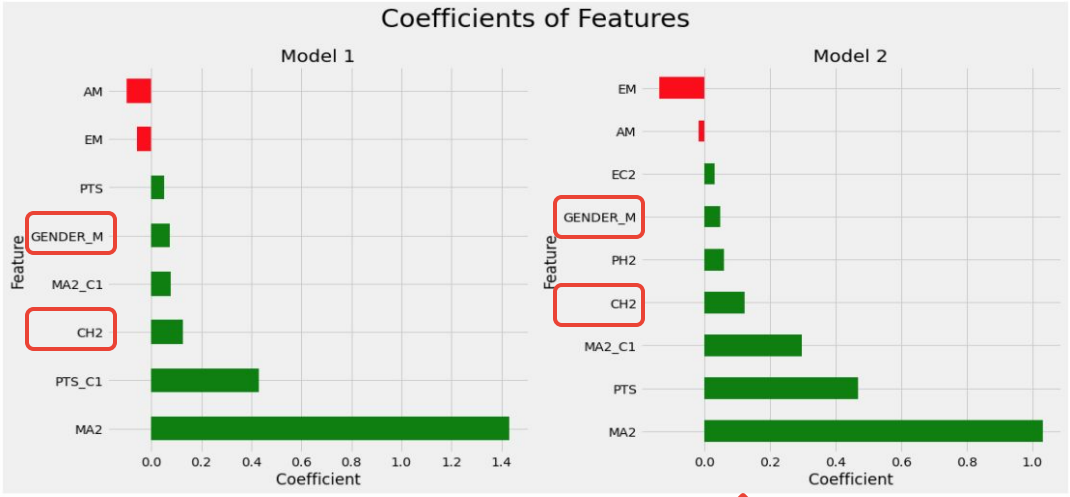
- Severe overfitting
- Better weighted kappa
- Too few ‘ESU’ predictions even with class weight of ‘ESU’ set to 1000 times higher

Extra Trees

- Severe overfitting
- Poor scores

Top Feature for Predicting ‘ESU’





Top Feature for Predicting ‘A’

THE RESULTS

Batch	Model	Weighted Cohen's Kappa	Recall for ESU	Macro Average Recall	Number Predicted as ESU
18	Logistic Regression	0.7681	1.0	0.5431	36
18	Benchmark	0.7581	1.0	0.5729	19
19	Logistic Regression	0.7130	1.0	0.5150	24
19	Benchmark	0.7204	1.0	0.5099	38
20	Logistic Regression	0.7876	0.8235	0.5216	25
20	Benchmark	0.7768	1.0	0.5482	35

Batch	Model	Correct Predictions	Within 1 Grade	2 or More Grades Away	Better Than Actual	Worse Than Actual
18	Logistic Regression	82.3%	97.0%	3.0%	9.9%	7.7%
18	Benchmark	81.0%	97.5%	2.5%	9.1%	9.9%
19	Logistic Regression	78.4%	96.6%	3.4%	12.0%	9.6%
19	Benchmark	77.6%	95.9%	4.1%	10.3%	12.1%
20	Logistic Regression	80.4%	97.3%	2.7%	12.4%	7.1%
20	Benchmark	80.7%	97.3%	2.7%	10.1%	9.1%

THE RESULTS

The Objectives

Objective 1

Benchmark is highly effective, with weighted kappa scores between 0.7 and 0.8.

Highly recommend schools to implement this method systematically.

Objective 2

Model gives higher proportion of predicted grades better than actual compared to benchmark.

For students at boundary of 'A' and 'B', teachers could focus more on the female students who do not take H2 Chemistry.

Objective 3

Final model only put 25 students on watchlist compared to 35 in benchmark for batch 20.

Failed to capture 100% of ‘ESU’ students.

Future Works

More batches

Limited to 3 batches

Better results may be obtainable with more years of data

Other subjects

e.g. General Paper

Taken by almost every A level student
Subjective marking making it more challenging

More features

Limited due to confidentiality concerns

e.g. Home-school distances, form teacher's comments

Thanks !