TELECOM ParisTech

Institut Mines-Télécom

Plate-forme d'évaluation par les pairs: cas d'usage en formation continue (Classgrade)

Joseph Salmon

université PARIS-SACLAY

« Mastère Spécialisé Big Data » (MSBGD)

Formation continue en plein essor :
statistiques/data science/ base de données, etc.

Effectif : **30** élèves 2013/14 -> **60** en 2017/18

Interventions personnelles :

- cours + TP en MDI 720 (Statistiques)
- cours + TP en MDI 343 (Machine Learning)
- TP en MDI 341 (Machine Learning avancé)

Notation par les pairs (*peer-grading*)

Concept : les élèves se notent les uns les autres de manière **anonyme** en s'appuyant sur une correction proposée par le professeur

Notation par les pairs (*peer-grading*)

Concept : les élèves se notent les uns les autres de manière **anonyme** en s'appuyant sur une correction proposée par le professeur

Motivation initiale :

- **homogénéiser** efficacement les pratiques numériques des élèves
- **plus de pédagogie** / moins de correction de copies
- augmenter la fréquence d'évaluation pour un **travail plus régulier**
- **renforcement du propos** : les élèves apprennent entre eux et passent plus de temps sur un même concept

Mise en place

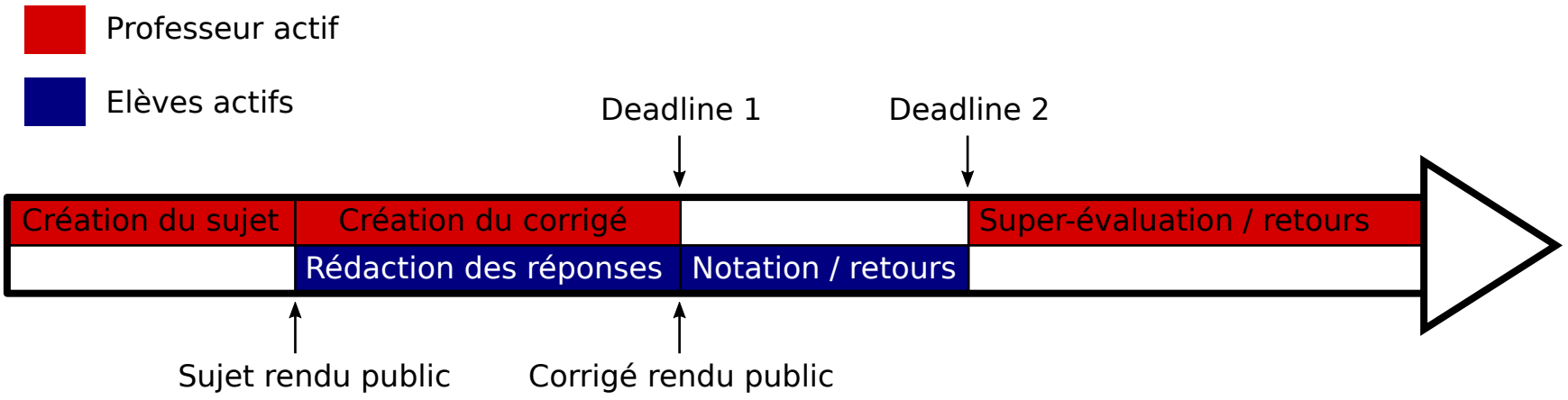
Démarrage :

- 2016 : MDI 720 (Statistiques) : 2 TP notés
- 2017 : MDI 720 (Statistiques) : 3 TP notés (dont un TP de “prise en main” ; coefficient 5% – 20% – 20% de la note finale)

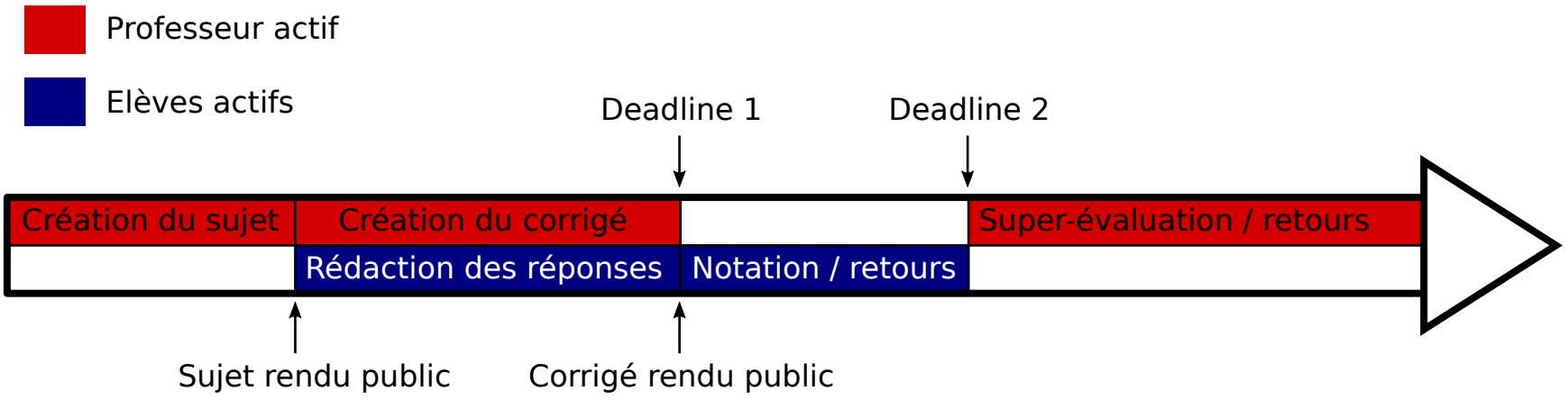
Solution proposée : **Classgrade**, logiciel libre financé par la Formation Continue (supervision : **C. Marini**, ingénieure de recherche CNRS), déployé sur un serveur interne **Peergrade**

Classgrade : `https://github.com/classgrade/`
Peergrade : `peergrade.enst.fr`

Organisation pratique

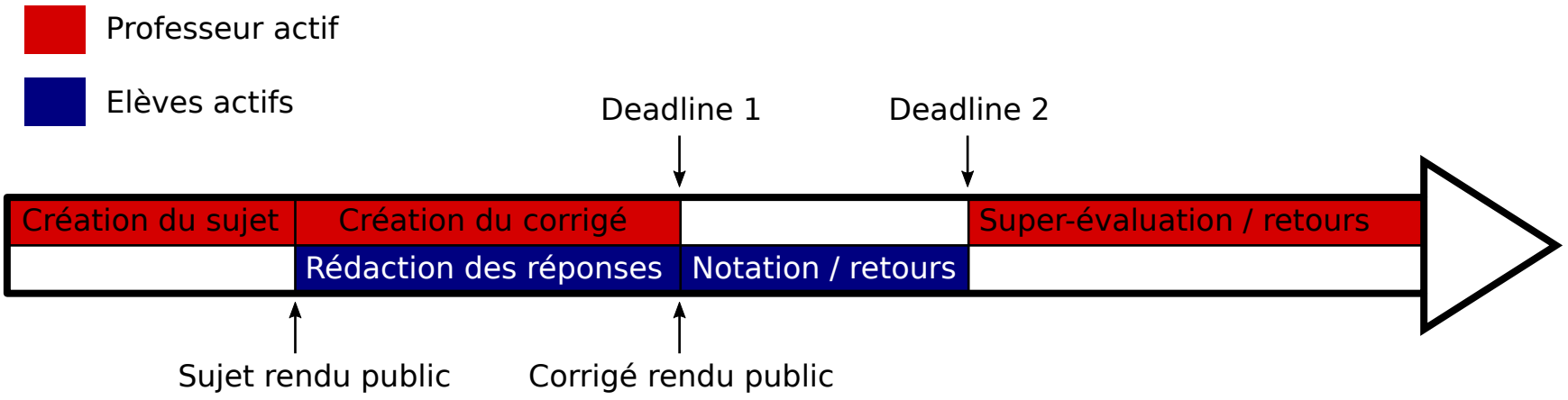


Organisation pratique



Déroulement des devoirs :

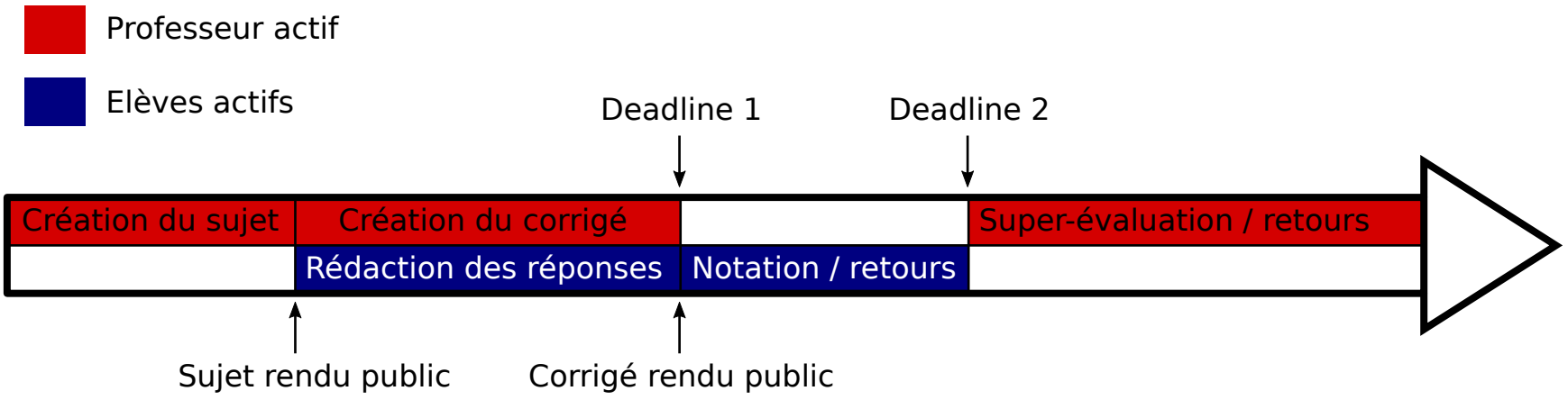
Organisation pratique



Déroulement des devoirs :

- Écriture du sujet par l'enseignant

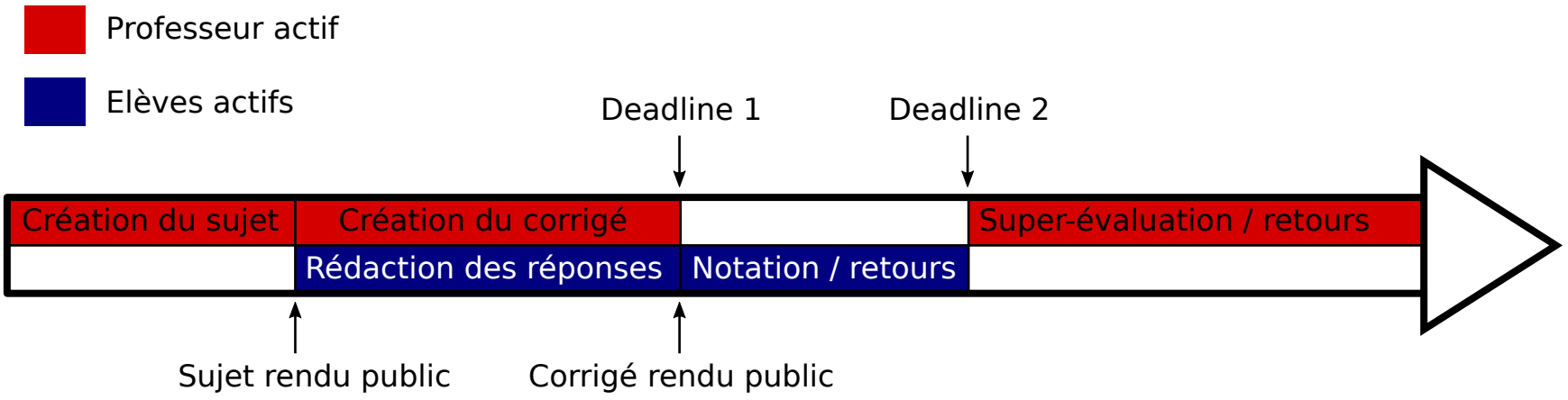
Organisation pratique



Déroulement des devoirs :

- Écriture du sujet par l'enseignant
- Collecte liste des étudiants (format `.csv` nom/prénom/email)

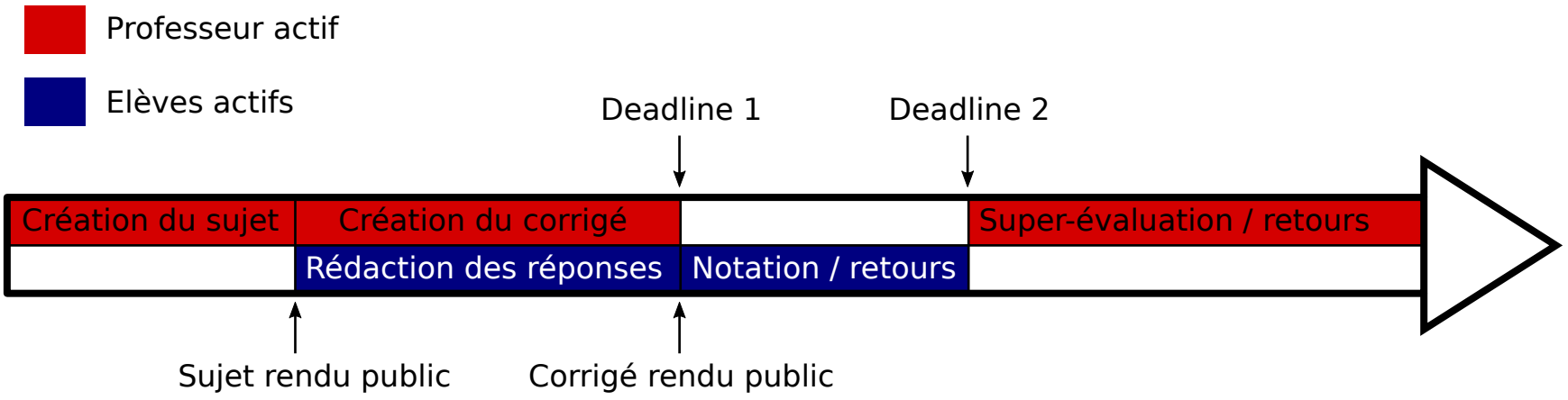
Organisation pratique



Déroulement des devoirs :

- Écriture du sujet par l'enseignant
- Collecte liste des étudiants (format `.csv` nom/prénom/email)
- Choisir le nombre de correcteurs par copie (défaut=3)

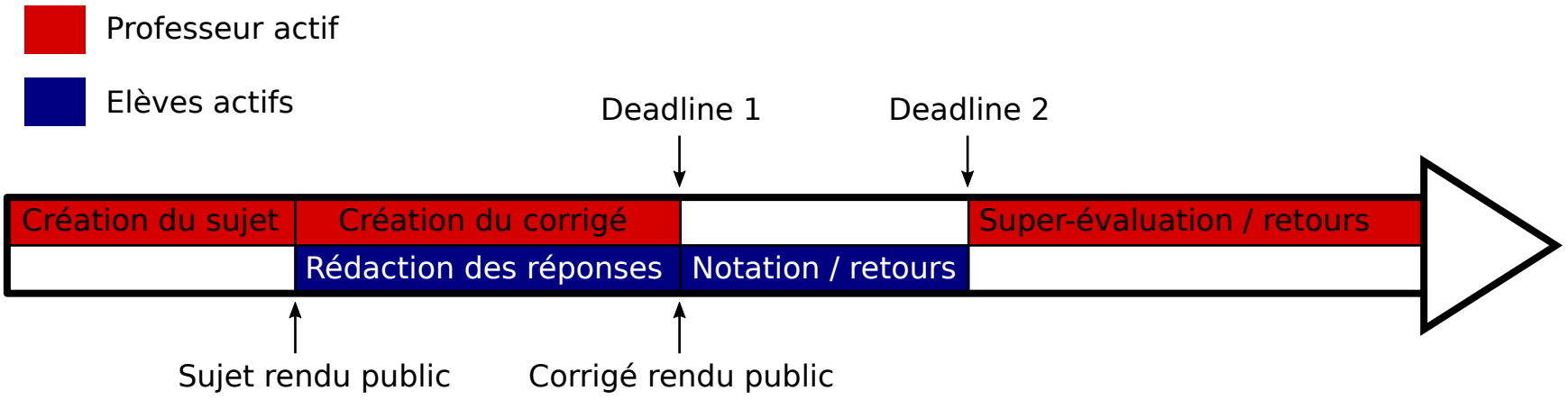
Organisation pratique



Déroulement des devoirs :

- Écriture du sujet par l'enseignant
- Collecte liste des étudiants (format `.csv` nom/prénom/email)
- Choisir le nombre de correcteurs par copie (défaut=3)
- Planifier : date/heure de rendu des devoirs/corrections

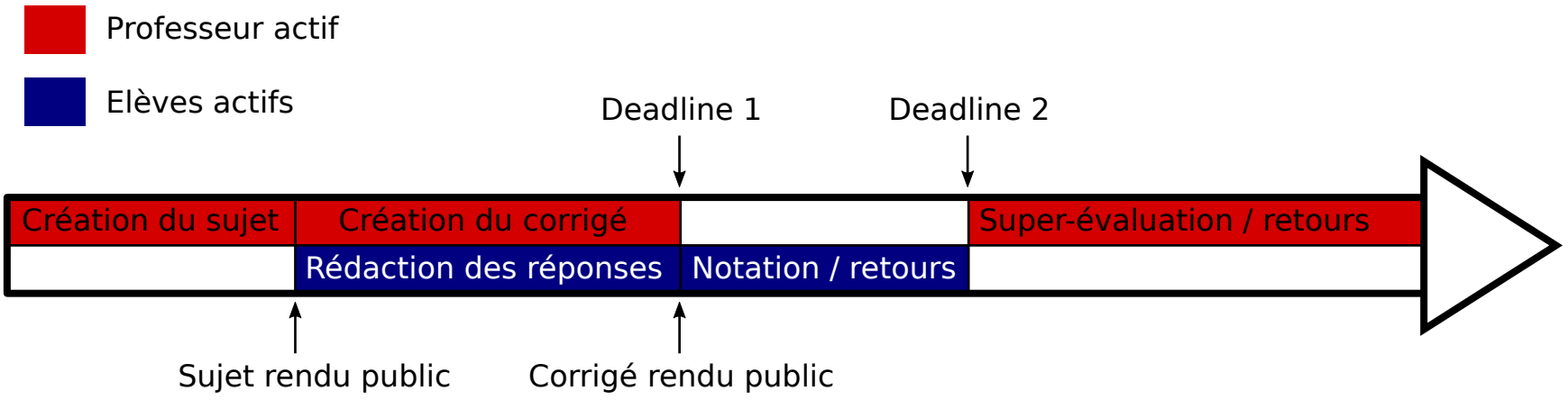
Organisation pratique



Déroulement des devoirs :

- Écriture du sujet par l'enseignant
- Collecte liste des étudiants (format `.csv` nom/prénom/email)
- Choisir le nombre de correcteurs par copie (défaut=3)
- Planifier : date/heure de rendu des devoirs/corrections
- Rédiger correction avant début phase de correction

Organisation pratique



Déroulement des devoirs :

- Écriture du sujet par l'enseignant
- Collecte liste des étudiants (format `.csv` nom/prénom/email)
- Choisir le nombre de correcteurs par copie (défaut=3)
- Planifier : date/heure de rendu des devoirs/corrections
- Rédiger correction avant début phase de correction
- Super-évaluation + feedback vers les élèves

Création des devoirs

Éléments requis pour l'enseignant :

- une liste (`.csv`) avec les emails des élèves (indépendant de synapse, flexibilité en formation continue)
- numéroter les questions d'un sujet de TP/examen
- barème modulable a posteriori (réponses simples : 0/1/2 points)

Classgrade | Create an assignment | Running assignments | Archived assignments | Hi joseph! | Logout

Your running assignment list:

Title	Number of graders	File type	Deadline submission	Deadline grading	Details	Modify	Reset	Archive	Delete
TEsting 1 MDI 7221	3	ipynb	Sept. 30, 2016, 5:05 p.m.	Sept. 30, 2016, 5:10 p.m.					
MDI720-DM1	3	ipynb	Oct. 10, 2016, 2:15 p.m.	Oct. 14, 2016, 11:59 p.m.					
MDI720-DM2	3	ipynb	Oct. 23, 2016, 11:59 p.m.	Oct. 28, 2016, 11:59 p.m.					
MDI720-TP1-2017-18	3	ipynb	Oct. 4, 2017, 11:59 p.m.	Oct. 9, 2017, 11:59 p.m.					
MDI720-TP2-2017-18	3	ipynb	Oct. 15, 2017, 11:59 p.m.	Oct. 23, 2017, 11:59 p.m.					
MDI720-TP3-2017-18	3	ipynb	Oct. 29, 2017, 11:59 p.m.	Nov. 6, 2017, 11:59 p.m.					

Création des devoirs (bis)

Nombre de correcteurs:

Deadlines:

Nombre de questions:

Liste des élèves:

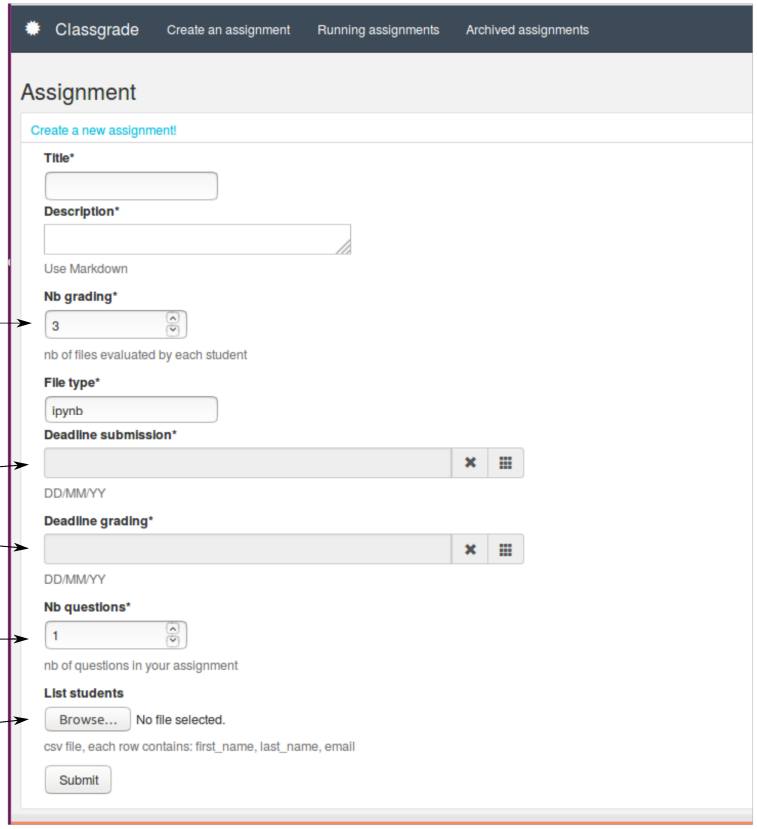


Tableau d'affichage des notes

Classgrade | Create an assignment | Running assignments | Archived assignments | Hi joseph! | Logout

Assignment MDI720-DM2

Description	Number of graders	Number of questions	Question coefficients	File type	Deadline submission	Deadline grading
Bootstrap	3	20	[1.0, 1.0, 2.0, 3.0, 2.0, 1.0, 1.0, 2.0, 2.0, 3.0, 3.0, 2.0, 2.0, 2.0, 2.0, 3.0, 3.0, 3.0, 2.0]	ipynb	Oct. 23, 2016, 11:59 p.m.	Oct. 28, 2016, 11:59 p.m.

Modify details (not the student list) | Set question coefficients | Set problem statements | Get csv with student grades | Get zip with student assignments | Show stats

Student	Assignment	Upload date	Peer-grade 0	Peer-grade 1	Peer-grade 2	Mean	Std	Grade
Victorien_GAUCH		10/23/16 23:40	GREGORY_FREYD 18.8	Yacine_HADDAM 17.38	Damien_RUSSIER 17.61	17.93	0.62	
Audrey_QUESSADA		10/23/16 21:09	Charlotte_ELI 20.0	Willie_DROUHET 19.76	Jean-Baptiste_KERBOUL 18.8	19.52	0.52	
GREGORY_FREYD		10/23/16 23:56	Maylis_COTADZE 18.8	Edwy_ZOONEKYND 18.33	Lorraine_NYEMBI_BINJENJ 17.85	18.33	0.39	
Guy_marcel_MBULA_BAKU		10/23/16 22:07	Dorian_BAGUR 15.95	Pierre-Yves_CASANOVA 17.14	Mahzad_KALANTARI 17.38	Supereval 16.19		
Arthur_OUAKNINE		10/23/16 19:10	Sidoine_KAKEUH_FOSSO 14.28	Mohamed_hamdi_JDAY 20.0	Clement_BEGOTTO 17.14	Supereval 17.61		

- notation : moyenne des trois notes
- super-évaluation : si mécontentement ou grand écart-type

Notations/retours des étudiants entre eux

Question 16: 1/2
Le résultat est erroné car le biais créé par le package PolynomialFeatures n'a pas été retiré (include_bias=False).

Question 17: 1/2
Dans la question on cherche à trouver l'origine la plus efficace en terme de consommation.

Question 18: 1/2
il manque un petit nettoyage sur les marques de voiture mal renseignées: car_brand = car_brand.replace('toyouta', 'toyota')\ .replace('vw', 'volkswagen')\ .replace('vokswagen', 'volkswagen')\ ...

Attention skl_linmod.coef_.max() ne prend pas la plus grande valeur en valeur absolue.

Question 19: 1/2
H doit aussi être symétrique.

Question 20: 1/2
Il manque une interprétation sur le caractère atypique des voitures. Attention à la présentation, pour un utilisateur final, il est plus facile d'identifier la voiture par son nom plutôt que par un index.

Question 21: 2/2
OK

Question 22: 1/2
Des efforts pourraient être faits sur la lisibilité du code. Pour PEP 8, je conseille d'installer jupyter-autopep8 qui te permettra de formatter les cellues sans effort.

Question 23: 2/2
OK

Get comments in a txt file Close

Sortie : notes etc. au format “.csv”

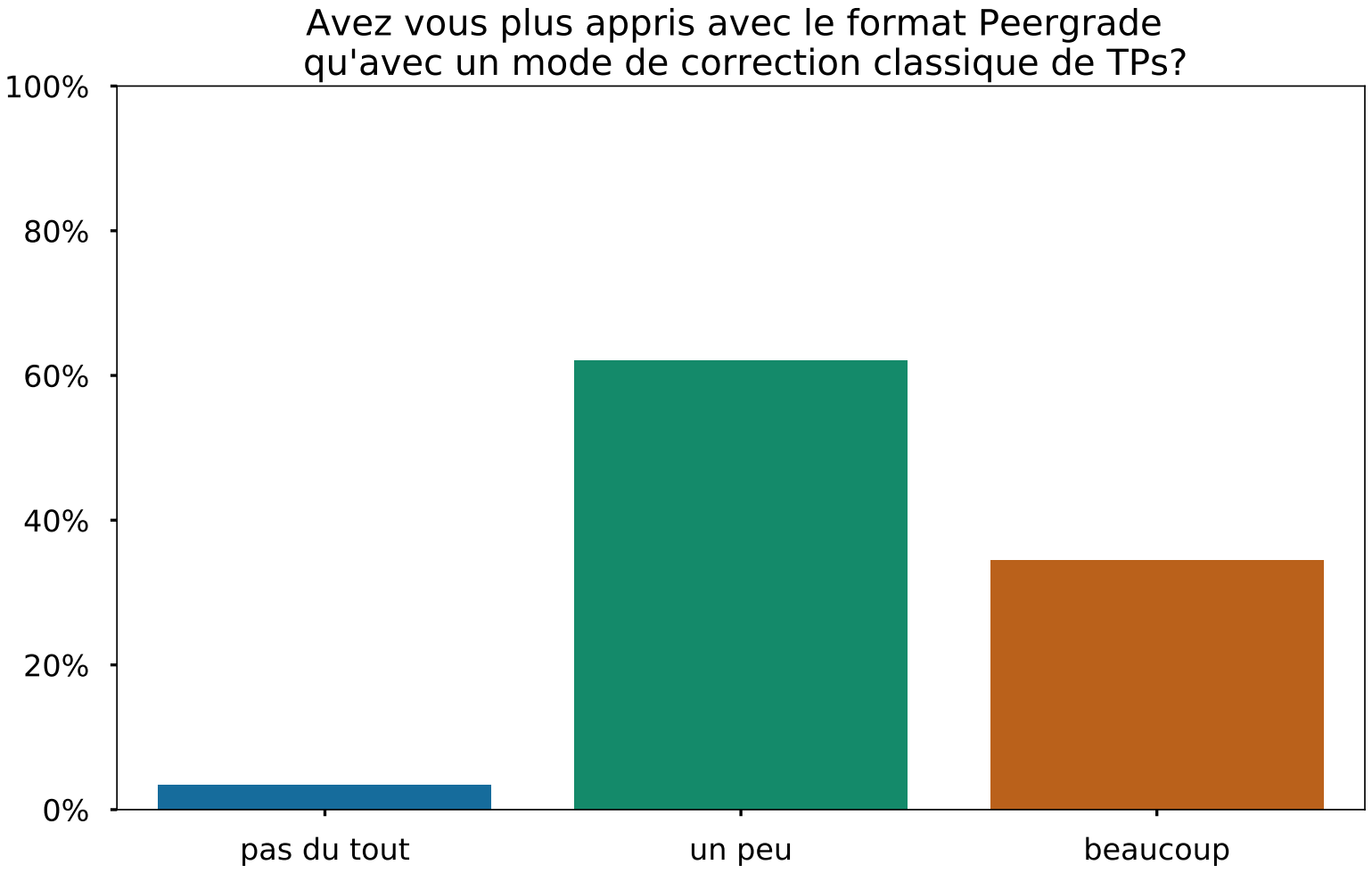
grades_TP3.csv - LibreOffice Calc

A1 | Coeff:

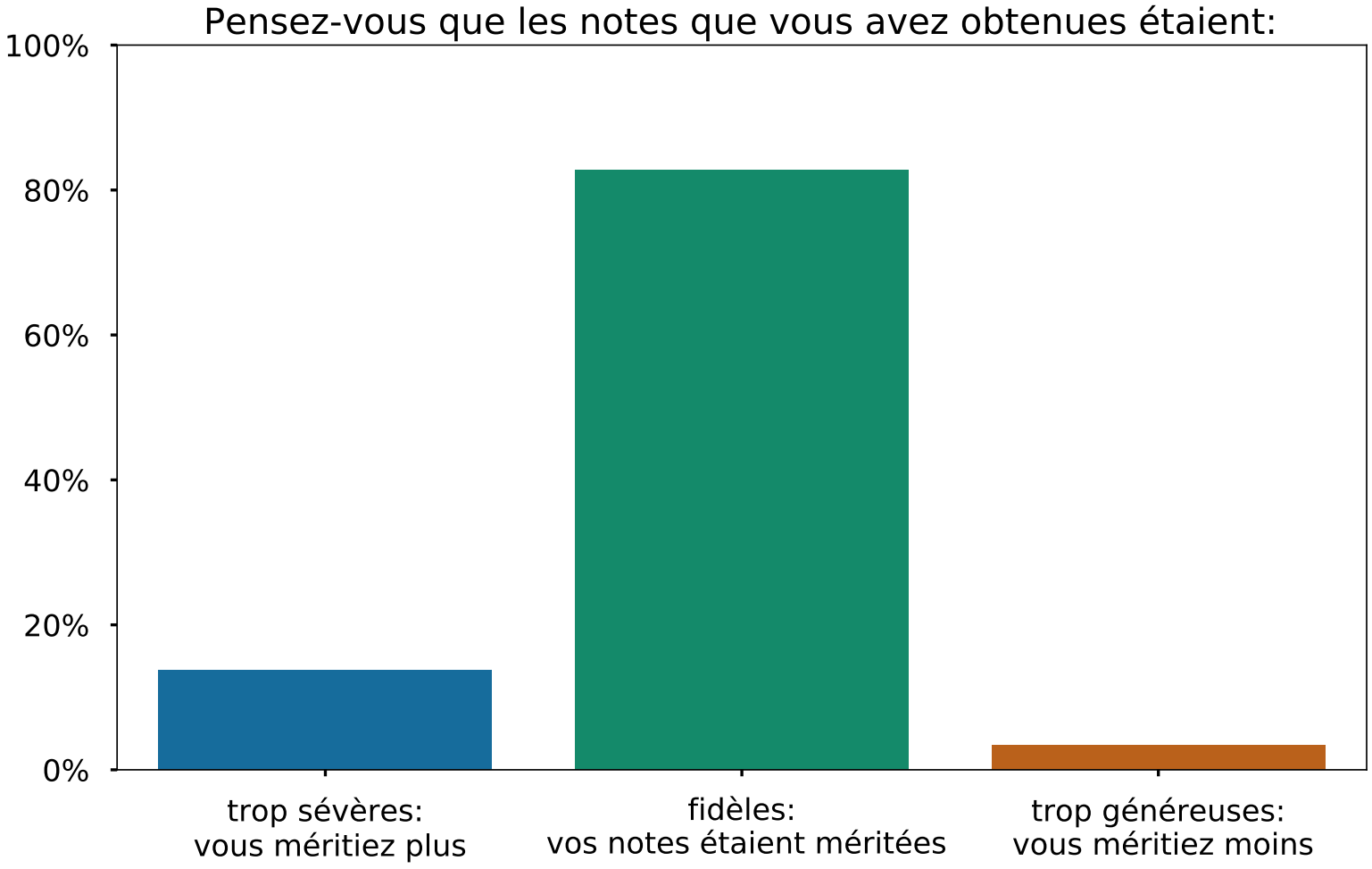
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Coeff	1	1	2	1	1	1	1	0.5	0.5	1	1	1	1	1	1	2	2	1	
2	[illegible]16213)	('mean', 18.4153439153439)																		
3	[illegible]	Mean	R1Name	R1Grade	R1Q1	R1Q2	R1Q3	R1Q4	R1Q5	R1Q6	R1Q7	R1Q8	R1Q9	R1Q10	R1Q11	R1Q12	R1Q13	R1Q14	R1Q15	R1Q16
4	[illegible]	18.5833333333	[illegible]	18.25	2	2	1	1	2	2	2	1	2	2	2	2	2	2	2	2
5	[illegible]	16	[illegible]	16.5	2	2	2	2	2	0	1	0	2	2	2	2	1	2	2	1
6	[illegible]	19.0833333333	[illegible]	19	2	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2
7	[illegible]	19.5	[illegible]	20	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
8	[illegible]	18.4166666667	[illegible]	18.5	2	2	2	2	2	2	2	2	2	2	2	1	2	2	0	2
9	[illegible]	16.8333333333	[illegible]	17	2	1	2	2	2	1	1	0	2	2	2	2	2	1	1	2
10	[illegible]	17.75	[illegible]	17.25	2	2	2	2	1	2	1	1	2	2	2	2	1	2	0	2
11	[illegible]	19.1666666667	[illegible]	19	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
12	[illegible]	18.75	[illegible]	19.5	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
13	[illegible]	19.8333333333	[illegible]	20	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
14	[illegible]	15.5	[illegible]	14	2	2	2	0	2	2	1	0	2	2	2	2	2	0	0	1
15	[illegible]	19.0833333333	[illegible]	18.25	2	2	2	2	0	2	2	1	2	2	2	2	2	2	1	2
16	[illegible]	15.3333333333	[illegible]	18.5	2	2	2	2	2	2	2	0	2	2	2	2	2	2	2	1
17	[illegible]	19.4166666667	[illegible]	19.5	2	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2
18	[illegible]	19.4166666667	[illegible]	19.5	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2
19	[illegible]	19.3333333333	[illegible]	20	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
20	[illegible]	19.5	[illegible]	19.75	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2
21	[illegible]	19.6666666667	[illegible]	20	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
22	[illegible]	19.6666666667	[illegible]	19	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
23	[illegible]	19.8333333333	[illegible]	19.5	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2
24	[illegible]	19.25	[illegible]	20	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
25	[illegible]	19.1666666667	[illegible]	19.25	2	2	2	2	2	2	1	1	2	2	2	2	2	2	2	2
26	[illegible]	19.5	[illegible]	18.5	1	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2
27	[illegible]	18.75	[illegible]	19	2	2	2	2	2	2	2	2	0	1	2	2	2	2	2	2
28	[illegible]	19.6666666667	[illegible]	20	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
29	[illegible]	18.9166666667	[illegible]	19	2	2	2	2	1	1	2	2	2	2	2	2	2	2	2	2
30	[illegible]	13.3333333333	[illegible]	14.75	2	2	2	2	2	1	2	1	2	2	1	0	0	0	0	2
31	[illegible]	19.5	[illegible]	19.5	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2
32	[illegible]	18.8333333333	[illegible]	18.5	2	2	2	2	1	2	1	0	2	2	2	2	2	2	2	2
33	[illegible]	19.3333333333	[illegible]	18	2	2	2	2	1	1	1	2	2	2	2	2	2	2	2	2
34	[illegible]	20	[illegible]	20	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
35	[illegible]	18.75	[illegible]	17.25	2	2	2	2	2	1	2	1	2	2	2	2	2	2	2	2
36	[illegible]	19	[illegible]	19	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	2
37	[illegible]	17.6666666667	[illegible]	18.5	2	2	2	2	2	2	2	0	2	2	2	2	2	2	0	2
38	[illegible]	18.6666666667	[illegible]	18.5	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	1
39	[illegible]	18.3333333333	[illegible]	18	1	2	2	0	2	2	2	0	2	2	2	2	2	2	2	2
40	[illegible]	18.9166666667	[illegible]	18.25	2	1	2	2	2	1	2	1	2	2	2	2	2	2	1	2

Sheet 1 of 1 | Default | Sum=0 | 100%

Feedback


Avez vous plus appris avec le format Peergrade
qu'avec un mode de correction classique de TPs?
100%
80%
60%
40%
20%
0%
pas du tout
un peu
beaucoup

Feedback


Pensez-vous que les notes que vous avez obtenues étaient:
100%
80%
60%
40%
20%
0%
trop sévères:
vous méritiez plus
fidèles:
vos notes étaient méritées
trop généreuses:
vous méritiez moins

Retours positifs

14% des élèves ne recommandent pas ce système

69% des élèves veulent plus de détails sur les corrections de TPs

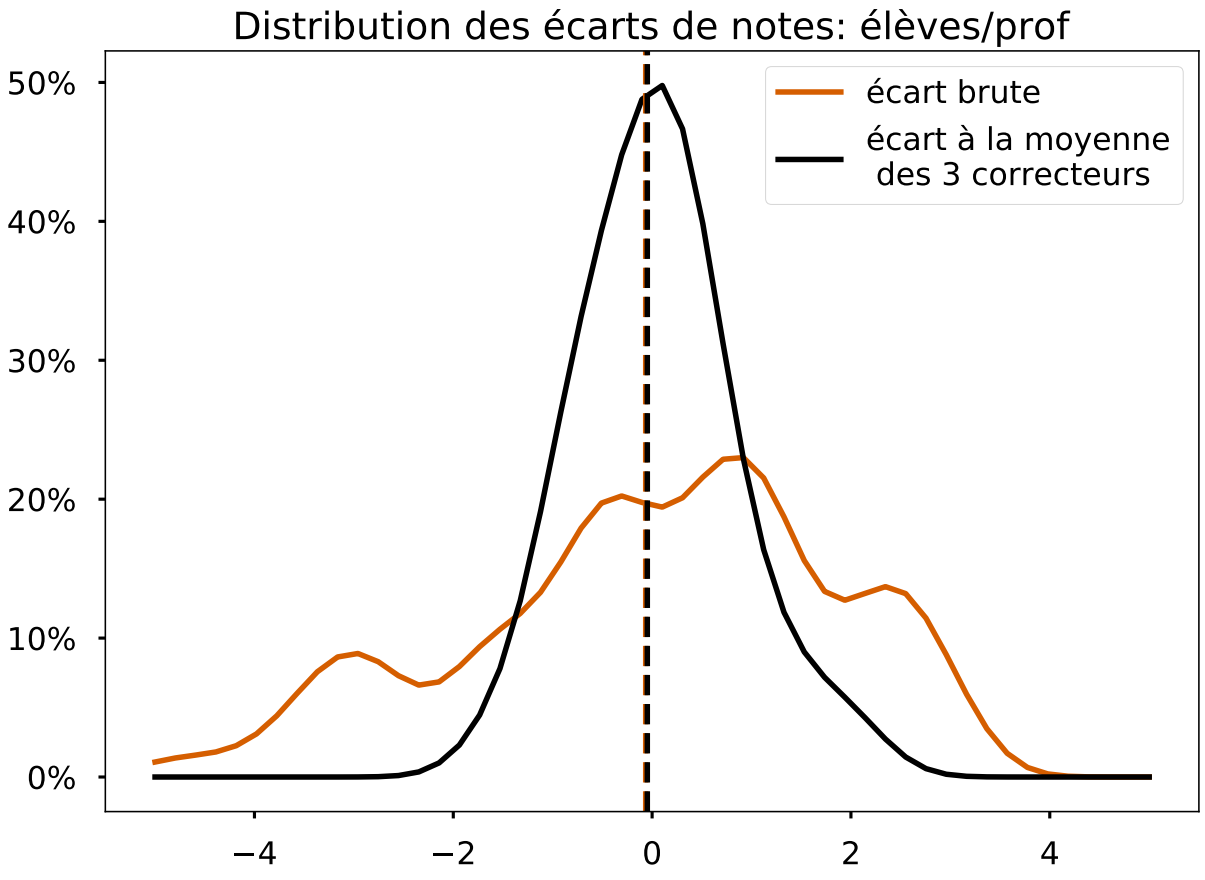
Point important : **facteur d'amélioration**

"L'utilisation de la plateforme permet de se rendre compte du niveau des autres et des bonnes (ou moins bonnes) pratiques de chacun. De plus, se corriger entre nous permet de 'prendre de la hauteur' sur l'énoncé."

"Concernant la notation, le fait de se noter entre élèves était très intéressant car cela permettait de voir les copies des autres et de comprendre en même temps comment l'on pouvait s'améliorer. C'était aussi un exercice pour faire ressortir des remarques constructives et permettre aux autres de s'améliorer."

Mon analyse : le TCL est vrai pour $n = 3$

- 248 copies (TPs)
- 29 super-évaluations ($\approx$ 11.69%) : mécontents/grands écart-type



Principales critiques

- Chronophage (3 copies = beaucoup pour certains)

“il faudrait 2 devoirs max pour une correction en profondeur”

“intéressant si tous les étudiants y consacrent la même quantité d’énergie pour la correction. Pour une copie, je passe quelquefois 1h, pour certains je pense qu’en 10 minutes c’est fait. ”

“pour bien corriger le temps passé à corriger peut parfois être très long (+3h si 3 copies difficiles). Je trouve dans l’ensemble intéressant ce système de notation mais y aurait-il un moyen de réduire ce temps ? (Je n’ai pas de réponse moi-même..)”

“ je pense que les gens, quand ils ne savent pas, ont tendance à dire que c’est faux (normal, nous sommes débutants, mais ca pénalise, si on tente des choses non expliquées en cours, recherchées sur internet par ex). Ou bien, au contraire, si on ne comprend pas, on peut avoir tendance à mettre la meilleure note, dans le doute... ”

Pistes d'améliorations : partager et échanger entre correcteurs

Les correcteurs aimeraient notamment

- **avoir accès aux autres corrections**
- pouvoir **avoir des discussions avec les autres** correcteurs (*cf.* conférences de Machine Learning type NIPS/ICML)

"Mettre en place une discussion anonyme entre les correcteurs si besoin de débattre. Cela augmenterait la plus-value pour tout le monde au niveau de ce que l'on apprend des corrections avec ce système !"

"il serait intéressant de pouvoir voir les notes/commentaires des autres correcteurs afin de se jauger / de se situer en tant que correcteur.

Pistes d'améliorations supplémentaires

- noter les correcteurs
- donner beaucoup (vraiment beaucoup !) de **détails**

"engager les élèves sur une correction juste et sérieuse avec par exemple 10% de la note du TP - faire un rapport de toutes les choses apprises en corrigeant une autre copie"

"+ de détails : permettre aux élèves de comprendre leurs erreurs et ensuite d'être en mesure de corriger les autres. les problèmes sur les notations viennent des ambiguïtés que peut contenir la copie corrigée."

Conclusion et perspectives

- Retour très positifs pour les élèves globalement

Conclusion et perspectives

- Retour très positifs pour les élèves globalement
- Retour très positif pour l'enseignant : expérience de correction beaucoup plus enrichissante et moins monotone

Conclusion et perspectives

- Retour très positifs pour les élèves globalement
- Retour très positif pour l'enseignant : expérience de correction beaucoup plus enrichissante et moins monotone
- nécessite plus de travail qu'une correction classique (en amont !)

Conclusion et perspectives

- Retour très positifs pour les élèves globalement
- Retour très positif pour l'enseignant : expérience de correction beaucoup plus enrichissante et moins monotone
- nécessite plus de travail qu'une correction classique (en amont !)
- améliorations possibles : ergonomie, simplicité, aide en ligne, etc. : bon projet d'élèves (en Python) ?

Conclusion et perspectives

- Retour très positifs pour les élèves globalement
- Retour très positif pour l'enseignant : expérience de correction beaucoup plus enrichissante et moins monotone
- nécessite plus de travail qu'une correction classique (en amont !)
- améliorations possibles : ergonomie, simplicité, aide en ligne, etc. : bon projet d'élèves (en Python) ?
- Extension en cours ⇒ **Peeramid** – avec CentraleSupelec (Co-PI : **V. Galtier**), AAP Paris-Saclay

Classgrade : `https://github.com/classgrade/`

Bonus

“J’ai aimer corriger les autres pour voir différentes façon de faire mais je n’ai pas trop aimé être corrigé”

“Le problème principale de cette notation (outre le temps que ça prends) c’est que les gens qui ont le moins compris ont du mal a comprendre les reponses justes mais differentes de la correction. Pour enlever ces problemes il faudrait que la correction donne systematiquement toutes les variations de resultat/ de figure acceptables. Bref **il faut donner une feuille de guide de notation détaillée** pour la correction qui donne pour chaque question un bareme precis en fonction d’un sous decoupage de la question. c’est ce que j’ai vu sur tous les moocs que j’ai fait et qui avaient des devoirs notés par les pairs et ca marchait plutot bien (enfin mieux que les notes ici je trouve)”