Informe del progreso de TFG, fase 1 (2017.10.04-2017.10-26)

- 1. Trabajo preparatorio:
 - Herramientas de programación:
 - Anaconda Python 3.6
 - Eclipse-Pydev
 - Mirar tutoriales de "Machine Learning"
 - Lectura del código del trabajo anterior (sobretodo el "grade prediction.ipynb")
 - Compresión de la estructura y flujo de ejecución
 - Compresión de las funciones y variables
 - Pasar el código des del "ipynb" al ".py" del Pydev (modularización y coherencia) para poder debuggear el código a profundidad.
 - Reutilización y adaptación del código del trabajo anterior
 - Se ha reutilizado la mayor parte del código que se corresponde a la carga y limpieza de datos. <u>También se aplica un reemplazo al</u> <u>dataframe resultante, de valores NaN por 5.0 de float.</u>
 - Se ha modificado el código reutilizado para que sea más legible y comprensible: estructuración, cambio de nombre de las variables, el añadido de comentario.
 - Se utiliza el editor Pydev de Eclipse para desarrollar el programa.
- 2. Implementación de funcionalidades:

Para esta vez, hay que implementar dos funcionalidades: un ranking de asignaturas por número ordinal y un clasificador binario de suspensos. Para empezar, vamos a tratar el caso de un sólo grado, pues se cargan los datos correspondientes a las calificaciones de los alumnos de "mates".

Dimensión de datos después de la limpieza:
 Se leen los datos des del fichero "qualifications_mates_info.csv", por el nombre se puede deducir que los datos del grado de mates e informática están en un fichero, posteriormente se hará una separación.

| Variable | Descripción | Dimensión |
|---------------|---|-----------|
| qual | Carga inicial | 19621 x 5 |
| qual2process | Eliminación de "id_assig"==NaN | 18383 x 5 |
| filtered_qual | Elección del grado a tratar | 9227 x 5 |
| pivot_year1 | Ordenación de la tabla por nota(value), | 516 x 10 |
| pivot_year2 | id_alumne(index), id_assig(column) | |
| pivot conc | Concatenación de pivot_year1 y pivot_year2; | 231 x 20 |
| | eliminación de filas thresh=11 de NaNs | |

| pivot_conc | Eliminación de filas cuyas sumas de las notas | 221 x 20 |
|------------|---|----------|
| | del segundo año = 0 | |
| df_year1 | pivot_conc[: , :10] | 221 x 10 |
| df_year2 | pivot_conc[: , 10:20] | 221 x 10 |

- Las dos aproximaciones de predicción para ambas funcionalidades
 - Ranking mediante predicción de notas: notas (input) -> predicción
 -> notas (predichas) -> conversión -> ranking (resultado final)
 - Predicción de ranking: notas (input) -> conversión -> ranking (inicial) -> clasificación -> ranking (resultado final)
 - Aprobado/Suspenso mediante predicción de notas: notas (input)
 -> predicción -> notas (predichas) -> conversión -> estado binario (resultado)
 - Clasificación binaria de aprobado/suspenso: notas (input) -> conversión -> estado binario (inicial) -> clasificación -> estado binario (resultado)
- Predictores empleados:

En total tenemos tres casos:

| Casos | Tipo de datos | Predictor |
|--------------------------|---------------|------------------------|
| Predicción de notas | Float64 | RandomForestRegressor |
| Clasificación de ranking | Int | RandomForestClassifier |
| Clasificación de | boolean | RandomForestClassifier |
| suspensos | | |

Ambos predictores son importados des del paquete "sklearn.ensemble".

- Los módulos (ficheros .py) del programa:
 - main.py: carga los datos y lanza en menú.
 - loadData.py: carga los datos y los procesa para que estén en condición para la predicción.
 - randomForest.py: implementa la predicción RandomForest de notas y clasificación de ranking; también tiene algunas funciones de soporte: pasar de notas al ranking, cálculo de puntuaciones.
 - binClassifier.py: implementa la predicción RandomForest de notas y clasificación de suspensos; también tiene algunas funciones de soporte: pasar de notas al estado binario, cálculo de puntuaciones.
- El output de predicciones y comparación de performances:
 - El output de predicciones:

Qp_t vs Qp_p

| | t | | | | | | | | |
|--|--|---|---|--|---|--|--|---|--|
| id_assig | CDDV | ESAL | . GELI | I GRAF | M | NU1 | CID | V GEPI | R \ |
| id_alumne | | | | | | | | | |
| 259 | 5.900000 | 4.550000 | 4.350000 | 7.70 | 4.2000 | 300 5 | .80000 | 0 5.00 | 0 |
| 361 | 5.000000 | 5.000000 | 5.000000 | 0 5.00 | 1.5000 | 900 5 | .00000 | 0 5.00 | 0 |
| 34 | 5.400000 | 6.200000 | 5.750000 | 8.10 | 5.8000 | 300 E | 5.10000 | 0 7.8 | 0 |
| 660 | 6.000000 | 3.750000 | 6.200000 | 0.50 | 5.0000 | 300 Z | 2.10000 | 0 5.00 | 0 |
| 326 | 5.700000 | 5.500000 | 5.200000 | 0 6.30 | 5.3000 | 900 5 | 5.30000 | 0 5.00 | 0 |
| 131 | 7.100000 | 8.000000 | 8.300000 | 9.00 | 5.0000 | 900 G | 5.00000 | 0 7.0 | 0 |
| 486 | 2.050000 | 0.000000 | 3.300000 | 2.65 | 3.1000 | 900 5 | .00000 | 0 5.00 | 0 |
| 679 | 3.500000 | 5.000000 | 5.000000 | 5.90 | 4.0000 | 900 5 | .00000 | 0 5.00 | 0 |
| 74 | 9.000000 | 7.500000 | 8.70000 | 9.10 | 7.0000 | 900 E | 3.60000 | 0 10.0 | 0 |
| 18 | 0.000000 | 0.000000 | 5.000000 | 7.60 | 5.0000 | 900 5 | .00000 | 0 5.00 | 0 |
| < | | | | | | | | | |
| >>> df_pr | edicted v | | | | | | | | |
| id assig | CDDV | ESAL | GELI | GRAF | MNU: | 1 | CIDV | GEF | PR \ |
| o | 5.255000 | 4.785000 | 5.120000 | 7.230 | 4.611667 | 7 4.4 | 160000 | 4.49000 | |
| 1 | 3.275000 | 1.940000 | 2.440000 | 5.260 | 3.850000 | | 335000 | 3.70000 | 30 |
| 2 | 5.580000 | 5.230000 | 5.810000 | 7.470 | 5.200000 | | 923333 | 5.60000 | |
| 3 | 5.030000 | 4.265000 | 5.000000 | 6.680 | 5.250000 | | 080000 | 5.60000 | |
| 4 | 5.490000 | 4.968333 | 5.240000 | 7.240 | 4.40166 | | 765000 | 5.50000 | |
| 5 | 6.095000 | 6.210000 | 6.845000 | 7.350 | 5.660000 | | 940000 | 6.32000 | |
| 6 | 3.450000 | 1.420000 | 2.275000 | 5.510 | 2.230000 | | 963333 | 1.03000 | |
| 7 | 3.710000 | 3.040000 | 3.205000 | 5.150 | 4.385000 | | 334167 | 3.95000 | |
| 8 | 7.260000 | 6.870000 | 7.380000 | 8.460 | 5.770000 | | 580000 | 6.48000 | |
| 9 | 4.240000 | 3.146667 | 4.400000 | 5.525 | 4.430000 | | 115000 | 3.72500 | |
| 10 | 6.430000 | 5.745000 | 6.330000 | 8.220 | 5.615000 | | 130000 | 6.52500 | |
| < | | | | | | | | | |
| | s Qp_rp | | | | | | | | |
| <pre>>>> y_te id_assig</pre> | st CDDV | ESAL GEL | I GRAF | MNU1 | CIDV G | EPR | HIMA | MMSD | ТОРО |
| >>> y_te id_assig id_alumn | st CDDV I | | | | | | | | |
| >>> y_te id_assig id_alumn 259 | cDDV I | 7 | 8 1 | 9 | 4 | 6 | 2 | 5 | 10 |
| >>> y_te id_assig id_alumn 259 361 | cDDV e | 7 2 | 8 1 3 4 | 9 10 | 4 5 | 6 | 2 7 | 5 8 | 10 9 |
| >>> y_te id_assig id_alumn 259 361 | cDDV e 3 1 9 | 7 2 4 | 8 1 3 4 7 1 | 9 | 4 | 6 | 2 | 5 | 10 |
| >>> y_te id_assig id_alumn 259 361 34 | cDDV e | 7 2 | 8 1 3 4 | 9 10 | 4 5 | 6 | 2 7 | 5 8 | 10 9 |
| >>> y_te id_assig id_alumn 259 361 34 660 | cDDV e 3 1 9 | 7 2 4 | 8 1 3 4 7 1 | 9 10 6 | 4 5 5 | 6 6 2 | 2 7 3 | 5 8 8 | 10 9 10 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 | st CDDV (e 3 1 9 5 | 7 2 4 9 | 8 1 3 4 7 1 3 1 | 9 10 6 7 | 4 5 5 10 | 6 6 2 8 | 2 7 3 2 | 5 8 8 4 | 10 9 10 6 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 | cDDV (e 3 1 9 5 4 | 7 2 4 9 | 8 1 3 4 7 1 3 1 8 2 | 9 10 6 7 6 | 4 5 5 10 7 | 6 6 2 8 9 | 2 7 3 2 1 | 5 8 8 4 | 10 9 10 6 10 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 | c CDDV e 3 1 9 5 4 5 | 7 2 4 9 5 4 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 | 9 10 6 7 6 10 7 | 4 5 5 10 7 8 | 6 6 2 8 9 6 | 2 7 3 2 1 9 | 5 8 8 4 3 7 4 | 10 9 10 6 10 2 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 679 | st CDDV e 3 1 9 5 4 5 9 9 | 7 2 4 9 5 4 10 2 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 3 1 | 9 10 6 7 6 10 7 8 | 4 5 5 10 7 8 1 | 6 6 2 8 9 6 2 | 2 7 3 2 1 9 3 6 | 5 8 8 4 3 7 4 10 | 10 9 10 6 10 2 5 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 679 74 | c CDDV e 3 1 9 5 4 5 9 9 3 3 | 7 2 4 9 5 4 10 2 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 3 1 5 2 | 9 10 6 7 6 10 7 8 | 4 5 5 10 7 8 1 4 | 6 6 2 8 9 6 2 5 | 2 7 3 2 1 9 3 6 7 | 5 8 8 4 3 7 4 10 4 | 10 9 10 6 10 2 5 7 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 679 74 | st CDDV e 3 1 9 5 4 5 9 9 | 7 2 4 9 5 4 10 2 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 3 1 | 9 10 6 7 6 10 7 8 | 4 5 5 10 7 8 1 | 6 6 2 8 9 6 2 | 2 7 3 2 1 9 3 6 | 5 8 8 4 3 7 4 10 | 10 9 10 6 10 2 5 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 679 74 | c CDDV e 3 1 9 5 4 5 9 9 3 9 9 | 7 2 4 9 5 4 10 2 9 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 3 1 5 2 | 9 10 6 7 6 10 7 8 | 4 5 5 10 7 8 1 4 | 6 6 2 8 9 6 2 5 | 2 7 3 2 1 9 3 6 7 | 5 8 8 4 3 7 4 10 4 | 10 9 10 6 10 2 5 7 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 679 74 18 < | st | 7 2 4 9 5 4 10 2 9 10 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 3 1 5 2 2 1 | 9 10 6 7 6 10 7 8 10 3 | 4 5 5 10 7 8 1 4 6 4 | 6 6 2 8 9 6 2 5 1 | 2 7 3 2 1 9 3 6 7 6 | 5 8 4 3 7 4 10 4 7 | 10 9 10 6 10 2 5 7 8 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 679 74 18 < | st | 7 2 4 9 5 4 10 2 9 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 3 1 5 2 2 1 | 9 10 6 7 6 10 7 8 10 3 | 4 5 5 10 7 8 1 4 6 4 | 6 6 2 8 9 6 2 5 | 2 7 3 2 1 9 3 6 7 | 5 8 8 4 3 7 4 10 4 7 | 10 9 10 6 10 2 5 7 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 679 74 18 days | st | 7 2 4 9 5 4 10 2 9 10 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 3 1 5 2 2 1 | 9 10 6 7 6 10 7 8 10 3 | 4 5 5 10 7 8 1 4 6 4 | 6 6 2 8 9 6 2 5 1 | 2 7 3 2 1 9 3 6 7 6 | 5 8 4 3 7 4 10 4 7 | 10 9 10 6 10 2 5 7 8 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 679 74 18 >>>> df_p id_assig 0 | st | 7 2 4 9 5 4 10 2 9 10 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 3 1 5 2 2 1 | 9 10 6 7 6 10 7 8 10 3 | 4 5 5 10 7 8 1 4 6 4 | 6 6 2 8 9 6 2 5 1 5 | 2 7 3 2 1 9 3 6 7 6 | 5 8 8 4 3 7 4 10 4 7 | 10 9 10 6 10 2 5 7 8 8 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 679 74 18 < | st CDDV e 3 1 9 5 4 5 9 9 3 3 9 Oredicted 5 CDDV E 4 5 5 | 7 2 4 9 5 4 10 2 9 10 2 9 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 3 1 5 2 2 1 I GRAF 5 1 8 1 | 9 10 6 7 6 10 7 8 10 3 MNU1 7 3 | 4 5 5 10 7 8 1 4 6 4 6 4 CIDV (| 6 6 2 8 9 6 2 5 1 5 | 2 7 3 2 1 9 3 6 7 6 7 6 | 5 8 8 4 3 7 4 10 4 7 | 10 9 10 6 10 2 5 7 8 8 TOPO 10 6 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 6679 74 18 < | st CDDV e 3 1 9 5 4 5 9 9 3 3 9 credicted CDDV E 5 6 | 7 2 4 9 5 4 10 2 9 10 ESAL GEL 6 9 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 3 1 5 2 2 1 I GRAF 5 1 8 1 | 9 10 6 7 6 10 7 8 10 3 MNU1 7 3 | 4 5 5 10 7 8 1 4 6 4 6 4 CIDV (| 6 6 2 8 9 6 2 5 1 5 | 2 7 3 2 1 9 3 6 7 6 7 6 HIMA 2 2 | 5 8 8 4 3 7 4 10 4 7 MMSD 3 10 3 | 10 9 10 6 10 2 5 7 8 8 TOPO 10 6 7 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 679 74 18 < | st CDDV e 3 1 9 5 4 5 9 9 3 3 9 credicted 4 5 6 7 | 7 2 4 9 5 4 10 2 9 10 ESAL GEL 6 9 8 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 3 1 5 2 2 1 I GRAF 5 1 8 1 4 1 | 9 10 6 7 6 10 7 8 10 3 MNU1 7 3 | 4 5 5 10 7 8 1 4 6 4 6 4 CIDV (9 7 10 6 | 6 6 2 8 9 6 2 5 1 5 | 2 7 3 2 1 9 3 6 7 6 7 6 HIMA 2 2 2 3 | 5 8 8 4 3 7 4 10 4 7 MMSD 3 10 3 | 10 9 10 6 10 2 5 7 8 8 TOPO 10 6 7 9 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 679 74 18 < | st CDDV e 3 1 9 5 4 5 9 9 3 3 9 coredicted 7 5 6 7 5 | 7 2 4 9 5 4 10 2 9 10 2 9 10 5 5 4 10 7 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 3 1 5 2 2 1 I GRAF 5 1 8 1 4 1 8 1 6 1 | 9 10 6 7 6 10 7 8 10 3 MNU1 7 3 9 5 | 4 5 5 10 7 8 1 4 6 4 4 CIDV (9 7 10 6 8 | 6 6 2 8 9 6 2 5 1 5 | 2 7 3 2 1 9 3 6 7 6 HIMA 2 2 2 2 3 | 5 8 8 4 3 7 4 10 4 7 MMSD 3 10 3 4 3 | 10 9 10 6 10 2 5 7 8 8 TOPO 10 6 7 9 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 679 74 18 < | st CDDV e | 7 2 4 9 5 4 10 2 9 10 2 9 10 5 5 4 10 7 6 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 3 1 5 2 2 1 I GRAF 5 1 8 1 4 1 8 1 6 1 2 1 | 9 10 6 7 6 10 7 8 10 3 MNU1 7 3 9 5 9 | 4 5 5 10 7 8 1 4 6 4 4 CCIDV (9 7 10 6 8 8 8 | 6 6 2 8 9 6 2 5 1 5 5 5 5 4 5 2 4 5 | 2 7 3 2 1 9 3 6 7 6 HIMA 2 2 2 2 3 4 | 5 8 8 4 3 7 4 10 4 7 MMSD 3 10 3 4 3 3 | 10 9 10 6 10 2 5 7 8 8 TOPO 10 6 7 9 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 679 74 18 >>> df_p id_assig 0 1 2 3 4 5 6 | st CDDV e 3 1 9 5 4 5 9 9 3 9 Predicted CDDV E 5 6 7 5 7 3 | 7 2 4 9 5 4 10 2 9 10 2 9 10 5 5 4 10 7 6 9 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 3 1 5 2 2 1 I GRAF 5 1 8 1 4 1 8 1 6 1 2 1 5 1 | 9 10 6 7 6 10 7 8 10 3 MNU1 7 3 9 5 9 | 4 5 5 10 7 8 1 4 6 4 4 CIDV 0 9 7 10 6 8 8 7 | 6 6 2 8 9 6 2 5 1 5 5 5 4 5 2 4 5 | 2 7 3 2 1 9 3 6 7 6 HIMA 2 2 2 2 3 2 4 2 | 5 8 8 4 3 7 4 10 4 7 MMSD 3 10 3 4 3 3 4 3 | 10 9 10 6 10 2 5 7 8 8 TOPO 10 6 7 9 10 9 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 679 74 18 < | st CDDV e 3 1 9 5 4 5 9 9 3 9 credicted CDDV E 5 6 7 5 7 3 5 | 7 2 4 9 5 4 10 2 9 10 2 9 10 5 5 4 10 7 6 9 8 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 3 1 5 2 2 1 I GRAF 5 1 8 1 4 1 8 1 6 1 2 1 5 1 | 9 10 6 7 6 10 7 8 10 3 MNU1 7 3 9 5 9 10 6 3 | 4 5 5 10 7 8 1 4 6 4 4 CIDV (9 7 10 6 8 8 7 9 | 6 6 2 8 9 6 2 5 1 5 5 5 5 4 5 2 4 5 | 2 7 3 2 1 9 3 6 7 6 HIMA 2 2 2 2 3 2 4 2 | 5 8 8 4 3 7 4 10 4 7 MMSD 3 10 3 4 3 3 | 10 9 10 6 10 2 5 7 8 8 8 TOPO 10 6 7 9 10 9 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 679 74 18 >>> df_p id_assig 0 1 2 3 4 5 6 | st CDDV e 3 1 9 5 4 5 9 9 3 9 Predicted CDDV E 5 6 7 5 7 3 | 7 2 4 9 5 4 10 2 9 10 2 9 10 5 5 4 10 7 6 9 8 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 3 1 5 2 2 1 I GRAF 5 1 8 1 4 1 8 1 6 1 2 1 5 1 | 9 10 6 7 6 10 7 8 10 3 MNU1 7 3 9 5 9 | 4 5 5 10 7 8 1 4 6 4 4 CIDV 0 9 7 10 6 8 8 7 | 6 6 2 8 9 6 2 5 1 5 5 5 4 5 2 4 5 | 2 7 3 2 1 9 3 6 7 6 HIMA 2 2 2 2 3 2 4 2 | 5 8 8 4 3 7 4 10 4 7 MMSD 3 10 3 4 3 3 4 3 | 10 9 10 6 10 2 5 7 8 8 TOPO 10 6 7 9 10 9 |
| >>> y_te id_assig id_alumn 259 361 34 660 326 131 486 679 74 18 < | st CDDV e 3 1 9 5 4 5 9 9 3 9 credicted CDDV E 5 6 7 5 7 3 5 | 7 2 4 9 5 4 10 2 9 10 2 9 10 5 5 4 10 7 6 9 8 10 7 6 | 8 1 3 4 7 1 3 1 8 2 3 1 6 8 3 1 5 2 2 1 I GRAF 5 1 8 1 4 1 8 1 6 1 2 1 5 1 | 9 10 6 7 6 10 7 8 10 3 MNU1 7 3 9 5 9 10 6 3 | 4 5 5 10 7 8 1 4 6 4 4 CIDV (9 7 10 6 8 8 7 9 | 6 6 2 8 9 6 2 5 1 5 5 5 4 5 2 4 5 4 5 4 5 4 6 6 6 7 7 8 7 8 7 8 7 8 7 8 7 8 7 8 7 8 | 2 7 3 2 1 9 3 6 7 6 HIMA 2 2 2 2 3 2 4 2 | 5 8 8 4 3 7 4 10 4 7 7 MMSD 3 10 3 4 3 3 4 6 | 10 9 10 6 10 2 5 7 8 8 8 TOPO 10 6 7 9 10 9 |

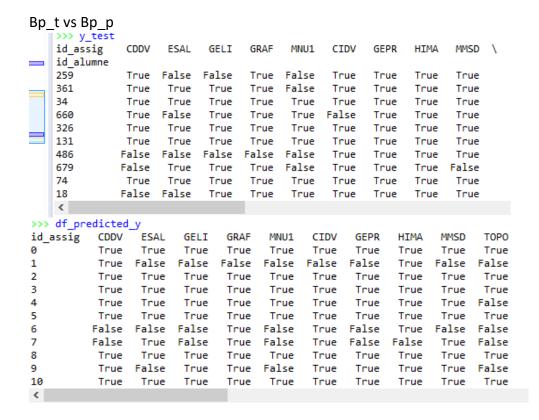
| Rp_t vs F | | | | | | | | | | |
|-----------------------|--------|------|------|------|------|------|------|------|------|------|
| id_assig id_alumne | CDDV | ESAL | GELI | GRAF | MNU1 | CIDV | GEPR | HIMA | MMSD | TOPO |
| 259 | 3 | 7 | 8 | 1 | 9 | 4 | 6 | 2 | 5 | 10 |
| 361 | 1 | 2 | 3 | 4 | 10 | 5 | 6 | 7 | 8 | 9 |
| 34 | 9 | 4 | 7 | 1 | 6 | 5 | 2 | 3 | 8 | 10 |
| 660 | 5 | 9 | 3 | 1 | 7 | 10 | 8 | 2 | 4 | 6 |
| 326 | 4 | 5 | 8 | 2 | 6 | 7 | 9 | 1 | 3 | 10 |
| 131 | 5 | 4 | 3 | 1 | 10 | 8 | 6 | 9 | 7 | 2 |
| 486 | 9 | 10 | 6 | 8 | 7 | 1 | 2 | 3 | 4 | 5 |
| 679 | 9 | 2 | 3 | 1 | 8 | 4 | 5 | 6 | 10 | 7 |
| 74 | 3 | 9 | 5 | 2 | 10 | 6 | 1 | 7 | 4 | 8 |
| 18 | 9 | 10 | 2 | 1 | 3 | 4 | 5 | 6 | 7 | 8 |
| < | | | | | | | | | | |
| >>> df_pr | edicte | d_y | | | | | | | | |
| id_assig | CDDV | ESAL | GELI | GRAF | MNU1 | CIDV | GEPR | HIMA | MMSD | TOPO |
| 0 | 1 | 9 | 8 | 3 | 4 | 6 | 1 | 3 | 2 | 9 |
| 1 | 3 | 2 | 3 | 4 | 5 | 6 | 10 | 1 | 6 | 7 |
| 2 | 4 | 9 | 8 | 1 | 6 | 6 | 1 | 3 | 5 | 9 |
| 3 | 3 | 9 | 7 | 1 | 10 | 5 | 8 | 1 | 4 | 9 |
| 4 | 2 | 6 | 6 | 1 | 2 | 9 | 7 | 4 | 4 | 10 |
| 5 | 3 | 9 | 7 | 1 | 8 | 6 | 1 | 8 | 2 | 10 |
| 6 | 2 | 2 | 6 | 2 | 5 | 4 | 4 | 1 | 3 | 10 |
| 7 | 5 | 9 | 4 | 2 | 8 | 2 | 1 | 8 | 9 | 10 |
| 8 | 6 | 9 | 10 | 1 | 9 | 6 | 1 | 1 | 5 | 8 |
| 9 | 6 | 10 | 9 | 1 | 4 | 4 | 3 | 2 | 7 | 6 |
| 10 | 8 | 8 | 3 | 1 | 10 | 9 | 5 | 1 | 2 | 10 |
| < | | | | | | | | | | |

Qp_bt vs Qp_bp

| >>> y_test | | | | | | | | | | |
|------------|-------|-------|-------|-------|-------|-------|------|------|-------|---|
| id_assig | CDDV | ESAL | GELI | GRAF | MNU1 | CIDV | GEPR | HIMA | MMSD | \ |
| id_alumne | | | | | | | | | | |
| 259 | True | False | False | True | False | True | True | True | True | |
| 361 | True | True | True | True | False | True | True | True | True | |
| 34 | True | True | True | True | True | True | True | True | True | |
| 660 | True | False | True | True | True | False | True | True | True | |
| 326 | True | True | True | True | True | True | True | True | True | |
| 131 | True | True | True | True | True | True | True | True | True | |
| 486 | False | False | False | False | False | True | True | True | True | |
| 679 | False | True | True | True | False | True | True | True | False | |
| 74 | True | True | True | True | True | True | True | True | True | |
| 18 | False | False | True | True | True | True | True | True | True | |
| < | | | | | | | | | | |

>>> df_predicted_y

| id_assig | CDDV | ESAL | GELI | GRAF | MNU1 | CIDV | GEPR | HIMA | MMSD | TOPO |
|----------|-------|-------|-------|------|-------|-------|-------|-------|-------|-------|
| 0 | True | False | True | True | False | False | False | True | True | False |
| 1 | False | False | False | True | False | False | False | False | False | False |
| 2 | True | True | True | True | True | False | True | True | True | True |
| 3 | True | False | True | True | True | True | True | True | True | False |
| 4 | True | False | True | True | False | False | True | True | True | False |
| 5 | True | True | True | True | True | True | True | True | True | True |
| 6 | False | False | False | True | False | False | False | False | False | False |
| 7 | False | False | False | True | False | False | False | True | False | False |
| 8 | True | True | True | True | True | True | True | True | True | True |
| 9 | False | False | False | True | False | False | False | False | False | False |
| 10 | True | True | True | True | True | True | True | True | True | False |
| < | | | | | | | | | | |



Comparación de performances:

| Tipo de predicción | Medida de Validación | Puntuación |
|-----------------------|----------------------------------|------------|
| Predicción de nota | Clf.score | 0.3055 |
| Predicción de nota | cross_validation.cross_val_score | 0.1759 |
| Predicción de | metrics.accuracy_score | 0.1375 |
| nota->ranking | | |
| Predicción de ranking | metrics.accuracy_score | 0.1393 |
| Predicción de | metrics.accuracy_score | 0.6804 |
| notas->binario | | |
| Clasificación binaria | metrics.accuracy_score | 0.7196 |

3. Tareas para la próxima cita:

- Apuntes en la cita
- Procesar datos de los grados "Informática" y "Derecho"
- Mejorar las precisiones de predictores (GridSearch), usar diferentes métricas.
- Etc

4. Dudas:

- ranking prediction: repeated class
- similarity/score computing type
- which approach is selected?
- guardar cada link?