

## Evaluación Final – Caso HESPE

### Contexto del Problema

Usted ha sido contratado como parte de un equipo de análisis del rendimiento de estudiantes en la facultad de ingeniería y de ciencias de la universidad de Wisconsin, para participar en el proyecto HESPE (Higher Education Student Performance Evaluation). Deberá, por lo tanto, poner a prueba todas las habilidades y conocimientos adquiridos a lo largo de este curso.

Su primer desafío consiste en predecir el resultado de los estudiantes al momento de cursar una asignatura en la universidad. Para esto, se ha construido un set de datos con 32 características, en donde las 10 primeras de ellas corresponden a información personal, de la 11 a la 16 corresponden a preguntas familiares, y el resto corresponde a hábitos de estudio. Se recolectaron 145 instancias para este estudio, que fue llevado a cabo durante 2019. Cada instancia corresponde a lo contestado por un estudiante en una encuesta.

A continuación, se describe el set de datos:

1. student\_id
2. age - Student Age (1: 18-21, 2: 22-25, 3: above 26)
3. sex - Sex (1: female, 2: male)
4. graduated\_h\_school\_type - Graduated high-school type: (1: private, 2: state, 3: other)
5. scholarship\_type - Scholarship type: (1: None, 2: 25%, 3: 50%, 4: 75%, 5: Full)
6. additional\_work - Additional work: (1: Yes, 2: No)
7. activity -Regular artistic or sports activity: (1: Yes, 2: No)
8. partner - Do you have a partner: (1: Yes, 2: No)
9. total\_salary - Total salary if available (1: USD 135-200, 2: USD 201-270, 3: USD 271-340, 4: USD 341-410, 5: above 410)
10. transport - Transportation to the university: (1: Bus, 2: Private car/taxi, 3: bicycle, 4: Other)
11. accomodation - Accommodation type in Cyprus: (1: rental, 2: dormitory, 3: with family, 4: Other)
12. mother\_ed - Mother's education: (1: primary school, 2: secondary school, 3: high school, 4: university, 5: MSc., 6: Ph.D.)
13. farther\_ed - Father's education: (1: primary school, 2: secondary school, 3: high school, 4: university, 5: MSc., 6: Ph.D.)
14. siblings - Number of sisters/brothers (if available): (1: 1, 2: 2, 3: 3, 4: 4, 5: 5 or above)
15. parental\_status - Parental status: (1: married, 2: divorced, 3: died - one of them or both)
16. mother\_occup - Mother's occupation: (1: retired, 2: housewife, 3: government officer, 4: private sector employee, 5: self-employment, 6: other)
17. father\_occup - Father's occupation: (1: retired, 2: government officer, 3: private sector employee, 4: self-employment, 5: other)
18. weekly\_study\_hours - Weekly study hours: (1: None, 2: <5 hours, 3: 6-10 hours, 4: 11-20 hours, 5: more than 20 hours)
19. reading\_non\_scientific - Reading frequency (non-scientific books/journals): (1: None, 2: Sometimes, 3: Often)

20. reading\_scientific - Reading frequency (scientific books/journals): (1: None, 2: Sometimes, 3: Often)
21. attendance\_seminars\_dep - Attendance to the seminars/conferences related to the department: (1: Yes, 2: No)
22. impact\_of\_projects - Impact of your projects/activities on your success: (1: positive, 2: negative, 3: neutral)
23. attendances\_classes - Attendance to classes (1: always, 2: sometimes, 3: never)
24. preparation\_midterm\_company - Preparation to midterm exams 1: (1: alone, 2: with friends, 3: not applicable)
25. preparation\_midterm\_time - Preparation to midterm exams 2: (1: closest date to the exam, 2: regularly during the semester, 3: never)
26. taking\_notes - Taking notes in classes: (1: never, 2: sometimes, 3: always)
27. listening - Listening in classes: (1: never, 2: sometimes, 3: always)
28. discussion\_improves\_interest - Discussion improves my interest and success in the course: (1: never, 2: sometimes, 3: always)
29. flip\_classrom - Flip-classroom: (1: not useful, 2: useful, 3: not applicable)
30. grade\_previous - Cumulative grade point average in the last semester (/4.00): (1: <2.00, 2: 2.00-2.49, 3: 2.50-2.99, 4: 3.00-3.49, 5: above 3.49)
31. grade\_expected - Expected Cumulative grade point average in the graduation (/4.00): (1: <2.00, 2: 2.00-2.49, 3: 2.50-2.99, 4: 3.00-3.49, 5: above 3.49)
32. course\_id
33. grade - OUTPUT Grade (0: Fail, 1: DD, 2: DC, 3: CC, 4: CB, 5: BB, 6: BA, 7: AA)

Su objetivo en este proyecto es elaborar un modelo de aprendizaje de máquina que permita realizar predicciones sobre el resultado académico (variable “grade”).

## Entregable

El entregable, al menos, debe considerar lo siguiente:

- Notebook Jupyter, ordenado, documentado y reproducible con su análisis.
- Planteamiento del problema y descripción de los datos y características
- Análisis exploratorio de datos
- Elaboración de un modelo supervisado, utilizando los conceptos, técnicas y herramientas vistos en clase que sean de utilidad para llevar a cabo este producto.
- Conclusiones de su análisis

Recuerde que Usted, como científico de datos, debe aplicar criterio para seleccionar los conceptos, técnicas y herramientas que requiera para llevar a cabo este propósito, tomando como base los conceptos entregados a lo largo de este módulo y los adquiridos en los módulos anteriores.

Recuerde que el contexto del problema es un estudio científico, por lo cual deberá considerar, dentro de otras cosas, orden, análisis y reproducibilidad en su análisis.

Recuerde, además, que la elaboración de conclusiones es vital en un estudio de estas características.

## Instrucciones

Considere las siguientes instrucciones para el desarrollo del trabajo:

- El trabajo es individual.
- Cualquier violación al código de ética podría significar la nota mínima en la evaluación.
- El plazo de entrega es de 7 días a partir de la entrega de este caso.
- Su trabajo deberá subirlo a la plataforma Moodle.
- Deberá entregar todos los códigos fuentes, notebooks y documentos que sean requeridos