

Predicción de calificaciones orientada a futuras adaptaciones



*Pablo García Fernández
Alejandro Luque Villegas
Javier Orbis Ramírez
Ignacio Sánchez Santatecla*

Índice

1. Introducción	(pág. 3)
2. Limpieza de datos	(pág. 7)
3. Experimentación y resultados	(pág. 9)
4. Conclusiones	(pág. 10)
5. Referencias	(pág. 11)
6. Anexo	(pág. 13)

Introducción

El trabajo de un profesor es un trabajo difícil, pues son los encargados de que el día de mañana sus pequeños aprendices sean personas que puedan aportar a la sociedad con su conocimiento y trabajo, pero al mismo tiempo también acaban teniendo un trabajo en valores debido a la gran cantidad de horas que conviven con sus pupilos.

Es por todo esto que la etapa escolar es bastante definitoria, y aunque haya opiniones diversas, al final esta etapa se superará y quedará marcada por el rendimiento que tenga el alumno en ella. Esto vuelve aún más complejo el trabajo de un profesor, pues debe ser capaz de hacer que sus alumnos tengan un buen rendimiento teniendo en cuenta que cada uno es distinto, y que cada uno tiene unas situaciones particulares. Ya no es algo solamente lógico, sino que también se ha notado que en muchas ocasiones los profesores no son capaces de prestar atención a aquel que más lo necesita, para poner especial empeño y adaptaciones para que este consiga el éxito en la etapa escolar.

Este estudio trata de echar una mano a estos profesores que no saben identificar a tiempo a aquellos alumnos que van a acabar necesitando más ayuda. Vamos a realizar un análisis de datos de una encuesta realizada en Bangladesh a ciertos alumnos. Esta encuesta contiene ciertos valores que nos parecen interesantes de revisar ya que pensamos que afectan o acabarán afectando al rendimiento escolar de un alumno. Estos valores son:

- Género del alumno
- Edad del alumno
- Horas de estudio al día
- Modalidad de aprendizaje
- Uso de las redes sociales
- Asistencia al aula
- ¿Con quién vive el alumno?
- Nota media anterior
- Ingresos mensuales de la familia

La explicación de porqué estos valores nos parecen interesantes la desarrollaremos posteriormente.

El objetivo final de este análisis será la realización de un sistema predictivo, que mediante estos valores nos permitirá saber qué alumno va a acabar necesitando más adaptaciones, ayuda y/o atención.

Género del alumno

De manera generalizada y durante mucho tiempo se ha indicado que los chicos son peores estudiantes que las chicas. Bueno, más correctamente lo que ocurre es que los chicos suelen sacar peores notas. Se ha demostrado que hay asignaturas que se les da mejor o que en ciertas áreas del conocimiento retienen mejor los conocimientos, pero aún así en las calificaciones las chicas acaban “ganando”. Esto no se trata de un tema de capacidad, sino que según estudios y profesores los chicos son menos estudiosos y se esfuerzan menos.

¿Habrá peores vaticinios para los alumnos chicos?

Edad del alumno

El dataset que hemos obtenido está formado por respuestas a un formulario hecho a universitarios. En esta etapa educativa no es raro ver todo tipos de mezclas, entre ellas la más llamativa que no se produce de manera normal en otras etapas, y es la mezcla entre edades.

Estudios demuestran que las calificaciones de los estudiantes de menor edad suelen ser menores que las de mayor edad. Comentan que esto suele deberse a una madurez mental y emocional.

¿Afectará la supuesta inmadurez mental de los estudiantes menores a sus notas?

Horas de estudio al día

Este quizás es uno de los puntos más complejos. Esto se debe a que no hace falta ni decirlo, pero es bien sabido por todos que cada persona es única. Esto influye a las relaciones humanas, gustos y a otros muchos ámbitos de la vida, en los cuales podemos incluir el estudio. Por ejemplo hay alumnos con mejor memoria, mejores para ciertas dinámicas, mejores para ciertas asignaturas... Lo que implica que las horas de estudio sean algo bastante volátil y dependiente de la asignatura y el alumno.

Sin embargo, también por consecuencia lógica, cuanto más estudie un alumno, más posibilidades tiene de aprenderse la lección. Ya no solo es un tema de lógica, sino que diversos estudios lo han respaldado.

¿Supondrá una mayor cantidad de horas de estudio una mejora para las calificaciones del estudiante? ¿O su influencia para la nota dependerá de otros factores?

Modalidad de aprendizaje

Desde la pandemia del Covid-19 hemos visto cómo ha evolucionado muchísimo el sector educativo. Uno de los factores más influyentes en el mismo fue el uso de la modalidad online. Esto conlleva unas ventajas y unas desventajas. Es una metodología que nos permite que se aprenda más en menos tiempo, estando disponible las 24h, sin embargo se pierde la experiencia del aula con el profesor, es más costoso si no se es disciplinado y no contiene interacción social (bastante importante en el proceso de aprendizaje).

¿Qué modalidad de aprendizaje beneficiará más a los alumnos?

Uso de las redes sociales

Este puede ser uno de los puntos que más debate genera en padres y profesores en los últimos años. Desde la aparición del Smartphone en la vida cotidiana del ser humano este instrumento ha acabado por ser una especie de prótesis.

El motivo de este debate se debe a que este elemento nos conlleva una serie de ventajas, pero también una serie de desventajas. Hay quienes destacan que este puede ser una gran herramienta para usar en las aulas. De hecho la Unesco desarrolla, en el Informe de

Seguimiento de la Educación en el Mundo, que la tecnología puede mejorar la calidad educativa, al ofrecer acceso a recursos educativos, fomentar la colaboración y el aprendizaje autónomo, entre otras ventajas.

Sin embargo, como comentamos anteriormente, el uso de los smartphones tiene también sus desventajas. Uno de los argumentos en contra de que los alumnos usen los smartphones en ámbito escolar y/o en su tiempo libre es el del uso de las redes sociales, ya que el uso de estas va fuertemente ligado con el uso de los smartphones. Mucha gente lo presenta como una distracción para el alumno. De hecho un estudio realizado por unos profesores de la Universidad de Kent en Ohio relacionaron de manera negativa el uso de horas en Facebook con la nota media de los alumnos.

¿Afectará negativamente el uso de las redes sociales en los alumnos?

Asistencia al aula

La falta de asistencia a las aulas siempre ha sido un problema. Aquellos alumnos que acaban faltando a clases ven como su nota disminuye. La falta de asistencia durante varios días enteros lleva a grandes disminuciones en las calificaciones, hasta el punto de suponer una distancia de hasta dos años de escolarización en algunas asignaturas. Es por tanto un problema que preocupa a los profesores y padres, pues afecta de manera negativa y directa sobre sus alumnos e hijos.

¿Cómo acabará afectando las “pellas” a los alumnos que las realizan?

¿Con quién vive el alumno?

Si es verdad que esto depende mucho del alumno, como comentamos en el apartado de horas de estudio, pues dependerá de la disciplina, la situación familiar, si trabaja o no mientras realiza sus estudios... Sin embargo, se han identificado ventajas y desventajas a la hora de vivir solo y vivir en familia.

¿Serán los alumnos que estudian más capaces de aprobar viviendo solos? ¿Estudian menos los alumnos que viven con su familia?

Nota media anterior

Las notas que los alumnos van sacando van definiendo su futuro. De hecho aquí se podría aplicar la ley física de la inercia. Es decir, las notas que se van sacando las sacarán en un futuro. Los motivos por el cual esto ocurre en los estudiantes con buenas notas es porque estos llegan a comprender lo que estudian. En el caso de los alumnos que obtienen notas más bajas tienden a seguir con esa nota, ya sea por abandonos de la asignatura o temas de desmotivación o similares.

¿Cómo terminará afectando a futuro el hecho de que los alumnos saquen malas notas?

Ingresos mensuales de la familia

Los niños de familias más pobres siempre han sido estigmatizados con el éxito académico. El punto es que cuando alguno acaba en su etapa adulta teniendo éxito esa historia se

destaca especialmente, debido a lo peculiar del caso. Todo esto aparte de venir de una idea popular es una cosa que se ha probado en distintos estudios.

¿Las familias con menor renta tendrán posibilidades de obtener el éxito académico?

Limpieza de datos

Como comentamos anteriormente en la introducción, el estudio que estamos realizando está basado en una encuesta realizada a alumnos de Bangladesh con los valores que también comentamos anteriormente. Previo a la experimentación hemos tenido que limpiarlos para poder hacer un estándar. A continuación explicaremos cómo hemos hecho esta limpieza y que lo ha motivado.

Primeramente escogimos las columnas que nos resultaban más interesantes, ya mencionadas anteriormente y con su explicación de porque fueron elegidas. Acto seguido distinguimos aquellas columnas que necesitaban una limpieza y las empezamos a limpiar.

Estandarización del Age

El problema que existía en el dataset con la columna Age era que algunos de los que habían respondido la encuesta se les había colado un carácter de más o indicaban los meses que tenían o similar. Estos valores los identificamos haciendo uso de la función `unique()`. Decidimos que lo más interesante para el estudio era quedarnos solamente con los años, así que procedimos a transformar aquellos datos que no cumplían nuestra condición y los cambiamos a numéricos.

Estandarización de tiempos

Hemos realizado una estandarización de los tiempos creando nuestra propia función para aplicarla en el tiempo de estudio y el tiempo que pasan los alumnos en las redes sociales.

Limpiamos el dataset de manera que el formato final que quedaba eran solamente las horas.

Estandarización de Attendance

La estandarización del attendance consiste en una función que limpia la cadena de caracteres con una expresión regular. Luego lo transforma en un número y finalmente comprueba que queda un valor válido. El formato final sería un porcentaje.

Estandarización de CGPA y SGPA

Esta estandarización es quizás la más importante pues son nuestros datos objetivos, ya que corresponden a las notas de los alumnos. Sencillamente transformamos los valores en un número.

Estandarización de Family Income

Como siempre usamos `data.unique()` para ver los valores que toman los datos y de ahí decidir cómo vamos a tratar los datos que consideramos "anómalos" o mal escritos.

1. Convertir de otras monedas y otras magnitudes a bdt 10E
2. Transformar los rangos a un valor numérico
3. Quitar el BDT que es la moneda de Bangladesh

4. Dejar números limpios. Alguna gente ha puesto valores como "Approximately 30000". Lo hemos convertido al valor numeric

Limpieza final

Finalmente convertimos los valores necesarios en numéricos (por ejemplo, terminamos de hacer la transformación de la edad, pues los valores que no tenían caracteres extras no habían sido convertidos para que dejen de ser cadenas de caracteres).

Experimentación y resultados

Para realizar la experimentación hemos elegido los siguientes algoritmos: KNN, SVM, árboles de decisión y redes neuronales.

Solamente explicaremos árboles de decisión y redes neuronales debido a que no hemos conseguido realizar la experimentación en KNN y SVM. Dicho de otra manera, no hemos conseguido hacer una predicción buena con dichos algoritmos. Sin embargo estos se encuentran disponibles en el notebook adjunto a esta entrega.

Árboles de decisión

Con el árbol de decisión hemos conseguido los resultados más cercanos posibles, demostrando así que permite una buena precisión. Una cosa a recalcar es que si hubiese mayor cantidad de datos la precisión sería mayor. Hemos observado que a mayor profundidad del árbol mejores resultados obtenemos, véase se acerca más a los datos reales. Esta vía de experimentación nos ha resultado la más útil y eficiente.

Redes neuronales

La parte principal de la red neuronal la hemos hecho con MLPRegressor de sklearn. Esto crea regresor de perceptrones multicapa inicializado con los parámetros como el número de iteraciones y el random_state. Seguido esto con la función fit hemos entrenado el modelo con los datos que hemos preparado anteriormente.

Aunque hemos conseguido un buen resultado con la red neuronal, no podemos fiarnos de ella al 100% ya que no hemos podido probarla con datos nuevos. Hemos probado también a cambiar las capas y las neuronas de la red neuronal, pero no hemos conseguido mejorar los resultados.

Conclusiones

Durante la experimentación de los datos que hemos tratado nos hemos dado cuenta de que esta experimentación hubiera sido más correcta con un dataset más grande y con los datos más ordenados. Además también sería interesante añadir otra serie de factores quizás no tan cuantificables como el entorno del alumno, la disciplina del alumno...

Sin embargo vemos como dentro de vernos más necesitados de recursos, la selección de datos que hemos elegido nos permite hacer una buena predicción.

Nuestra propuesta futura sería realizar el estudio completa, desde la parte de la encuesta, y buscando también la forma de cuantificar todos aquellos factores que influyen en el alumno para realizar una predicción aún más óptima.

Referencias

- Cáceres, D. (2021, May 30). *Cuántas horas debes estudiar cada día para tener éxito en el examen (y consejos para que tus estudios cundan)*. Business Insider España.
<https://www.businessinsider.es/cuantas-horas-debes-estudiar-cada-dia-tener-exito-examen-871681>
- ¿Cómo afectan las notas a la autoestima de los alumnos? (2024, February 19). Centre GRAT. <https://www.centregrat.com/es/suspensos-fracaso-escolar-autoestima/>
- Edad de alumnos influye en su rendimiento escolar. (n.d.). PlanoInformativo.com.
<https://planoinformativo.com/425478/edad-de-alumnos-influye-en-su-rendimiento-escolar/>
- El 23,6% de los madrileños de 15 años reconocen hacer pellas. (n.d.). ABC.
<https://www.abc.es/madrid/20140825/abci-absentismo-escolar-madrid-201408241802.html?ref=https%3A%2F%2Fwww.abc.es%2Fmadrid%2F20140825%2Fabci-absentismo-escolar-madrid-201408241802.html>
- El absentismo escolar a análisis. (n.d.). ABC.
https://www.abc.es/familia/educacion/abci-cosl-201204170000_noticia.html#vca=modulo&vso=abc&vmc=noticias-rel&vli=madrid
- Fuente, O. (2023, January 20). *Ventajas e inconvenientes de la formación online, virtual y a distancia - Noticias y Recursos Gratuitos sobre formación, grados, masters, maestrías y postgraduados*. Iberestudios.
<https://www.iberestudios.com/noticias/ventajas-e-inconvenientes-de-estudiar-a-distancia/>
- La pobreza afecta el rendimiento académico de los niños. (n.d.).
<https://www.universia.net/es/actualidad/orientacion-academica/pobreza-afecta-rendimiento-academico-ninos-1119385.html>

Las chicas sacan mejores notas que los chicos en todos los niveles. (n.d.). Diario de Navarra.

<https://www.diariodenavarra.es/archivo/actualidad/20110206/navarra/las-chicas-sacan-mejores-notas-chicos-niveles.html>

Parada, R. (2024, January 17). *Computadores y smartphones: herramientas tecnológicas clave en el regreso a clases.* Infobae.

<https://www.infobae.com/colombia/2024/01/17/computadores-y-smartphones-herramientas-tecnologicas-clave-en-el-regreso-a-clases/>

The Relationship Between Cell Phone Use and Academic Performance in a Sample of U.S. College Students. (2024, March 5).

<https://journals.sagepub.com/doi/10.1177/2158244015573169>

Simons, C. (2019, December 18). *Vivir con tus padres durante la universidad es lo contrario al sueño universitario.* VICE.

<https://www.vice.com/es/article/epggxw/vivir-con-tus-padres-universidad-espana>

Tafur, W. (2021, January 21). *Los pros y los contras de vivir solo.* Study in the USA.

<https://www.studyusa.com/es/a/1849/los-pros-y-los-contras-de-vivir-solo>

Ventajas de vivir cerca de los padres - Panamá. (2021, July 23). Pacific Developers.

<https://pacificdeveloperspanama.com/ventajas-de-vivir-cerca-de-los-padres/>

Anexo

Para la realización de este trabajo las distintas tareas han sido divididas como indicamos posteriormente. Sin embargo, cabe indicar que aunque nos hayamos dedicado cada uno a una serie de tareas, hemos participado en todas de manera general.

- Elección del dataset y limpieza de datos para estandarizarlos: Alejandro Luque Villegas
- Experimentación, modelos y conclusiones de los mismo: Javier Orbis Ramírez e Ignacio Sánchez Santatecla
- Realización de la memoria y motivaciones del proyecto: Pablo García Fernández