



Uso de técnicas predictivas para la planificación de grupos en Secundaria y FP

Universidad Rey Juan Carlos

Autor: Abel de Andrés Gómez

Tutor: Aurelio Berges García

14 de junio de 2019

Índice

- 1 Introducción
 - Introducción
 - Objetivos
- 2 Justificación teórica
- 3 Propuesta de intervención
- 4 Diseño de la investigación
- 5 Analisis de Resultados
 - Análisis exploratorio
 - Análisis predictivo
- 6 Conclusiones
 - Aportaciones del TFM
 - Conclusiones
 - Lineas de Trabajo Futuro
- 7 Referencias

Introducción

Introducción: Introducción I

¿Que es la Sobrepoblación en el aula?

Es el exceso del número de estudiantes que se encuentran en un espacio determinado cuya capacidad no es adecuada para acogerlos ni cuenta con las condiciones adecuadas para el buen desenvolvimiento de los mismos. (Racancoj, 2013)

¿Que es la ratio?

Es la relación entre el número de alumnos y profesores, es un factor importante a la hora de realizar la planificación de los recursos.

Introducción: Introducción II

- La ratio creció en España en 2012 un 20 %, ahorrándose así 464 millones de euros.
- Las ratios en España son superiores a la media de la OCDE y la UE22. (Ministerio de Educación y Formación Profesional, 2018)
- Los recursos de las administraciones públicas no son infinitos.
- **¡Se debe planificar!**

Introducción: Objetivos I

El **objetivo general** es proponer un modelo para contribuir a la óptima planificación de los grupos escolares para los nuevos cursos, evitando así el gasto innecesario de recursos y controlando la sobrepoblación en el aula.

Este objetivo se divide en los siguiente sub objetivos:

- Seleccionar variables de interés, relativas a la resolución de la necesidad anteriormente expuesta por Unidad de Planificación, que aporten valor en el desarrollo de este TFM.

Introducción: Objetivos II

- Estudiar la relación entre dichas variables con el propósito de comprender el contexto de la sobrepoblación en el aula y la planificación de grupos.
- Probar distintos modelos predictivos y seleccionar aquellos que aporten mayor precisión en la predicción de los grupos.
- Obtener y utilizar el modelo de mayor precisión para realizar predicciones.

Justificación teórica

Justificación teórica

Objeto de búsqueda

Documentación científica acerca de la minería de datos en el ámbito educativo, concretamente, en la gestión de la educación.

- ScienceDirect, Scopus, Google Academics, etc.
- ¿Existen artículos relacionados con la minería de datos en la educación?
- ¿Existen artículos relacionados con la planificación educativa?
- ¿Que metodologías se siguen?
- ¿Que modelos predictivos contemplan?
- ¿Que variables se utilizan?
- ¿Que herramientas se usan?

Propuesta de intervención

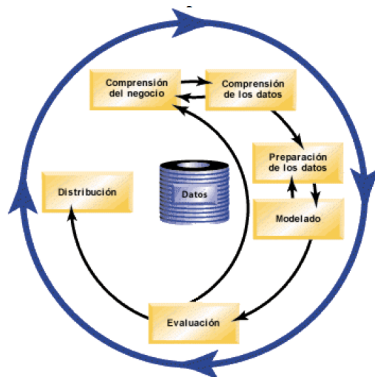
Propuesta de intervención

- Actualmente, la Unidad de Planificación de Centros Públicos utiliza herramientas poco automatizadas para conocer el número de alumnos y unidades, y no disponen de algoritmos predictivos que faciliten y mejoren esta labor.
- Diseñar un sistema que sea capaz de ayudar en la predicción a la Unidad de Planificación de Centros Públicos.
 - Global: que se pueda utilizar en distintos dispositivos y evite la transferencia de archivos.
 - Flexible: que permita la inclusión y eliminación de nuevas variables.
 - Fiable: que otorgue garantía en la planificación de grupos.
 - Evitar y detectar la sobrepoblación en el aula.
- Además, con los datos se pretende identificar los factores que impliquen mayor demanda en un determinado centro o curso.

Diseño de la investigación

Diseño de la investigación: I

Figura: Fases del ciclo de vida de CRISP-DM. Recuperado de *Manual CRISP-DM de IBM SPSS Modeler (2012)*.



Diseño de la investigación: II

- Comprensión del negocio: Obtener información acerca de los objetivos de la investigación. Estudio de la situación de la educación en la Comunidad de Madrid.
- Comprensión de los datos:
- Preparación de los datos: Se realiza un tratamiento de los datos, se eliminan valores y variables vacías, se separan o se juntan variables a conveniencia de la investigación. Se separan los datos en dos subconjuntos (entrenamiento y pruebas)
- Modelado: Se utilizan y comparan los modelos predictivos. Para ello se utilizan métricas de selección de variables

Diseño de la investigación: III

- Evaluación: Se utilizan métricas de precisión para evaluar los modelos. Además, también se cuenta el tiempo que tardan en entrenarse. A partir de estas métricas se toma una decisión.
- Distribución: Toda la información recogida en la investigación se documenta. En esta parte se indica el modelo, las variables, herramientas y metodología que deben utilizarse para el control del proceso de la minería de datos.

Analisis de Resultados

Analisis de Resultados: Análisis exploratorio

- La media de la ratio de la Comunidad de Madrid es de 0,88, por lo tanto, los centros educativos no están sobrepoblados.
- Las variables de número de alumnos y número de grupos son importantes a la hora de predecir el número de grupos finales.
- La DAT que sufre mayor sobrepoblación en las aulas es la DAT-Centro, seguida de la DAT-Sur.
- Las variables que mayor correlacionan con la **ratio** son: número de alumnos, servicio de comedor, naturaleza del centro y código genérico del centro.
- Las variables que mayor correlacionan con el **número de grupos a planificar** son: número de alumnos, unidades, naturaleza, código genérico y servicio de comedor.

Analisis de Resultados: Análisis predictivo

- Utilizando algoritmos se observan que las variables más importantes para realizar la predicción son: naturaleza del centro, número de curso, número de unidades, nivel de enseñanza y ratio.
- El algoritmo que mejores predicciones obtiene es el Árbol de Decisión, seguido de la Regresión Lineal.
- Se observa que, en los resultados de la predicción, de 4436 datos existentes para grupos, únicamente 52 grupos se modifican. De estos 52 grupos, 30 de ellos disminuyen en el número de unidades y 22 aumentan. Se observa para este curso 2017/2018, que **se reduce la población en el aula.**

Conclusiones

Conclusiones: Aportaciones del TFM

- Descubrimiento de relaciones entre variables.
- Uso de herramientas de minería de datos para la planificación académica.
- Comparación de modelos predictivos en el entorno educativo para la planificación de grupos.
- Investigación relativa a la gestión y planificación en el entorno educativo.

Conclusiones: Conclusiones I

- ➊ **Selección de variables:** Se utilizan algunas de las variables propuestas a partir de la instrucción de la Unidad de Planificación. (Consejería de Educación e Investigación, 2018)
 - Se parte de 27 variables, finalmente se eligen 11 para el estudio de predicción. Se eliminan variables descriptivas (nombre del centro, código de este, etc.)
- ➋ **Estudio de la relación entre variables para comprender el contexto de sobrepoblación en el aula:** Para ello se utiliza la matriz de correlación. Se han obtenido las variables que hacen que la ratio en las aulas aumente o disminuya. Se han obtenido también las variables con más importancia en la predicción de grupos.

Conclusiones: Conclusiones II

- ③ **Pruebas con distintos modelos predictivos y selección de los óptimos:** De todos los modelos propuestos, se han elegido dos: Árbol de decisión y Regresión Linear. Ambos modelos son los que mejores resultados (en cuanto a precisión y tiempo de entrenamiento han obtenido).
- ④ **Se utiliza el algoritmo de mayor precisión en la predicción con datos existentes:** A partir del modelo de árbol de decisión se obtiene la predicción esperada para el curso 2017/2018 a partir del modelo entrenado a partir de los datos de 2016/2017.

Conclusiones: Líneas de Trabajo Futuro

- Inclusión a la investigación de nuevas variables como aspectos físicos del centro, acceso a este, comunicaciones, estaciones de metro cercana, paradas de autobús, etc.
- Número y nuevas enseñanzas.
- Nuevos modelos de predicción y parámetros de ajustes.
- Realización de Software usando modelos investigados.

Referencias

Referencias

Consejería de Educación e Investigación. (2018). *Instrucciones de la dirección general de educación infantil, primaria y secundaria sobre la planificación del próximo curso escolar 2018/2019 en los centros públicos que imparten eso y bachillerato, creación de nuevos centros y modificación de la red, implantación y autorización de enseñanzas y propuesta de grupos.*

Manual crisp-dm de ibm spss modeler. (2012).

Ministerio de Educación y Formación Profesional. (2018). *Panorama de la educación indicadores de la ocde 2018.*

Racancoj, L. V. M. (2013). *Sobrepoblación estudiantil y desempeño docente en el aula* (Thesis).