

TÍTULO: Análisis de datos y predicción en variables del entorno educativo.

Autor: Abel de Andrés Gómez

Director: Aurelio Berges García

Resumen

En los últimos años, gracias al gran desarrollo tecnológico que se ha vivido tanto a nivel de cómputo (mejorando la eficiencia y el uso de los recursos disponibles) como a nivel de transmisión de datos (mejorando las comunicaciones), ha permitido a las organizaciones el almacenamiento de una gran cantidad de información.

Como se puede observar en las siguientes gráficas: las millones de instrucciones por segundo (MIPS) que realiza un procesador (relacionado con el tiempo de cómputo) y la velocidad de transmisión de datos en bits por segundo (BPS) han crecido a lo largo de los últimos años (Nielsen, 2018).

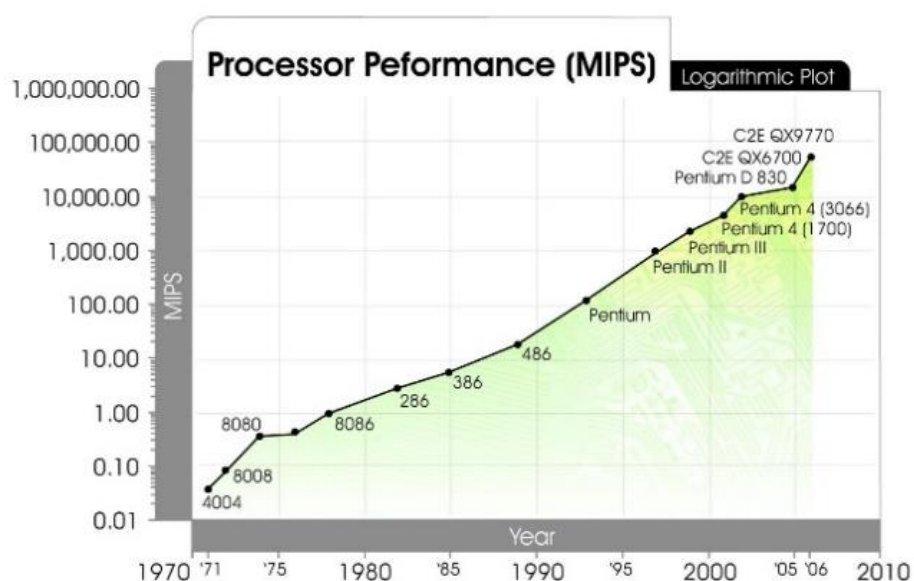


Ilustración 1. Rendimiento procesador (MIPS) a lo largo del tiempo. (Fuente: Kurzweil <http://www.kurzweilai.net>)

Análisis de datos y predicción en variables del entorno educativo

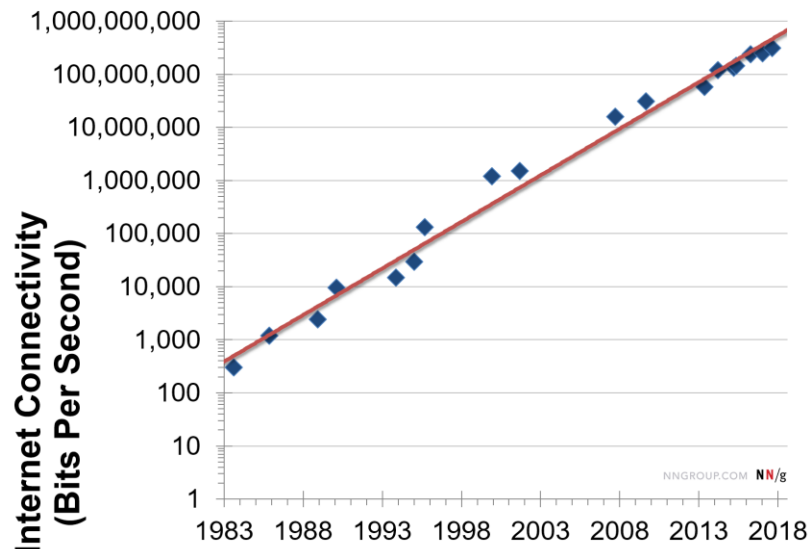


Ilustración 2. Velocidad de transferencia a lo largo del tiempo. (Fuente: Nielsen Norman Group)

Se hace por tanto necesario el uso de métodos, técnicas, herramientas además de personas con conocimientos formando un vínculo estrecho que permita y ayude a explotar, investigar, predecir y obtener información relevante para tomar decisiones de forma adecuada, utilizando este gran volumen de datos que se encuentra en continuo crecimiento.

En este Trabajo Fin de Máster (TFM), en primer lugar se aborda el estudio de la ciencia de datos, que se define como el campo de la estadística y de las ciencias de computación que intenta descubrir patrones (mediante modelos matemáticos) en grandes volúmenes de datos, que puedan aportar información valiosa en la toma de decisiones. Y seguidamente, se estudia el concepto de análisis predictivo, que consiste en el empleo de datos pasados o históricos para predecir eventos futuros. (MathWorks, 2018).

En segundo lugar y una vez definidos los conceptos más relevantes, se aplican las técnicas estudiadas en el ámbito educativo sobre el que trata este TFM.

Según la OCDE (Organización para la Cooperación y el Desarrollo Económicos), la duración y la calidad de la enseñanza tienen un impacto importante en su transición desde la educación al trabajo. Asimismo, se puede observar una estrecha correlación entre el nivel de desarrollo de los países con la fortaleza de sus sistemas educativos y de investigación científica y tecnológica. Por lo que la educación, es uno de los factores que más influye en el avance y progreso de las personas y las sociedades. Por lo tanto, la educación se hace necesaria en todos los sentidos, ya sea para alcanzar los mejores niveles de bienestar social y de crecimiento económico o para nivelar las desigualdades económicas y sociales entre otros numerosos ámbitos.

Por esta razón, se justifica la investigación planteada en este TFM sobre cómo la ciencia de datos junto con las técnicas predictivas, que son disciplinas emergentes en este ámbito, puede mejorar la educación. (Educational Data Mining, 2018).



La metodología seguida en el desarrollo de la investigación que se va a llevar a cabo en este TFM está basada en la norma UNE 166006:2018 Gestión de la I+D+I: Sistemas de vigilancia e inteligencia. Dicha norma tiene como objeto facilitar la formación y estructuración del proceso de recogida, análisis y comunicación de la información sobre el entorno de una organización. (UNE 166006,2018)

Respecto al origen de los datos, existen numerosas bases de datos donde se puede localizar información acerca de datos educativos, entre ellas se encuentran la OCDE y MECD. Por una parte, se pretende utilizar estas bases de datos como referencia sobre los posibles datos a analizar y estudiar. Y por otra, los datos anonimizados y agregados sobre escolarización en secundaria en la Comunidad de Madrid durante los últimos 10 años.

Para el tratamiento de datos, se utilizan técnicas ETL (extracción, transformación y carga) que consisten básicamente en obtener los datos de la fuente de origen (bases de datos, ficheros Excel, etc.), seleccionar aquellos que convengan al estudio, transformarlos y depurarlos (evitando así datos erróneos y modificándolos a conveniencia del investigador) y cargarlos de nuevo en otra fuente de información. (Sharma, 2014) (Guillermo Matos, 2006) (Prakash, 2017).

Para finalizar, y una vez que se realice el proceso ETL y con los datos ya preparados para poder aplicar técnicas de análisis de datos, se va a probar y comparar distintos modelos para conseguir la mejor precisión y obtener información para la posible toma de decisiones. (Archanco, 2018).

Bibliografía Básica

Aluja, T. (2001). La minería de datos, entre la estadística y la inteligencia artificial. *QÜESTIÓ*.

Archanco, R. (12 de 12 de 2018). *Listado de técnicas de análisis de datos útiles para tu empresa*. Obtenido de <https://papelesdeinteligencia.com/tecnicas-de-analisis-de-datos/>

Eduardo Fernandes, M. H. (7 de Febrero de 2018). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *ELSEVIER*.

Educational Data Mining. (2018). Obtenido de <http://educationaldatamining.org>

Guillermo Matos, R. C. (2006). Metodología para la Extracción del Conocimiento Empresarial a partir de los Datos. *Información Tecnológica*, 81-88.

MathWorks. (13 de 12 de 2018). Obtenido de <https://es.mathworks.com/discovery/predictive-analytics.html>

MECD. (s.f.). *Estadísticas Nacionales*. Obtenido de <http://www.mecd.gob.es/servicios-al-ciudadano-mecd/estadisticas/educacion.html>

Nielsen, J. (6 de Enero de 2018). *NN/g Nielsen Norman Group*. Obtenido de <https://www.nngroup.com/articles/law-of-bandwidth/>

Análisis de datos y predicción en variables del entorno educativo

OCDE. (s.f.). *Estadística OCDE*. Obtenido de <https://stats.oecd.org>

Prakash, G. H. (2017). ETL Data conversion: Extraction, transformation and loading Data Conversion. *International Journal of Engineering And Computer Science*.

Sharma, S. (2014). Modeling ETL Process in Data warehouse: An Exploratory Study. 2014 *Fourth International Conference on Advanced Computing & Communication Technologies*.

UNE. (Abril de 2018). UNE 166006. Gestión de la I+D+I: Sistema de vigilancia e inteligencia.