

Ensayo de artículos sobre Data Science

En el artículo llamado "Statistical Modeling: The Two Cultures" escrito por Breiman, nos habla acerca de que existen diferentes maneras de crear conclusiones con los datos pero que desde su perspectiva puede destacar dos distintas culturas; Una de ellas usa el modelado estocástico de datos y la otra utiliza un modelado algorítmico en el cual a priori no sabemos que modelo implementar a nuestro conjunto de datos.

El artículo cuestiona a los Estadísticos y el uso generalizado de modelos estocásticos para tratar de resolver cualquier problemática. El autor menciona que los investigadores se enfocan más en el modelo que se usará, que en resolver verdaderamente el problema.

Breiman también nos habla de varios proyectos en los cuales él ha estado involucrado y como es que al usar esta cultura de modelado algorítmico se puede llegar a obtener mejores resultados que con solo usar modelos estocásticos.

Por último él nos muestra los pro y los contra entre el modelado clásico y el modelado algorítmico, indicando de manera extensiva la forma en como estos problemas aparecen además agregando la experiencia que él ha tenido con estos.

Al final del artículo comienzan a presentarse una serie de comentarios, algunos de ellos se oponen a las ideas propuestas por Leo Breiman. En contraste a lo anterior existen dos comentarios a favor de lo que dice el autor. Es fácil saber la razón por la cual el artículo termina siendo tan discutido entre grandes estudiosos de la estadística, como Breiman menciona en su escrito, en aquella época más del noventa por ciento de las personas enfocadas en la estadística simplemente prefería usar modelos estocásticos ya bien conocidos, es decir, aquellas personas pensaban que el problema a resolver "debería" cumplir con alguna distribución que ellos ya tenían bien estudiada.

Para finalizar Leo Breiman responde a sus opositores de una manera concisa y ultra detallada, simplemente es una genialidad de artículo que describió lo que ahora mismo está ocurriendo en el mundo de la ciencia de datos.

En cuanto al artículo llamado "Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics" por William S. Cleveland nos muestra un plan grande rasgos con posibles cambios según las necesidades y los cambios en centros educativos enfocados en el análisis de datos. Cleveland hace énfasis en seis áreas técnicas. Las áreas son: investigaciones multidisciplinarias, modelos y métodos para los datos, computo con datos, pedagogía, evaluación de herramientas y teoría. Siendo la más importante para él las investigaciones multidisciplinarias.

La finalidad de su plan es poder enseñar de una manera conjunta ciencias computacionales y matemáticas como áreas primordiales para el campo de la ciencia de datos y el autor nos muestra como podríamos comenzar a lograr esto.