

Propuesta de Trabajo Final del Curso

Clasificar y agrupar: Machine Learning aplicado a un caso de encuesta de opinión política en Argentina

Manuel Miller

Propuesta

Distintas sociedades occidentales parecen atravesar un proceso a primera vista contradictorio que es materia de estudio relevante para la Ciencia Política: la difusión y creciente liquidez en el contenido de las identidades, al mismo tiempo que la radicalización y exacerbación de la violencia en el debate publico en torno a estas etiquetas identitarias aumentan.

Este problema lleva a preguntarse como distintas ideas culturales y políticas se relacionan con otras y su mayor o menor grado de correspondencia con la identificación política que una persona hace de sí misma en Argentina. Para ello, diseñé una encuesta de opinión que aportará una muestra no probabilística que servirá como marco de aproximación al problema. Los datos fueron recogidos mediante Google Forms con un diseño muestral de bola de nieve.

Se relevaron al momento 110 sujetos / unidades análisis a quienes se les pregunto sobre su posicionamiento frente a ciertos dilemas y frases típicas del debate publico

argentino, su identificación ideológica y su opinión acerca de otros países, aparte de información que permita posicionarlos demográfica y socioeconómicamente en distintos grupos.

Luego se realizará un análisis exploratorio mediante el lenguaje de procesamiento estadístico R para conocer información certera de la muestra.

El segundo paso será la aplicación de algoritmos de clasificación como XGBoost o Random Forest en python para buscar predecir la identificación política de la persona a través de sus respuestas previas. Secundariamente evaluar la efectividad el modelo en la predicción de otras de las variables disponibles.

Finalmente, implementar un modelo de aprendizaje no supervisado de agrupamiento como K-means en Python nos permitirá saber en que medida la autopercepción política de las personas es congruente con los clusters obtenidos.

Relación con el curso:

La relación y pertinencia con los contenidos de la asignatura "Sistemas de Información y Bases de Datos Estadísticos" viene dada por la adopción de los métodos de Machine learning supervisado y no supervisado trabajados.

En este trabajo se combinan el conocimiento del experto en Ciencia Política que escoge a conciencia las preguntas del cuestionario para someter los resultados a una comparación frente a un modelo de clasificación y entender así si dos personas que se ubican en el mismo espacio del espectro político comparten o no un patrón de ideas comunes y congruentes.

Por otro lado, las técnicas de agrupamiento no supervisadas permiten someter los mismos conocimientos del experto a prueba y ofrecer una imagen de como la muestra se distribuye en grupos, independientemente de las presunciones del investigador.