Proposición de trabajo

Procesamiento de datos textuales EthicApp con algoritmos de procesamiento de lenguaje natural

Camilo Carvajal Reyes

30 de marzo de 2023

Descripción de la problemática

En el marco de la enseñanza de la formación ética en la FCFM, se ha utilizado la aplicación EthicApp para la obtención y posterior análisis de las decisiones morales de estudiantes ante un dilema ético. Estos casos de estudio consisten en una problemática, que se plantea en forma de pregunta. Como respuesta a esta pregunta, los estudiantes manifiestan una preferencia en la forma de un número entre 1 y 6, donde los extremos de esta escala representan posturas dispares en cuanto a la decisión a tomar. Luego de una primera respuesta se realiza una instancia de deliberación grupal, donde se toma una nueva decisión en conjunto. Finalmente, los estudiantes otorgan nuevamente una calificación, que puede diferir de la señalada en las dos instancias anteriores.

El Área de Ética de la FCFM ha llevado a cabo un completo análisis de las posturas de los estudiantes. No obstante, un aspecto difícil de procesar son las justificaciones que deben colocar luego de cada instancia de decisión. Pese a que análisis cualitativos de algunas respuestas han permitido plantear hipótesis preliminares respecto a los juicios morales y justificaciones de las decisiones, la gran magnitud de datos presentes dificultan la tarea de tomar conclusiones acerca de las competencias éticas exprimidas en la instancia, y como consecuencia la verificación de la adquisición de esta durante los cursos formativos de la escuela.

Proposición de solución

La ciencia de datos es una disciplina en constante crecimiento, en particular en nuestra facultad. En particular, los métodos basados en aprendizaje de máquinas han experimentado un aumento considerable en sus capacidades, ayudado por la mejora en capacidad de cómputo en las últimas décadas. Los algoritmos de procesamiento han sido una demostración de aquello, con modelos conversacionales como *chatGPT* tomando protagonismo entre los medios y el público.

Dentro de esta rama se encuentran los modelos de lenguaje, que intentan aproximar modelar uno o más lenguajes humanos al inducir una probabilidad a secuencias

de palabras (frases, oraciones o documentos). Estos modelos se implementan usando redes neuronales profundas y siendo entrenados en grandes volúmenes de datos textuales (córpuses). Finalmente, un modelo sirve para resolver variadas tareas de procesamiento de texto, incluyendo clasificación, pregunta-respuesta e categorización/identificación de elementos relevantes de una secuencia. Entre las desventajas que presentan estos modelos están su costo de entrenamiento y capacidad limitada de interpretabilidad, ambas consecuencias de su gran tamaño.

Se propone la implementación de estos modelos pero también el uso de algoritmos más simples y más interpretables, para procesar los argumentos escritos por estudiantes en sus decisiones éticas. Más precisamente, se procederá a:

1. Explorar las justificaciones textuales de las respuestas usando técnicas de minería de datos.
2. Implementar modelos estadísticos para texto, que sean interpretables, para predecir la respuesta (número en la escala de 1 a 6) de los estudiantes utilizando el texto de justificación.
3. Implementar los modelos anteriormente descritos para la predicción del cambio de respuesta de una etapa a otra, usando las justificaciones de la etapa intermedia y última etapa.
4. Identificar, a través del texto, elementos semánticos que justifiquen los argumentos dados por los estudiantes. Esto usando tanto el análisis exploratorio de datos como los algoritmos.
5. Identificar, del mismo modo que el punto anterior, elementos semánticos en cambios de valoraciones entre distintas etapas de la actividad, tanto con elementos diferentes como comunes entre ambas justificaciones.
6. Implementar modelos de lenguaje entrenados con aprendizaje profundo para las dos tareas de predicción anteriores. Comparar la capacidad de predicción tanto con los algoritmos básicos como con la capacidad humana.
7. Utilizar modelos predictivos de texto para predecir el grado de competencia ética en las justificaciones, utilizando tanto técnicas simples como avanzadas de procesamiento de lenguaje natural.(*)
8. Utilizar modelos de reconocimiento de entidades para la identificación automática de elementos textuales que denoten elementos positivos y negativos en cuanto a la calidad de la respuesta otorgada.(*)

Si es que los modelos muestran una buena capacidad de predicción, se pueden usar como herramienta que a la larga servirá para evaluar la progresión de competencia ética de los estudiantes con menor inversión humana. Esto es de particular relevancia para los objetivos finales del área de ética. Por otro lado, existe un interés en verificar hasta que punto los algoritmos pueden modelar relaciones semánticas complejas como lo son las justificaciones morales de estudiantes ante a una problemática. Este es un objetivo complementario y que logrará plantear nuevas perspectivas de investigación cualquiera sea el resultado, tanto del punto de vista computacional como del estudio de la ética.

(*) Notar que estas tareas requieren la creación de un dataset con etiquetas especiales.

Metodología de trabajo

El tiempo de trabajo contempla aproximadamente 21 semanas (dedicación aproximada de 6 horas semanales), las cuales estarán distribuidas como se muestra acontinuación:

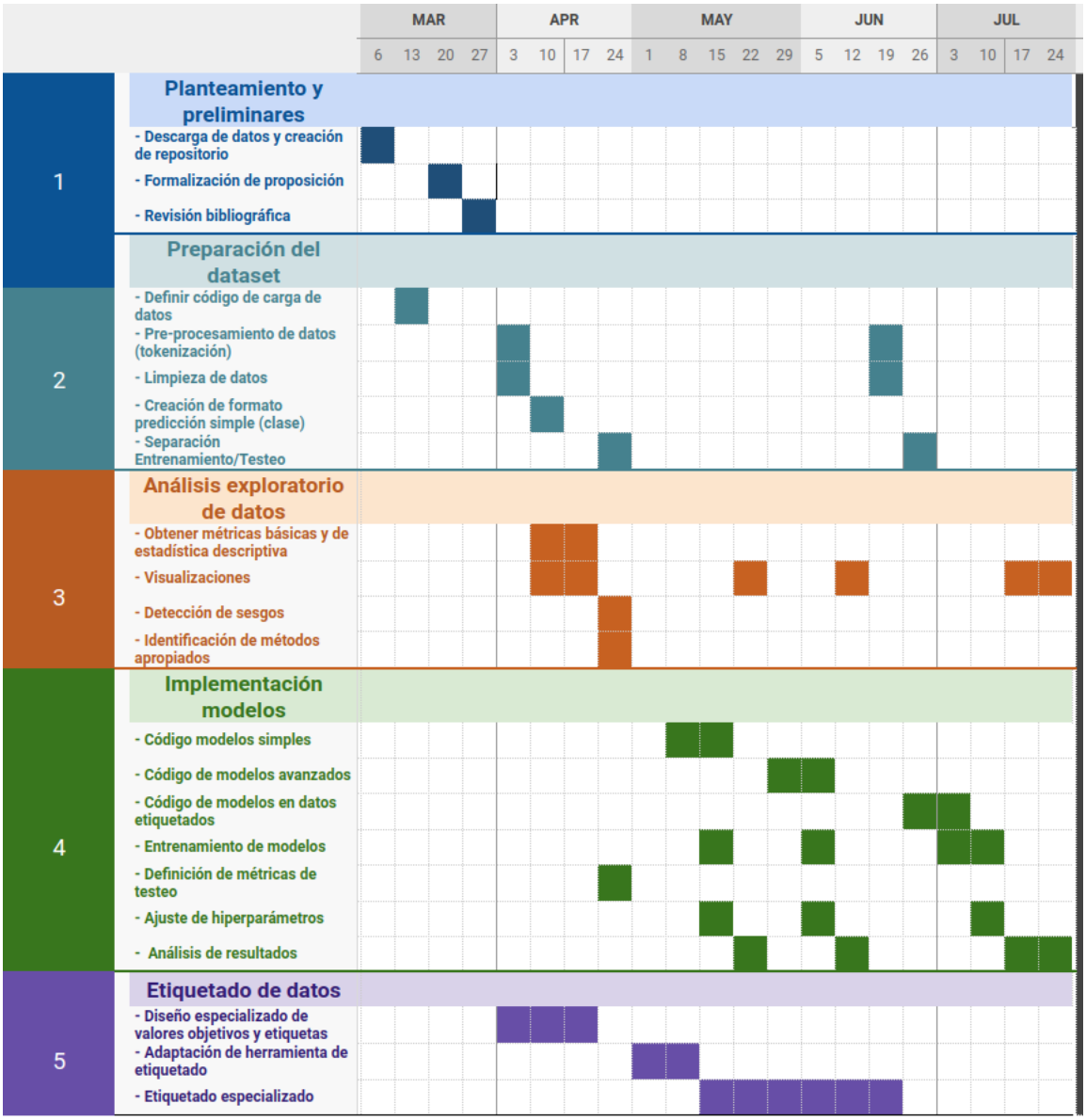

Figura 1: Planificación de trabajo.