**Proyecto Análisis Avanzado de Datos para el sistema judicial de EEUU**

**CONTEXTO DEL PROYECTO**

A lo largo de los años, diversas metodologías para predecir riesgo de reincidencia en el ámbito judicial han sido objeto de polémica, revelando que, en general, su validez predictiva es moderada. Hasta hace poco, el uso del sistema Compass en EE.UU. generaba una preocupación particular por la ausencia de investigaciones exhaustivas que examinarán si las puntuaciones de riesgo presentaban algún tipo de sesgo hacia ciertos grupos étnicos. La cuestión de si estos resultados reflejan un sesgo inherente a las herramientas o discrepancias más amplias dentro del sistema de justicia sigue siendo un tema de debate.

Este contexto plantea importantes cuestiones éticas y metodológicas, especialmente en lo que respecta al uso de algoritmos y herramientas de análisis de datos en la toma de decisiones judiciales. La identificación y comprensión de posibles sesgos en los datos y en las herramientas analíticas son cruciales para garantizar un tratamiento justo y equitativo en el sistema de justicia, y subrayan la importancia de abordar estas problemáticas en proyectos de análisis de datos.

**DESCRIPCIÓN DEL PROYECTO**

Analizar el conjunto de datos de reincidencia en delincuentes del condado de Broward en EEUU para entender los factores que pueden contribuir a un mayor riesgo de reincidencia, con foco en los sesgos potenciales que existan en los datos. Tenemos disponibles los scores de reincidencia de Compass para cada persona, y disponemos también de la información sobre si una persona reincide o no reincide en un plazo de dos años. Otro de los objetivos del proyecto será evaluar la fiabilidad de los scores devueltos por Compass y sus sesgos. Compararemos esa fiabilidad con la de un modelo simple que tendremos que implementar nosotros (sin sesgos), y acabaremos presentando los hallazgos más interesantes del proyecto en un dashboard.

**TAREAS**

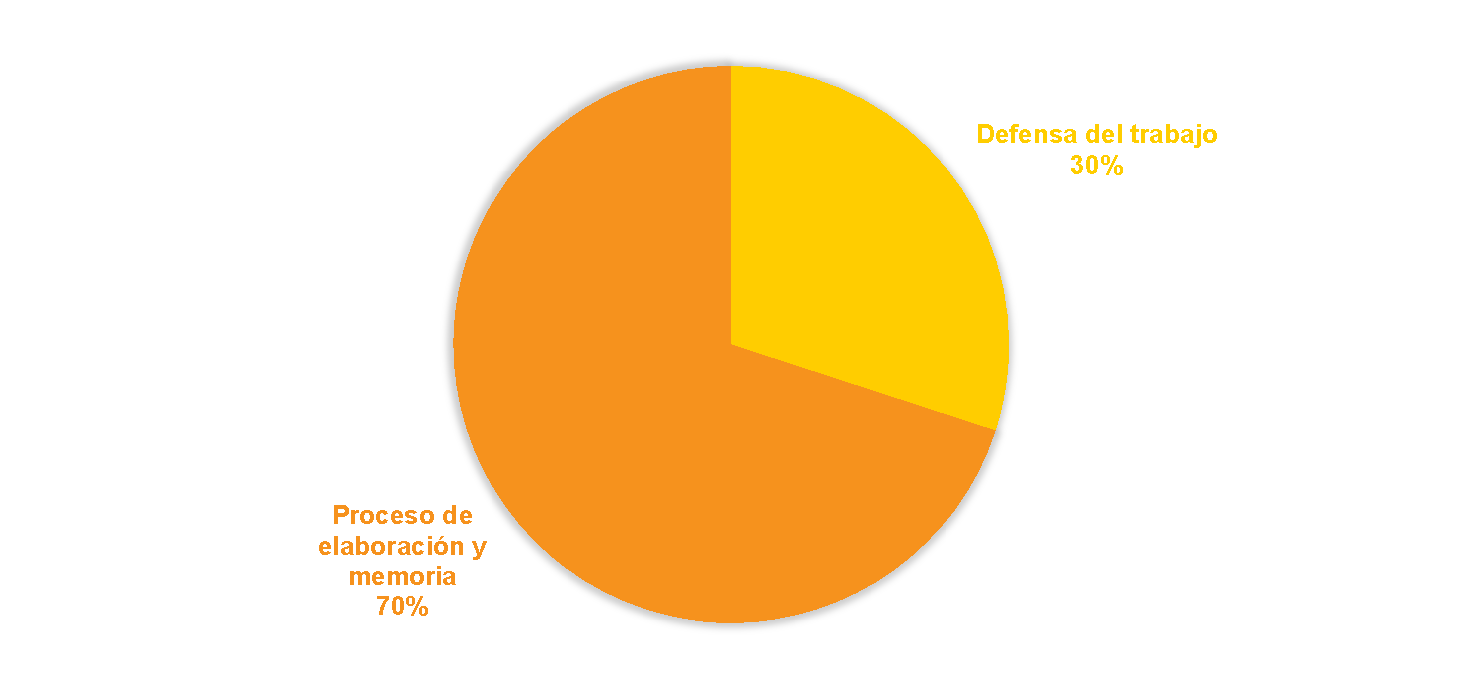
1. **Análisis de Datos (EDA):**
   * Familiarización con el conjunto de datos, incluyendo variables demográficas, puntajes de Compass y resultados de reincidencia.
   * Limpieza y preprocesamiento de datos para análisis.
   * Estadísticas descriptivas para comprender la composición de los datos.
   * Análisis de variables demográficas y su relación con los puntajes de Compass y la reincidencia.
2. **Identificación de sesgos:**
   * Investigación de posibles sesgos en los datos, particularmente relacionados con variables demográficas como raza y género.
   * Discusión sobre la importancia de identificar y abordar sesgos en el análisis de datos.
3. **Evaluación de los Puntajes de Compass:**
   * Análisis de la eficacia de los puntajes de Compass para predecir la reincidencia.
   * Comparación de tasas de verdaderos positivos y falsos positivos entre diferentes grupos demográficos.
   * Evaluación de otras posibles métricas o KPIs para el análisis. Será necesaria una familiarización previa con esas métricas a través de los recursos que se consideren oportunos.
4. **Desarrollo de un Modelo Predictivo Simple:**
   * Implementación de un modelo predictivo simple, como una regresión logística, para predecir la reincidencia basándose en variables seleccionadas (evitando sesgos).
   * Evaluación y comparación de la precisión del modelo desarrollado por los estudiantes con los puntajes de Compass.
5. **Visualización de Datos y Creación de Dashboards:**
   * Integración de resultados del análisis en un dashboard para mostrar resultados clave del análisis.
   * Incluir distribución de los puntajes de Compass, comparaciones de precisión entre modelo y los puntajes de Compass, especialmente en áreas identificadas de sesgo potencial.
6. **Interpretación y Recomendaciones:**

* Análisis crítico de los hallazgos, discusión sobre los sesgos identificados y su impacto en las predicciones de reincidencia.
* Propuesta de estrategias para mitigar los sesgos en futuros análisis predictivos.

**ENTREGABLES**

1. **Documento técnico** con metodología y resultados (memoria de proyecto).
2. **Entregables solicitados para cada tarea:** notebooks y códigos de cada tarea, y dashboard
3. **Presentación final para congreso de EEUU**

**CÁLCULO DE LA NOTA**



**INFORMACIÓN DE LOS DATOS**

Disponemos de datos de dos años de puntuaciones Compass de la Oficina del Sheriff del Condado de Broward en Florida, que fueron obtenidos por ProPublica. ProPublica es una organización de periodismo de investigación sin fines de lucro con sede en Estados Unidos, conocida por su análisis y reportajes sobre temas de interés público, incluidas las prácticas y políticas dentro del sistema de justicia.

Los datos corresponden a personas evaluadas en 2013 y 2014. Las puntuaciones de riesgo de reincidencia de Compass se basan en las respuestas de un acusado a la encuesta de cribado Compass. Esa encuesta se completa por el acusado después de su arresto, bien el mismo día o el día después de que una persona es encarcelada.

Para los más de 11 mil acusados en espera de juicio en este conjunto de datos, ProPublica se encargó de recopilar los datos de sus posibles arrestos futuros hasta finales de marzo de 2016, con el objetivo de estudiar cómo la puntuación Compass predice la reincidencia para estos acusados. En los datos con los que trabajaremos para el proyecto contamos con ese indicador de si una persona ha sido arrestada de nuevo por un crimen cometido en un intervalo de dos años posteriores al crimen original, y con el período de tiempo entre arrestos.

Cada acusado en la etapa previa al juicio recibió al menos tres puntuaciones Compass: "Riesgo de Reincidencia", "Riesgo de Violencia" y "Riesgo de No Comparecencia". Las puntuaciones varían de 1 a 10, siendo 10 el riesgo más alto. Compass clasifica las puntuaciones del 1 al 4 como "Bajo", del 5 al 7 como "Medio" y del 8 al 10 como "Alto".