**[Diapositiva 1]**

**¡Buenos días!** Voy a presentar mi trabajo de final de máster titulado:

“Análisis de redes sociales dinámicas de aprendizaje colaborativo”

dirigido por el profesor Antonio Rodriguez Anaya.

Este trabajo pretende inicialmente estudiar como el análisis de redes sociales puede ayudarnos a predecir el abandono estudiantil en el ámbito educativo.

**[Diapositiva 2]**

**Aquí les muestro el índice de contenidos de la presentación**

**[Diapositiva 3]**

En un mundo cada vez más conectado, el aprendizaje colaborativo online se ha convertido en una herramienta poderosa para adquirir conocimientos y desarrollar habilidades.

Las redes sociales de aprendizaje virtual se convierten en espacios donde los estudiantes pueden interactuar, colaborar y aprender unos de otros.

Los foros de una asignatura son mucho más que simples espacios para hacer preguntas y respuestas. Son entornos sociales dinámicos donde construimos conocimiento juntos.

El análisis de redes sociales aplicado a los foros proporciona información valiosa sobre la participación de los estudiantes, que puede ser utilizada para monitorizar su progreso de aprendizaje y predecir su rendimiento académico

**[Diapositiva 4]**

Por eso, me planteé la siguiente pregunta de investigación: ¿cómo podemos utilizar el análisis de redes sociales de aprendizaje colaborativo para predecir el abandono de los estudiantes? Esta pregunta me llevó a investigar y desarrollar modelos de aprendizaje automático supervisado que me permitieran comprender las interacciones entre los estudiantes y predecir su abandono en una asignatura.

La motivación para realizar este estudio fue doble. Por un lado, quería analizar las interacciones y la estructura social en una comunidad educativa en línea utilizando técnicas de análisis de redes sociales aplicadas a los foros de discusión de una asignatura.

Por otro lado, quería desarrollar modelos de aprendizaje automático supervisado que utilizaran medidas de centralidad de la red social y medidas del sentimiento expresado en los mensajes para predecir el abandono estudiantil.

**(CONTINÚA)**

**[Diapositiva 5]**

Para evaluar el uso del análisis de redes sociales dinámicas en la predicción del abandono estudiantil. Se proponen tres objetivos principales:

* El **primer objetivo** es analizar la eficacia de las predicciones basadas en el análisis de redes sociales, considerando diferentes **rangos de tiempo** desde el inicio del curso. Esto nos permitirá entender cómo varía la precisión de las predicciones a medida que ampliamos o reducimos el período de tiempo analizado.
* El **segundo objetivo**, es evaluar la eficacia del análisis de las redes temporales al subdividir el rango temporal en **bloques de días seriados secuencialmente**. Queremos determinar si esta subdivisión en bloques secuenciales tiene alguna influencia en la precisión de las predicciones.
* Por último, el **tercer objetivo** consiste en evaluar la eficacia del análisis de las redes temporales al analizar **bloques seriados dinámicamente o encadenados**. Nos preguntamos si esta subdivisión en bloques encadenados puede mejorar la precisión de las predicciones en comparación con la subdivisión secuencial.

**[Diapositiva 6]**

Para operativizar estos objetivos, hemos considerado cuatro **rangos de tiempo**: 30, 60, 90 y 120 días. Estos rangos representan segmentos específicos de interacción en los foros de los estudiantes desde el inicio del curso.

Por ejemplo, el rango de 30 días cubre el 25% de la interacción estudiantil, mientras que el rango de 120 días cubre el 100%.

**[Diapositiva 7]**

En estas figuras se ilustra el conjunto de datos y atributos, como las medidas de centralidad de grado, intermediación y cercanía, simbolizadas con **GCI**, que serían analizados predecir la variable de resultado, en este caso, el abandono de la asignatura, simbolizada como **Y**, utilizando diferentes rangos de tiempo.

Es interesante destacar que, en términos prácticos, obtener una predicción precisa con un rango temporal más corto brinda una ventaja al equipo docente al permitirles tomar medidas preventivas del abandono de manera más eficaz durante el curso.

**[Diapositiva 8]**

Siguiendo estudios previos, hemos adoptado la **subdivisión en bloques** como una estrategia para realizar un análisis más detallado de la red social en segmentos específicos de tiempo. Esta subdivisión nos permite obtener una perspectiva más granular que puede capturar la continuidad de las interacciones a lo largo del tiempo.

Para ello, hemos considerado los diversos rangos temporales y hemos determinado las opciones de subdivisión en bloques siguiendo una regla estricta: los bloques deben ser de días múltiplos de 5 y divisores del total del rango temporal analizado.

En la tabla presentada, podemos ver la subdivisión en bloques para cada rango temporal. Por ejemplo, para un rango de 30 días, consideramos opciones como 6 bloques de 5 días, 3 bloques de 10 días y 2 bloques de 15 días. De manera similar, se han establecido opciones de subdivisión para los rangos de 60, 90 y 120 días.

**[Diapositiva 9]**

En la figura, se ilustra la subdivisión en 6 bloques secuenciales para un total de 60 días, donde cada bloque de 10 días se sucede en orden consecutivo.

Esta subdivisión nos permite generar redes sociales temporales en cada bloque y obtener las diferentes medidas de centralidad en cada una de ellas.

**[Diapositiva 10]**

Finalmente, examinamos la subdivisión en **bloques dinámicos o encadenados**. En esta estrategia, cada bloque temporal se superpone con el bloque anterior y el siguiente.

Esto implica que hay nodos y conexiones compartidos entre bloques adyacentes, lo que nos permite capturar la continuidad y los cambios graduales en la red a lo largo del tiempo.

**[Diapositiva 11]**

En la figura, se muestra la subdivisión en bloques encadenados para un total de 60 días con 6 bloques de 10 días. Cada par de bloques consecutivos de 10 días se superponen para generar las redes sociales temporales.

Al superponer los bloques, se establece una relación de continuidad que captura la evolución y los cambios dinámicos en la red a lo largo del tiempo.

**[Diapositiva 12]**

Para lograr mis objetivos principales, he planteado tres subobjetivos basados en **la teoría del aprendizaje social**. El propósito de estos subobjetivos es evaluar cómo las configuraciones de tiempo y subdivisión en bloques pueden afectar a la precisión de las predicciones utilizando distintos tipos de atributos empleados en la medición de la red y los mensajes.

* El **primer subobjetivo** se basa en **la teoría del capital social de la red**. Aquí, he planteado la hipótesis de que las medidas de centralidad ponderadas serían más efectivas para capturar la importancia de las interacciones entre los estudiantes y predecir el abandono de la asignatura. Se espera que estas medidas ponderadas proporcionen una visión más precisa de la influencia de los estudiantes en la red y ayuden a identificar patrones relevantes para la predicción.
* El **segundo subobjetivo** se basa **en la teoría de la estructura social de la red**. En este caso, he planteado la hipótesis de que los modelos que consideran medidas de centralidad globales, que miden la importancia de los nodos vecinos, serían más eficaces para predecir el abandono estudiantil en comparación con los modelos que solo consideran medidas locales.
* Por último, el **tercer subobjetivo** se basa en **la teoría del balance social de la red**. Aquí, he planteado la hipótesis de que las medidas de centralidad, junto con las medidas de sentimiento expresadas en los mensajes, serían más precisas para predecir el abandono estudiantil. Se espera que la combinación de medidas de centralidad y sentimiento proporcione una visión más completa de las interacciones sociales y los vínculos emocionales en la red, lo que permitirá una mejor comprensión de las dinámicas que conducen al abandono.

**[Diapositiva 13]**

Derivado de estos subobjetivos, y siguiendo los estudios previos sobre el análisis de redes sociales dinámicas, para evaluar la importancia y la posición de los nodos en la red. En este estudio, consideramos tres **medidas de centralidad principales**: el grado, la cercanía y la intermediación.

También consideramos estas **medidas de centralidad ponderadas**. Es decir, podemos distinguir entre medidas de centralidad sin pesos y con pesos. Las medidas sin pesos consideran solo la cantidad de nodos vecinos, mientras que las medidas con pesos también incluyen la frecuencia de interacción con los nodos vecinos.

Además, las medidas de centralidad mencionadas anteriormente se obtuvieron **como medidas de centralidad local y global.** Las medidas de centralidad local se centran en la relación de un nodo con sus vecinos directos. Por otro lado, las medidas de centralidad global o híbridas tienen en cuenta las interacciones de los nodos vecinos a los que un nodo está conectado.

**[Diapositiva 14]**

En cuanto a las **medidas de sentimiento y emoción**, utilizamos métricas específicas para evaluar el tono emocional y la carga afectiva de los mensajes de texto.

Finalmente, utilizamos una clasificación emocional basada en la región de Russell, que nos permite categorizar los textos en emociones específicas, como neutro, relajado, feliz, triste o enfadado.

**[Diapositiva 15]**

En cuanto al procesamiento y **análisis de datos**, se utilizó los principales algoritmos de aprendizaje supervisado como la Regresión Logística, SVM (Support Vector Machine), Decision Tree Classifier, KNN (K-Nearest Neighbors) y MLP Classifier (Multi-Layer Perceptron).

Además, se evaluó la predicción utilizando **métricas diversas** como la precisión, exhaustividad, puntuación F1 y el área bajo la curva ROC.

**[Diapositiva 16]**

También se desarrolló **índices de evaluación** específicos para analizar el efecto de las configuraciones temporales de las redes sociales, como el **índice de amplitud o cobertura** de la información que mide la proporción del tiempo total analizado, el **índice de unidad** de información que evalúa el grado de subdivisión en bloques de distinto tamaño y **el índice de cantidad** de información que es una medida que combina el grado de subdivisión en bloques y la proporción de cobertura temporal.

**[Diapositiva 17]**

El propósito de estos índices es realizar una evaluación exhaustiva de los objetivos e hipótesis propuestos en mi investigación. Me permitieron comparar la precisión de los algoritmos de aprendizaje en diferentes configuraciones de subdivisión en bloques y según los distintos tipos de atributos utilizados.

**(CONTINÚA)**

**[Diapositiva 18]**

Por cuestiones de tiempo, en esta presentación me enfocaré en mostrar los resultados clave de la investigación. Sin embargo, es importante destacar que el conjunto de configuraciones de bloques temporales y atributos generó un total de **352 posibles modelos**. Considerando los 5 algoritmos de aprendizaje supervisado utilizados, se entrenaron y evaluaron un total de **1760 modelos** de datos para realizar los resultados.

**[Diapositiva 19]**

Entre los hallazgos más destacados, en relación al **Objetivo 1,** se constató que al emplear un rango de tiempo **de 60 a 90 días** se lograban predicciones precisas sobre el abandono de la asignatura.

**[Diapositiva 20]**

Además, se evidenció que a medida que se aumentaba **el porcentaje de cobertura temporal**, también se incrementaba la efectividad de todos los tipos de algoritmos utilizados.

**[Diapositiva 21]**

Sin embargo, se identificó un **punto óptimo** de eficacia en la predicción que **no** se ubicaba en el máximo nivel **de unidad de información**, lo cual sugiere la existencia de un equilibrio entre la cantidad de información analizada y la capacidad predictiva del modelo.

La precisión mejora contra mayor es el rango de información utilizada pero la subdivisión en bloques puede optimizar la precisión mientras no se reduzca demasiado la unidad de información o tamaño de la red.

**[Diapositiva 22]**

En relación con al tipo de subdivisión de la red social en **bloques temporales, Objetivos 2 y 3,** observamos una ligera ventaja en la precisión de los algoritmos cuando utilizamos **bloques encadenados**, especialmente en los rangos de tiempo de **60 y 90** días.

**[Diapositiva 23]**

La subdivisión en **bloques encadenados** aumenta la precisión a medida que aumenta el rango de tiempo y el tamaño de la red.

**[Diapositiva 24]**

Además, la **subdivisión en bloques encadenados** muestra una eficacia superior en comparación con la subdivisión en bloques secuenciales a medida que aumenta la **cantidad de información**.

Esto indica que la forma en que se segmenta el tiempo en el análisis tiene un impacto significativo en la precisión de las predicciones.

**(CONTINÚA)**

**[Diapositiva 25]**

En relación con el **subobjetivo 1** de nuestro estudio, que se basa en la hipótesis del Capital Social Cognitivo, el propósito era analizar la eficacia de las predicciones utilizando **medidas centrales ponderadas**, es decir, considerando pesos asignados a las interacciones en la red social.

Los resultados muestran que estas medidas de centralidad con pesos tienden a ser igualmente efectivas para todos los algoritmos analizados. Sin embargo, se observó una ligera superioridad de las medidas con pesos en **rangos de tiempo superiores al 50%**, es decir de 60 a 120 días.

**[Diapositiva 26]**

Sin embargo, al aumentar los **índices de amplitud y tamaño de la unidad** de información se observa un incremento en la precisión similar entre las medidas de centralidad sin pesos y las medidas ponderadas.

Esto sugiere que, para nuestro análisis, la inclusión de pesos en las interacciones **no** proporciona una ventaja significativa en la precisión de las predicciones.

**[Diapositiva 27]**

Igualmente, a medida que se dispone de una **cantidad** de información superior **al 50 por ciento**, se observa que las medidas de centralidad se vuelven más precisas, siendo superiores solo en ocasiones las medidas ponderadas.

**[Diapositiva 28]**

Por otro lado, en relación con el **subobjetivo 2**, se hipotetizó que **las medidas de centralidad globales,** las cuales evalúan la posición de los nodos vecinos en la red, resultarían ser más efectivas para capturar la estructura de la red del estudiante y predecir el abandono de la asignatura.

Sin embargo, se observó que las medidas de centralidad globales son igualmente precisas para los distintos tipos de algoritmos de aprendizaje supervisado, pero ligeramente superiores a las medidas locales cuando se considera el **rango de tiempo máximo**.

**[Diapositiva 29]**

Aunque, se constató que tanto las medidas locales como las globales aumentan su precisión con la amplitud y tamaño de la unidad de información, sin embargo, las medidas globales suelen ser más precisas.

**[Diapositiva 30]**

En resumen, se ha observado que las medidas de centralidad globales tienden a ser más beneficiosas cuando se dispone de una mayor cantidad de información.

**(CONTINÚA)**

**[Diapositiva 31]**

Finalmente, en relación con el **subobjetivo 3,** se hipotetizó que la incorporación de **medidas emocionales**, como el análisis de sentimiento, en conjunto con las medidas de centralidad mejoraba la capacidad predictiva del modelo.

Confirmando esta hipótesis, se observó que la precisión al utilizar medidas de centralidad y sentimiento de manera conjunta era similar o superior en la mayoría de los algoritmos de aprendizaje, y que mejoraba especialmente en los **rangos de tiempo intermedios** (50-75%).

**[Diapositiva 32]**

Al incrementar el **rango de tiempo** analizado y **el tamaño de la unidad** de análisis de la red social, el uso conjunto de medidas emocionales y de centralidad mejora la capacidad predictiva del modelo.

**[Diapositiva 33]**

Además, las medidas de centralidad y de sentimiento conjuntamente mejoran la predicción cuanto mayor **cantidad de información** es analizada.

Esto indica que las emociones desempeñan un papel importante en las decisiones de los estudiantes y pueden influir en su probabilidad de abandonar una asignatura.

**[Diapositiva 34]**

Si bien estos resultados son prometedores, es importante tener en cuenta las **limitaciones** del estudio.

* El tamaño limitado y sesgado de los datos analizados plantea la necesidad de realizar investigaciones futuras en conjuntos de datos más amplios y diversos.
* Asimismo, se sugiere explorar diferentes medidas de cohesión de la red y algoritmos de detección de comunidades con el fin de obtener una comprensión más completa de la estructura de la red estudiantil.
* Además, para comprender mejor la relación entre las emociones y la posición de los nodos en la red, se sugiere integrar medidas de centralidad y análisis de sentimiento.
* Dado que los conjuntos de datos pueden presentar desequilibrios en las clases objetivo, se propone utilizar técnicas de balanceo de datos, como SMOTE.
* Igualmente, para comprender qué características son más relevantes en el proceso de predicción, se recomienda utilizar técnicas como SHAP.
* Finalmente, las técnicas de ensamblaje pueden mejorar la precisión de las predicciones al combinar múltiples modelos de aprendizaje.

**(CONTINÚA)**

**[Diapositiva 35]**

En conclusión, este estudio nos ha proporcionado valiosos conocimientos sobre la predicción del abandono de asignaturas en entornos educativos.

Los resultados obtenidos tienen importantes implicaciones para mejorar los sistemas de apoyo a los estudiantes y diseñar intervenciones tempranas para prevenir el abandono.

Esperamos que este estudio sirva como punto de partida para futuras investigaciones y avances en el campo de la educación y la predicción del comportamiento estudiantil.

Algunas aplicaciones posibles son: la identificación temprana de estudiantes en riesgo, el diseño de estrategias de enseñanza personalizadas, la mejora de la experiencia del estudiante en entornos de aprendizaje en línea, y el desarrollo de sistemas de recomendación personalizados.

**[Diapositiva 36]**

**Muchas gracias** por su atención, y quedo disponible a sus preguntas y comentarios.

**(FIN)**