



## **PROPUESTAS DE TRABAJO FIN DE GRADO Y TRABAJO FIN DE MASTER**

**CURSO 2019-2020**

### **Listado:**

- **SABERMED**
- **Evaluación de modelos de clasificación binaria y multiclase utilizando diferentes métricas.**
- **Clustering genéticos.**
- **Métricas para medir el solapamiento entre clases en problemas de clasificación.**
- **Datos simbólicos y su uso para la representación y clasificación de textos.**
- **Red neuronal para análisis de sentimiento en español.**
- **Soluciones software para datos desequilibrados.**
- **Sistema de recomendación de animes basado en grafos de conocimiento.**
- **Clasificación en tiempo real de patrones de comportamiento en ganado mediante análisis de series temporales.**
- **Métodos de reducción de la dimensionalidad como herramienta de diagnosis y evaluación de modelos de predicción.**
- **Time Series for Drift Detection.**
- **Métodos de reducción de la dimensionalidad y su proyección en el plano.**
- **Sistema de detección y clasificación de oportunidades en el sector viajes.**
- **User Behavior Analytics para detectar comportamientos anómalos y su implantación en sistemas de gestión de identidades.**
- **Red Neuronal para transferencia de estilos en imágenes de anime / manga.**
- **Plataforma de tráfico rodado.**

<b>Título:</b>	SABERMED
<b>Descripción:</b>	Cálculo de la reputación de webs en un ámbito determinado. Detección de fraude.
<b>Pérfil:</b>	Informático con conocimientos deseable en Python. Web scraping, minería de texto.
<b>Objetivos:</b>	Desarrollar un prototipo capaz de realizar predicciones aceptables sobre sitios web fraudulentos.
<b>Tecnologías:</b>	Python.

<b>Título:</b>	Evaluación de modelos de clasificación binaria y multiclase utilizando diferentes métricas.
<b>Descripción:</b>	Los modelos de clasificación son una parte importante del Machine Learning, cuyo objetivo es hacer una división perfecta en diferentes clases. Algunos de estos modelos son: Regresión logística, Árboles de decisión o Random Forest. Además, existen varios tipos de clasificación, como son la clasificación binaria o la clasificación multiclase. Para evaluar este tipo de modelos, existen distintas métricas, las más utilizadas son F1-score o Accuracy; sin embargo, existe otra métrica no tan conocida pero que en el caso de datos desbalanceados es más informativa que las anteriores, dicha métrica es Matthews Correlation Coefficient (MCC). El objetivo del trabajo sería evaluar diferentes modelos sobre unos datos, balanceados o desbalanceados, usando distintas métricas.
<b>Pérfil:</b>	TFG/TFM - Matemáticas
<b>Objetivos:</b>	Septiembre 2020
<b>Tecnologías:</b>	Python o R

<b>Título:</b>	Clustering genéticos.
<b>Descripción:</b>	Desarrollar algoritmos genéticos distribuidos para la automatización de agrupaciones.
<b>Pérfil:</b>	Informático con conocimientos en Python.
<b>Objetivos:</b>	Demostración empírica de los resultados obtenidos. Mejora de algoritmos genéticos tradicionales.
<b>Tecnologías:</b>	Python.

<b>Título:</b>	Métricas para medir el solapamiento entre clases en problemas de clasificación.
<b>Descripción:</b>	Uno de los problemas a los que se enfrentan las técnicas de clasificación es el solapamiento entre las clases que se quieren diferenciar. Existen diversas técnicas para cuantificar el solapamiento entre datos aunque la mayor parte de ellas requieren que los datos sigan una distribución de probabilidad. No obstante, también existen otros índices

	como R-value (Oh, S. (2011). A new dataset evaluation method based on category overlap. Computers in Biology and Medicine, 41(2), 115-122.) y Constraint-based measure (Adam, A., & Blockeel, H. (2017). Constraint-based measure for estimating overlap in clustering. In Benelearn 2017: Proceedings of the Twenty-Sixth Benelux Conference on Machine Learning (pp. 54-61).) que estiman el solapamiento sin necesidad de una distribución. Se pretende evaluar la calidad de estos dos índices comparando sus valores con el valor de solapamiento real que se puede obtener cuando se trabaja con distribuciones normales. Habría que programar todos los métodos, simular datos y compararlos.
<b>Péfil:</b>	Matemático, informático o matemático+informático.
<b>Objetivos:</b>	Comparar y evaluar la calidad de los dos índices de solapamiento nombrados.
<b>Tecnologías:</b>	R o Python (preferiblemente python)

<b>Título:</b>	Datos simbólicos y su uso para la representación y clasificación de textos
<b>Descripción:</b>	<p>El trabajo pretende estudiar y revisar la representación de textos a través de variables simbólicas intervalo-evaluadas. También se quiere aplicar y probar métodos de clasificación de textos para la representación simbólica de los mismos. Se desea comparar el método estudiado con los métodos de representación y clasificación a través de datos clásicos no simbólicos. Existen varios artículos que tratan este tema que aseguran que el modelo propuesto reduce la dimensión de los vectores de características para guardar la información de los textos y aumenta la eficiencia computacional del posterior algoritmo de clasificación aplicado.</p> <p>Por tanto, el trabajo consistiría primero en estudiar los datos simbólicos y en concreto las variables intervalo-evaluadas. También se realizaría una revisión del estado del arte para la representación y clasificación de textos a través de datos clásicos.</p> <p>El foco de la investigación se centraría en una revisión de los métodos de representación y clasificación de textos a través de datos simbólicos ya existentes en la literatura, y la revisión completa de uno de ellos, a poder ser el más actual.</p> <p>Para valorar las posibles mejoras del método estudiado, se realizarían diversos experimentos con los que evaluar la precisión de clasificación y eficiencia computacional.</p>
<b>Péfil:</b>	Matemático/Informático
<b>Objetivos:</b>	Colaboración proyecto SABERMED.
<b>Tecnologías:</b>	Python/R.

<b>Título:</b>	Red neuronal para análisis de sentimiento en español.
<b>Descripción:</b>	A partir de alguna de las redes neuronales para NLP que aparecen en el estado del arte (BERT, RoBERTa, XLNet, ...) y haciendo fine-tuning con ellos, entrenar una red que sea capaz de analizar el sentimiento en texto en español. Habría que estudiar las diferentes

	redes ya preentrenadas, elegir la mejor para este objetivo, buscar dataset que puedan ser usados para el proceso de fine-tuning y realizar el entrenamiento. Además de los experimentos necesarios.
<b>Pérfil:</b>	Informático.
<b>Objetivos:</b>	Publicación.
<b>Tecnologías:</b>	Python. Si han trabajado anteriormente con TF o PyTorch mejor.

<b>Título:</b>	Soluciones software para datos desequilibrados
<b>Descripción:</b>	<p>En el ámbito de los datos desbalanceados o desequilibrados, la técnica de remuestreo SMOTE (Chawla, N. V., Bowyer, K. W., Hall, L. O., &amp; Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. <i>Journal of artificial intelligence research</i>, 16, 321-357.) es una de las más conocidas. Desde su publicación en el 2002, se han creado más de 90 versiones basadas en SMOTE (Fernández, A., Garcia, S., Herrera, F., &amp; Chawla, N. V. (2018). SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. <i>Journal of artificial intelligence research</i>, 61, 863-905.). Algunas de ellas como BorderlineSMOTE, SVMSMOTE y KmeansSMOTE están programadas en Python (también hay versiones en R, hay un paquete llamado smotefamily). Otra técnica de remuestreo famosa que nació con el boom de SMOTE fue ADASYN, también programa en Python y R. Se propone:</p> <ul style="list-style-type: none"> <li>• Realizar una revisión de las variantes de SMOTE y seleccionar las más famosas o más llamativas para el alumno</li> <li>• Escoger 2 ó 3 conjuntos de datos con distinto ratio de desequilibrio (en Kaggle o en UCI repository)</li> <li>• Aplicar en los conjuntos de datos las distintas técnicas de remuestreo (nótese que las técnicas de sobremuestreo de la clase minoritaria funcionan mejor si se acompañan de un bajomuestreo de la clase mayoritaria)</li> <li>• Aplicar el mismo modelo de clasificación (por ejemplo SVM o árbol de decisión) tras el remuestreo</li> <li>• Comparar los resultados</li> </ul>
<b>Pérfil:</b>	<p>Matemático, informático o matemático+informático.</p> <ul style="list-style-type: none"> <li>• El perfil matemático se puede centrar más en la interpretación de los resultados y realizar gráficos donde se muestren claramente las muestras sintéticas generadas por cada método para poder comparar las diferentes maneras de generar las muestras.</li> <li>• El perfil informático puede programar una variedad de SMOTE que no esté ya programada en R o Python.</li> </ul>
<b>Objetivos:</b>	Comparar las distintas técnicas de remuestreo basadas en SMOTE
<b>Tecnologías:</b>	R o Python (preferiblemente python)

<b>Título:</b>	Sistema de recomendación de animes basado en grafos de conocimiento.
<b>Descripción:</b>	A partir de la información sobre los diferentes animes que se puede extraer de la página MyAnimeList (tiene API y también existe un dataset en kaggle), construir un grafo de conocimiento a partir del cual poder ofrecer recomendaciones en base a los gustos e histórico de un usuario. Sería interesante no solo basarse en la opinión de otros usuarios, sino también en la descripción aportada sobre el anime en la sinopsis. Para poder buscar la similitud entre los diferentes animes y grupos de usuarios, se usarán distintos algoritmos pertenecientes al estado del arte de teoría de grafos. Este sistema ofrecerá además un sistema web de visualización donde se podrá ver el grafo e interactuar con él.
<b>Pérfil:</b>	Informático.
<b>Objetivos:</b>	Este trabajo puede ser suficiente tanto para un TFG como un TFM.
<b>Tecnologías:</b>	Python, Base de datos de grafos.

<b>Título:</b>	Clasificación en tiempo real de patrones de comportamiento en ganado mediante análisis de series temporales.
<b>Descripción:</b>	Se desea crear una herramienta que permita clasificar, en tiempo real, la conducta de animales monitorizados mediante dispositivos IoT dentro de una serie de comportamientos predefinidos. Para ello, se ha de aplicar técnicas de análisis de series temporales a un conjunto de datos obtenidos mediante acelerómetros. Para la clasificación, se dispone de un conjunto de etiquetas temporales en base al comportamiento del animal en todo momento.
<b>Pérfil:</b>	Matemático y/o informático.
<b>Objetivos:</b>	si se realiza la parte de procesamiento en tiempo real, puede ser trabajo suficiente como para un TFM o para dos TFG (matemáticas + informática).
<b>Tecnologías:</b>	Para la parte de clasificación, con emplear python/R sería suficiente. Para el procesamiento en tiempo real: ¿spark+kafka?

<b>Título:</b>	Métodos de reducción de la dimensionalidad como herramienta de diagnosis y evaluación de modelos de predicción.
<b>Descripción:</b>	Utilizar los distintos métodos de reducción de la dimensionalidad para la diagnosis de modelos. De manera que puedas ver la localización gráficamente los datos. Así puedes comprobar si el modelo clasifica los nuevos datos de test de manera correcta por su distribución o se trata de predicciones espúreas, tanto en aciertos como en fallos.
<b>Pérfil:</b>	Matemático/Informático.
<b>Objetivos:</b>	TFM que conozca los métodos de reducción de la dimensionalidad.
<b>Tecnologías:</b>	R o python, indistinto.


<b>Título:</b>	Time Series for Drift Detection
<b>Descripción:</b>	Aplicar modelización de series temporales para crear un nuevo método que detecte la necesidad de actualizar un modelos de streaming machine learning. De esta manera se pretende detectar la naturaleza de la falta de adecuación del modelo actual mediante modelos ARIMA, principalmente modelos AR.
<b>Pérfil:</b>	Matemático/estadístico.
<b>Objetivos:</b>	TFM que conozca los métodos de reducción de la dimensionalidad o TFG que conozca modelos sencillos de series temporales.
<b>Tecnologías:</b>	R o python, indistinto.

<b>Título:</b>	Métodos de reducción de la dimensionalidad y su proyección en el plano.
<b>Descripción:</b>	Utilizar los distintos métodos de reducción de la dimensionalidad para la representación en un plano. Comprobar las diferencias, ventajas y desventajas de los métodos. Un estado del arte de métodos de reducción de la dimensionalidad.
<b>Pérfil:</b>	Matemático/Informático.
<b>Objetivos:</b>	TFM que conozca los métodos de reducción de la dimensionalidad.
<b>Tecnologías:</b>	R o python, indistinto.

<b>Título:</b>	Sistema de detección y clasificación de oportunidades en el sector viajes.
<b>Descripción:</b>	Crear un sistema que clasifique las oportunidades disponibles en el sector viajes. Además, establecer un ranking en base a los beneficios proporcionados por las mismas.
<b>Pérfil:</b>	Matemático.
<b>Objetivos:</b>	TFM con conocimientos de machine learning. Deseable teoría de la decisión, funciones de utilidad, etc. pero no imprescindible.
<b>Tecnologías:</b>	Python

<b>Título:</b>	User Behavior Analytics para detectar comportamientos anómalos y su implantación en sistemas de gestión de identidades.
<b>Descripción:</b>	User Behavior Analytics esta presente en muchos aspectos de la ciberseguridad. Algunos ejemplos son: Detectar ciberataques o en sistemas de control de accesos para detectar suplantaciones de identidad o implementar la autenticación continua. Actualmente se ha desarrollado un chat en el que se se han recogido información de comportamiento de 10 usuarios. Esta información contiene todos los eventos que es posible recoger desde

	Javascript de dichos usuarios, entre los que cabe destacar: Pulsaciones de teclado (KeyUp, KeyDown, Keypress), movimientos de ratón (MouseMove, MouseClick, MouseDown) o en el entorno móvil pulsaciones de pantalla (Touchscreen) y sensores (Giroscopio y acelerometro). Con esta información se pretende modelar el comportamiento de los usuarios con el objetivo de saber si un usuario está sufriendo una suplantación de identidad. Este proceso puede llevarse a cabo tanto analizando bloques de datos a posteriori o en Real Time.
<b>Pérfil:</b>	TFG/TFM - Informática/Matemáticas
<b>Objetivos:</b>	Agosto/o Diciembre
<b>Tecnologías:</b>	En el caso de enfocarlo en matemáticas sería únicamente Python o R. En el caso de informática se podría meter el real time con ELK - Kafka - Spark/python

<b>Título:</b>	Red Neuronal para transferencia de estilos en imágenes de anime / manga.
<b>Descripción:</b>	<p>Al igual que en otras disciplinas artísticas, cada autor de anime / manga tiene su propio estilo, y sus obras son fácilmente reconocidas. Una red neuronal (¿CycleGAN?) podría ser entrenada para que sea capaz de, dada la imagen del personaje de un autor, transformarla para que pareciera realizada por otro. La siguiente imagen muestra un ejemplo, hecho por alguien, de la idea (el personaje original es el de la esquina superior izquierda):</p>  <p>En este ejemplo hay ciertos cambios, como de vestimenta, que no se espera que realice la red, pero sí el estilo de la cara. En una primera aproximación se podría usar como referencia el estilo de dibujo de Hayao Miyazaki de estudio Ghibli ya que es muy representativo, he intentado transformar otras imágenes a su estilo.</p>
<b>Pérfil:</b>	Matemático e Informático.
<b>Objetivos:</b>	Preparar los datos, crear la red y realizar experimentos, supera con creces el alcance de un TFM o TFG. En caso de tener resultados prometedores, de aquí puede salir al menos un paper. Tecnologías: Buen manejo de Python (sin esta base, se alargaría demasiado) y deseable saber algo de Redes Neuronales.

<b>Tecnologías:</b>	Python.
---------------------	---------

<b>Título:</b>	Plataforma de tráfico rodado.
<b>Descripción:</b>	Mejora de herramienta de simulación de tráfico. Integración con big data y distribución de cálculo de rutas. Puesta a punto de la plataforma y actuación. Mejoras abiertas según capacidades del desarrollador.
<b>Pérfil:</b>	Informático con conocimientos en Java J2SE.
<b>Objetivos:</b>	Elaboración de una nueva versión actualizada y potente de una plataforma de simulación de tráfico.
<b>Tecnologías:</b>	Java última versión.