

LISTA DE PROBLEMAS A RESOLVER PARA EL PERCEPTRON MULTICAPA:

1. Predicción del Tipo de Vino según Características Químicas:

Dominio: Química/Alimentación.

Dataset: Utilizar el dataset "Wine Quality" (disponible en UCI Machine Learning Repository o Kaggle). Contiene características químicas de diferentes vinos (acidez, azúcar residual, alcohol, etc.) y una etiqueta de calidad (o tipo de vino si usas la versión de clasificación).

Tarea: Clasificar el tipo de vino, basándose en sus características químicas.

Utilidad: Demuestra cómo MLPs pueden analizar datos químicos para clasificación, relevante en industrias como la alimentaria, la agricultura, y el control de calidad.

2. Detección de Fraude en Tarjetas de Crédito (Simplificado):

Dominio: Finanzas/Seguridad.

Dataset: Buscar un dataset de transacciones de tarjetas de crédito etiquetadas como fraudulentas o no fraudulentas (hay datasets públicos simplificados en Kaggle o repositorios de datos). Importante: usar un dataset simplificado para que sea manejable para un MLP básico.

Tarea: Clasificar transacciones como fraudulentas o no fraudulentas basándose en características como monto de la transacción, ubicación, hora, frecuencia, etc. (las características dependerán del dataset).

Utilidad: Muestra la aplicación de MLPs en detección de anomalías y clasificación binaria en un contexto de seguridad financiera. Es un problema muy relevante en la industria.

3. Clasificación de Géneros Musicales a partir de Características de Audio Simplificadas:

Dominio: Música/Procesamiento de Señales.

Dataset: Utilizar un dataset de música con etiquetas de género (puede hacerse para 2 o más géneros de música: clásica, rock, pop, jazz, etc.). Para simplificar, puedes usar un dataset con características de audio ya extraídas (como "features" espectrales básicos, o incluso características muy simples como energía y ritmo promedio). Si es más avanzado, podrían

extraer características simples de audio ellos mismos usando bibliotecas de procesamiento de audio.

Tarea: Clasificar el género musical de una canción basándose en sus características de audio.

Utilidad: Introduce al procesamiento de audio y clasificación en un dominio creativo y atractivo como la música. Muestra cómo MLPs pueden aprender patrones en datos de audio.

4. Predicción de la Calidad del Aire en Base a Datos de Sensores Ambientales:

Dominio: Medio Ambiente/Ciencia de Datos.

Dataset: Buscar un dataset de calidad del aire que contenga mediciones de varios contaminantes (PM2.5, PM10, CO, SO2, NO2, O3, etc.) y posiblemente datos meteorológicos (temperatura, humedad, viento). Hay datasets públicos disponibles de estaciones de monitoreo de calidad del aire de diferentes ciudades o regiones.

Tarea: Predecir un índice de calidad del aire (o clasificar la calidad del aire en categorías: buena, moderada, mala, etc.) basándose en las mediciones de contaminantes y datos meteorológicos.

Utilidad: Aplica MLPs a un problema ambiental crucial, mostrando cómo pueden analizar datos ambientales para predicción y clasificación, útil para alertas de salud pública y políticas ambientales.

5. Reconocimiento de Dígitos Manuscritos Simplificado (con un dataset menos común):

Dominio: Visión por Computadora/Reconocimiento de Patrones.

Dataset: En lugar del clásico MNIST, buscar un dataset de dígitos manuscritos menos común que incluye letras y dígitos, o mejor aún crear un dataset pequeño de dígitos manuscritos propios o de compañeros.

Tarea: Clasificar imágenes de dígitos manuscritos.

Utilidad: Aunque es un problema clásico, usar un dataset menos común puede hacerlo más interesante. Permite entender los fundamentos de la clasificación de imágenes con MLPs.

6. Clasificación de Noticias por Categoría Temática (Simplificado):

Dominio: Procesamiento del Lenguaje Natural/Información.

Dataset: Utilizar un dataset de titulares o noticias cortas etiquetadas por categoría temática (deportes, política, tecnología, etc.). Datasets más pequeños y simplificados son ideales para este proyecto inicial. Podrían incluso crear un dataset pequeño recopilando titulares de noticias de diferentes fuentes.

Tarea: Clasificar una noticia en su categoría temática basándose en el texto del titular o la noticia corta. Utilizar técnicas de vectorización de texto sencillas como Bag of Words o TF-IDF para representar el texto como entrada al MLP.

Utilidad: Introduce al NLP y la clasificación de texto en un contexto informativo. Muestra cómo MLPs pueden entender el tema general de un texto.

7. Predicción de la Demanda de Bicicletas Compartidas (Bike Sharing):

Dominio: Transporte/Ciudades Inteligentes.

Dataset: Utilizar un dataset de sistemas de bicicletas compartidas (hay datasets públicos como "Bike Sharing Dataset" en UCI o Kaggle). Contiene información sobre el uso de bicicletas compartidas por hora o día, junto con datos meteorológicos, día de la semana, festivos, etc.

Tarea: Predecir la cantidad de bicicletas que se alquilarán en una hora o día específico, basándose en las características temporales y ambientales. Esto puede ser un problema de regresión (predecir el número exacto) o clasificación (clasificar la demanda en categorías: baja, media, alta).

Utilidad: Aplica MLPs a un problema de planificación y optimización de recursos en un contexto de movilidad urbana sostenible.

8. Detección de Spambots en Redes Sociales (Simplificado):

Dominio: Redes Sociales/Seguridad/Detección de Anomalías.

Dataset: Buscar un dataset de perfiles de usuarios de redes sociales etiquetados como "bot" o "no bot". Estos datasets pueden ser más difíciles de encontrar, pero existen datasets académicos o simplificados para fines educativos. Alternativamente, podrían crear un dataset simplificado de características de perfiles de usuario simulados o reales (anonimizados) y etiquetarlos manualmente.

Tarea: Clasificar perfiles de usuario como bots o no bots basándose en características del perfil (número de seguidores, actividad, ratio seguidores/seguídos, contenido de los posts, etc.).

Utilidad: Muestra la aplicación de MLPs en la detección de cuentas automatizadas y spam en redes sociales, un problema relevante para la seguridad y la calidad de la información en línea.

9. Clasificación de Emociones Básicas a partir de Texto Corto (Ej. Tweets, Frases Cortas):

Dominio: Psicología/Procesamiento del Lenguaje Natural/Análisis de Sentimientos (más allá de positivo/negativo).

Dataset: Buscar o crear un dataset de frases cortas o tweets etiquetados con emociones básicas (alegría, tristeza, ira, miedo, sorpresa, neutral). Existen datasets de emociones en texto disponibles públicamente, o se puede crear uno simplificado.

Tarea: Clasificar la emoción expresada en un texto corto.

Utilidad: Extiende el análisis de sentimientos a un espectro más amplio de emociones, mostrando cómo MLPs pueden capturar matices emocionales en el lenguaje.

10. Predicción de Fallos en Maquinaria Industrial (Mantenimiento Predictivo Simplificado):

Dominio: Industria/Mantenimiento/Internet de las Cosas (IoT).

Dataset: Buscar un dataset de datos de sensores de maquinaria industrial (temperatura, vibración, presión, corriente, etc.) etiquetados con información sobre fallos (o tiempo hasta el fallo). Hay datasets públicos simplificados de mantenimiento predictivo.

Tarea: Predecir si una máquina fallará en un futuro cercano (clasificación binaria: fallo/no fallo) o predecir el tiempo restante hasta el fallo (regresión), basándose en los datos de los sensores.

Utilidad: Introduce al mantenimiento predictivo, una aplicación importante del aprendizaje automático en la industria para reducir costos y mejorar la eficiencia operativa.