**Documento Técnico: Selección de Características y Evaluación de Modelos con Algoritmos Genéticos y Knowledge Base**

**Introducción**

Este documento describe el proceso de selección de características y evaluación de modelos utilizando algoritmos genéticos y una base de conocimiento (Knowledge Base) en el contexto de un problema de clasificación. El objetivo es identificar las características más relevantes para predecir el estado de un préstamo (Loan\_Status) y seleccionar el mejor modelo de clasificación basado en el F1-score.

**Descripción del Problema**

Se dispone de un conjunto de datos que contiene información sobre préstamos, incluyendo características como el género del solicitante, estado civil, ingresos, historial crediticio, entre otros. La variable objetivo es Loan\_Status, que indica si el préstamo fue aprobado (Y) o no (N). El objetivo es predecir esta variable utilizando un modelo de clasificación.

**Proceso de Selección de Características**

**1. Preprocesamiento de Datos**

El primer paso consiste en preprocesar los datos para manejar valores faltantes y codificar variables categóricas. Se utiliza un ColumnTransformer para aplicar diferentes transformaciones a las características numéricas y categóricas:

* **Características Numéricas**: Se imputan los valores faltantes utilizando la mediana.
* **Características Categóricas**: Se imputan los valores faltantes con la moda y se codifican utilizando OrdinalEncoder.

**2. División de Datos**

Los datos se dividen en conjuntos de entrenamiento y prueba, manteniendo la proporción de la variable objetivo mediante stratify.

**3. Algoritmo Genético para Selección de Características**

Se implementa un algoritmo genético para seleccionar las características más relevantes. El algoritmo genético sigue los siguientes pasos:

1. **Inicialización de la Población**: Se genera una población inicial de cromosomas, donde cada cromosoma representa una combinación de características.
2. **Evaluación de la Aptitud (Fitness)**: Se evalúa la aptitud de cada cromosoma utilizando una función de fitness que entrena un modelo de clasificación (en este caso, un RandomForestClassifier) y calcula el F1-score en el conjunto de prueba.
3. **Selección**: Se seleccionan los cromosomas más aptos utilizando la selección por ruleta.
4. **Cruce**: Se aplica un cruce de un punto para generar descendencia.
5. **Mutación**: Se aplica una mutación bit a bit con una tasa de mutación dada.
6. **Actualización de la Población**: La población se actualiza con la descendencia mutada.

**4. Selección de Características**

Después de ejecutar el algoritmo genético, se obtienen las características seleccionadas:

**Evaluación de Modelos con Knowledge Base**

**1. Knowledge Base Agent**

Se implementa una base de conocimiento (KnowledgeBase) que almacena los modelos y sus respectivos F1-scores. La base de conocimiento permite "decir" (tell) un modelo y su puntuación, y "preguntar" (ask) por el mejor modelo.

**2. Evaluación de Modelos**

Se evalúan tres modelos de clasificación (RandomForestClassifier, XGBoost, y LogisticRegression) utilizando las características seleccionadas. Cada modelo se entrena y se evalúa en el conjunto de prueba, y su F1-score se almacena en la base de conocimiento.

**3. Selección del Mejor Modelo**

Finalmente, se consulta la base de conocimiento para obtener el mejor modelo basado en el F1-score.

**Resultados**

El algoritmo genético seleccionó las siguientes características: [1, 2, 4, 5, 7, 8, 9]. El mejor modelo fue LogisticRegression con un F1-score de 0.8627.

**Conclusión**

Se ha presentado un enfoque para la selección de características y evaluación de modelos utilizando algoritmos genéticos y una base de conocimiento. El algoritmo genético permitió identificar las características más relevantes, mientras que la base de conocimiento facilitó la selección del mejor modelo basado en el F1-score. Este enfoque puede ser aplicado a otros problemas de clasificación donde la selección de características y la evaluación de modelos son críticas en diferentes ámbitos como finanzas, salud, etc.