Imagen que contiene Forma

Descripción generada automáticamente

**Roberto Matilla Augustinus**

**Inteligencia**

**Artificial**

**Clasificadores con Weka**

**Parte avanzada**

Texto

Descripción generada automáticamenteImagen que contiene Texto

Descripción generada automáticamente

**Roberto Matilla Augustinus**

# Parte 1: Predicción

# Datos originales

Tabla

Descripción generada automáticamente

# 1.2 Modelo K-nn

Tabla

Descripción generada automáticamente

Con la implementación de este modelo, podemos ver que se han cometido **dos fallos** respecto a los datos originales. Como podemos ver en la imagen superior, el **caso 2 y el caso 7** son resultados erróneos, por lo que se han clasificado un 80% de los casos de forma correcta y un 20% de forma incorrecta.

# Modelo de Bayes

Tabla

Descripción generada automáticamente

Con la implementación de este modelo, podemos ver que se han cometido **dos fallos** respecto a los datos originales, al igual que en el modelo anterior. Como podemos ver en la imagen superior, el **caso 6 y el caso 9** son resultados erróneos, por lo que se han clasificado un 80% de los casos de forma correcta y un 20% de forma incorrecta.

# Modelo de árboles

Tabla

Descripción generada automáticamente

Con la implementación de este modelo, podemos ver que se ha cometido **solamente un fallo**, a diferencia de los dos anteriores modelos. Como podemos ver en la imagen superior, el erróneo ha sido el **caso** **9**, haciendo así que se consiga un 90% de resultados correctos y un 10% de resultados erróneos.

Podemos deducir que este modelo es el más adecuado de los tres probados en esta práctica.

# Parte 2: Parámetros

# Visualización de todos los atributos. Archivo diabetes.arff

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

# InfoGainAttributeEval y Ranker

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Como sabemos, el método **Ranker** ordena los atributos según su importancia, y nos muestra dos resultados diferenciados: **Average Merit** y **Average Rank**, como podemos ver en la imagen superior.

Comparando los datos de la imagen con la imagen del Apartado 1, fijándonos en **Average Rank**, podemos relacionar los resultados de la siguiente manera:

* El atributo **plas** tiene una desviación típica de 0 y está situado en la posición 1. Esto significa que en los 5 ciclos realizados, su posición no ha variado nada, considerando entonces que es el atributo con más importancia. Comparando con su gráfica, podemos corroborar estos datos ya que es la que mejor diferenciadas tiene las clases.
* Después del atributo plas, están los atributos **mass**, **age** e **insu**, que le siguen en cuanto a posición. En cuanto a la desviación típica de estos atributos, podemos ver que ronda cercana al uno, y que su nivel de importancia sigue siendo bastante alto. Esto tiene sentido comparándolo con las gráficas, ya que después del atributo **plas**, desde mi punto de vista hay gran diferenciación en cuanto a las clases en estas.
* Finalmente tenemos los atributos **preg**, **pedi**, **pres** y **skin**, que no consiguen un buen resultado, al igual que su diferenciación de clases en las gráficas.

# CfsSubsetEval y GreedyStepwise

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

En este caso, sabemos que **CfsSubsetEval** evalúa el valor de un subconjunto de atributos considerando la capacidad predictiva individual de cada característica junto con el grado de redundancia entre ellas, y nos muestra el porcentaje de veces que fueron seleccionados en los ciclos seleccionados.

Como podemos ver, los atributos **plas**, **mass** y **age** han sido seleccionados en todos los ciclos realizados, dando a entender que estos atributos son los mejores en cuanto a su criterio de funcionamiento. Estos resultados reafirman lo que se mostró en el método anterior, siendo también estos tres atributos los mejores.

Después de estos atributos, vemos que les siguen el **pedi** (60%) e **insu** (20%), como los mejores. Podemos observar que su porcentaje de selección es bastante inferior, y de nuevo, vuelve a reafirmar lo visto anteriormente y comparándolo con las gráficas obtenidas.

Finalmente, los atributos **preg**, **pres** y **skin** no han sido seleccionados en ninguno de los ciclos, queriendo decir que estos atributos no tienen la suficiente importancia según el criterio.