|  |  |  |
| --- | --- | --- |
| **Nombre**: Ruben Alejandro Deambrossi | | **Matrícula**: 3012348 |
| **Nombre del curso:**  Aprendizaje Automático | **Nombre del profesor**:  Rolando Evelio Pérez Versón | |
| **Módulo**: 2 | **Actividad**: 6 | |
| **Fecha**: 04-Jul-2022 | | |
| **Bibliografía**: | | |

**Título**: Regresión logística y árboles de decisión.

**Introducción**:

El siguiente reporte corresponde a la actividad 6 de Aprendizaje Automático.

**Desarrollo**:

1. Generación de modelo de regresión logística

Notebook Link:

1. Métricas de evaluación del modelo.

Notebook Link:

1. Análisis de la problemática del punto 1.
2. ¿Se puede resolver mediante un modelo de árbol de decisión?

Si, un modelo de árbol de decisión presenta una gran flexibilidad que lo hace apto para resolver el problema del punto 1.

1. ¿Consideras posible extender el análisis inicial, a una situación donde pudieran estar involucradas nuevas variables de salida? ¿Cómo lo harías?

Si, es posible, actualmente se realizaron dos etiquetas donde una corresponde a niveles de glucosa aceptable dentro del rango 70-180 y otra etiqueta para cualquier nivel fuera de ese rango. Se podría rediseñar el etiquetado para tener tres categorías, niveles aceptables 70-180, hipoglucemia <70 e hiperglucemia >180.

En dicho caso de agregar una etiqueta más, se debería usar un árbol de clasificación.

1. Evaluación del modelo de árbol de clasificación del punto 3.

Evaluación del modelo de regresión logística:

**Precision**: 0.8210922787193974, **Recall**: 0.7771836007130125, **F1**: 0.7985347985347985, **Accuracy**: 0.8023360287511231

Evaluación del modelo de árbol de decisión:

**Precision**: 1.0, **Recall**: 1.0, **F1**: 1.0, **Accuracy**: 1.0

**Conclusión**: Si bien los resultados de la evaluación del modelo de regresión logística son muy buenos ~ >0.80, no logra hacer una correcta predicción cuando el nivel de glucosa esta por debajo del rango normal. La matriz de confusión me indica que hay un gran porcentaje de predicción de clase 0 cuando realmente son 1.

En el caso de establecer 3 clases (normal, hipoglucemia e hiperglucemia) y usar un árbol de clasificación, los resultados nos indican que el modelo puede estar sobre ajustado ya que los diferentes resultados de la evaluación del modelo nos dan un perfecto 1.0. Si bien esto puede ser cierto, he realizado diferentes predicciones de valores que no están en el conjunto de datos de prueba y para todos ellos el resultado predicho fue correcto. Además el grafico del árbol de clasificación muestra una lógica sencilla pero evidente de predecir la clase correcta. Por lo cual puedo asumir que este modelo se adapto mejor al conjunto de datos y puede realizar una correcta predicción.

Personalmente siento que ejemplo se pudo adaptar mejor al utilizar un árbol de clasificación, tal vez los datos por su origen no se adaptaron bien al modelo de regresión logística, es decir describir los datos de manera binaria.