Logotipo, nombre de la empresa

Descripción generada automáticamente

Introducción a la Ciencia de Datos

Propuesta Trabajo Final 2024

**Contexto:**

Ustedes son un equipo de científicos de datos contratados por un prestigioso hospital para ayudar a predecir y comprender mejor los trastornos de la tiroides, en particular el hipertiroidismo y el hipotiroidismo. Los trastornos de la tiroides son afecciones que afectan el funcionamiento de la glándula tiroides, que es fundamental para regular el metabolismo, la energía y la salud en general. La predicción y comprensión efectivas de estas condiciones pueden mejorar significativamente los resultados de los pacientes al permitir un diagnóstico temprano y un tratamiento personalizado.

**Descripción del proyecto:**

El hospital ha proporcionado un conjunto de datos que contiene registros médicos detallados de los pacientes, incluidos resultados de análisis de sangre, historial médico y otros indicadores de salud relevantes. Su función como científico de datos es aprovechar este conjunto de datos para crear un modelo predictivo de trastornos de la tiroides y proporcionar información sobre los factores que contribuyen a estas afecciones. Al hacer esto, ayudará a los médicos a tomar decisiones más informadas, comprender los factores de riesgo y potencialmente mejorar la salud y la vida de los pacientes.

**Puntos a tener en cuenta**

Algunos lineamientos basicos del proyecto incluyen de manera no exhaustiva:

1. Exploración de datos: Realice un extensivo análisis de datos exploratorio (EDA) para obtener información sobre el conjunto de datos.

2. Predicción de la calidad: Desarrollar un modelo de machine learning que prediga y logre clasificar entre hipertiroidismo, hipotiroidismo y ningun diagnostico.

3. Interpretabilidad: Identifique los factores más críticos que influyen en la presencia de la enfermedad. Esto ayudará al hospital formar diagnosticos mas precisos.

4. Visualización: Cree visualizaciones e informes para comunicar de manera efectiva sus hallazgos y recomendaciones a las partes interesadas del hospital. La presentación debe constar de un máximo de 10 diapositivas.

Este proyecto requerirá que aplique varias técnicas vistas en el curso, incluida la limpieza de datos, feature engineering, entrenamiento de modelos y la visualización de datos. Es una excelente oportunidad para demostrar sus habilidades en ciencia de datos y generar un impacto real en la prediccion de enfermedades.

**Descripcion del dataset:**

* age: edad del paciente (int)
* sex: sexo con el que se identifica el paciente (str)
* on\_thyroxine: si el paciente está tomando tiroxina (bool)
* query\_on\_thyroxine: si el paciente está en consulta sobre tiroxina (bool)
* on\_antithyroid\_meds: si el paciente está tomando medicamentos antitiroideos (bool)
* sick: si el paciente está enfermo (bool)
* pregnant: si el paciente está embarazado (bool)
* thyroid\_surgery: si el paciente ha sido sometido a cirugía de tiroides (bool)
* I131\_treatment: si el paciente está recibiendo tratamiento con I131 (bool)
* query\_hypothyroid: si el paciente cree tener hipotiroidismo (bool)
* query\_hyperthyroid: si el paciente cree tener hipertiroidismo (bool)
* lithium: si el paciente toma litio (bool)
* goitre: si el paciente tiene bocio (bool)
* tumor: si el paciente tiene un tumor (bool)
* hypopituitary: si el paciente tiene hipopituitarismo (float)
* psych: si el paciente tiene un trastorno psicológico (bool)
* TSH\_measured: si se midió TSH en la sangre (bool)
* TSH: nivel de TSH en la sangre según análisis de laboratorio (float)
* T3\_measured: si se midió T3 en la sangre (bool)
* T3: nivel de T3 en la sangre según análisis de laboratorio (float)
* TT4\_measured: si se midió TT4 en la sangre (bool)
* TT4: nivel de TT4 en la sangre según análisis de laboratorio (float)
* T4U\_measured: si se midió T4U en la sangre (bool)
* T4U: nivel de T4U en la sangre según análisis de laboratorio (float)
* FTI\_measured: si se midió FTI en la sangre (bool)
* FTI: nivel de FTI en la sangre según análisis de laboratorio (float)
* TBG\_measured: si se midió TBG en la sangre (bool)
* TBG: nivel de TBG en la sangre según análisis de laboratorio (float)
* referral\_source: fuente de derivación (str)
* target: diagnóstico médico de hipertiroidismo (str)
* patient\_id: id único del paciente (str)

**Descripcion del target**

El diagnóstico consiste en una cadena de letras que indican las condiciones diagnosticadas. Un diagnóstico "-" indica que no hay ninguna condición que requiera comentario. Un diagnóstico de la forma "X|Y" se interpreta como "consistente con X, pero más probable Y". Las condiciones se dividen en grupos donde cada grupo corresponde a una clase de comentarios.

|  |  |
| --- | --- |
| **Letra** | **Diagnóstico** |
| **Condiciones de hipertiroidismo** | |
| A | hipertiroidismo |
| B | T3 tóxico |
| C | bocio tóxico |
| D | tóxico secundario |
| **Condiciones de hipotiroidismo** | |
| E | hipotiroidismo |
| F | hipotiroidismo primario |
| G | hipotiroidismo compensado |
| H | hipotiroidismo secundario |
| **Proteína de unión** | |
| I | proteína de unión aumentada |
| J | proteína de unión disminuida |
| **Salud general** | |
| K | enfermedad concurrente no tiroidea |
| **Terapia de reemplazo** | |
| L | consistente con terapia de reemplazo |
| M | subreemplazo |
| N | sobre-reemplazo |
| **Tratamiento antitiroideo** | |
| O | medicamentos antitiroideos |
| P | tratamiento con I131 |
| Q | cirugía |
| **Misceláneo** |  |
| R | resultados discordantes de ensayos |
| S | TBG elevado |
| T | hormonas tiroideas elevadas |

Notese que no todas las condiciones mencionadas en la tabla estan alineadas con su objetivo: predecir hipertiroidismo, hipotiroidismo o ningun diagnostico.