

Proyecto Filtro para HackaTec Local 2024

Siguiendo la línea de los proyectos de transformación nacional dirigidos a la salud, se necesita hacer un proceso de clasificación que facilite a los médicos especialistas para evaluar el riesgo que pueden tener los pacientes de adquirir una enfermedad.

Los datos fueron recolectados en una clínica de salud del sector público, en un periodo de un año, usando una encuesta con los siguientes puntos:

1. **Sexo:** M (masculino), F (femenino).
2. **Edad:** Edad del paciente.
3. **Fumador:** SI=2, NO=1.
4. **Dedos amarillos:** SI=2, NO=1.
5. **Ansiedad:** SI=2, NO=1.
6. **Presión de grupo:** SI=2, NO=1.
7. **Enfermedad crónica:** SI=2, NO=1.
8. **Fatiga:** SI=2, NO=1.
9. **Alergia:** SI=2, NO=1.
10. **Sibilancias:** SI=2, NO=1.
11. **Alcohol:** SI=2, NO=1.
12. **Tos:** SI=2, NO=1.
13. **Dificultad para respirar:** SI=2, NO=1.
14. **Dificultad para tragar:** SI=2, NO=1.
15. **Dolor de pecho:** SI=2, NO=1.
16. **Cáncer de pulmón:** YES, NO.

Las respuestas fueron almacenadas en el archivo **respuestas.csv**.

Para iniciar el proceso de generación de clasificación, primero se debe hacer un análisis exploratorio de cada una de las variables, para dar una idea clara de cómo es el comportamiento de cada una de estas. Para los médicos es muy importante encontrar dependencias directas e indirectas a todas estas variables, sus distribuciones o cualquier grafica que aporte significado.

Una vez finalizado el proceso exploratorio, se pretende generar el proceso de clasificación probando diferentes técnicas de Inteligencia Artificial (mínimo 5 modelos diferentes), para generar el mejor modelo de clasificación.

Proyecto Filtro para HackaTec Local 2024

Las consideraciones que se pudieran de tener para la generación de los modelos son:

- Se puede trabajar con menos columnas, siempre y cuando se ocupen técnicas adecuadas de selección de ellas.
- Para el proceso de entrenamiento se debe usar el archivo **respuestas.csv** y para el proceso de validación se debe usar el archivo **test_data.csv**.

Se debe escoger el mejor de los modelos propuestos con base al mejor rendimiento (**valor 100% o cualquier otra técnica o técnicas que ayude a clarificar el proceso de selección**), para ser puesto en producción, y usado a través una interfaz del servicio de comunicación en línea *Telegram*, donde las personas contestarán a cada una de las preguntas de la encuesta mencionada en la parte superior, y entregándole un resultado. **(ACLARACIÓN: EL RESULTADO ENTREGADO POR EL MODELO NO SIGNIFICA QUE EL PACIENTE PRESENTE DE MANERA DEFINITIVA LA ENFERMEDAD, EL RESULTADO ES UNA PROBABILIDAD DE PADECERLA).**

En adición, cada una de las respuestas dadas por los pacientes, se deberán almacenar en una base de datos NoSQL que tenga posibilidad de escalabilidad horizontal a futuro.

Así también de contar una interfaz Web para los médicos especialistas donde pueden observar reportes ejecutivos de esas respuestas dadas de los pacientes.

La interfaz Web debe de contar también con servicios web autocontenidos, para poder capturar datos de la encuesta antes mencionada pero que vengan de clínicas del sector privado.



Tecnológico Nacional de México
Instituto Tecnológico de Zacatepec



Proyecto Filtro para HackaTec Local 2024

Entregables

- Código fuente del análisis exploratorio de los datos (**nombre: exploratorio.ipynb**).
- Código fuente de los modelos de clasificación en proceso de entrenamiento (**nombre: modelos.ipynb**).
- Código fuente del modelo de mejor rendimiento en producción (**nombre: encuesta_modelo.py**).
 - Comunicación con *Telegram*.
 - Comunicación con base de datos NoSQL.
- Código Fuente de interfaz Web mostrando reportes ejecutivos (**nombre carpeta: web**).
- Código fuente de servicios web autocontenidos con comunicación con base de datos NoSQL (**nombre: ms_encuesta.py**).

Plataforma de Entrega

Microsoft Teams: **Proyecto Filtro para Hackaton Local 2024**

Código Microsoft Teams: **636fagd**

Fecha de Desarrollo y Entrega

Todos los documentos entregables deben ser desarrollados del **17 al 25 de Abril** y cargados por **un solo miembro** del equipo a Microsoft Teams el día **25 de Abril antes de la 14:00 hrs.**, y se notificará el resultado entre los días **26 al 29 de Abril de continuar a la fase del concurso local a celebrarse en el ITZ.**