**Documento de sesiones Proyecto 3**

**Fecha:** 29/04/2024

**Participantes:** Manuel, Jezabel, Yassmina & Evgeny

**Agenda:**

**Acciones de seguimiento:**

revisar el autoclean

considerar interacciones multiplicando el 0 1 de tipo de operación con todas las demás (nuevas variables)

Se tiene que hacer como un problema de regresión (Todos).

Scatter plot para ver si el modelo se ajusta bien a la línea.

PLS-DA para separar implantología y peri-implantitis.

Hacer Random Forest con todas las variables.

Representar los valores predichos frente a los reales en un gráfico. La línea del gráfico representará la diferencia entre ambos valores o su igualdad si x=y.

Podemos ahorrarnos el AFC porque hay modelos de regresión que ya te dan las variables más esenciales. AFC es útil como exploración previa. Mejor pillar árbol de decisión y ver relevancia en las subdivisiones. El de regresión sí mantiene el orden y toda la info. Con Random Forest se pueden poner pesos.

**Para el martes:**

Acabar Resampling y análisis estadístico pacientes más completo(Jezabel).

**Para el viernes:**

Tener la demo de la aplicación para los médicos (Evgeny y Manuel).

Buscar la estructura de un paper, donde publicarlo, donde escribirlo, más artículos a partir del análisis (Carla, Yassmina)

Aplicar autoclean a las bases de datos separadas, hacer el random forest de cada base de datos y guardar las 10 características más importantes de cada base de datos (Carla).

Probar random forest sin AFC (YASSMINA)

pls discriminante, ver probabilidades de q sea de una u otros y si es ambiguo sospechosos.

hacer test estadísticos para comparar las variables. (JEZABEL)

hacer estructura de paper (JEZABEL)

Pesos con random forest (MANUEL)

Modificar los modelos y métricas para poner que es mejor overestimating than underestimating.

**Próxima reunión:**

Cumplir con los objetivos de cada uno

**IDEAS**

* REdu

**Reunión con Rui (30/04/2024):**

Los 17 minutos ya es algo que está razonablemente bien. Interesa saber a qué variables van asociadas las intervenciones que se retrasan más. 17 minutos ya está bien según él.

Estudiar si las intervenciones más largas están asociadas a una variable.

**1. Agrupación tiempos > 120.**

Interesante agrupar de 90 para arriba, porque es lo poco normal. De hecho, más de una hora ya empieza a ser una cosa poco usual.

**2. Las 10 variables.**

El alcohol le sorprende un poco, la medicación podría ser y otras drogas.

Las características del implante tampoco le convencen.

El resto tiene lógica.

pero porq esto es lo q afecta a la variabilidad de los datos, no a al duración, no?

**3. La app web (qué podríamos poner).**

para los gráficos:

* El tipo de intervención tendría que estar.
* El tipo de defecto, el número de implantes, el tipo de prótesis, la edad, el sexo del paciente.

Quizás no hace falta predecir el tiempo exacto, un semáforo puede bastar si es difícil.

4. El paper.

Fumadores es un factor de riesgo muy relevante.

Hipercolesterolemia le sorprende un poco, podríamos pasarle el estudio.

Hipertensión arterial dice que puede ser, porque hay algunos trabajos que lo asocian a la periodontitis y con las enfermedades cardiovasculares. Tenemos que verlo más.

Los fármacos también le parecen interesantes y sí puede afectar.

cuidado con la relación ‘causa-efecto’

artículo : enfocarlo a patología sistémicas o a un tema en concreto porq si no es muy difícil publicar. siempre matizando y con precaución

5. Los objetivos.

Le parecen interesantes para la correcta gestión.

Y quiere ver si lo de la patología sistémica se puede publicar en algún lado interesante.