**Resumen**

Se propone un estudio utilizando técnicas de machine learning para determinar la importancia diagnóstica de algunas variables en la predicción de padecer diabetes. Esto se desarrollaría en función de un modelo random forest (o una red neural?) entrenado sobre distintos subconjuntos escogidos mediante k-means.

Hay demasiados aspectos poco claros en la propuesta por lo que se requiere una reunión con el profesor para definir los alcances concretos del proyecto.

**Comentarios mayores**

La idea me parece súper interesante. Lamentablemente, la redacción hace muy difícil comprender los detalles.

Hay una combinación de técnicas discutidas en el curso, donde no es clara la función que tendrían o cómo se utilizarían dentro del proceso (e.g., PCA).

No es claro que se espera obtener con k-means, es claro que grupos, pero ¿cómo quedarían conformados esos grupos? ¿Es mejor eso que establecer ciertos límites en las variables usando el criterio experto de un especialista? ¿se podrían obtener límites para establecer estos grupos de otra manera? (discutir con el profesor)

No es claro si se usará random forest o una red neuronal. Mi propuesta es que sea sólo uno y se ajusten los hiperparámetros utilizando alguna validación cruzada.

Dado que se harán grupos y se calibraran modelos de manera independiente, ¿cómo se hará el Split de los datos? ¿Cómo se cuantificará la precisión del modelo?

La idea general me parece muy buena, pero hay muchos aspectos metodológicos que definir. Propongo una discusión con el profesor para definir estos detalles.

**Comentarios menores**

Se debió usar una redacción más precisa para poder entender de mejor manera los detalles de la metodología.

**Ortografía, redacción y estilo**

La redacción es confusa.

**Evaluación**

Buena idea, pero la redacción hace muy difícil entender el objetivo concreto y la metodología.

**Nota: 4.5**