



Técnicas estadísticas aplicadas a la estimación de cumplimientos en recursos humanos

Justificación del resumen

Debido a que este Trabajo de Fin de Máster fue realizado en el seno de una entidad financiera como ABANCA, contiene datos sensibles que la entidad consideró conveniente no publicar.

A razón de esto se optó por elaborar este resumen que contiene una explicación menos detallada del trabajo realizado.

Introducción

Este Trabajo de Fin de Máster se desarrolló en el área de Capital Humano de ABANCA. Este área es la encargada de la gestión y administración del personal de la empresa y, entre otras tareas, se ocupa de la evaluación del rendimiento de los empleados mediante una puntuación denominada "Cumplimiento del Panel de Productividad Individual" o "Cumplimiento del PDPI", que mide en que medida los empleados cumplen los objetivos que se les asignan.

La meta que se persiguió en este proyecto fue elaborar un modelo estadístico que permitiese estimar dicha puntuación. Para ello se establecieron tres objetivos:

1. Estimar el "Cumplimiento del PDPI" tomando como variables explicativas el cumplimiento de cada objetivo que se le asigna a un empleado.

Estas se obtienen a la vez que la variable respuesta y suponen una medida directa del "Cumplimiento del PDPI", por lo que solo permiten ofrecer un análisis descriptivo.

2. Estimar el "Cumplimiento del PDPI" usando variables explicativas que ya están disponibles antes de obtener los datos de la variable respuesta y que supondrían una medida indirecta de esta, como son las características del empleado, de la oficina donde trabaja, la carga de trabajo del empleado o la evaluación que hacen de él sus superiores.

Una ventaja de este modelo es que, además de ofrecer una descripción de los cumplimientos, también abriría la posibilidad de elaborar un modelo predictivo según se fuesen obteniendo datos de años venideros.

3. Clasificar a los empleados en base a si alcanzan o no los objetivos que se les asignan usando las mismas variables explicativas que en el punto anterior.

Al igual que en el caso anterior, este modelo ofrecería la posibilidad de hacer predicciones según se consiguiesen nuevos datos.

Estimación del cumplimiento en base a medidas directas

Para tratar de alcanzar el primer objetivo se optó por estimar un modelo de regresión lineal que tuviese como variable respuesta el "Cumplimiento del PDPI" y como variables explicativas el cumplimiento de cada objetivo asignado a los empleados.

Este modelo resultaba significativo para un p-valor < 0.05 y arrojaba un R^2 ajustado de 0.6468, lo que se consideró lo bastante alto como para aceptar el modelo.

Sin embargo, cuando se procedió a validar los supuestos se halló que se incumplían el de linealidad y el de homocedasticidad. Por ello, se decidió estimar un modelo aditivo generalizado, ya que este permitiría ajustar los datos a parámetros no lineales.

El GAM resultaba significativo para un p -valor < 0.05 y ofrecía un porcentaje de la deviance explicada del 65.90 %.

En cuanto a los supuestos, al no tener que cumplir el de linealidad y, al asumir que los problemas en la homocedasticidad de los residuos se debían al bajo tamaño muestral para los valores pequeños, estos se dieron por válidos, lo que llevó a aceptar el modelo.

Estimación del cumplimiento en base a medidas indirectas

Para el segundo objetivo intentó ajustarse un modelo de regresión lineal que estimase el "Cumplimiento del PDPI" en base a medidas indirectas como son las características de los empleados, de la oficina donde trabajan, la carga de trabajo del empleado o la evaluación que hacen de él sus superiores.

El modelo resultaba significativo para un p -valor < 0.05 , pero explicaba un 34,4 % de la varianza e incumplía los supuestos de homocedasticidad y normalidad, por lo que se decidió descartar el modelo y ajustar un modelo aditivo generalizado para ver si ofrecía mejores resultados.

El GAM resultó significativo para un p -valor < 0.05 , no obstante arrojaba un porcentaje de la deviance explicada del 41.1 %, lo que se consideró un ajuste demasiado bajo para aceptar el modelo como válido.

En cuanto a los supuestos, se aceptaban todos como válidos salvo el de normalidad, lo que suponía una mejora con respecto al modelo de regresión lineal.

Clasificación de los empleados según su cumplimiento

En este objetivo, antes de estimar ningún modelo, lo primero que se hizo fue clasificar a los empleados entre los que cumplían el mínimo de objetivos establecidos por la empresa y los que no.

Una vez clasificados los empleados se procedió a ajustar un modelo de regresión logística usando las mismas variables regresoras que en el caso anterior.

El modelo ajustado mediante regresión logística resultaba significativo para un p -valor < 0.05 .

Como el modelo de regresión logística estima la probabilidad de que se dé un evento, una vez ajustado clasificó a los empleados con una probabilidad estimada igual o mayor a 0.5 como que cumplían y a aquellos con una probabilidad estimada menor a 0.5 como que no.

Una vez clasificados, se compararon las estimaciones frente a los datos reales y se encontró que el modelo clasificaba correctamente al 72.7 % de los empleados.

Además se comprobó que los predictores lineales se ajustaban a una función logit y se cumplían los supuestos del modelo, por lo que se decidió aceptarlo como válido.

Aún así, se decidió ajustar un GAM logístico a los datos para ver si mejoraba el ajuste.

El modelo aditivo generalizado logístico resulta significativo para un p -valor < 0.05 .

Como en el caso anterior, se clasificaron a los empleados con una probabilidad estimada igual o mayor a 0.5 como que cumplían y aquellos con una probabilidad estimada menor a 0.5 como que no.

Al comparar las estimaciones frente a los datos reales se encontró que el modelo clasificaba correctamente el 74.3 % de los empleados.

Como en el modelo de regresión logística, los predictores lineales se ajustaban a una función logit y se aceptaban los supuestos como válidos, por lo que se consideró que el modelo GAM logístico resultaba útil a la hora de clasificar a los empleados.

Dado que ambos modelos se aceptaron como válidos y que, aunque el GAM logístico presentaba un ajuste algo superior, ambos arrojaban resultados muy semejantes, se optó por elegir el modelo de regresión logística para estimar la variable respuesta, ya que era más sencillo de explicar y representar.

Conclusiones

Una vez estimados ambos modelos y analizados sus resultados se pueden extraer varias conclusiones de este trabajo:

- Los modelos aditivos generalizados se muestran útiles a la hora de estimar una variable respuesta de la familia exponencial cuando otros modelos como los de regresión lineal o los modelos lineales generalizados no cumplen determinados supuestos, como es el de linealidad.
- Los GAMs también suponen una ventaja en cuanto a que solo deberían igualar o mejorar el ajuste con respecto a los modelos lineales, esto se debe a que en el peor de los casos un modelo aditivo generalizado debería ajustar un modelo igual al lineal.
- Al tratar de estimar el "Cumplimiento del PDPI" mediante el cumplimiento en los distintos objetivos que obtiene el trabajador, el modelo aditivo generalizado se muestra como el único válido frente al de regresión lineal al no tener que cumplir el supuesto de linealidad.
- No se obtiene un ajuste lo suficientemente alto cuando se intenta estimar el "Cumplimiento del PDPI" haciendo uso de las usando variables explicativas que ya están disponibles antes de obtener los datos de la variable respuesta, como las características de los empleados, de la oficina donde trabajan, la carga de trabajo del empleado o la evaluación que hacen de él sus superiores.
- Usando estas mismas variables, el modelo de regresión logística y el GAM logístico se encuentran útiles a la hora de clasificar los trabajadores en aquellos que cumplen y aquellos que no. Aunque se opta por el modelo de regresión logística por simplicidad, el GAM logístico obtendría unos resultados semejantes.