

Evaluation

David Guzmán Leyva - A01706417
Enrique Santos Fraire - A01705746
Leonardo Alvarado Menéndez - A01705998
Oscar Enrique Delgadillo Ochoa - A01705935

Índice

Índice	2
Bitácora de cambios	3
Evaluación de los resultados	3
Evaluación de los objetivos de negocio	3
Modelos aprobados	3
Proceso de revisión	4
Diagrama del seguimiento de las fases CRISP-DM	4
Revisión del proceso	4
Siguientes pasos	5
Posibles acciones	5
Decisión	6

Bitácora de cambios

Versión	Fecha	Autor(es)	Modificaciones
1.0	18/11/2022	David Guzmán Leyva Enrique Santos Fraire Leonardo Alvarado Menéndez Oscar Enrique Delgadillo Ochoa	Línea base

Evaluación de los resultados

Evaluación de los objetivos de negocio

En parte de nuestro proceso, obtenemos la información del promedio de home-office realizado en cada comuna, así como el número de usuarios que se detectó que realizaban home-office y las que no, conociendo así las comunas con mayor presencia de home-office.

[HomeOffice.csv](#)

Según estos datos, podemos intuir que:

- La comuna con mayor presencia de home-office es San Pedro con 73.49%
- La comuna con menor porcentaje de home-office es Alhué con 46.66%
- Debido a que el porcentaje de home-office mayor a 46% en todas las comunas, los datos deben ser en tiempos de pandemia.

Aparte, contamos con un modelo, el cual basándose en datos de cada comuna como la cantidad de empresas micro, chicas, medianas y grandes, número de empleados y otros datos, predice el porcentaje de home-office que se realiza en cada comuna.

Tomando en cuenta nuestros objetivos de negocio, planteados en el documento [Business Understanding](#), los cuales son:

- Determinar las comunas con mayor presencia de home-office.
- Identificar los factores determinantes que sean indicadores porcentuales de presencia de home office.

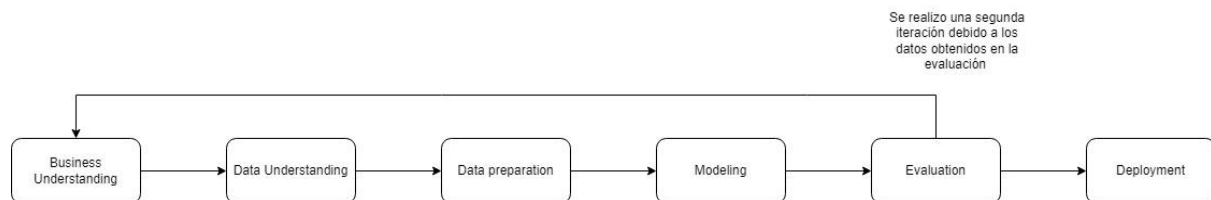
Podemos ver que los objetivos se cumplen y estamos en espera de la aprobación de la socia, para poder decir que cumplimos exitosamente con los objetivos de negocio.

Modelos aprobados

En este caso, de los modelos creados, finalmente se evaluará el GBTs, ya que este fue el modelo que presentó mejores resultados sobre la predicción del porcentaje de home-office de las comunas de Santiago de Chile.

Proceso de revisión

Diagrama del seguimiento de las fases CRISP-DM



Revisión del proceso

Dentro de las fases de CRISP-DM se llegaron a las siguientes conclusiones:

- Business Understanding
 - 1° iteración: Se ejecutó de manera adecuada, sin embargo el objetivo planteado carecía de profundidad, lo que llevaría a que más tarde en la fase de modelado nuestra solución tuviera poco o nulo impacto.
 - 2° iteración: Luego de la validación con nuestros profesores y socia, llegamos a un objetivo más claro, no completamente alejado de la primera iteración, sino derivada de la misma, por lo que nuevamente se ejecutó con éxito esta fase pero esta vez con un enfoque significativo.
 - Conclusiones: Algo que tuvimos muy en claro al momento de partir nuevamente desde el inicio, es que no debemos tomar a la ligera el enfoque del proyecto, y debemos validar realmente lo que queremos hacer, sólo así podremos dar soluciones correctas a las preguntas correctas.
- Data Understanding
 - 1° iteración: Durante esta fase entendimos los datos con los que vamos a trabajar, realizando de forma efectiva este análisis, por lo que pudimos avanzar pronto a la siguiente fase.
 - 2° iteración: Al cambiar de iteración no hubo realmente cambios significativos, pues ya teníamos un entendimiento claro de los datos que se nos habían brindado y las posibles fuentes de información que nos podían proporcionar.
 - Conclusiones: Debido a que estos datos se nos brindaron limpios y anonimizados, esta etapa fue sencilla y directa. Sin embargo, en casos más cercanos a la vida real, esta fase representaría uno de los más grandes retos, pues tendríamos que realizar un análisis a profundidad, así como tomar decisiones más contundentes respecto a la dirección del proyecto.

- Data Preparation

- 1° iteración: Esta fase fue especialmente engañosa, pues es la que más tiempo ha requerido, dejando en claro que llegar a el dataset deseado puede ser más complicado de lo que parece.
 - Debido a que no tomamos en cuenta el tamaño del dataset original, cometimos varios errores durante la estimación del proyecto, por lo que se tuvo que ajustar el plan de trabajo original.
 - Aún con los imprevistos, esta fase demostró ser una de las de más alto valor, pues logramos adaptar la metodología a nuestra forma de trabajo, además que nos ayudó a mantener todo el tiempo la trazabilidad de los archivos, siendo legible aún para aquellos miembros externos al equipo.
- 2° iteración: Dado que esta fase fue ejecutada de manera óptima, y que nuestros objetivos más que cambiar se expandieron, de igual manera se hizo en esta fase. Aún con los cambios el flujo de trabajo continuó con normalidad.
- Conclusiones: Ciertamente debemos tomar en consideración las limitaciones técnicas que puedan repercutir en nuestros tiempos de trabajo y explorar alternativas eficientes, así como son imprescindibles los recursos visuales para un mejor entendimiento del proyecto.

- Modeling

- 1° iteración: Realmente lo más laborioso a realizar durante esta fase es la investigación previa al modelado, donde debemos entender que se quiere realmente predecir, qué modelos son los más óptimos y por qué, siendo una inversión un tanto costosa pero necesaria en el tiempo del proyecto. Realmente el código no es tan complejo de realizar y, en nuestro caso, al tener pocos datos de entrada el modelo, este es bastante rápido de entrenar.
- 2° iteración: En casos distintos, donde el modelo necesita más datos para entrenar, o no es tan óptimo el modelo, el tiempo de entrenamiento puede aumentar mucho, que igualmente no fue algo que consideramos. Durante la segunda iteración sucedió algo similar a la anterior y la creación de los modelos fue relativamente rápida.
- Conclusiones: Algo que se puede cambiar, es realizar la investigación previa a la par que fases anteriores, de manera que mientras exploramos nuestros datos sea fácil direccionar nuestro tiempo y recursos en el estudio necesario para llegar a la solución adecuada.

- Evaluation

- Aunque se cumplían los objetivos durante la primera iteración, no se estaba utilizando realmente una solución de utilidad aplicando machine learning. Además que el valor otorgado no era significativo, por lo que se optó por iniciar una nueva iteración en busca de una mejor solución con alternativas más profundas.

Siguientes pasos

Posibles acciones

Llegados a este punto, contamos con las siguientes posibilidades:

- Definir el modelo como validado y lanzarlo a producción
- Hacer una nueva iteración con un modelo que contenga más variables

Decisión

Debido a pláticas con la socia y profesores, notamos que los objetivos previos no eran de tanta relevancia y no se podía aplicar un modelo de IA útil, por lo que cambiamos nuestros objetivos de negocio y de minería de datos, se realizó una nueva iteración.

Cambiamos los objetivos de únicamente predecir si un usuario realizaba home-office a calcular el porcentaje de home-office realizado dentro de una comuna según los datos de esta.