Predicción de Costos de Proyectos en Construcción

Seminario de Analítica y Ciencia de Datos

# Resumen Descriptivo del Proyecto

El proyecto tiene como objetivo desarrollar un modelo de Machine Learning para estimar con mayor precisión y eficiencia los costos de proyectos de construcción de vivienda VIS y NO VIS. Este modelo permitirá a las empresas del sector:

* Reducir los sobrecostos en proyectos de construcción.
* Mejorar la planificación financiera y la toma de decisiones.
* Aumentar la eficiencia y la rentabilidad de los proyectos.
* Reducir el riesgo de retrasos y problemas durante la ejecución de los proyectos

Basado en el Data Product Canvas1 resuelve:

## Problema.

* 1. ¿Cuál es mi problema alrededor de mi proyecto?

En los últimos cinco años la industria de la construcción en Colombia se vio afectada por diversos factores, tales como; planeación ineficiente de proyectos, altas tasas de interés bancario, disminución en el número de subsidios, aumento precio de materiales, entre otros. Lo anterior conllevó a que aproximadamente 150 constructoras liquidaran completamente, y 366 constructores en proceso de reestructuración por parte de la SuperSociedades 2. Por esta razón, se entiende que la estimación de costos en la construcción es un proceso complejo y sujeto a una gran cantidad de incertidumbres, lo que genera los siguientes problemas: sobrecostos (el costo real supera lo presupuestado), retrasos (la falta de precisión en la estimación de costos puede provocar retrasos en la ejecución del proyecto, lo que genera pérdidas de tiempo y dinero), Problemas durante la ejecución (la imprecisión en la estimación de costos puede dificultar la planificación y gestión del proyecto, lo que aumenta el riesgo de problemas durante la ejecución)

* 1. ¿Por qué se manifiesta el problema?

Los métodos tradicionales de estimación de costos, como las estimaciones manuales o basadas en hojas de cálculo, suelen ser imprecisos y estar sujetos a errores humanos. Además, la complejidad de los proyectos de construcción, la diversidad de variables involucradas y la dinámica del mercado dificultan la estimación precisa de los costos.

* 1. ¿De quién es el problema?

Este problema afecta a todos los actores involucrados en el sector de la construcción, incluyendo:

*Empresas constructoras:* son las directamente afectadas por los sobrecostos,

retrasos y problemas durante la ejecución de los proyectos.

*Clientes:* asumen el riesgo de pagar más de lo presupuestado por el proyecto.

*Proveedores:* pueden verse afectados por cambios en las estimaciones de costos que afecten sus márgenes de ganancia.

## Datos.

* 1. Describa la fuente de datos (Sea más específico que en documento pasado)

Las fuentes de datos para el proyecto se pueden clasificar en dos categorías:

*Datos históricos de proyectos de construcción:* estos datos van a provenir de una constructora en especifica, la constructora se llama “Vértice Ingeniería y Construcción”. Los datos van a incluir información sobre las características del proyecto, los costos directos, costos indirectos y otros variables relevantes.

*Datos de mercado:* estos datos se refieren a los precios de materiales, equipos y otros insumos utilizados en la construcción. Los datos de mercado pueden provenir de fuentes públicas, privadas y plataformas especializadas en la información de precios.

* 1. ¿Cuál es la calidad que posee dicha fuente?

La calidad de los datos de la fuente va a ser verídico, dado que se está analizando la información de los proyectos de una empresa constructora que opera actualmente. Además, los datos van a ser precisos, completos, consistentes y actualizados. Sin embargo, es importante realizar un proceso de limpieza y preprocesamiento de los datos para eliminar errores, inconsistencias y valores atípicos.

* 1. ¿Es accesible y disponible?

Con respecto a la accesibilidad y disponibilidad de los datos no existirá ningún problema, dado que se tiene consentimiento por parte de la organización utilizar los datos para realizar el proyecto.

## Hipótesis.

* 1. ¿Qué es lo que pretende probar?

Las hipótesis principales del proyecto son:

* Es posible desarrollar un sistema de Machine Learning que permita predecir con mayor precisión los costos de proyectos de construcción de vivienda VIS y NO VIS.
* La precisión del sistema de predicción será superior a la de los métodos tradicionales de estimación de costos.
* El uso del sistema de predicción permitirá a las empresas del sector de la construcción reducir los sobrecostos, mejorar la planificación financiera y aumentar la rentabilidad de los proyectos.
  1. ¿Cuáles pueden ser las respuestas esperadas a lo que pretende probar?

Las posibles respuestas esperadas a las hipótesis son:

Respuesta positiva: los resultados del proyecto demuestran que es posible desarrollar un modelo de Machine Learning que permita predecir con mayor precisión los costos de proyectos de construcción de vivienda VIS y NO VIS. El modelo tiene una precisión superior a la de los métodos tradicionales de estimación de costos y su uso genera beneficios tangibles para las empresas del sector.

Respuesta negativa: los resultados del proyecto demuestran que no es posible desarrollar un modelo de Machine Learning que permita predecir con mayor precisión los costos de proyectos de construcción de vivienda VIS y NO VIS. El modelo no tiene una precisión superior a la de los métodos tradicionales de estimación de costos o su uso no genera beneficios tangibles para las empresas del sector.

* 1. ¿Qué acciones debo tomar para cada respuesta anterior?

Respuesta positiva:

* Implementar el modelo de predicción de costos en la constructora que habilito la utilización de los datos.
* Capacitar a los usuarios del modelo en su uso e interpretación de los resultados (cuando el modelo ya este implementado en la organización-postproyecto).
* Monitorear el rendimiento del modelo y realizar ajustes según sea necesario (cuando el modelo ya este implementado en la organización-postproyecto).
* Difundir los resultados del proyecto y los beneficios del uso del modelo.

Respuesta negativa:

* Analizar las causas del fracaso del proyecto e identificar las áreas de mejora.
* Redefinir las hipótesis del proyecto y rediseñar el sistema de predicción de costos.
* Mejorar la calidad de los datos existentes.
* Evaluar la viabilidad de utilizar otras técnicas de Machine Learning o algoritmos.

## Solución.

* 1. Tipo. (¿Qué tipo de algoritmo de ML es? ¿Qué técnica es?

El tipo de algoritmo de Machine Learning a utilizar dependerá de las características de los datos y se va a tener más claro, cuando se trabaje netamente en el modelo. A medida que se vaya avanzando con las materias de la especialización, se va a tener defino la técnica a implementar, hasta el momento lo visto en los cursos se consideran las siguientes técnicas que podrían ser utilizadas:

* Regresión lineal: esta técnica es útil para predecir valores continuos, como el costo total del proyecto.
* Árboles de decisión: esta técnica es útil para identificar las variables que más influyen en el costo del proyecto.
  1. ¿Cómo debe ser implementado?

El modelo de predicción de costos debe ser implementado en una plataforma que sea accesible y fácil de usar para los usuarios del sector de la construcción. La plataforma debe permitir a los usuarios ingresar las características del proyecto y obtener una predicción del costo. El desarrollo del modelo se implementará utilizando Python y bibliotecas como scikit-learn y TensorFlow.

Cabe destacar que el alcance del proyecto no va hasta el deployment por ejemplo en una pagina web, solo abarcará el desarrollo y ejecución del código del modelo en notebooks en Colab o Databricks.

* 1. ¿Qué resultados se esperan?

Los resultados esperados del proyecto son:

* Un modelo de predicción de costos de construcción de vivienda VIS y NO VIS que sea preciso, eficiente y fácil de usar.
* Una reducción de los sobrecostos en proyectos de construcción.
* Una mejora en la planificación financiera y la toma de decisiones en las empresas del sector de la construcción.
* Un aumento en la eficiencia y rentabilidad de los proyectos de construcción.

## KPI’s

* 1. ¿Cómo espera evaluar el modelo?

El modelo de Machine Learning será evaluado utilizando una serie de métricas, como:

* Error medio absoluto (MAE): esta métrica mide la diferencia promedio entre el costo real del proyecto y el costo predicho por el modelo.
* Raíz cuadrada del error medio cuadrático (RMSE): esta métrica es similar al MAE, pero penaliza más los errores grandes.
* R cuadrado (R²): esta métrica indica la proporción de la variabilidad del costo total del proyecto que es explicada por el modelo.
  1. ¿Qué métricas debe usar?

Además de las métricas mencionadas anteriormente, también se puede utilizar otras métricas, como:

* Precisión: esta métrica indica el porcentaje de proyectos en los que el costo predicho por el modelo está dentro de un rango de error predefinido.
* Recall: esta métrica indica el porcentaje de proyectos en los que el modelo identifica correctamente que el costo real está por encima del presupuesto.
* F1-score: esta métrica es una combinación de la precisión y el recuerdo.

## Actores.

* 1. ¿Quién es su cliente?

Los clientes del proyecto son las empresas constructoras de proyectos de vivienda VIS y NO VIS, que necesitan una solución para estimar con mayor precisión los costos de sus proyectos. Sin embargo, el cliente principal es la constructora Vértice Ingeniería & Construcción, quien nos esta facilitando los datos.

* 1. ¿Quiénes son los interesados (Stakeholders)?

Los interesados del proyecto son:

* *Empresas constructoras:* son las directamente beneficiadas por la implementación del modelo de predicción de costos, en especial, la constructora Vértice Ingeniería & Construcción, quien nos está facilitando los datos.
* *Proveedores:* se beneficiarán de una mayor estabilidad en los proyectos y una menor probabilidad de cambios en los presupuestos.
  1. ¿Quién usará la solución?

La solución será utilizada por los siguientes usuarios:

* *Estimadores de costos*: son los responsables de realizar las estimaciones de costos de los proyectos dentro de la organización.
* *Gerentes de proyecto:* son los responsables de la planificación y ejecución de los proyectos.
* *Directivos de empresas:* son los responsables de la toma de decisiones estratégicas en las empresas.
  1. ¿A quién impactará?

El proyecto impactará a los siguientes actores:

*Empresas del sector de la construcción:* la implementación del modelo de predicción de costos de vivienda VIS y NO VIS les permitirá reducir los sobrecostos, mejorar la planificación financiera y aumentar la rentabilidad de sus proyectos.

*Clientes/inversionistas:* se beneficiarán de una mayor precisión en la estimación de los costos de los proyectos, lo que les permitirá tomar mejores decisiones de inversión.

*Proveedores:* se beneficiarán de una mayor estabilidad en los proyectos y una menor probabilidad de cambios en los presupuestos.

1

[https://medium.com/@leandroscarvalho/data-product-canvas-a-practical-framework-for-building-high-](https://medium.com/%40leandroscarvalho/data-product-canvas-a-practical-framework-for-building-high-) performance-data-products-7a1717f79f0

2

https://www.las2orillas.co/las-constructoras-que-no-aguantaron-la-crisis-y-terminaron-en-la-quiebra/

# Rúbricas

R01: Completitud y consistencia: el estudiante realiza una entrega del momento evaluativo con todos los elementos desarrollados, manteniendo una relación coherente entre los elementos solicitados. El estudiante usa una consistencia argumentativa que incluye el uso de referencias cuando hay mérito de autoría entre otros.

R02: Pertinencia: el proyecto del estudiante está dentro del ámbito de la analítica y la ciencia de datos y es evidente el desarrollo de cada uno de los elementos del Data Product Canvas

R03: Solución: el estudiante es capaz de responder a las preguntas del Data Product Canvas con la consistencia necesaria para relacionarla con el proyecto que desarrolla. Hay una relación evidente entre ellos. Las herramientas de MAchine Learning son una solución viable dentro de la descripción de la necesidad y objetivos del proyecto.

R04: Datos: los datos sugeridos son pertinentes al proyecto y al alcance delimitado en el momento evaluativo anterior.