

TÍTULO DE LA TESIS O TRABAJO DE INVESTIGACIÓN

Autor(es)

Nombres y apellidos completos del autor

Universidad de Antioquia

Facultad de Ingeniería

Ciudad, Colombia

Año

INSTRUCCIONES

* Copia este documento en tu drive
* Usa Heading 1,2,3, etc para estructurar las secciones en tu texto y que así el índice se actualice automáticamente
* Usa [[Cross Reference](https://gsuite.google.com/marketplace/app/cross_reference/269114033347)] para referenciar tablas, figuras, bibliogrfía, etc. en tu texto (es un add-on de Google docs). Lee las [instrucciones](https://github.com/davidrthorn/cross_reference/blob/master/README.md) detenidamente
* Elimina esta página de tu documento.
* No hay un requisito máximo ni mínimo de longitud de este documento. A modo de guía, entre 15 y 30 páginas es una longitud promedio.
* Respeta las secciones de primer nivel. Modifica las subsecciones si lo consideras necesario.
* Sé breve y preciso en tus descripciones. El objetivo no es que es este documento sea largo, sino que describa adecuadamente el trabajo que has hecho.

NOTEBOOKS

* Has de acompañar este documento con un repositorio Github con los notebooks que implementen los procesos e iteraciones de tu proyecto.
* Incluye un notebook llamado ‘**00-INSTRUCCIONES.ipynb’** en el que describes cómo ejecutar tus notebooks.
* Los notebooks del curso han de estar publicados en un repositorio Github y, si posible, han de ser ejecutables desde Colab.
  + Si desarrollaste tu proyecto en otro entorno (por ejemplo Kaggle) incluye en el notebook de instrucciones cómo ejecutar tus notebooks en Kaggle
* Si puedes, incluye en tu Github los datos de tu proyecto. Si no puedes (porque son muy grandes, porque ya están en Kaggle, o porque son privados) incluye en el notebook con instrucciones sobre cómo obtenerlos.
  + Si tus notebooks son privados (p.ej. por restricciones de la empresa para la cual se hizo la monografía) haz el repositorio privado.
  + Si tus datos son privados incluye en las instrucciones tu dirección de correo como contacto para solicitar los datos.
* Adjunta un zip del repositorio (usa *download zip* desde Github) en tu entrega en la biblioteca.

TABLA DE CONTENIDOS

[**1. RESUMEN EJECUTIVO**](#_sqyu3rma9krf) **4**

[**2. DESCRIPCIÓN DEL PROBLEMA**](#_izg59o76q0tg) **5**

[2.1 PROBLEMA DE NEGOCIO](#_44h0p4n9zx4i) 5

[2.2 APROXIMACIÓN DESDE LA ANALÍTICA DE DATOS](#_9o7v8dc76xwz) 5

[2.3 ORIGEN DE LOS DATOS](#_a5n2wgo7nljd) 5

[2.4 MÉTRICAS DE DESEMPEÑO](#_pe45y91q0o9b) 5

[**3. DATOS**](#_fhbsfoidppif) **6**

[3.1 DATOS ORIGINALES](#_ptjtanjwkxtn) 6

[3.2 DATASETS](#_y6lmtaszmdl1) 6

[3.3 DESCRIPTIVA](#_p7yt2mukxk7u) 6

[**4. PROCESO DE ANALÍTICA**](#_df0j799h8ltp) **7**

[4.1 PIPELINE PRINCIPAL](#_wkofic467dkv) 7

[4.2 PREPROCESAMIENTO](#_4jcu5pefxe00) 7

[4.3 MODELOS](#_ewd55quzl840) 7

[4.4 MÉTRICAS](#_20it13y28c81) 7

[**5. METODOLOGÍA**](#_jhy4beqp3v5) **8**

[5.1 BASELINE](#_tnrsguuau124) 8

[5.2 VALIDACIÓN](#_ges19o90yutn) 8

[5.3 ITERACIONES y EVOLUCIÓN](#_hmxkpq3c5m7l) 8

[5.4 HERRAMIENTAS](#_iuz6agv9ll9z) 8

[**6. RESULTADOS**](#_3g2rb07ral3b) **9**

[6.1 MÉTRICAS](#_5wktvixcazwi) 9

[6.2 EVALUACIÓN CUALITATIVA](#_n0xcyecyddq1) 9

[6.3 CONSIDERACIONES DE PRODUCCIÓN](#_ug54okirh7ae) 9

# 1. RESUMEN EJECUTIVO

Esto es un abstract. Realiza una descripción general de tu proyecto: qué se persigue, qué datos se tiene, qué estrategia se siguió para las iteraciones, que obstáculos hubo, qué resultados se obtuvieron, etc. **La longitud máxima es de una página**

Incluye al final de dicha página la dirección de tu repositorio github

La finalidad del proyecto es la de predecir las ventas para el mes siguiente para una cadena de supermercados haciendo uso de algoritmos de inteligencia artificial. Se cuenta con 4 datasets (productos , categoría de los productos, tiendas y ventas diarias) tomados de la competencia de kaggle “Predict Future Sales” que proporcionan información detallada de las ventas históricas para 22.171 productos en 34 meses a partir del mes de enero de 2013. A nivel metodológico el trabajo esta conformado por las siguientes etapas: limpieza y preparación de los datos, entrenamiento y selección del modelo final. La primera etapa inicia con un análisis exploratorio de los datos que permitió detectar outliers y garantizar que nuestros datasets se encuentran libres de datos nulos, seguidamente ………

# 2. DESCRIPCIÓN DEL PROBLEMA

Describe de manera general el problema de negocio que se quiere abordar, cómo se puede resolver con modelos predictivos, cómo

## 2.1 PROBLEMA DE NEGOCIO

Describe el contexto en el cual se desarrolla tu proyecto. (p.ej. Tal empresa quiere mejorar sus procesos de marketing)

## 2.2 APROXIMACIÓN DESDE LA ANALÍTICA DE DATOS

Explica para qué servirían los modelos predictivos desarrollados (p.ej. Se desarrollaran modelos para recomendar productos a clientes)

## 2.3 ORIGEN DE LOS DATOS

Explica qué representan los datos y cómo se obtuvieron. P.ej. tenemos datos clínicos e imágenes de pacientes de X patologías, obtenidos en tal hospital, en tales condiciones y durante tal periodo de tiempo. No describas aquí los datos, eso lo harás más adelante.

## 2.4 MÉTRICAS DE DESEMPEÑO

Incluye

* las métricas de machine learning (desempeño de los modelos)
* las métricas de negocio (desempeño del proceso de negocio que usaría tus modelos)
* un juicio sobre qué valor mínimo de las métricas son necesarias en el contexto de aplicación
  + P.ej. a partir de x% de acierto (métrica de ML) , se estima un ROI de USD yy (métrica de negocio) que es lo mínimo para cubrir la operación, despliegue, etc. de los modelos.

# 3. DATOS

## 3.1 DATOS ORIGINALES

Describe los datos crudos de los que se dispone, incluyendo formatos, cómo se distribuyen los datos en ficheros, descripción de las columnas, formatos y resoluciones de imágenes, etiquetado de los mismos (para aprendizaje supervisad), tamaño de los datos (número de registros, tamaño total en MB o GB), etc. etc.

Describe también el modo de acceso a los datos y restricciones que pudiera haber sobre los mismos.

## 3.2 DATASETS

Describe cómo, a partir de los datos, se construyen los datasets de entrenamiento y validación

## 3.3 DESCRIPTIVA

Muestra algunos ejemplos de tu dataset (imágenes, registros), incluyendo una breve analítica descriptiva (distribuciones de datos, etc.)

# 4. PROCESO DE ANALÍTICA

## 4.1 PIPELINE PRINCIPAL

Describe con un gráfico el flujo de trabajo general de los datos en tu proyecto. Añade o modifica secciones si lo consideras necesario

## 4.2 PREPROCESAMIENTO

Describe las alternativas de preprocesamiento de datos que consideraste, aumentación de datos, etc.

## 4.3 MODELOS

Describe los distintos modelos que consideraste, sus configuraciones, etc.

## 4.4 MÉTRICAS

Describe cómo calculas las métricas de desempeño ML y de negocio (p.ej. Con qué funciones de sklearn, tensorflow, etc.)

# 5. METODOLOGÍA

## 5.1 BASELINE

Describe tu primera iteración, sus resultados y los problemas técnicos que tuviste que resolver

## 5.2 VALIDACIÓN

Describe tu proceso de validación, particiones de train/test/validación, etc.

## 5.3 ITERACIONES y EVOLUCIÓN

Describe en qué te enfocaste en las iteraciones subsiguientes. No es necesario hacer un listado exhaustivo de las iteraciones, pero sí has de describir qué perseguías y qué obtuviste con cada grupo de iteraciones. P.ej. primero hicimos un conjunto de iteraciones para definir las opciones de preprocesamiento, luego hicimos unas iteraciones probando distintos modelos de tal tipo, etc. etc.

## 5.4 HERRAMIENTAS

Menciona las herramientas que usaste para tu proyecto

# 6. RESULTADOS

## 6.1 MÉTRICAS

Describe los resultados numéricos de las métricas de las iteraciones que consideres más relevantes, junto con las configuraciones. Usa tablas o gráficas según consideres.

## 6.2 EVALUACIÓN CUALITATIVA

Realiza una evaluación cualitativa de los resultados, indicando casos de overfitting, underfitting, utilidad de los resultados, relación obtenida entre la métrica de ML y la de negocio, etc. etc.

## 6.3 CONSIDERACIONES DE PRODUCCIÓN

Menciona consideraciones técnicas de una posible puesta en producción. P.ej. condiciones de monitoreo del desempeño de los modelos, integración con streams de datos, servicios en la nube, etc.

# 7. CONCLUSIONES

Haz un breve juicio crítico de tu proyecto y del posible trabajo futuro que podría surgir a partir de tus resultados.