

77.50 Introducción a los sistemas inteligentes

## Trabajo Práctico Final

### Integrantes:

Alumno	padron
Llauró, Manuel Luis	95736
Rial, Sebastián Andrés	90309
Blanco, Sebastian Ezequiel	98539

**Fecha de Entrega:** 24/06/2019

**GitHub:**

<https://github.com/BlancoSebastianEzequiel/7750-TPFinal>

# Índice

<b>1. Fase 1: Comprensión del negocio</b>	<b>1</b>
1.1. Determinar los objetivos del negocio	1
1.1.1. Escenario actual	1
1.1.2. Objetivos del negocio	1
1.1.3. Criterios de éxito del negocio	1
1.2. Evaluación de la situación	1
1.2.1. Inventario de recursos	1
1.2.2. Requisitos, supuestos y restricciones	1
1.2.3. Riesgos y contingencias	2
1.2.4. Terminología	2
1.2.5. Costos y beneficios	3
1.3. Determinar objetivos de Minería de Datos	3
1.3.1. Objetivo de minería de datos	3
1.3.2. Criterios de éxito de minería de datos	3
1.4. Realizar el Plan del Proyecto	4
1.4.1. Plan de proyecto	4
1.4.2. Validación inicial de las herramientas	4
<b>2. Fase 2: Comprensión de los datos</b>	<b>5</b>
2.1. Recolectar los datos Iniciales	5
2.1.1. Reporte de la recolección de datos iniciales	5
2.1.2. Descubrir datos	5
2.2. Reporte de descripción de datos	5
2.3. Exploración de los datos	5
2.3.1. Reporte de exploración de datos	5
2.4. Verificación de calidad de datos	5
2.4.1. Reporte de calidad de datos	5
<b>3. Fase 3: Preparación de los datos</b>	<b>6</b>
3.1. Seleccionar los datos	6
3.1.1. Inclusión/Exclusión de datos	6
3.2. Limpiar los datos	6
3.2.1. Reporte de limpieza de datos	6
3.3. Estructurar los datos	6
3.3.1. Derivación de atributos	6
3.3.2. Generación de registros	6
3.4. Integrar los datos	6
3.4.1. Unificación de los datos	6
3.5. Formato de los datos	6
3.5.1. Reporte de formato de los Datos	6
<b>4. Fase 4: Modelado</b>	<b>7</b>

4.1.	Seleccionar una técnica de modelado . . . . .	7
4.1.1.	Técnica de modelado . . . . .	7
4.1.2.	Supuestos de modelado . . . . .	7
4.2.	Generar el diseño de las pruebas . . . . .	7
4.2.1.	Diseño de las pruebas . . . . .	7
4.3.	Construir el modelo . . . . .	7
4.3.1.	Configuración de parámetros . . . . .	7
4.3.2.	Modelos . . . . .	7
4.3.3.	Descripción del modelo . . . . .	7
4.4.	Evaluar el modelo . . . . .	7
4.4.1.	Evaluación del modelo . . . . .	7
4.4.2.	Revisión de la configuración de parámetros . . . . .	7
<b>5.</b>	<b>Fase 5: Evaluación . . . . .</b>	<b>8</b>
5.1.	Evaluar Resultado . . . . .	8
5.1.1.	Valoración de los resultados de minería de datos . . . . .	8
5.1.2.	Modelo aprobado . . . . .	8
5.2.	Proceso de revisión . . . . .	8
5.2.1.	Revisión del proceso . . . . .	8
5.3.	Determinar Próximos pasos . . . . .	8
5.3.1.	Listado de posibles acciones . . . . .	8
<b>6.</b>	<b>Fase 6: Implementación . . . . .</b>	<b>9</b>
6.1.	Plan de Implementación . . . . .	9
6.1.1.	Plan de Implementación . . . . .	9
6.2.	Plan de monitoreo y mantenimiento . . . . .	9
6.2.1.	Plan de monitoreo y mantenimiento . . . . .	9
6.3.	Informe Final . . . . .	9
6.3.1.	Informe Final . . . . .	9
6.3.2.	Presentación final . . . . .	9
6.4.	Revisión del proyecto . . . . .	9
6.4.1.	Documentación de la experiencia realizada . . . . .	9
<b>7.</b>	<b>Conclusiones . . . . .</b>	<b>10</b>

# **1. Fase 1: Comprensión del negocio**

## **1.1. Determinar los objetivos del negocio**

### **1.1.1. Escenario actual**

Al momento de escribir este informe, la empresa en cuestión tiene interés en desarrollar un canal de youtube, y requiere medir de alguna forma las probabilidades de éxito de llevar a cabo videos populares.

### **1.1.2. Objetivos del negocio**

El objetivo es en este caso poder cuáles son las condiciones que un canal de youtube y sus videos debe cumplir para que los mismos sean exitosos en la plataforma.

### **1.1.3. Criterios de éxito del negocio**

El proyecto se considerará exitoso si se llegan a detectar las variables clave que influyen en que los videos del canal de youtube sean exitosos o no, y en qué nivel influye cada una de ellas. Se considerará que los videos del canal de youtube si los mismos obtienen un porcentaje de X likes y más de 2 millones de vistas por video.

## **1.2. Evaluación de la situación**

### **1.2.1. Inventario de recursos**

Para desarrollar el presente proyecto se cuenta con una amplia gama de recursos que asegura un desarrollo de calidad y confianza del mismo. Se cuenta con un set de datos extraído de la pagina de kaggle sobre videos de youtube que fueron subidos a la plataforma. Además se cuenta con herramientas de software de análisis y visualización de datos líderes en el mercado, como Pandas. Por último, se cuenta con personal altamente calificado para la correcta interpretación de los mismos.

### **1.2.2. Requisitos, supuestos y restricciones**

**1.2.2.1. Requisitos** Contar con datos suficientes y sobre todo representativos de la plataforma youtube

**1.2.2.2. Supuestos** Los datos en estudio son lo suficientemente correctos como para poder sacar conclusiones confiables a partir de ellos.

**1.2.2.3. Restricciones** Se cuenta solamente con datos de videos subidos a youtube Estados Unidos. No se asegura que los resultados sean válidos para otros países.

### 1.2.3. Riesgos y contingencias

Si bien se puede llevar a cabo un análisis lo más riguroso posible, siempre existe la posibilidad de que un video pueda no ser exitoso porque hay que tener en cuenta que el éxito de cada video puede depender de muchos factores de incertidumbre, y si bien se cumplen patrones, no hay nada que asegure al 100%. En caso de detectar que un video no está teniendo el éxito esperado, habrá que recurrir a las diversas métricas que pueda ofrecer Youtube a desarrolladores y analizar la situación.

### 1.2.4. Terminología

#### 1.2.4.1. Glosario de términos del negocio

**Youtube:** es una plataforma que te permite subir videos a partir de la creación de un canal.

**Canal de youtube:** Es el equivalente a crearse una cuenta en cierta aplicación, en la cual se almacenan los videos subidos.

**Categoría:** define el nombre de un grupo de videos con cualidades comunes.

**likes/dislikes:** Es un atributo de un video que muestra la cantidad de usuarios que le dieron like al video en cuestión.

**Vistas:** Es la cantidad de veces que el video fue visto.

**Video exitoso:** Aquel cuya cantidad de vistas es mayor a 200000 y promedio de likes versus dislikes es mayor al 80 %.

#### 1.2.4.2. Glosario de términos de la minería de datos

**Atributo:** dato sobre alguna característica de las observaciones.

**Atributo relevante:** atributo que juega un papel principal en la clasificación, por lo que la clase dependerá en alguna medida de qué valor tenga.

**Registro:** fila que representa una observación, está compuesto de atributos.

**Dataset:** conjunto de datos a ser utilizados para la ejecución de los algoritmos de Data Mining, está compuesto de registros.

**Filtrado de atributos:** especifica al dataset formado considerando sólo los atri-

butos relevantes.

**Regla:** es una implicación, que representa una acción mediante una condición. Sigue la estructura “Si..., entonces...”.

**Soporte:** es la relación entre la cantidad total de registros del dataset que cumplen la regla y la cantidad de observaciones procesadas.

**Confianza:** es la relación entre la cantidad total de observaciones de la clase mayoritaria que cumplen la regla y la cantidad de observaciones que fueron afectadas por esa misma regla.

**Captura:** es la relación entre la cantidad de observaciones de la clase mayoritaria que cumplen la regla y la cantidad de observaciones procesadas pertenecientes a esa misma clase.

**Coefficiente de correlación de Pearson:** es la estadística de prueba que mide la relación estadística, o asociación, entre dos variables continuas. Es conocido como el mejor método para medir la asociación entre variables de interés porque se basa en el método de covarianza. Da información sobre la magnitud de la asociación, o correlación, así como la dirección de la relación

### 1.2.5. Costos y beneficios

El beneficio del proyecto es detectar las características que hacen a un video de youtube sea exitoso. De esta manera se pueden tener en cuenta ciertos parametros para poder desarrollar un video que exitoso basandose en la historia. Al ser un trabajo final educativo, no hay costos.

## 1.3. Determinar objetivos de Minería de Datos

### 1.3.1. Objetivo de minería de datos

El objetivo de minería de datos es el análisis de los datos obtenidos a partir de la información disponible, buscando obtener así información relevante que permita predecir las condiciones bajo las cuales una aplicación es exitosa.

### 1.3.2. Criterios de éxito de minería de datos

Selección de al menos 4 reglas con soporte mayor o igual al 20

## **1.4. Realizar el Plan del Proyecto**

### **1.4.1. Plan de proyecto**

- Recolección de datos: 5 horas.
- Preparación de datos: 5 horas.
- Ejecución del algoritmo de Inducción: 4 horas.
- Análisis de resultados de algoritmo de Inducción: 4 horas.
- Combinación de resultados: 3 horas.
- Elaboración de reporte: 7 horas.

### **1.4.2. Validación inicial de las herramientas**

Se utilizarán las siguientes herramientas:

- Python
- Pandas
- Weka
- Jupiter Notebook
- Numpy

## **2. Fase 2: Comprensión de los datos**

### **2.1. Recolectar los datos Iniciales**

#### **2.1.1. Reporte de la recolección de datos iniciales**

#### **2.1.2. Descubrir datos**

### **2.2. Reporte de descripción de datos**

### **2.3. Exploración de los datos**

#### **2.3.1. Reporte de exploración de datos**

### **2.4. Verificación de calidad de datos**

#### **2.4.1. Reporte de calidad de datos**



### **3. Fase 3: Preparación de los datos**

#### **3.1. Seleccionar los datos**

##### **3.1.1. Inclusión/Exclusión de datos**

#### **3.2. Limpiar los datos**

##### **3.2.1. Reporte de limpieza de datos**

#### **3.3. Estructurar los datos**

##### **3.3.1. Derivación de atributos**

##### **3.3.2. Generación de registros**

#### **3.4. Integrar los datos**

##### **3.4.1. Unificación de los datos**

#### **3.5. Formato de los datos**

##### **3.5.1. Reporte de formato de los Datos**

## **4. Fase 4: Modelado**

### **4.1. Seleccionar una técnica de modelado**

#### **4.1.1. Técnica de modelado**

#### **4.1.2. Supuestos de modelado**

### **4.2. Generar el diseño de las pruebas**

#### **4.2.1. Diseño de las pruebas**

### **4.3. Construir el modelo**

#### **4.3.1. Configuración de parámetros**

#### **4.3.2. Modelos**

#### **4.3.3. Descripción del modelo**

### **4.4. Evaluar el modelo**

#### **4.4.1. Evaluación del modelo**

#### **4.4.2. Revisión de la configuración de parámetros**

## **5. Fase 5: Evaluación**

### **5.1. Evaluar Resultado**

5.1.1. Valoración de los resultados de minería de datos

5.1.2. Modelo aprobado

### **5.2. Proceso de revisión**

5.2.1. Revisión del proceso

### **5.3. Determinar Próximos pasos**

5.3.1. Listado de posibles acciones

## **6. Fase 6: Implementación**

### **6.1. Plan de Implementación**

#### **6.1.1. Plan de Implementación**

### **6.2. Plan de monitoreo y mantenimiento**

#### **6.2.1. Plan de monitoreo y mantenimiento**

### **6.3. Informe Final**

#### **6.3.1. Informe Final**

#### **6.3.2. Presentación final**

### **6.4. Revisión del proyecto**

#### **6.4.1. Documentación de la experiencia realizada**

## 7. Conclusiones