

PROYECTO FINAL MÁSTER  
CLASIFICADOR DOCUMENTOS MÉDICOS HOPE  
2020 - 2021

Ruben Vasallo Gonzalez

9 de junio de 2020



# Índice general

<b>I</b>	<b>Objetivo del Proyecto</b>	<b>5</b>
<b>1.</b>	<b>Introducción</b>	<b>7</b>
1.1.	Definición del proyecto y objetivo . . . . .	7
1.2.	Que es el proyecto HOPE . . . . .	7
<b>2.</b>	<b>Extracción y procesamiento de datos</b>	<b>9</b>
2.1.	Extracción . . . . .	9
2.2.	Procesado de datos . . . . .	10
<b>3.</b>	<b>Análisis de los datos</b>	<b>11</b>
3.1.	Análisis de componentes principales . . . . .	11
<b>4.</b>	<b>Aproximación de resultados</b>	<b>13</b>
4.1.	Aproximación por Vecinos más próximos (K-NN) . . . . .	13
<b>II</b>	<b>Objetivo del Proyecto</b>	<b>15</b>
<b>5.</b>	<b>Modelos Predictivos</b>	<b>17</b>
5.1.	Regresión logística . . . . .	17
<b>III</b>	<b>Conclusiones</b>	<b>19</b>
<b>6.</b>	<b>Resultados obtenidos</b>	<b>21</b>



## **Parte I**

### **Objetivo del Proyecto**



# Capítulo 1

## Introducción

### 1.1. Definición del proyecto y objetivo

El proyecto de fin de máster nace de la necesidad de clasificar y recomendar resultados sobre estudios clínicos de confianza y actualizados. En Internet existe muchísima información sobre medicina y salud y no siempre toda es de fiar.

Actualmente existen bases de datos de confianza en donde los científicos y el público en general puede buscar informes sobre estudios clínicos desarrollados anteriormente, pero no siempre es fácil o rápido encontrar estos resultados.

Actualmente existe el proyecto HOPE, donde médicos de todo el mundo puede consultar en una base de datos informes médicos relacionados con los síntomas que puedan tener sus pacientes y ver que otros tratamientos han dado resultado. Todo y con eso, el sistema no siempre devuelve los artículos más relevantes o actualizados por lo que, no siempre la información consultada es útil.

En este ámbito, los profesionales médicos pueden valorar si la información recibida ha sido útil o no respecto a la búsqueda que han realizado, por lo que con ese *feedback*, se pretende crear un recomendador que aproxime mejor los resultados a las búsquedas realizadas.

Con todo esto, el objetivo que se pretende alcanzar en este proyecto es el de entrenar un modelo predictivo que sea capaz de discriminar en base a los términos buscados por el médico, cuales son los mejores artículos útiles que pueden ayudar a resolver el problema.

### 1.2. Que es el proyecto HOPE

El proyecto HOPE (que significa *Health Operations for Personalized Evidence* en inglés) nace de la necesidad de ayudar a los científicos médicos a encontrar la información que necesitan de la manera más rápida y fácil posible. Existe infinidad de información médica en Internet de miles de proyectos investigados y esto hace que, muchas veces sea complicado encontrar la información sobre ensayos médicos para tratar información. En el ámbito de la medicina el tiempo perdido puede costar vidas y es un precio demasiado elevado a pagar, tanto a nivel económico como emocional.

El proyecto HOPE es un sistema basado en inteligencia artificial para identificar los datos claves de casos clínicos registrados en la Historia Clínica Electrónica, en base a los cuales realiza una búsqueda única por paciente para proporcionar al profesional de la salud recomendaciones de tratamientos, estudios de investigación, información para el paciente, todo en base a registros de fuentes científicas de información.





## Capítulo 2

# Extracción y procesamiento de datos

### 2.1. Extracción

El Origen de los datos se encuentra en una Base de datos SQL distribuida en dos tablas, que pasamos a detallar a continuación:

- Tabla *fed\_hope\_sugerencia*

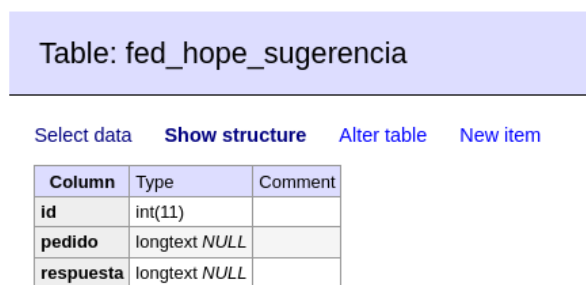


Table: fed_hope_sugerencia		
<a href="#">Select data</a> <a href="#">Show structure</a> <a href="#">Alter table</a> <a href="#">New item</a>		
Column	Type	Comment
id	int(11)	
pedido	longtext NULL	
respuesta	longtext NULL	

Figura 2.1: Visualización de los atributos de la tabla *fed\_hope\_sugerencia*

En esta tabla encontraremos la sugerencia que dio el programa HOPE en base a los parámetros que introdujo el usuario, almacenado en el atributo pedido y la respuesta que dio el programa, almacenado en el atributo respuesta. Todos los datos son almacenados en formato json.

**Atributo pedido:** Si analizamos el formato de datos que tenemos del atributo pedido observamos los siguientes atributos:

```
{
  "data": {
    "type": "emr--em",
    "attributes": {
      "name": null,
      "affected_organ": "",
      "age": "75",
      "diagnostic_main": "FISTULA PERITONEAL",
      "gender": "male",
      "medical_history": "Paciente de 75 años con antecedentes de gastrostomía total por adenocarcinoma gástrico que intercurrió con eventración y posterior formación de fístula entero-atmosférica. Actualmente cursa postoperatorio de resección intestinal mu001s eventroplástica con colocación de malla. Al examen impresiona en regular estado general, lúcido, hemodinámicamente estable, sin signos de falla de bomba. Regular entrada de aires, ruidos crepitantes bilaterales. Abdomen blando, depresible, levemente doloroso a la palpación profunda. Herida cubierta por apósitos estériles. Catarsis positiva. Diuresis positiva.\n\nPROBLEMAS ACTIVOS:\n- POP resección intestinal mu001s eventroplástica: Paciente clínicamente estable, hemodinámicamente compensado. Refiere buena tolerancia al dolor. Afebril hace 72 hs. Cumple 4to día de tratamiento con piperacilina tazobactam por neumonía broncoaspiratoria con aislamiento de E.coli BLEE. Hemocultivos vienen negativos. TAC de tórax y abdomen informan: Consolidación bilateral mu001s derrame pleural bilateral. Colección mu001s laminar posterior a ambos subfascículos rectos de 10x2x0.4 cm, y otra colección mu001s en herida quirúrgica de pared de 10x2.8x1.2 cm. Se da aviso a cirujano tratante. Persiste con estado nauseoso, por lo que continúa con antieméticos reglados. \n\nEn aislamiento de contacto por germen multirresistente. \nSe da informe. Control evolutivo."
    }
  }
}
```

Podemos ver varios atributos haciendo referencia a los síntomas que consulta el profesional médico.

**Atributo respuesta:** En el atributo respuesta obtenemos entre otros datos, el listado de artículos científicos sugeridos relacionados con los síntomas descritos por el profesional médico. Esta respuesta es muy amplia pero entre todos los atributos, podemos observar un listado de ids de artículos, con sus fechas de revisión de estos artículos y unas claves descriptivas para esos artículos.

**Nota:** Mostramos una pequeña parte del contenido de una observación del atributo respuesta.

```
{
  "data": {
    "type": "emr--rpt",
    "id": "ef6d3fb-afef-4659-ba8d-4a7f5e0d4f5",
    "attributes": {
      "id": "376",
      "uuid": "ef6d3fb-afef-4659-ba8d-4a7f5e0d4f5",
      "language": "es",
      "name": null,
      "status": "true",
      "created": "20200224",
      "changed": "20200224",
      "affected_organ": null,
      "age": "75",
      "diagnostic_main": "FISTULA PERITONEAL",
      "diagnostic_main_terms": null,
      "gender": "male",
      "histopathology": null,
      "medical_history": "Paciente de 75 años con antecedentes de gastrostomía total por adenocarcinoma gástrico que intercurrió con eventración y posterior formación de fístula entero-atmosférica. Actualmente cursa postoperatorio de resección intestinal mu001s eventroplástica con colocación de malla. Al examen impresiona en regular estado general, lúcido, hemodinámicamente estable, sin signos de falla de bomba. Regular entrada de aires, ruidos crepitantes bilaterales. Abdomen blando, depresible, levemente doloroso a la palpación profunda. Herida cubierta por apósitos estériles. Catarsis positiva. Diuresis positiva.\n\nPROBLEMAS ACTIVOS:\n- POP resección intestinal mu001s eventroplástica: Paciente clínicamente estable, hemodinámicamente compensado. Refiere buena tolerancia al dolor. Afebril hace 72 hs. Cumple 4to día de tratamiento con piperacilina tazobactam por neumonía broncoaspiratoria con aislamiento de E.coli BLEE. Hemocultivos vienen negativos. TAC de tórax y abdomen informan: Consolidación bilateral mu001s derrame pleural bilateral. Colección mu001s laminar posterior a ambos subfascículos rectos de 10x2x0.4 cm, y otra colección mu001s en herida quirúrgica de pared de 10x2.8x1.2 cm. Se da aviso a cirujano tratante. Persiste con estado nauseoso, por lo que continúa con antieméticos reglados. \n\nEn aislamiento de contacto por germen multirresistente. \nSe da informe. Control evolutivo."
    }
  },
  "response": {
    "articles": [
      {
        "id": "279542",
        "title": "Indications and Results of Reconstructive Techniques with Flap Transposition in Patients Requiring Complex Thoracic Surgery: A 12-Year Experience",
        "abstract": "Abstract: The aim of this retrospective study was to report our experience in a referral unit of general thoracic surgery analyzing the early results after flap transposition.",
        "attributes": {
          "label": "RECONSTRUCTION",
          "value": "Flap Transposition is an infrequent but far from exceptional thoracic surgical procedure. The aim of this retrospective study was to report our experience in a referral unit of general thoracic surgery analyzing the early results after flap transposition."
        },
        "date": "2013-11",
        "get_dates": {
          "label": "RECONSTRUCTION",
          "value": "Overall, a surgical approach adopting flap reconstruction techniques was performed in 83 patients (54 males, 27 females) with a median age of 62.60years (range 26-87). Flap transposition was necessary to reconstruct chest wall after resection for malignancy (27 patients), to repair postoperative viscera perforation (15 patients), and to fill residual cavities secondary to"
        }
      }
    ]
  }
}
```

Figura 2.2: Ejemplo de contenido del atributo respuesta de una observación

## 2.2. Procesado de datos

TODO

## Capítulo 3

# Análisis de los datos

### 3.1. Análisis de componentes principales

TODO



## Capítulo 4

# Aproximación de resultados

### 4.1. Aproximación por Vecinos más próximos (K-NN)

TODO



## **Parte II**

### **Objetivo del Proyecto**





## Capítulo 5

# Modelos Predictivos

### 5.1. Regresión logística

TODO



### **Parte III**

### **Conclusiones**



## Capítulo 6

# Resultados obtenidos

TODO