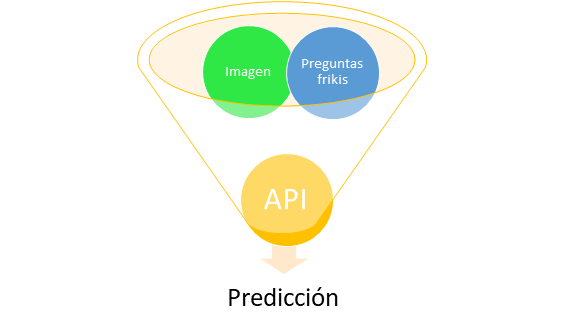
BRAINSTORMING

**\*github**

**Bases de Datos** -> Combinar bases ya creadas con elaboración propia.

* Podemos sacar alguna información mediante Webscraping

API para que pueda interactuar el usuario.

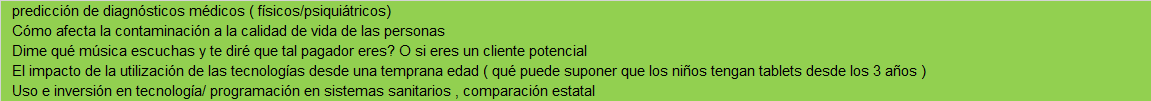


* Hacerlo por transfer learning para poder utilizar una programación más elaborada
* Elegir un tema orginal para diferenciarnos
* Además de una predicción, hacer análisis de esa bbdd mediante otros lenguajes como sql/mongoDB...

**TEMAS**:

* Maleta abandonada/no abandonada a través de una imagen
* mascotas - ¿ antecedentes?
* ojos azules - ¿ delincuentes?
* predicción de diagnósticos médicos ( físicos/psiquiátricos)
* Comparación de bbdd estudios de distintos países para explicar la diferencia entre ellos
* Evolución histórica
* Audios de emergencia para saber tu situación
* Cómo afecta la contaminación a la calidad de vida de las personas
* Problemas mentales a edades tempranas. Causas
* Características que pueden llevar a una persona a la delincuencia.
* Con una foto ( por los rasgos faciales ) saber si es un delincuente potencial. ¿ Está relacionado ?
* Rasgos físicos-> estimación del tipo de trabajo que tienes. Posibles usos : clientes potenciales
* Personas propensas a sufrir una estafa en internet
* Según ciertas características de una persona, saber qué precio estaría dispuesto a pagar.
* Dime qué música escuchas y te diré que tal pagador eres? O si eres un cliente potencial
* Qué variables llevan a tener una mayor esperanza de vida ( incluyendo varios países)
* El impacto de la utilización de las tecnologías desde una temprana edad ( qué puede suponer que los niños tengan tablets desde los 3 años )
* La evolución de los requerimientos de las tecnologías en la demanda laboral
* ( intentar predecir también la obsolescencia de las tecnologías tb)
* Estudio sobre la transición energética (muy de empresa)
* Estudio sobre el valor añadido en la economía de la formación tecnológica frente a otros sectores
* Estudios sobre casos de ineficiencia en automatización de procesos. Búsqueda del punto de corte de beneficio para la utilización de la tecnología o mano de obra manual.
* Uso e inversión en tecnología/ programación en sistemas sanitarios , comparación estatal

TEMAS SELECCIONADOS:



**OTROS TEMAS:** propuestas

1. Predicciones relacionadas con la población dependiente mayor de 65 años a nivel ciudad / región / comunidad / país, etc. con la intención de conocer la necesidad de nuevos centros especializados para personas mayores dependientes en distintas zonas del país.
2. Crear un modelo para el diagnóstico, evaluación y gestión de la reputación de una o varias universidades de España (podríamos enviar un formulario Google Forms a los amigos, familiares, etc. para obtener información relevante - hacer preguntas que nos ayudan entender mejor la experiencia y evolución de los alumnos de distintas universidades). Con la idea de crear un ÍNDICE de la reputación universitaria, entre otras.
3. Predicción del precio de cotización de empresas del IBEX 35
   1. identificar las variables que pueden afectar al precio de cotización (e.g. precio del petróleo, del oro, otros índices bursátiles, materias primas, etc.)
   2. crear un sistema de web scrapping que extraiga y estructure los datos, y que pasan a ser el dataset de input para crear modelos de ML o DL
   3. una vez extraídos los datos se realizan varios modelos de predicción, para predecir el valor futuro de las acciones de empresas del IBEX 35.

**Páginas web de interés:**

1. [**https://keras.io/api/applications/**](https://keras.io/api/applications/)- Modelos de DL pre-entrenados - Deep Learning models that are made available alongside pre-trained weights. These models can be used for **prediction**, **feature extraction**, and **fine-tuning**.
2. [**https://www.tensorflow.org/datasets/catalog/overview**](https://www.tensorflow.org/datasets/catalog/overview)- Datasets de audio, graphs, image, image classification, object detection, text, video, etc.
3. [**https://tfhub.dev/**](https://tfhub.dev/)- Modelos ML pre-entranados - trained, ready-to-deploy machine learning models in one place.
4. <https://openaq.org/#/locations?page=1> . Hay bastantes bases de datos
5. https://stats.oecd.org/

BASES DE DATOS QUE SE PODRÍAN UTILIZAR

* Perfección de diagnósticos médicos

Necesitamos datasets relacionados con técnicas utilizadas para los diagnósticos - mejora en la detección de los casos- inversión en tecnología médica…

De esta página podemos sacar varios:

[World Health Statistics 2020|Complete|Geo-Analysis | Kaggle](https://www.kaggle.com/utkarshxy/who-worldhealth-statistics-2020-complete)

* [Deaths by causes (reduced list), sex, place of residence and age . ECM (API identifier: /t15/p417/a2018/l0/01002c.px) - Dataset | datos.gob.es](https://datos.gob.es/en/catalogo/ea0010587-defunciones-por-causas-lista-reducida-sexo-lugar-de-residencia-y-edad-ecm-identificador-api-t15-p417-a2018-l0-01002c-px) Muertes por causas y franjas de edad
* ¿ Cómo afecta la contaminación a la calidad de vida de las personas?
* Life expectancy at birth: [World Health Statistics 2020|Complete|Geo-Analysis | Kaggle](https://www.kaggle.com/utkarshxy/who-worldhealth-statistics-2020-complete?select=lifeExpectancyAtBirth.csv)
* [World Health Statistics 2020|Complete|Geo-Analysis | Kaggle](https://www.kaggle.com/utkarshxy/who-worldhealth-statistics-2020-complete?select=airPollutionDeathRate.csv) Air Pollution Death rate
* [World Health Statistics 2020|Complete|Geo-Analysis | Kaggle](https://www.kaggle.com/utkarshxy/who-worldhealth-statistics-2020-complete?select=30-70cancerChdEtc.csv) 30-70cancerChdEtc ( probability of dying of cardiovascular disease)
* [Environment | Data (worldbank.org)](https://data.worldbank.org/topic/6) Dataset ambiental, con muchos indicadores chulos como especies amenazadas ( PODEMOS HACER MAPA EN TABLEAU!)
* [World's Air Pollution: Real-time Air Quality Index (waqi.info)](https://waqi.info/#/c/38.398/-3.42/4.3z) índice de la calidad del aire en tiempo real, ( por si nos puede servir)
* <https://datos.bancomundial.org/tema/medio-ambiente>
* <https://www.kaggle.com/brendan45774/countries-life-expectancy> . Esperanza de vida por países y años. Podemos utilizarlo para añadirlo al dataset que usemos de contaminación,etc.
* Relación entre los géneros musicales y el perfil del consumidor / pagador/inversor etc
* El impacto de la utilización de las tecnologías desde una temprana edad

Podemos demostrar que esto es tanto positivo como negativo: negativo por el punto de vista sanitario y mental y positivo desde el punto de vista educativo, que tiene acceso a muchos más recursos y más predisposición a elegir una vía profesional con carga tecnológica importante.

* [myopia: Myopia data set. in emilelatour/purposeful: Purposeful Selection (Hosmer, Lemeshow) (rdrr.io)](https://rdrr.io/github/emilelatour/purposeful/man/myopia.html) - va de la mano de este estudio, igual el dataset lo podemos utilizar para algo [Predicting myopia in a child population👓 – Jose Luis Fernández Data Science (joseluisfernandez.me)](https://joseluisfernandez.me/predicting-child-myopia/)
* [Proporción de personas entre 16 y 74 años que han utilizado Internet en los últimos tres meses por territorio histórico y sexo. Indicadores de la Agenda 2030. C.A. de Euskadi. - Dataset | datos.gob.es](https://datos.gob.es/en/catalogo/a16003011-proporcion-de-personas-entre-16-y-74-anos-que-han-utilizado-internet-en-los-ultimos-tres-meses-por-territorio-historico-y-sexo-indicadores-de-la-agenda-2030-c-a-de-euskadi)
* [Proportion of people who used the Internet in the last three months by autonomous community (API identifier: 45877) - Dataset | datos.gob.es](https://datos.gob.es/en/catalogo/ea0010587-proporcion-de-personas-que-utilizan-internet-en-los-ultimos-tres-meses-por-comunidades-autonoma-identificador-api-45877)
* Uso e inversión en tecnología/ programación en sistemas sanitarios , comparación estatal
* [Research and development (R&D) - Gross domestic spending on R&D - OECD Data](https://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm) Gasto por países en tecnología
* [Scientific and technical journal articles | Data (worldbank.org)](https://data.worldbank.org/indicator/IP.JRN.ARTC.SC) Cantidad de publicaciones científicas por países, igual lo podemos relacionar con el gasto

[**https://www.mapa.gob.es/es/desarrollo-rural/estadisticas/Incendios\_default.aspx**](https://www.mapa.gob.es/es/desarrollo-rural/estadisticas/Incendios_default.aspx)

[**Incendios Forestales (miteco.gob.es)**](https://www.miteco.gob.es/es/biodiversidad/servicios/banco-datos-naturaleza/informacion-disponible/incendios-forestales.aspx)

**Propuestas de temas**

**He estado investigando y creo que la mejor fuente para obtener la información es Kaggle sin duda ( leyendo bien al principio que los datos son reales y no son sintéticos, que lo he visto en algunos casos).**

**Aquí hay varias propuestas de temas que podríamos interrelacionar , combinar , incluír mapas en algunos casos y un poco de todo.**

**A ver qué os parecen ( podéis completarlos / matizarlos también) → algunos me parecen la leche!! :)**

**Propuesta para TFM**

**1) Relacionar la producción de energía con los incendios en Estados Unidos ( falta buscar dataset de incendios)**

[**Industrial Production: Electric and gas utilities | Kaggle**](https://www.kaggle.com/sadeght/industrial-production-electric-and-gas-utilities)

**Industrial production of electric and gas utilities in the United States, from the years 1985–2018.**

**2) Relacionar películas y sus géneros con las preferencias de comida ( o cualquier otro tipo de preferencias)**

[**TMDb Top 10,000 Popular Movies Dataset | Kaggle**](https://www.kaggle.com/balaka18/tmdb-top-10000-popular-movies-dataset)

**This is dataset of the** 10,000 most popular movies across the world, ***irrespective of language and recency***

[**Food Preferences | Kaggle**](https://www.kaggle.com/vijayashreer/food-preferences)

**Food Preference Survey 2019**

**3) Relación de las empresas que han prosperado en el Covid con ( relacionar esto con algo, un fenómeno o lo que sea)**

[**Market capital addition during pandemic | Kaggle**](https://www.kaggle.com/parulpandey/market-capital-addition-during-pandemic)

**Financial Times recently shared a very interesting article on the** [**health of the companies**](https://www.ft.com/content/844ed28c-8074-4856-bde0-20f3bf4cd8f0#comments-anchor) **during the current pandemic.**

**4) Hacer un análisis de las charlas TED en distintos países**

[**TED – Ultimate Dataset | Kaggle**](https://www.kaggle.com/miguelcorraljr/ted-ultimate-dataset)

**Podemos unir datasets de varios países, hacer clusters por sectores y categorizarlos, agrupándolos por temáticas. Ver si la mayoría de speakers son hombres o mujeres. Ver las visualizaciones y alas temáticas más populares. Podemos combinar también este dataset con el de las películas y ver si coinciden sde agluna manera las charlas de años concretos con las películas más populares producidas en esos años.**

**5) Relacionar la producción de energía la distribución de las lunas llenas- relacionar la energía con los fenómenos lunares**

[**Full Moon Calendar 1900-2050 | Kaggle**](https://www.kaggle.com/lsind18/full-moon-calendar-1900-2050)

**The full moon is the lunar phase when the Moon appears fully illuminated from Earth's perspective. This occurs when Earth is located between the Sun and the Moon and calls *syzygy*. The full moon occurs roughly once a month.**

[**Industrial Production: Electric and gas utilities | Kaggle**](https://www.kaggle.com/sadeght/industrial-production-electric-and-gas-utilities)

**Industrial production of electric and gas utilities in the United States, from the years 1985–2018.**

**6) Comparar dataset de emisiones de co2 con dataset de producción energética.**

**Podemos incluso hacer un mapa bonito con los territorios con más emisiones y ver si está relacionado con otras cosas, como los sectores o países de las compañías líderes en bolsa en el covid. Sobretodo los países:**

[**CO2\_GHG\_emissions-data | Kaggle**](https://www.kaggle.com/yoannboyere/co2-ghg-emissionsdata)

**This dataset contains CO2 and GHG emissions for countries since 1750 until 2017.**

[**Industrial Production: Electric and gas utilities | Kaggle**](https://www.kaggle.com/sadeght/industrial-production-electric-and-gas-utilities)

**Industrial production of electric and gas utilities in the United States, from the years 1985–2018.**

**7) Relacionar la violencia de género con el género de las películas ( en común están las fechas)**

[**Violence Against Women and Girls | Kaggle**](https://www.kaggle.com/andrewmvd/violence-against-women-and-girls)

**This dataset aggregates agreement with key questions across gender, education level and many other socioeconomic variables from 70 different countries.**

[**TMDb Top 10,000 Popular Movies Dataset | Kaggle**](https://www.kaggle.com/balaka18/tmdb-top-10000-popular-movies-dataset)

**This is dataset of the** 10,000 most popular movies across the world, ***irrespective of language and recency***