**PASOS DEL PROYECTO**

1. **ELEGIR TEMATICA**

**¿Qué características tiene un país europeo que apuesta por la energía solar fotovoltaica?**

1. **OBTENCION DE DATOS**

[**https://datacommons.org/**](https://datacommons.org/)

[**https://globalsolaratlas.info/map?c=36.385913,-18.720703,5&r=PRT**](https://globalsolaratlas.info/map?c=36.385913,-18.720703,5&r=PRT)

**https://globalsolaratlas.info/global-pv-potential-study**

[**Global Photovoltaic Power Potential by Country | Data Catalog (worldbank.org)**](https://datacatalog.worldbank.org/search/dataset/0038379)

1. **DEFINIR HIPÓTESIS**

¿Qué características tiene un país europeo que apuesta por la energía solar fotovoltaica?

¿Dependerá del tipo de población?

* Habitantes / km2
* Nivel de educación
* Esperanza de vida
* Actividad profesional

¿Dependerá del clima?

* Precipitaciones medias

¿Dependerá de la superficie del país?

* Superficie km2
* Superficie verde

¿Dependen exclusivamente de la irradiación horizontal global (GHI)? – información de contraste para los 10 países con mayor capacidad de producción

1. **PROCESADO DE DATOS**

Juntar los datos y ponerlos en formato entendible

1. **LIMPIEZA DE DATOS**

Eliminar columnas inútiles, duplicados…

1. **EDA**

Resumen del EDA en la exposición del proyecto EDA.

1. **CONCLUSIONES**

Coste de construcción: 0.10-0.13 USD / kWh

PIB > 42 300 USD

Consumo energético > 5800 kWh per cápita

IGH > 2.75 kWh/m2/día

Capacidad instalada Wp per cápita

1. **HIPOTESIS A ESTUDIAR**

**Comprobación mediante ML de las conclusiones obtenidas con el EDA.**

Clasificación de los países según su capacidad instalada (inversión): MUY BAJA, BAJA, MEDIA, MUY ALTA, ALTA

**SUPERVISADO CATEGORICO**

Clasificación de los países en Europa según las 5 features obtenidas en el EDA. Para ver en qué países deberíamos poner el ojo. Ver si coincide con la clasificación anterior.

INVERTIR SI / NO

Crear un target nuestro en los países europeos, en función al criterio obtenido en el EDA. Extenderlo al resto de países del mundo. Basandonos en que el mercado europeo es el más asentado.

Extender este estudio al resto del mundo.

**Train test**

**SUPERVISADO NUMÉRICO**

Predecir países donde poner la vista para invertir.

MODELOS:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Simple imputer | Escalado | Clustering | Features | Modelo | TEST – r2 |
| Media | Standard Scaler | KMEANS – 6 gen | Todas |  | 4% |
| Mediana | Standard Scaker | Kmeans – 6 gen | Todas |  | 6.6% |
| Mediana | MinMax Scaler | Kmeans- 6 gen | Todas | Elastic net | 22% |
| Media | MinMax | Kmeans – 6 gen | todas | Elastic net | 22% |
| Media | Min max | Kmeans – 2 + 2 modelos diferentes | todas | + 2 modelos diferentes | 17% |
| Media | Min Max | K Means – 2 | Totas | Elastic | 24% |
| Media | Min Max | K Means – 3 | Totas | Elastic | 18% |
| Media | Min Max | Pca KMEANS – 1 | Todas | CatBoost | 30% |
| Media | Min max | - | Todas | Catboost | 30% |
| knn 15 | Min max |  | todas | elastic | 23% |
| Knn 5 | Min max |  |  | Elastric | 26% |
| Knn5 | Satandar scaler |  |  |  | 8% |
| Knn5 | - |  | Todas | MLP Regressor | 44% |
| Simple imputer most frequent | Estándar scaler / min max / - |  | Todas | Catb | 36% |
| Simple imputer most frequent | Estándar scaler / min max / - | Tratando outlayers antes del simple imputer | todas |  | 6% |
| Simple most fre |  | Tratando los outlayers después del simple imputer | Todas |  | 5% |
| Simple most fre | Min max | Tratando los outlayers después del simple imputer | Todas | cat | 36% |
| Simple most fre | Min max |  | Sin desarrollo | Cat | 33% |
| Simple most fre | Min max |  | Sin pib | MLP regr | 26% |
| Knn 5 | * / min max |  | Sin pib | Mlp | 19% |
| Knn5 | - |  | Todas | Knn | -12% |
| Knn5 | - |  | Solo desarrollo | Ridge | 19% |
| Knn5 | - |  | Desarro+pib | Rdige | 18% |
| Knn5 |  |  | Desa+pib+irradiancia | Catboost | 45% |
| Knn5 |  |  | Desa+pib+irradiancia+ coste constr | Catboosr | 55% |
| Knn3 |  |  | Desa+pib+irradiancia+ coste constr | Cat /árbol | 51% /42% |
| Knn7 |  |  | Desa+pib+irradiancia+ coste constr | Cat /árbol | 46&/42% |
| Knn5 |  |  | Pib+consumo+irra+coste con | gradient | 50% |
| Knn5 | Min max |  | Desa+pib+irradiancia+ coste constr | Catboosr | 55% |
| Most frequert | Min max |  | Desa+pib+irradiancia+ coste constr | Cat boost | 48% |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

El dataset está desbalanceado y nos interesa conocer los que si invierten en solar. Quiza hay que disminuir los países que no invierten mucho. ¿

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Knn 5 | Min max | Todas | perceptron | Recall 71% |
| Knn 10 | Min max | Todas | Perceptrón/árbol decisión /ada boost | 43% / 57% / 57% |
| Knn5 + undersampling | Min max | Todas | Ridge lightgbm | 71% y 45% de precision |
| Knn5 + undersampling |  | Todas | Ridge lightgbm | 71% y 50% de precision |
| Mean |  | Todas |  | Menos valor |
| Knn5 + undersampling |  | Con las 4 primeras | Adaboost | 71% y 56% precisión . 85% |