



Proyecto 3

Sistema de Predicción Meteorológico

Metodología CRISP-DM

Aakriti Guerrero Bertolín
Oriol Fernández Font
Joan Pau Grau Homs

ÍNDICE

1. Entendimiento del Negocio	3
1.1. Definición de los Objetivos del Negocio	3
Antecedentes	3
Objetivos de Negocio	3
Criterios de Negocio Exitoso	4
1.2. Evaluar la Situación	4
Inventario de Recursos	4
Requisitos	5
Supuestos y Restricciones	5
Riesgos y Contingencias	5
Terminología	5
1.3. Definición de Objetivos de la Minería de Datos	6
Objetivos de Minería de Datos	6
Criterios de Éxito en la Minería de Datos	6
2. Entendimiento de los datos	7
2.1. Descripción de los datos	7
2.2. Verificación de la Calidad de los Datos	8
2.3. Exploración de los Datos	8
3. Preparación de los Datos	12
3.1. Seleccionar Datos	12
3.2. Limpieza de los Datos	13
3.3. Construcción de los Datos	13
4. Modelación	15
4.1. Selección de la Técnica de Modelado	15
4.2. Construir y Evaluar el Modelo	15
5. Evaluación	19
5.1 Evaluación de los Resultados	19
5.2 Propuestas de mejora	20
6. Despliegue	21
7. GITHUB	22

1. Entendimiento del Negocio

1.1. Definición de los Objetivos del Negocio

Antecedentes

En PREMETEO S.L. estamos comprometidos con el almacenamiento y análisis de millones de datos meteorológicos. Estos datos se utilizan exclusivamente para desarrollar sistemas avanzados de predicción meteorológica, el producto principal de nuestra empresa, diseñado para satisfacer la creciente demanda de diversas industrias.

Desde PREMETEO somos conscientes de que operamos en un mercado con más de 100 años de historia, donde existen grandes competidores cuya experiencia y capacidad valoramos profundamente. Reconocemos su veteranía y el alto nivel de los productos que ofrecen, lo que nos motiva a innovar y destacar con nuestras propias soluciones.

PREMETEO nace con el propósito de proporcionar un sistema de predicciones meteorológicas de alta precisión, superando significativamente a las opciones disponibles en el mercado. Además, buscamos satisfacer plenamente las necesidades de las empresas, ofreciendo resultados en un tiempo más reducido que otros productos similares.

Objetivos de Negocio

Con el fin de cumplir con los objetivos de PREMETEO S.L., a partir de un análisis de los datos recopilados, se analizarán las tendencias y características del comportamiento meteorológico, y así lograr predecir la tendencia a futuro. Esto no ayudará a cumplir con los objetivos de la empresa a la hora de desarrollar el sistema de predicción.

Objetivos de Negocio	KPI (Éxito empresarial)
Precisión del sistema	Aumentar la precisión de los datos y predicciones del sistema en comparación a los sistemas actuales.
Velocidad de procesamiento	Disminuir el tiempo de espera a la hora de obtener un resultado y el tiempo de ejecución de los modelos entrenados.
Gestión de datos	Debido a la gran cantidad de datos que se reciben, es importante saber tratar los datos y obtener variables significativas.
Visualización de datos	Crear una interfaz user friendly a la vez que los datos se muestren de forma detallada.

Criterios de Negocio Exitoso

Desde PREMETEO S.L. creemos que para lograr estos objetivos y dar como exitoso el proyecto, el sistema debería:

- Ofrecer predicciones meteorológicas con al menos un 90% de precisión en intervalos de tiempo de 24 a 72 horas.
- Ser capaz de procesar grandes volúmenes de datos a tiempo real.
- Presentar una interfaz accesible y generar informes o visualizaciones fáciles de interpretar para los usuarios finales.

1.2. Evaluar la Situación

Inventario de Recursos

- Datos históricos meteorológicos (mínimo 10 TB)
- Infraestructura computacional escalable (ej. clústeres Hadoop, GPU)
- Equipo multidisciplinar (científicos de datos, meteorólogos, ingenieros)
- Presupuesto inicial de 500,000€

Requisitos

- Precisión mínima del 90% en predicciones a 24-72h
- Procesamiento en tiempo real de flujos de datos meteorológicos
- Interfaz de usuario intuitiva y visualizaciones efectivas

Supuestos y Restricciones

Supuestos	Restricciones
Datos históricos de calidad	Plazo ajustado para el desarrollo e implementación
Equipo con habilidades técnicas requeridas	Presupuesto limitado para infraestructura y personal
Demanda de mercado para predicciones meteorológicas	

Riesgos y Contingencias

Riesgos	Contingencias
Datos históricos faltantes o poco fiables	Controles rigurosos de calidad y procesos de limpieza de datos
Complejidad técnica	Desarrollo del modelo por fases con validaciones intermedias
Restricciones de recursos computacionales	Provisión de infraestructura en la nube escalable
Competencia	Monitoreo continuo del mercado y actualización ágil de la hoja de ruta del producto

Terminología

- Predicción meteorológica a corto plazo: Pronósticos a 24-72 horas.
- Modelo predictivo: Sistema de inteligencia artificial para generar predicciones.

- Datos climáticos históricos: Series de datos meteorológicos pasados.
- Procesamiento en tiempo real: Análisis de flujos de datos a medida que se generan.
- Interfaz de usuario: Aplicación o plataforma para visualizar los pronósticos.
- Precisión: Nivel de exactitud de las predicciones meteorológicas.
- Escalabilidad: Capacidad de la infraestructura para manejar grandes volúmenes de datos.
- Margen de error: Diferencia entre la predicción y las condiciones meteorológicas reales.
- Modelo de suscripción: Esquema de ingresos recurrentes por el uso del servicio.
- Licenciamiento: Esquema de ingresos por permisos de uso del modelo predictivo.

1.3. Definición de Objetivos de la Minería de Datos

Objetivos de Minería de Datos

Para lograr los objetivos de negocio que se han expuesto anteriormente de manera eficiente, hemos acordado que el objetivo de la minería de datos es desarrollar un modelo de predicción de datos (meteorológicos). Dicho modelo permite presentar los datos con un gran porcentaje de acierto y rápidamente.

Criterios de Éxito en la Minería de Datos

Para cumplir con los objetivos de nuestra minería de datos, hemos definido unos criterios de éxito:

- Precisión > 90% en test
- Recall > 75% en test
- Accuracy > 80% en test

2. Entendimiento de los datos

2.1. Descripción de los datos

Variables	Description	Data Type
cloudiness_id	Identificador único que representa un estado específico de nubosidad	int64
cloudiness	Descripción del nivel de nubosidad	object
date_id	Identificador único para una fecha específica	int64
date	Fecha para identificar cuándo se realizó la observación	object
weather_id	Identificador único asociado a un tipo específico de condición meteorológica	int64
weather	Descripción de las condiciones meteorológicas observadas	object
season_id	Identificador único para la estación del año	int64
season	Estación del año	object
observation_id	Identificador único para cada registro individual de observación	int64
precipitation	Cantidad de precipitación registrada durante el periodo (mm)	float64
temp_max	Temperatura máxima registrada durante el periodo, medida en grados Celsius (°C)	float64
temp_min	Temperatura mínima registrada durante el periodo, medida en grados Celsius (°C)	float64
wind	Velocidad del viento registrada (km/h)	float64
humidity	Nivel de humedad relativa en el aire	float64
pressure	Presión atmosférica registrada (hPa)	float64
solar_radiation	Cantidad de radiación solar recibida en la superficie (W/m ²)	float64
visibility	Distancia máxima a la que se puede ver un objeto claramente (km)	float64

2.2. Verificación de la Calidad de los Datos

Para poder garantizar rendimientos adecuados del modelo a realizar, se analiza la base de datos de la empresa, donde se evidencia que es necesario realizar ajustes para tener una buena calidad de los datos. Por ello, inicialmente se excluyen las variables mencionadas a continuación:

'cloudiness_id', 'date_id', 'weather_id', 'season_id', 'observation_id', 'cloudiness', 'weather', 'season'

Una vez excluidas, se analizan el resto de variables para identificar los valores vacíos, duplicados, perdidos o nulos. Analizando la base de datos, encontramos algunas variables con nulos que se deberán tratar. A continuación las variables en mención:

'precipitation', 'temp_max', 'temp_min', 'wind'

2.3. Exploración de los Datos

En primer lugar, se realiza un análisis exploratorio de la base del Dataset. Primero se describirán las tablas cloudiness, weather y seasons ya tan solo tienen dos variables, el tipo de nubosidad, tiempo y estación y sus respectivos identificadores.

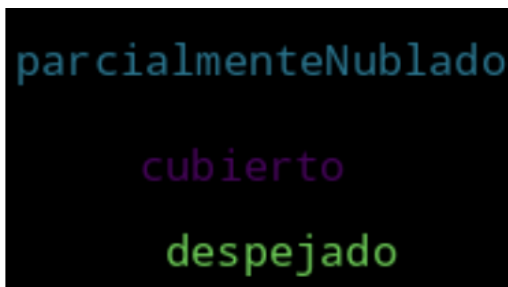


Figura 1- Cloudiness



Figura 2- Weather



Figura 3- Seasons

A continuación describiremos las variables categóricas, en la Figura 4 se puede observar que los climas que toman más presencia son la lluvia, la tormenta y el nublado, notablemente por encima de la niebla y el sol. Cuando observamos la Figura 5 vemos en detalle que el tipo de cielo más común es el parcialmente nublado, seguido del cubierto (la mitad de valores que el primero) y el despejado como el que menos aparece. Finalmente, en la Figura 6, observamos que todas las estaciones tienen unos valores de presencia prácticamente iguales, algo que nos parece lógico.

Proyecto 3: BGD-Predicción Meteorológica

Aakriti Guerrero Bertolín

Oriol Fernández Font

Joan Pau Grau

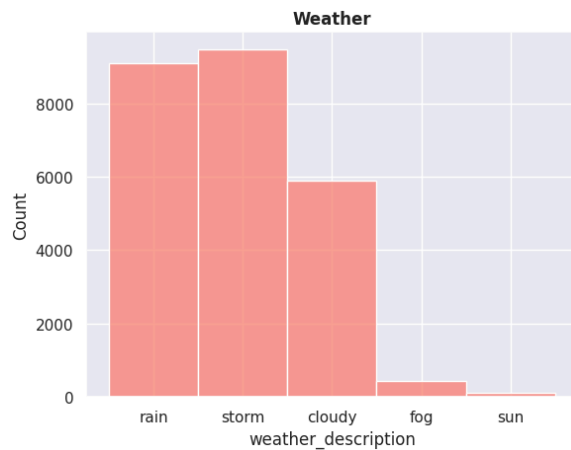


Figura 4- Weather

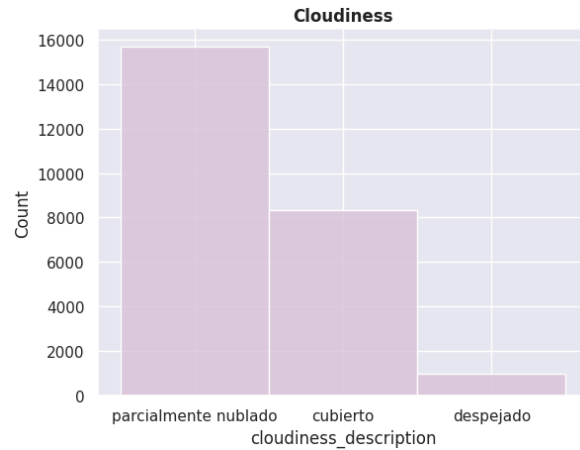


Figura 5- Cloudiness

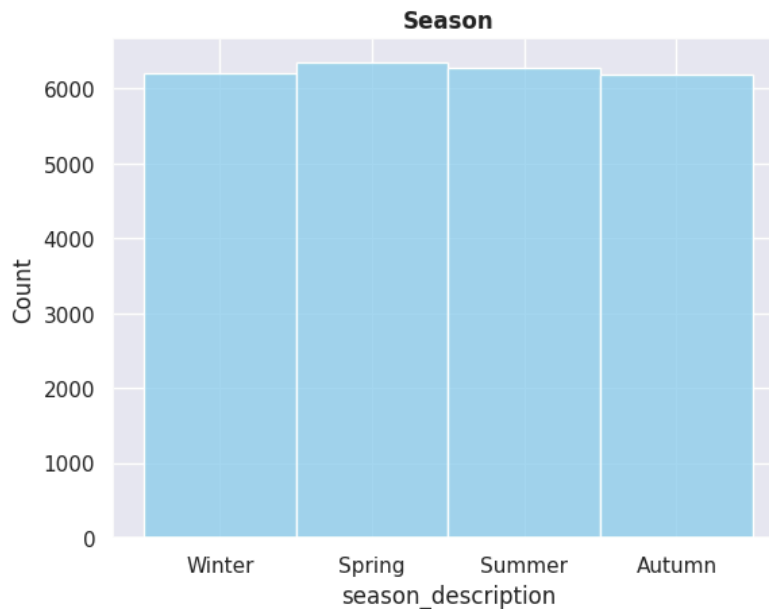


Figura 6- Seasons

En segundo lugar, revisamos las variables numéricas, observando que todas tienen una distribución simétrica o casi simétrica, excepto la temperatura mínima que tiene un sesgo positivo o a la derecha.

Los histogramas muestran que la mayoría de variables presenta distribuciones relativamente uniformes, lo que implica la recopilación de datos en diferentes condiciones estacionales o geográficas.

La temp_max se distribuye entre 0 y 35 °C sin un pico predominante, mientras que la temp_min concentra la mayor parte de observaciones en torno a 2 y -5 °C.

La precipitation y la humidity también se distribuyen de manera regular en el gráfico, sin un pico predominante.

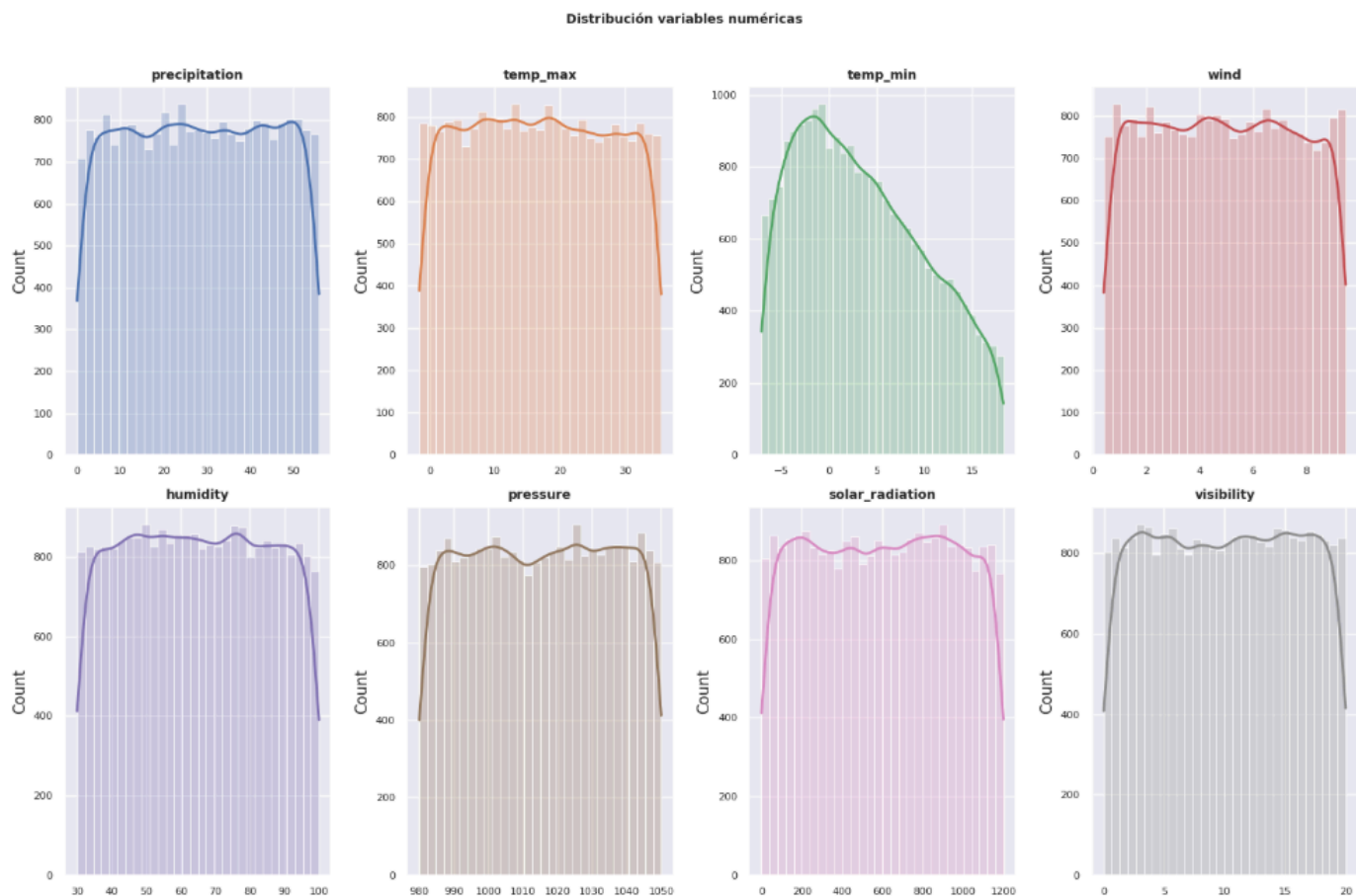


Figura 7- Gráficas numéricas

Analizando los datos vemos que en las columnas precipitation, temp_max, temp_min y wind tenemos 2500 valores nulos de 25000 datos (10%), esto nos plantea la duda de como tratar estos datos:

- **Eliminar los nulos**, aunque simplifica el proceso, **no** es la mejor opción porque tenemos una cantidad considerable de datos nulos.
- **Imputar los nulos con la media**, es una **buena** opción porque la mayoría de las distribuciones son simétricas y no debería sesgar demasiado los datos.
- **Limpiar y Normalizar los datos**, es la **mejor** opción para que todas las variables tengan un rango similar. Esto implica, imputar los nulos y luego normalizar. (Estandarización (Z-score), escalado max-min)

observation_date	0
precipitation	2500
temp_max	2500
temp_min	2500
wind	2500
humidity	0
pressure	0
solar_radiation	0
visibility	0
weather_description	0
cloudiness_description	0
season_description	0

Figura 8- Total de nulos

3. Preparación de los Datos

3.1. Seleccionar Datos

En este apartado exponemos y argumentamos la selección de los datos más relevantes e importantes para usar en nuestro modelo de predicción meteorológica. Los datos son:

- Viento
- Humedad
- Temperatura Mínima
- Temperatura Máxima

A partir del análisis de estos datos podremos clasificar e identificar las variables clima y tipo de cielo.

Hemos hecho una matriz de correlación de estos datos para que sea más visual nuestra selección.

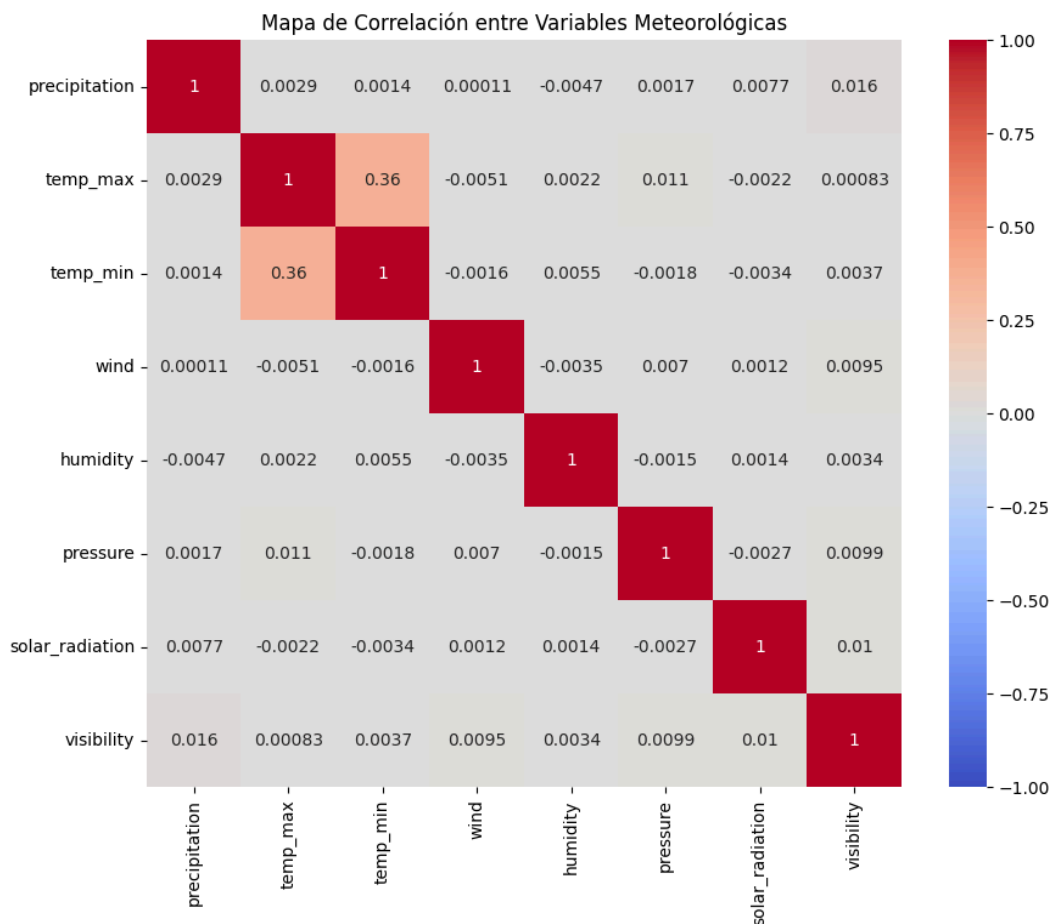


Figura 9- Correlaciones

3.2. Limpieza de los Datos

Antes de pasar a la fase de creación del modelo es necesario realizar una limpieza y/o construcción de la base de datos para elevar o garantizar la calidad de los mismos. Esto a través de eliminación de valores innecesarios, valores duplicados, conversión de tipos de datos, imputación de datos faltantes y balanceo de la data para poder entrenar de una mejor manera el modelo. Para este caso, y como se mencionó anteriormente, se evidenció que algunas variables contaban con el 10% de sus datos vacíos, por lo que se utilizaron técnicas de imputación de datos para poder completar la información faltante.

3.3. Construcción de los Datos

Como ya comentamos en puntos anteriores, nos encontramos con el 10% de los datos que son nulos en las columnas precipitation, temp_max, temp_min y wind. Calculamos la media e imputamos los datos:

```
Media calculada para temp_max: 16.892393173582324
Media calculada para temp_min: 3.580655338596054
Media calculada para wind: 4.931800815971483
Media calculada para precipitation: 28.093320743142307
```

Figura 10- Valores medias

Tras realizar este paso no quedan valores nulos en nuestros datos:

```
observation_date    0
precipitation       0
temp_max            0
temp_min            0
wind                0
humidity            0
pressure            0
solar_radiation     0
visibility           0
weather_description 0
cloudiness_description 0
season_description  0
```

Figura 11- Total nulos limpios

Así quedan los gráficos de distribución de las variables numéricas:

Proyecto 3: BGD-Predicción Meteorológica

Aakriti Guerrero Bertolín

Oriol Fernández Font

Joan Pau Grau

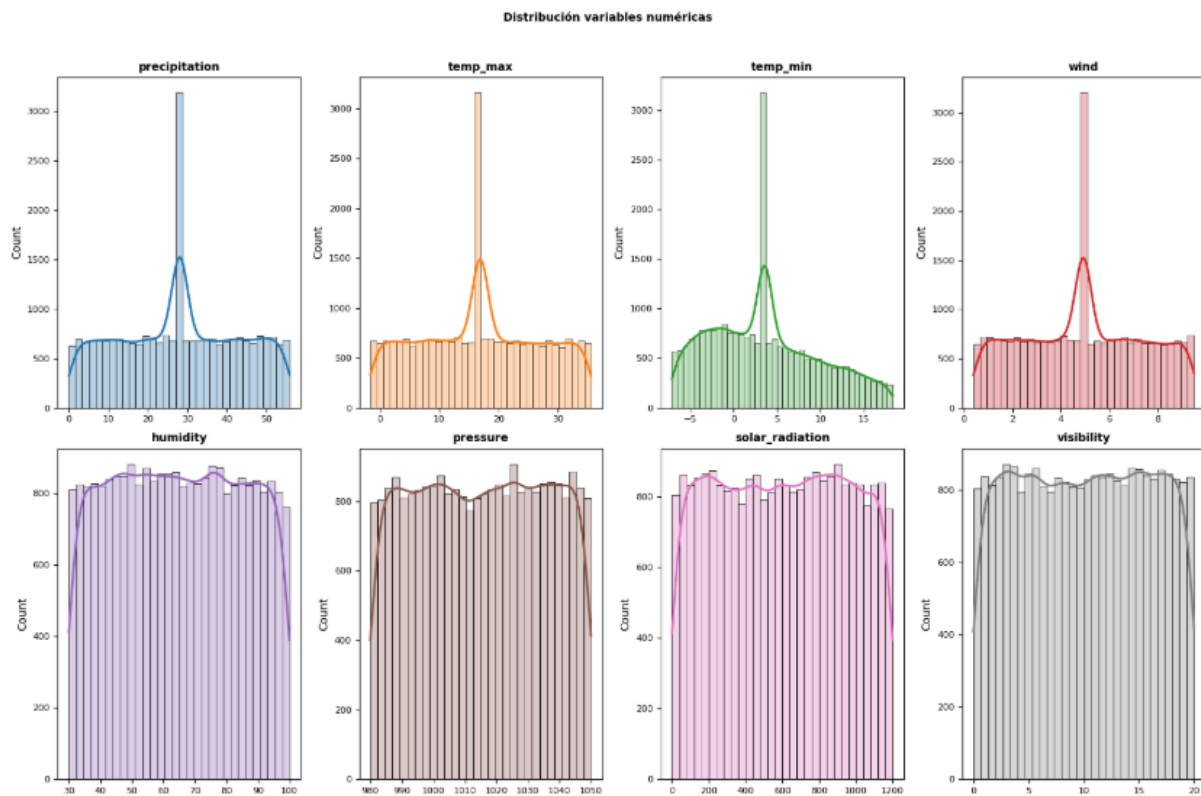


Figura 12- Gráficas numéricas medias añadidas

Una vez tenemos todos los datos sin nulos, procedemos a normalizarlos y transformarlos con tal de que no estén desbalanceados (como se puede observar en los gráficos de las columnas donde se han substituido los nulos por los nuevos datos). Para este proceso se ha realizado lo siguiente (en el orden que se detalla):

- Precipitation: Transformación logarítmica
- Wind: Transformación de raíz cuadrada
- Temp_max, Temp_min, Precipitation y Wind: Escalar min-max

	temp_max	temp_min	precipitation	wind
count	25000.000000	25000.000000	25000.000000	25000.000000
mean	0.497122	0.420459	0.778948	0.612188
std	0.272907	0.248287	0.195149	0.253493
min	0.000000	0.000000	0.000000	0.000000
25%	0.276552	0.221942	0.696589	0.437180
50%	0.497122	0.420459	0.833931	0.648349
75%	0.716804	0.588774	0.922420	0.816774
max	1.000000	1.000000	1.000000	1.000000

Figura 13- Descripción variables medias añadidas

4. Modelación

4.1. Selección de la Técnica de Modelado

El objetivo del presente es poder construir un modelo predictivo que pueda mostrar datos meteorológicos futuros a partir de los que ya tenemos recopilados. Para esta predicción, encontramos diferentes tipos de modelos como lo son Decision Trees o SVM (Support Vector Machines). Hemos decidido utilizar este último aprovechando su eficacia en memoria y su facilidad para utilizar diversas funciones del núcleo para la función de decisión. A pesar de que actualmente se utilizan para resolver problemas de tipo regresión, agrupamiento y multi clasificación, su origen fue pensado para resolver problemas de clasificación binaria, como lo es el presente.

Las máquinas de vectores de soporte realizan la clasificación encontrando un hiperplano separador óptimo de las clases. Este se obtiene por medio de un problema de optimización, en donde se seleccionan las observaciones de una muestra de entrenamiento que van a ser fundamentales a la hora de clasificar el conjunto total de observaciones. Estas observaciones son llamadas vectores de soporte.

Para darle un enfoque adecuado, el algoritmo debe entrenarse con datos anteriormente clasificados o datos de entrenamiento como lo hemos llamado anteriormente, con los cuales se busca crear reglas para la clasificación para luego clasificar los datos de prueba. Los hiperparámetros son fundamentales para este tipo de algoritmos en búsqueda de obtener resultados más precisos y es por esto que uno de los modelos a utilizar en el presente es Grid Search. Los hiperparámetros a sintonizar son el valor C y el valor Gamma. El parámetro C representa la penalización que se da a las muestras mal clasificadas, el parámetro Gamma representa la mayor distancia entre las observaciones. Si se establece un C muy elevado, el modelo tenderá a minimizar el número de errores y podría llegar a perder capacidad.

El rendimiento del modelo o técnica propuesta se evalúa a través de parámetros conocidos como viento, humedad, temperatura mínima y temperatura máxima, los cuales hacen parte de los criterios de éxito de este proyecto.

4.2. Construir y Evaluar el Modelo

Preprocesamiento de los Datos

El propósito de esta sección, es realizar una selección de variables en un conjunto de datos con el objetivo de identificar las características más relevantes para entrenar un modelo de clasificación (en este caso, un modelo SVM). Para ello, se implementan dos procesos principales:

1. **Balanceo del conjunto de datos** (mediante submuestreo y sobremuestreo).

Los datos se ajustan para abordar problemas de desbalance de clases, lo cual es común en problemas de clasificación. Este paso es crucial para evitar que el modelo sea sesgado hacia la clase mayoritaria. Se emplean las siguientes técnicas:

- **Submuestreo** (Undersampling):
 - Se utiliza **RandomUnderSampler** para reducir la cantidad de ejemplos de la clase mayoritaria, equilibrando así el conjunto de datos.
 - También se aplica el método **NearMiss**, que selecciona ejemplos basados en la proximidad de las clases en el espacio de características.
- **Sobremuestreo** (Oversampling):
 - Se utiliza **RandomOverSampler** para aumentar la cantidad de ejemplos de la clase minoritaria.
 - También se emplea SMOTE (Synthetic Minority Oversampling Technique), que genera ejemplos sintéticos a partir de los vecinos más cercanos de la clase minoritaria.

Estos métodos aseguran que las clases estén equilibradas antes de proceder con la selección de variables y el entrenamiento del modelo.

2. **Selección de características** utilizando pruebas estadísticas y análisis de varianza.

Tras balancear las clases, el código evalúa las características para identificar las más relevantes mediante pruebas estadísticas. Esto reduce la dimensionalidad del conjunto de datos y mejora la eficiencia y precisión del modelo SVM. Los métodos utilizados son:

- **Chi-cuadrado (Chi²):**
 - Se aplica el método **SelectKBest** con la prueba chi-cuadrado para evaluar la dependencia entre las características y la variable objetivo.
 - Las características con mayores puntajes son seleccionadas como las más relevantes.
- **F-Value (ANOVA):**
 - Se aplica el análisis de varianza (ANOVA) para evaluar la relación lineal entre cada característica y la variable objetivo.
 - Las características con mayor F-Value son seleccionadas como las mejores para el modelo.
- **Correlación por pares:**
 - Se aplica la correlación por pares para reducir la dimensionalidad. Al eliminar variables menos relevantes, se mejora la eficiencia del modelo.
 - Las variables con mayor correlación son candidatas ideales para ser incluidas en el modelo, ya que tienen una fuerte relación con la variable objetivo.
 - Al excluir características con baja correlación, se reduce la posibilidad de que el modelo aprenda patrones irrelevantes o redundantes.

Después los datos se han normalizado utilizando el método MinMaxScaler, con esto las variables se escalan a un rango entre 0 y 1. Este paso es crucial para algoritmos como SVM.

```
#FEATURE SCALING
scaler = MinMaxScaler()
X_train_std = scaler.fit_transform(X_train)
X_test_std = scaler.fit_transform(X_test)
```

	humidity	season_id	visibility	temp_max	solar_radiation	cloudiness_id	precipitation	wind
0	0.521172	3	0.202249	0.983715	0.204287	1	0.877906	0.923093
1	0.373589	3	0.577739	0.740827	0.928256	2	0.780796	0.863055
2	0.089378	2	0.766157	0.634852	0.426436	1	0.912617	0.810687
3	0.652486	3	0.706856	0.507192	0.517923	1	0.998198	0.944479
4	0.684179	1	0.527809	0.812584	0.401363	3	0.474521	0.160393
...
17495	0.943632	1	0.561549	0.054852	0.012841	1	0.770353	0.027713
17496	0.225910	4	0.635207	0.776229	0.810232	2	0.873815	0.878617
17497	0.886430	2	0.255769	0.595736	0.400323	1	0.987596	0.975468
17498	0.578403	2	0.261375	0.791962	0.043457	1	0.951711	0.959375
17499	0.895916	4	0.323742	0.769788	0.483868	1	0.687292	0.630175

17500 rows x 8 columns

Figura 14- Escalado de valores con MinMaxScaler

Selección del modelo

Para construir el algoritmo SVM, se ha utilizado el kernel Polinomial, el cual es adecuado para capturar relaciones más complejas que pueden aparecer en interacciones entre variables meteorológicas.

Para encontrar la mejor combinación de hiperparámetros para un modelo, se ha utilizado el GridSearchCV de la biblioteca scikit-learn, utilizando validación cruzada y optimizando el puntaje F1 ponderado.

Se ha definido un diccionario param_grid que contiene las combinaciones posibles de hiperparámetros a evaluar:

```
param_grid = {  
    'C': [0.1, 1, 10],  
    'gamma': [0.01, 0.1, 'scale'],  
    'kernel': ['rbf', 'poly', 'linear']  
}
```

Finalmente los hiperparámetros principales configurados han sido:

1. `class_weight='balanced'`: Este ajuste compensa los desequilibrios en las clases del conjunto de datos.
2. `C=10`: Un valor más alto de C prioriza minimizar los errores de clasificación, buscando un margen más ajustado.
3. `gamma='scale'`: Este valor automático ajusta el parámetro gamma en función de las características de entrada, mejorando la adaptabilidad del modelo.
4. `kernel='poly'`: El kernel polinómico permite capturar relaciones no lineales en los datos, modelando interacciones complejas entre características.

Entrenamiento del Modelo

El modelo SVM se ha entrenado utilizando el conjunto de datos de entrenamiento estandarizado: `svm_modelM3.fit(X_train_std, y_train)`.

Dónde los datos han sido divididos en un 70% para el entrenamiento (X_train_std) y un 30% (y_train)

```
# Dividir los datos en entrenamiento (70%) y prueba (30%)  
X_train, X_test, y_train, y_test = train_test_split(X, y,  
test_size=0.3, random_state=42)
```

La división final de los datos queda con un X_train_std = (17500, 8) y un y_train = (7500, 8).

5. Evaluación

5.1 Evaluación de los Resultados

El rendimiento del modelo se ha evaluado en el conjunto de prueba utilizando las siguientes métricas clave:

- **Reporte de Clasificación:** Proporciona métricas como precisión, recall, F1-score y soporte para cada clase.
- **Exactitud (Accuracy):** Se utilizó como una métrica global de evaluación para medir el porcentaje de predicciones correctas.

Después de entrenar el modelo y probarlo, estos han sido los resultados:

Reporte de Clasificación:				
	precision	recall	f1-score	support
1	0.97	0.89	0.93	2704
2	0.97	0.85	0.91	2852
3	0.74	0.88	0.80	1773
4	0.38	0.93	0.54	142
5	0.39	0.90	0.55	29
accuracy			0.87	7500
macro avg	0.69	0.89	0.75	7500
weighted avg	0.90	0.87	0.88	7500
Accuracy: 0.8741333333333333				

Figura 15- Reporte de clasificación

El modelo ha mostrado un rendimiento sólido y con una exactitud global del **87.41%**.

El modelo muestra un buen desempeño en las clases con mayor cantidad de datos (Clases 1, 2 y 3), con valores de *F1-score* superiores al **80%**. La alta exactitud global indica que el modelo clasifica correctamente la mayoría de los eventos meteorológicos.

No obstante, hay ciertos puntos dónde el modelo falla. En las clases con menos datos (Clases 4 y 5), el desempeño es inferior. Por ejemplo, la Clase 4 tiene un *F1-score* de **0.54**, lo que refleja la necesidad de un mejor manejo del desequilibrio de clases. Aunque el recall es alto para estas clases, la precisión es más baja, indicando que el modelo produce falsos positivos.

Proyecto 3: BGD-Predicción Meteorológica

Aakriti Guerrero Bertolín

Oriol Fernández Font

Joan Pau Grau

Una vez la página ha sido lanzada, podemos ver el resultado de las predicciones y basándonos en el dataset comprobamos que acierta.

Prueba 1

Dataset

Radiación solar	Viento	Visibilidad	Tem_max	Estación	Precipitación	Humedad	Tipo Cielo	Tiempo
206.294	8.27884	5.26442	4.626	Winter	40.0418	70.8672	partly cloudy	storm

Resultado

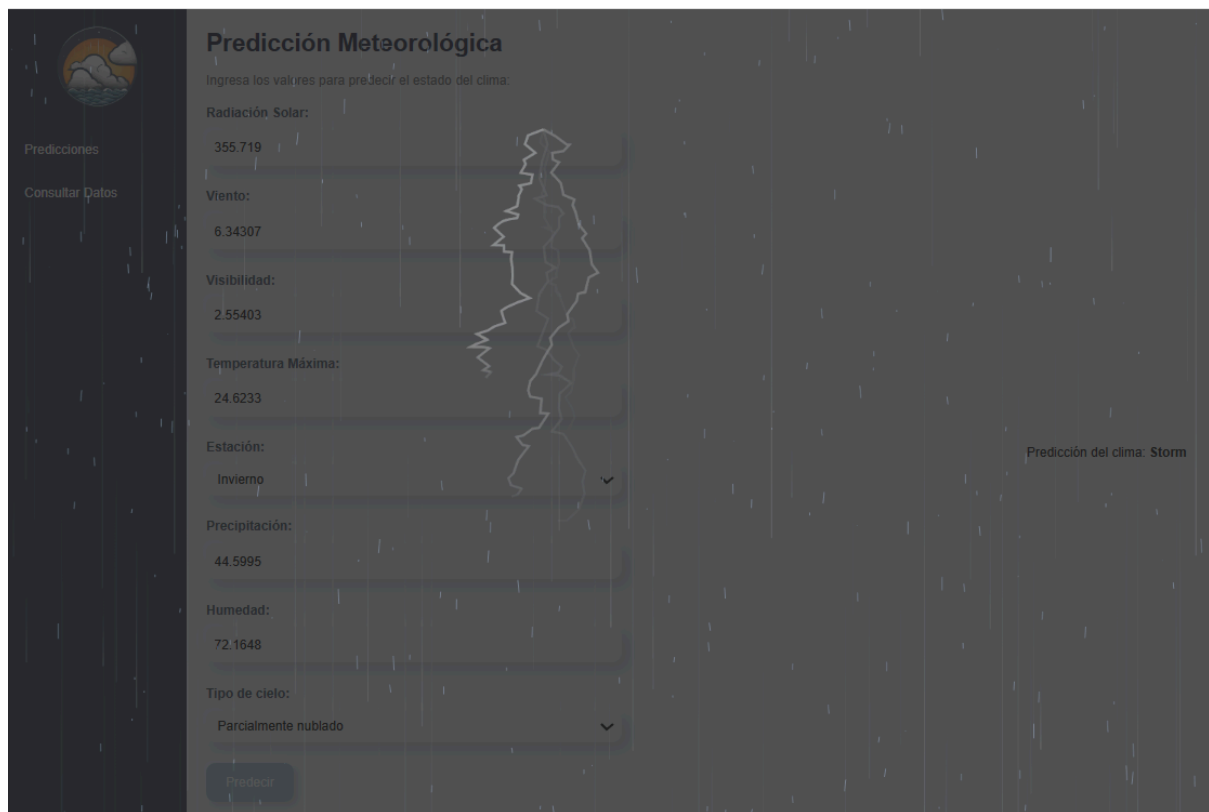


Figura 16- Predicción Storm

Proyecto 3: BGD-Predicción Meteorológica

Aakriti Guerrero Bertolín

Oriol Fernández Font

Joan Pau Grau

Prueba 2

Dataset

Radiación solar	Viento	Visibilidad	Tem_max	Estación	Precipitación	Humedad	Tipo Cielo	Tiempo
1108.25	7.27928	2.95978	34.8902	Spring	0.939341	51.9892	covered	sun

Resultado

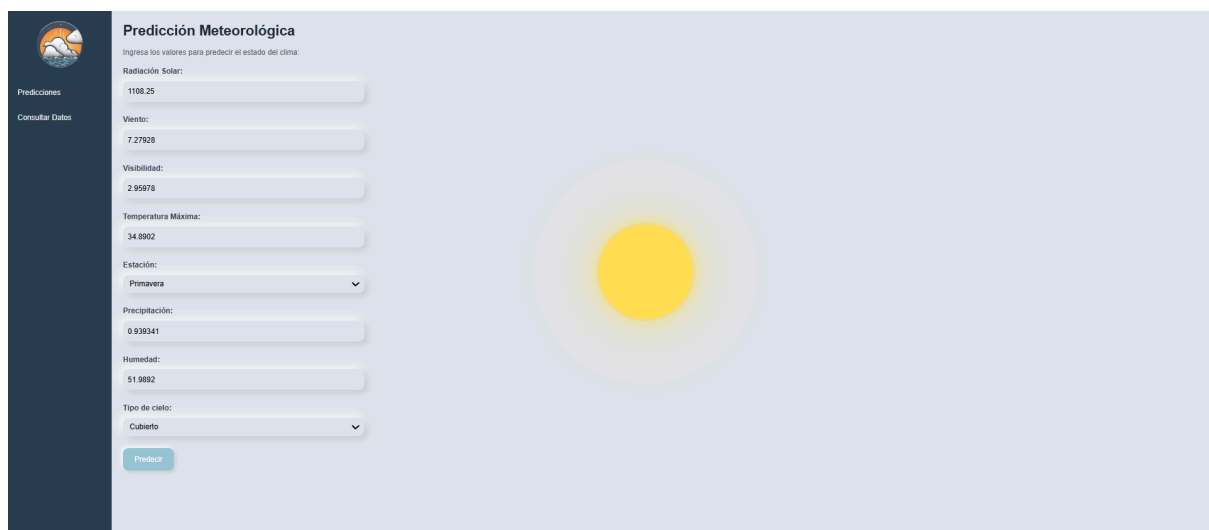


Figura 17- Predicción Sun

Prueba 3

Dataset

Radiación solar	Viento	Visibilidad	Tem_max	Estación	Precipitación	Humedad	Tipo Cielo	Tiempo
0.204287	0.923093	0.202249	0.983715	Winter	0.877906	0.521172	parcialmente nublado	rain

Proyecto 3: BGD-Predicción Meteorológica

Aakriti Guerrero Bertolín

Oriol Fernández Font

Joan Pau Grau

Resultado

Predicción Meteorológica

Ingresa los valores para predecir el estado del clima:

Radiación Solar: 0.204287

Viento: 0.923093

Visibilidad: 0.202249

Temperatura Máxima: 0.983715

Estación: Invierno ▼

Precipitación: 0.877906

Humedad: 0.521172

Tipo de cielo: Parcialmente nublado ▼

Predecir

Predicción del clima: Rain

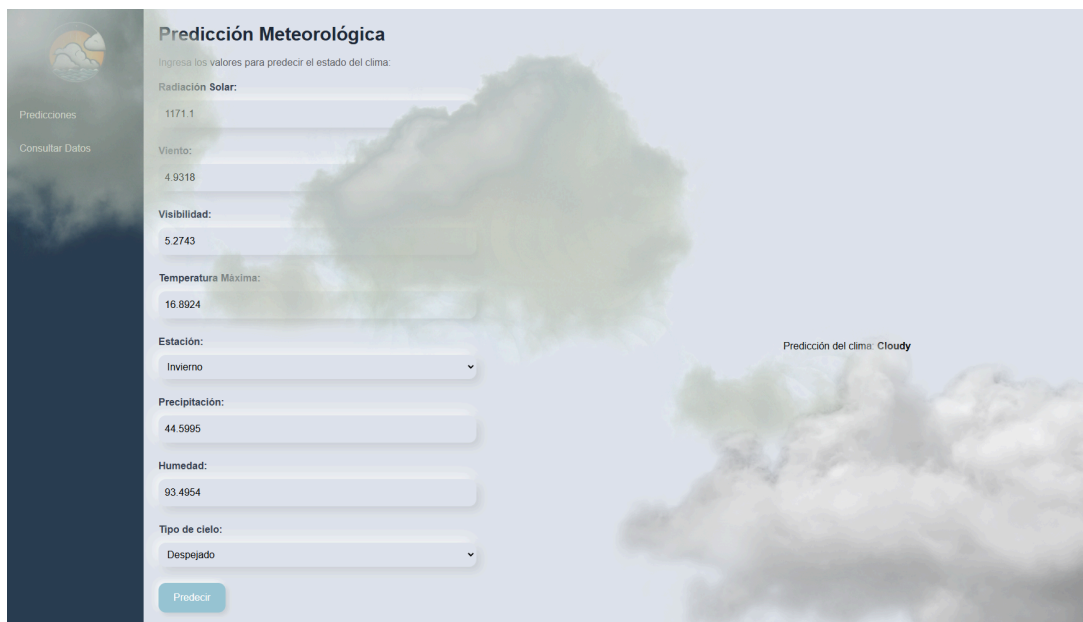
Figura 17- Predicción Rain

Prueba 4

Dataset

Radiación solar	Viento	Visibilidad	Tem_max	Estación	Precipitación	Humedad	Tipo Cielo	Tiempo
1171.1	4.9318	5.2743	16.8924	Winter	28.0933	93.4954	clear	cloudy

Resultado



Predicción Meteorológica

Ingresar los valores para predecir el estado del clima:

Radiación Solar:
1171.1

Viento:
4.9318

Visibilidad:
5.2743

Temperatura Máxima:
16.8924

Estación:
Invierno

Precipitación:
44.5995

Humedad:
93.4954

Tipo de cielo:
Despejado

Predecir

Predicción del clima: Cloudy

Figura 18- Predicción Cloudy

5.2 Propuestas de mejora

Para mejorar los puntos en los que falla el modelo se pueden considerar técnicas como explorar ajustes adicionales en los hiperparámetros, especialmente en C y γ , mediante una búsqueda más exhaustiva con *GridSearchCV* o *RandomizedSearchCV*. O bien, incrementar la cantidad de datos para las clases minoritarias (Clases 4 y 5) para mejorar la robustez del modelo.

6. Despliegue

Una vez el modelo ha sido entrenado, este se ha implementado en la interfaz del usuario. La interfaz se ha creado con Django y al ejecutarla despliega una página web dónde el usuario podrá interactuar y ver los resultados. Esta sección está explicada de manera más amplia en el Manual de usuario.

Figura 19- Vista Predicción Meteorológica

observation_id	observation_date	precipitation	temp_max	temp_min	wind	humidity	pressure	solar_radiation	visibility	weather_name	cloudiness_name	season_name
1	2016-01-01	44.0995	24.0233	14.1948	6.34307	72.1648	1047.16	355.719	2.55403	storm	partly cloudy	Winter
2	2016-01-02	42.805	5.42469	4.73406	1.36506	61.8476	1015.71	320.129	16.96	rain	partly cloudy	Winter
3	2016-01-03	40.0418	4.626	2.75272	8.27884	70.8672	1032.35	206.294	5.26442	storm	partly cloudy	Winter
4	2016-01-04	11.7333	33.9087	18.5858	4.36902	82.6927	1018.33	340.83	1.65044	rain	partly cloudy	Winter
5	2016-01-05	28.0933	16.8924	3.58066	4.9318	93.4954	1040.78	1171.1	5.2743	cloudy	clear	Winter
6	2016-01-06	16.1717	4.65227	0.902042	7.37924	92.7704	1005.37	968.706	17.4834	storm	partly cloudy	Winter
7	2016-01-07	4.22043	5.96706	-6.15683	6.00404	34.2203	1032.6	72.9699	1.51714	cloudy	partly cloudy	Winter
8	2016-01-08	14.1172	26.3022	-1.95847	3.74957	50.7973	984.014	444.794	0.411192	rain	partly cloudy	Winter
9	2016-01-09	32.8687	19.3676	11.542	0.972101	54.5463	1041.29	853.303	5.21402	rain	clear	Winter
10	2016-01-10	11.9768	22.158	-4.77025	1.03535	92.1622	969.783	624.903	8.55745	rain	covered	Winter
11	2016-01-11	15.7791	30.4892	8.01089	4.76448	38.2312	998.442	1053.18	17.0491	rain	clear	Winter
12	2016-01-12	20.9391	21.9665	6.48077	4.36245	33.458	991.469	1067.99	2.84958	rain	clear	Winter
13	2016-01-13	28.0933	16.8924	3.58066	4.9318	78.6112	1048.78	1184.54	18.3769	cloudy	clear	Winter
14	2016-01-14	28.0933	16.8924	3.58066	4.9318	32.111	1011.95	860.531	19.3961	cloudy	clear	Winter
15	2016-01-15	28.0933	16.8924	3.58066	4.9318	78.5264	1020.90	852.963	5.4192	cloudy	clear	Winter
16	2016-01-16	33.7006	2.30618	-6.40889	6.70937	51.167	986.45	963.007	8.51818	storm	clear	Winter
17	2016-01-17	28.0933	16.8924	3.58066	4.9318	46.0605	983.897	680.65	7.95497	cloudy	partly cloudy	Winter
18	2016-01-18	43.0534	-1.42652	-5.36943	6.23668	94.8835	1033.66	678.028	15.2183	storm	partly cloudy	Winter
19	2016-01-19	25.5804	14.0088	-3.01885	8.1232	57.7319	1049.82	42.3936	17.8945	storm	partly cloudy	Winter
20	2016-01-20	15.0571	22.6369	-5.52859	0.53404	79.7123	1040.77	43.3864	1.0867	rain	partly cloudy	Winter

Figura 20- Vista Consulta de Observaciones

7. GITHUB

<https://github.com/Aakriti313/SistemaMeteorologico>