**Presentación del Proyecto de Aprendizaje Automático**

**Alumno: Mariano Buet**

**Proyecto: Predicción de alertas térmicas mediante aprendizaje automático a partir de datos climatológicos nacionales**

**Introduccion**

El presente proyecto propone el desarrollo de un modelo de aprendizaje automático basado en el conjunto de datos “Estadísticas Climatológicas (1981–2010)” del Servicio Meteorológico Nacional (SMN). Dicho dataset contiene información mensual de distintas estaciones meteorológicas del país, con variables como temperatura media, temperatura máxima, temperatura mínima, humedad relativa, velocidad del viento, nubosidad total y precipitación.  
A partir de estas variables y considerando solo 3 o 4 de ellas como entradas principales el modelo buscará predecir tres estados climáticos de alerta térmica: riesgo de helada, clima de confort o riesgo de sobrecalor. El objetivo práctico es anticipar condiciones extremas para permitir la toma de decisiones preventivas, como proteger cultivos ante heladas, evitar el congelamiento de cañerías en viviendas o implementar medidas frente a olas de calor, como el riego adicional o la protección de personas y animales. De esta forma, el sistema se plantea como una herramienta predictiva aplicable a distintos contextos agrícolas y domésticos dentro del territorio argentino.

**Objetivo General**

Desarrollar un modelo de aprendizaje automático de clasificación capaz de predecir el estado térmico esperado (riesgo de helada, clima de confort o sobrecalor) a partir de variables meteorológicas seleccionadas, utilizando datos históricos del Servicio Meteorológico Nacional (SMN).

**Objetivos Específicos**

* Procesar y estructurar el conjunto de datos “Estadísticas Climatológicas (1981–2010)” del SMN para su uso en el modelo predictivo.
* Seleccionar y analizar variables relevantes como temperatura mínima, temperatura máxima, temperatura media, humedad relativa, nubosidad total y velocidad del viento.
* Entrenar un modelo de clasificación que permita anticipar tres categorías de alerta térmica:
  + Riesgo de helada (temperaturas mínimas inferiores a 3 °C)
  + Clima de confort (temperaturas entre 5 °C y 30 °C)
  + Riesgo de sobrecalor (temperaturas máximas superiores a 30–35 °C)
* Evaluar el desempeño del modelo mediante métricas de precisión, recall y matriz de confusión.
* Aplicar los resultados al desarrollo de un sistema de alerta temprana, útil para la toma de decisiones preventivas en el ámbito agrícola, doméstico e industrial.

**Contexto y relevancia**

El proyecto se enmarca en el contexto del monitoreo climático y la gestión de riesgos meteorológicos en Argentina, especialmente en regiones australes como Tierra del Fuego, o las más calurosas como el centro norte del país, donde las condiciones térmicas extremas afectan tanto la producción agrícola como la infraestructura.  
Anticipar eventos de heladas o sobrecalor permite reducir pérdidas económicas y daños materiales, por ejemplo: proteger cultivos de bajas temperaturas, evitar el congelamiento de cañerías en viviendas o activar sistemas de riego y ventilación ante temperaturas elevadas.  
La disponibilidad de un dataset climatológico nacional, con registros históricos de temperatura, humedad, viento, nubosidad y precipitación, brinda la base ideal para entrenar un modelo predictivo con aplicación práctica y escalable.

**Tipo de problema**

Se trata de un problema de clasificación multiclase, donde el modelo debe asignar cada registro meteorológico a una de las tres categorías de alerta térmica (helada, confort o sobrecalor) en función de las variables de entrada.

**Modelos que se podrían utilizar**

Para abordar este problema se propone utilizar distintos algoritmos de clasificación supervisada, con el fin de comparar su rendimiento:

*Vistos en clases:*

Árbol de Decisión (Decision Tree Classifier) → por su fácil interpretación.

K-Nearest Neighbors (KNN) → como alternativa simple y comparativa.

Máquinas de Vectores de Soporte (SVM) → para probar límites no lineales entre clases.

*Modelo aún no conocido pero recomendado para estos caso*s

Bosques Aleatorios (Random Forest Classifier) → para mejorar la precisión y evitar sobreajuste.

El modelo final se seleccionará según su desempeño en métricas de clasificación y su capacidad de generalizar los patrones climáticos observados.