**A Weather Prediction Model With Big Data**

En este proyecto, me ocupo de la predicción del clima utilizando herramientas de Aprendizaje Automático y Big Data, con el propósito de demostrar si es factible realizar pronósticos valiosos de las condiciones meteorológicas basándose exclusivamente en datos meteorológicos previamente registrados. El objetivo de la clasificación es, por lo tanto, predecir qué condición meteorológica debería ocurrir dada una serie de mediciones del clima.

Para realizar este proyecto me baso en 3 papers:

* **Big Data Analytics in Weather Forecasting: A Systematic Review**

Marzieh Fathi · Mostafa Haghi Kashani · Seyed Mahdi Jameii · Ebrahim Mahdipour

* **A Review of Weather Data Analytics using Big Data**

Priyanka Chouksey · Abhishek Singh Chauhan

* **A Weather Prediction Model with Big Data**

Ning Yang · Lewis Westfall · Ms. Preeti Dalvi

Descripción del proyecto:

El proyecto se titula "A Weather Prediction Model With Big Data" y se basa en la predicción del clima utilizando técnicas de Big Data. Se utilizarán datos históricos del clima y algoritmos de aprendizaje automático (machine learning) para realizar las predicciones.

Importación de librerías y configuración del entorno:

En las primeras celdas, se importan las librerías necesarias, como Pandas, NumPy, Matplotlib y Scikit-learn. También se establecen algunas configuraciones del entorno, como el estilo de los gráficos y la opción de mostrar todas las columnas de un DataFrame.

Carga de datos:

En esta sección, se carga el conjunto de datos históricos del clima desde un archivo CSV utilizando Pandas. El archivo se denomina "weather.csv" y contiene información sobre la temperatura, humedad, presión y otras variables relacionadas con el clima.

Análisis exploratorio de datos (EDA):

Se realiza un análisis exploratorio de datos para comprender la estructura y las características del conjunto de datos. Se muestran algunas estadísticas descriptivas, como la media, la desviación estándar y los valores mínimos y máximos de las variables.

Preprocesamiento de datos:

En esta sección, se lleva a cabo el preprocesamiento de datos antes de entrenar los modelos de predicción. Se eliminan las columnas innecesarias, se convierten las fechas a un formato adecuado y se realiza una imputación de valores faltantes utilizando la media de cada columna.

División del conjunto de datos:

El conjunto de datos se divide en conjuntos de entrenamiento y prueba utilizando la función train\_test\_split de Scikit-learn. Se utiliza una proporción del 80% para el conjunto de entrenamiento y del 20% para el conjunto de prueba.

Modelado y entrenamiento:

Se aplican algoritmos de regresión para predecir la temperatura. En el notebook, se utilizan tres modelos: Regresión Lineal, Support Vector Regression (SVR) y Random Forest Regression. Cada modelo se entrena utilizando el conjunto de entrenamiento y se evalúa utilizando el conjunto de prueba.

Evaluación de los modelos:

Se evalúa el rendimiento de los modelos utilizando métricas como el error medio absoluto (MAE), el error cuadrático medio (MSE) y la raíz del error cuadrático medio (RMSE). Se comparan los resultados de los tres modelos para determinar cuál tiene un mejor rendimiento.

Visualización de resultados:

Se muestran gráficos de dispersión para comparar las predicciones de temperatura realizadas por los modelos con los valores reales del conjunto de prueba. También se muestra un gráfico de barras que representa la importancia de las características utilizadas por el modelo de Random Forest Regression.

En general, este proyecto presenta un enfoque básico pero sólido para la predicción del clima utilizando técnicas de Big Data. Se abarcan aspectos clave, como el análisis exploratorio de datos, el preprocesamiento de datos, el model