

|  |
| --- |
| INFORME TÉCNICO |
|  |
| 29 enero 2025  Análisis Meteorológico de Australia.  Creado por: Claudio Sanchez Torres.  Javier Astorga. |



|  |
| --- |
| Indice 1. Introducción  1.1. Descripción del contexto  1.2. Propósito y justificación del proyecto  2. Carga y Exploración Inicial de Datos  2.1. Estructura del Dataset  2.2. Análisis Inicial  2.3. Estadísticas descriptivas  3. Preprocesamiento de Datos  3.1. Herramientas utilizadas  3.2. Procesamiento de datos  3.3. Modelado predictivo  4. Razonamiento y Justificación  4.1. Evaluación de resultados  4.2. Comparación con métodos alternativos  5. Planificación del Proyecto  5.1. Fases y entregables  5.2. Consideraciones futuras  6. Recomendaciones Estratégicas  7. Conclusión  8. Glosario de Términos |
| ***1. Introducción***  ***1.1. Descripción del Contexto***  *Australia, siendo uno de los países más extensos del mundo y el continente más seco del planeta, presenta un caso de estudio fascinante para el análisis meteorológico. Esta investigación se centra en el análisis de observaciones meteorológicas diarias recopiladas de múltiples ubicaciones a lo largo del territorio australiano, utilizando datos proporcionados por la Oficina de Meteorología de la Commonwealth de Australia.*  *El estudio examina un conjunto de datos completo que incluye dos variables fundamentales: 'RainTomorrow', un predictor binario de la ocurrencia de lluvia para el día siguiente (Sí/No), y 'RISK\_MM', que registra la cantidad de lluvia en milímetros. Este análisis cobra especial relevancia considerando la diversidad climática de Australia, que abarca desde zonas tropicales en el norte hasta regiones templadas en el sureste, junto con sus características geográficas únicas, incluyendo la Gran Cordillera Divisoria y las extensas regiones desérticas que cubren el 40% de su territorio.* |

1.2. Propósito y Justificación del Proyecto

El propósito principal de este proyecto es desarrollar un análisis predictivo de las precipitaciones en Australia utilizando técnicas de minería de datos, con el objetivo de establecer patrones y modelos que permitan anticipar con precisión la ocurrencia de lluvia para el día siguiente.

La justificación del proyecto se fundamenta en varios aspectos clave:

Impacto Económico: La predicción precisa de lluvias es vital para sectores como la agricultura, minería y turismo, que representan pilares fundamentales de la economía australiana.

Gestión de Recursos: El análisis permitirá una mejor administración de los recursos hídricos en un continente donde el agua es un bien escaso y valioso.

Planificación Urbana: Con el 90% de la población concentrada en zonas urbanas, la predicción de lluvias contribuye a la planificación eficiente de infraestructuras y servicios públicos.

Prevención de Riesgos: El modelo ayudará a anticipar eventos climáticos extremos, permitiendo implementar medidas preventivas para proteger a la población y los recursos.

Valor Científico: El estudio aporta conocimiento valioso sobre patrones climáticos en una de las regiones más diversas climatológicamente del planeta.

Este proyecto representa una oportunidad significativa para aplicar técnicas avanzadas de minería de datos en la resolución de un problema real con impacto directo en la sociedad y economía australiana.

2 Carga y Exploración Inicial de Datos.

2.1 Estructura del Dataset

* El conjunto de datos incluye las siguientes variables principales:
* Fecha: Fecha de observación
* Ubicación: Estación meteorológica
* Variables climáticas:
  + Temperatura (mínima y máxima)
  + Precipitación
  + Viento (dirección y velocidad)
  + Humedad
  + Presión atmosférica
* Variable objetivo: Lluvia al día siguiente (RainTomorrow)

2.2 Análisis Inicial

Examinamos:

Dimensiones del dataset

* + Número de registros
  + Número de variables

Tipos de datos

* + Numéricos
  + Categóricos
  + Temporales

Calidad de datos

* + Valores faltantes
  + Duplicados
  + Consistencia
  1. Estadísticas descriptivas
  + Medidas de tendencia central
  + Dispersión
  + Distribución

Muestra:

Tabla, Calendario

Descripción generada automáticamente

3. Preprocesamiento de Datos

3.1 Limpieza de Datos

Conversión de Tipos de Datos

* Transformación de fechas a formato datetime
* Extracción de características temporales:
  + Mes
  + Día de la semana
  + Estación
  + Año
    - Manejo de Valores Faltantes
* Variables numéricas: imputación con media
* Variables categóricas: categoría 'Desconocido'

3.2 Transformación de Variables

Variables Numéricas

* Identificación de columnas numéricas
* Tratamiento de valores atípicos
* Normalización de datos

Variables Categóricas

* Location (ubicación)
  + - Características Temporales
* Extracción de patrones estacionales
* Análisis de tendencias temporales

Muestra:

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

4. Análisis Exploratorio de Datos (EDA)

4.1 Análisis Univariado

Distribución de Variables Clave

* Precipitaciones
  + Distribución por predicción del día siguiente
  + Promedio mensual de lluvia
  + Variación por día de la semana
* Temperatura
  + Temperatura máxima por estación
  + Rangos de temperatura
  + Evolución temporal

4.2 Análisis Bivariado

Relaciones entre Variables

* Temperatura y Lluvia
  + Correlación entre temperatura mínima y máxima
  + Influencia en predicción de lluvia
  + Patrones de temperatura por estación

4.3 Patrones Temporales

Análisis Estacional

* Variaciones Mensuales
  + Precipitación promedio mensual
  + Temperatura máxima por estación

Tendencias a Largo Plazo

* Evolución Anual
  + Tendencias de temperatura mínima
  + Tendencias de temperatura máxima
  + Cambios en patrones de lluvia

Muestra:

|  |  |
| --- | --- |
|  |  |

5.1 Selección de Características

* Variables Meteorológicas
  + Temperaturas (MinTemp, MaxTemp, Temp9am, Temp3pm)
  + Precipitación (Rainfall)
  + Viento (WindSpeed9am, WindSpeed3pm)
  + Humedad (Humidity9am, Humidity3pm)
  + Presión (Pressure9am, Pressure3pm)

5.2 Preparación de Datos

* Variable Objetivo
  + RainTomorrow (convertida a binaria: 0/1)

5.3 División del Dataset

* Conjunto de Entrenamiento (80%)
  + Para entrenar el modelo
* Conjunto de Prueba (20%)
  + Para evaluar el rendimiento
* Random State: 42
  + Garantiza reproducibilidad

Muestra:

Texto

Descripción generada automáticamente

6. Modelado Predictivo

Análisis Estadístico de Datos Meteorológicos

A. Análisis por Estación

* Cálculo de promedios estacionales para:
  + Precipitaciones
  + Temperaturas máximas
  + Temperaturas mínimas
* Identificación de patrones estacionales

B. Análisis de Precipitaciones Mensuales

* Conteo de días lluviosos por mes
* Distribución temporal de precipitaciones
* Identificación de meses con mayor frecuencia de lluvia

C. Análisis de Temperaturas Anuales

Métricas Calculadas:

* Temperatura máxima:
  + Valor máximo absoluto
  + Promedio anual
* Temperatura mínima:
  + Valor mínimo absoluto
  + Promedio anual

D. Análisis de Correlaciones

Variables Analizadas:

* Precipitación (Rainfall)
* Temperatura Máxima (MaxTemp)
* Temperatura Mínima (MinTemp)

Visualización:

* Mapa de calor (heatmap)
* Escala de colores:
  + Rojo: correlación positiva
  + Azul: correlación negativa
* Valores numéricos con precisión de 2 decimales

Interpretación:

* Correlaciones significativas (>0.3)
* Relaciones entre variables meteorológicas
* Patrones identificados en los datos

Metodología

* Agrupación por períodos temporales
* Cálculo de estadísticas descriptivas
* Análisis de correlaciones bivariadas
* Visualización mediante seaborn y matplotlib

Muestra:

Diagrama

Descripción generada automáticamente

7. Análisis de Patrones Climáticos usando K-means

Este análisis utiliza el algoritmo K-means para identificar patrones naturales en los datos meteorológicos de Australia, agrupando condiciones climáticas similares en clusters.

Variables Utilizadas

* MinTemp: Temperatura mínima
* MaxTemp: Temperatura máxima
* Rainfall: Precipitaciones
* Humidity9am: Humedad a las 9am
* Humidity3pm: Humedad a las 3pm

Proceso de Análisis

A. Preparación de Datos

* Selección de características relevantes para el clustering
* Normalización de datos usando StandardScaler para equilibrar las escalas

B. Aplicación de K-means

* Número de clusters: 4 (representando patrones climáticos principales)
* Random state: 42 (para reproducibilidad)
* Asignación de clusters a cada observación

C. Visualización

* Gráfico de dispersión 2D
* Ejes: Temperatura Mínima vs Temperatura Máxima
* Colores: Representan diferentes clusters climáticos
* Paleta: 'viridis' para mejor distinción visual

D. Interpretación de Resultados

Los clusters resultantes pueden interpretarse como:

1. Cluster 0: Condiciones frías y secas
2. Cluster 1: Condiciones templadas
3. Cluster 2: Condiciones cálidas y húmedas
4. Cluster 3: Condiciones extremas

Aplicaciones Prácticas

* Planificación de recursos según patrones climáticos
* Identificación de condiciones climáticas típicas
* Apoyo a decisiones agrícolas y de gestión ambiental

Muestra:

Imagen que contiene Texto

Descripción generada automáticamente

8. Modelo Predictivo de Lluvia usando Random Forest

Este análisis implementa un modelo de Random Forest para predecir la probabilidad de lluvia del día siguiente, utilizando variables meteorológicas clave.

Variables Predictoras

* Temperaturas (MinTemp, MaxTemp, Temp9am, Temp3pm)
* Viento (WindSpeed9am, WindSpeed3pm)
* Humedad (Humidity9am, Humidity3pm)
* Presión atmosférica (Pressure9am, Pressure3pm)
* Precipitación (Rainfall)
* Metodología

A. Preparación de Datos

* División del dataset: 80% entrenamiento, 20% prueba
* Variable objetivo: RainTomorrow (binaria 0/1)
* Random state: 42 para reproducibilidad

B. Entrenamiento del Modelo

* Algoritmo: Random Forest Classifier
* Hiperparámetros por defecto
* Entrenamiento sobre datos de entrenamiento

C. Evaluación del Modelo

* Métricas Principales
* Matriz de Confusión: Visualización de predicciones correctas vs incorrectas

Reporte de Clasificación:

* Precisión
* Recall
* F1-Score
* Support

Importancia de Variables

* Ranking de variables más influyentes
* Visualización mediante gráfico de barras
* Interpretación de impacto relativo

Aplicaciones Prácticas

* Predicción de lluvia para planificación
* Identificación de variables climáticas clave
* Apoyo a decisiones agrícolas y urbanas

Muestra:

Gráfico

Descripción generada automáticamente

9. RECOMENDACIONES ESTRATÉGICAS:

A. Gestión de Recursos Hídricos:

* + La estación 1 presenta mayor precipitación con 2.87mm
  + Recomendación: Implementar sistemas de recolección de agua en esta temporada

B. Planificación Urbana:

* + Ubicación con mayor variación térmica: Woomera
  + Rango de temperatura: 0.7°C a 48.1°C
  + Recomendación: Implementar sistemas de climatización adaptativa

C. Energía Renovable:

* Top 3 ubicaciones para energía eólica:
  + GoldCoast: 553 días con vientos óptimos
  + SydneyAirport: 541 días con vientos óptimos
  + Wollongong: 417 días con vientos óptimos
* Total días con condiciones ideales: 6979

Gráfico, Gráfico de líneas

Descripción generada automáticamente

10. Conclusiones y Recomendaciones

Conclusiones Principales

A. Patrones Climáticos

* Se identificaron 4 clusters climáticos distintos mediante análisis no supervisado
* Las temperaturas muestran patrones estacionales consistentes
* La precipitación varía significativamente por ubicación y estación

B. Predicción de Lluvia

* El modelo Random Forest alcanza una precisión significativa
* Las variables más predictivas son:
  + Humedad (9am y 3pm)
  + Presión atmosférica
  + Temperatura máxima
  + Velocidad del viento

C. Tendencias Geográficas

* Las zonas costeras muestran mayor variabilidad climática
* Se identificaron patrones de viento consistentes por región
* Las temperaturas extremas se concentran en ubicaciones específicas

Recomendaciones Estratégicas

A. Gestión de Recursos Hídricos

* Implementar sistemas de recolección de agua en estaciones lluviosas
* Desarrollar infraestructura de almacenamiento en zonas con alta variabilidad
* Establecer protocolos de gestión según predicciones del modelo

B. Planificación Urbana

* Adaptar diseños arquitectónicos según clusters climáticos
* Implementar sistemas de climatización eficientes
* Desarrollar espacios verdes estratégicos

C. Energía Renovable

* Instalar parques eólicos en ubicaciones con vientos óptimos
* Aprovechar la alta radiación solar en zonas áridas
* Diversificar fuentes de energía según patrones climáticos

D. Agricultura

* Planificar cultivos según clusters climáticos
* Implementar sistemas de riego inteligente
* Utilizar predicciones de lluvia para optimizar recursos

Próximos Pasos

* Integrar datos en tiempo real
* Desarrollar sistemas de alerta temprana
* Expandir el análisis a más variables climáticas
* Implementar monitoreo continuo de patrones