



DEPARTAMENTO DE INGENIERÍA INDUSTRIAL
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE
CD2201-7 MÓDULO INTERDISCIPLINARIO

INFORME FINAL

MACHINE LEARNING

Integrantes:	Pedro Escobar Sebastián González Constanza Leiva Leonardo Rikhardsson
Profesor:	Carlos Flores
Profesor Auxiliar:	Francisco Santibáñez

Fecha de entrega: 16 de Diciembre de 2022
Santiago, Chile

Índice de Contenidos

1. Contexto	1
1.1. Problema	1
1.2. Análisis de Datos	1
2. Matriz de características	3
3. Modelos	4
4. Resultados	5
4.1. Validación cruzada	5
4.2. Logistic Regression	5
4.3. LDA	5
4.4. AdaBoost	5
4.5. Gradient Boosting	6
4.6. Modelo Ensamble	6
5. Conclusiones	7

1. Contexto

El proyecto a desarrollar se enfoca en la predicción del clima en Australia, el cual, en rasgos generales, corresponde a cálido y templado, sin embargo, debido a su gran tamaño, existen zonas con sus propias características climáticas. La predicción del clima es fundamental al momento de tomar decisiones en ámbitos económicos, tales como el turismo, mercado, viajes, agricultura, industria ganadera, y en ámbitos sociales, como el transporte, actividades, gestión hídrica, etc.

1.1. Problema

Como se mencionó anteriormente, el sector agrícola es uno de los que se ven afectados por el cambio del tiempo a lo largo del año. Por ende, es importante la predicción del clima en zonas agrícolas como lo es Wagga Wagga y el estudio de las variables Humedad, Precipitaciones, Presión, Evaporación, etc.

Con todo lo anterior se puede definir el problema a modo de predecir si lloverá el día de mañana o no en Wagga Wagga, Australia, con el fin de ayudar a la industria agricultora de la zona.

1.2. Análisis de Datos

Las variables a trabajar en el análisis de datos y en nuestra MC de características son: Fechas, Wagga Wagga (zona), Temperatura máxima, Temperatura Mínima, Evaporación, Caída de agua (mm), Humedad, Presión y Nubosidad.

Las variables más influyentes en la predicción de lluvia son las fechas, ubicación, temperatura máxima y mínima, y cantidad de lluvia caída.

Los datos recopilados del sector de Wagga Wagga son los siguientes:

1. Mediana de temperatura máxima en días sin lluvias se encuentra en el intervalo de 20°C a 25°C.
2. Mediana de temperatura máxima en días con lluvia se encuentra en el intervalo de 15°C a 20°C.
3. Mediana de temperatura mínima en días sin lluvia se encuentra en el intervalo de 10°C a 15°C.
4. Mediana de temperatura mínima en días con lluvia se encuentra en el intervalo de 10°C a 15°C.
5. Existen casos extremos donde en días sin lluvia la temperatura máxima se eleva hasta los 40°C, y en días con lluvia hasta los 35°C
6. Se difiere que una temperatura bajo los 20°C conlleva una alta probabilidad de precipitaciones en la zona.

7. La mediana de nubosidad en días de lluvia corresponde al menos al 75 por ciento de los datos que presenta la nubosidad cuando no llueve, por lo tanto, una nubosidad mayor a 6 es mucho más probable que hayan precipitaciones en el día.

Para terminar esta sección, se debe aclarar que el enfoque principal del proyecto lo tiene la zona de Wagga Wagga, ubicado al sur de Australia. Esto debido a que presenta grandes cantidades de industrias agricultoras las cuales para crecer necesitan las predicciones de lluvia en la región.

2. Matriz de características

A continuación se adjunta MC de características con 5 filas con nuestras columnas finales y objetivos.

	Fecha	Ubicacion	Temp. Min	Temp. Max	Caída de agua (mm)	Evaporacion	Humedad(9am)	Humedad(3pm)	Presion(9am)	Presion(3pm)	Nubosidad(9am)	Nubosidad(3pm)	Lluvia hoy	Lluvia mañana
36529	2009-01-01	WaggaWagga	9.9	28.4	0.0	9.8	37.0	24.0	1005.2	1003.5	1.0	1.0	0	0
36530	2009-01-02	WaggaWagga	8.9	25.3	0.0	14.8	37.0	10.0	1013.9	1012.9	1.0	0.0	0	0
36531	2009-01-03	WaggaWagga	11.0	30.7	0.0	9.8	43.0	18.0	1017.8	1014.3	0.0	1.0	0	0
36532	2009-01-04	WaggaWagga	14.7	35.2	0.0	9.4	46.0	19.0	1014.8	1010.6	1.0	4.0	0	0
36533	2009-01-05	WaggaWagga	11.8	36.5	0.0	11.6	23.0	5.0	1013.1	1009.8	1.0	1.0	0	0

En el siguiente link puede ver todos los datos y análisis realizados para la matriz. [Datos MC](#)

Destacar que nuestras variables finales son las siguientes:

- Fecha.
- Ubicación.
- Temp. Mín.
- Temp. Máx.
- Caída de agua (mm).
- Evaporación.
- Humedad (9am), (3pm).
- Presión (9am), (3pm).
- Nubosidad (9am), (3pm).
- Lluvia hoy.
- Lluvia mañana.

Por otro lado, se catalogan las variables entre variables objetivos y variables explicativas. Las variables objetivos son "Lluvia hoy" y "Lluvia mañana", debido a que son los escenarios que se tiene como finalidad predecir.

También, el resto de variables son las variables explicativas, debido a que apoyan la predicción de lluvia deseada agregando e indicando la información necesaria para hacerlo.

3. Modelos

Para estudiar los datos de manera correcta, se utilizan distintos métodos predictivos. Para utilizar estos métodos se utilizan variables para medir las características que corresponden a: Falso negativo, falso positivo, verdadero negativo y verdadero positivo. Con la ayuda de estas variables, se obtienen indicadores, los cuales miden la exactitud del problema:

- Correctitud: Porcentaje de acierto del modelo.
- Sensibilidad: Porcentaje de acierto sobre lo realmente Éxito.
- Especificidad: Porcentaje de acierto sobre lo realmente No Éxito.
- Precisión: Porcentaje de acierto sobre lo predicho como Éxito.
- Tasa real de éxito: Porcentaje de acierto de los Éxitos al azar (sin modelo).
- F1-Score: Medida que relaciona la sensibilidad con la precisión.

Considerando lo anterior, se aplicaron los siguientes modelos:

En primer lugar, se aplicó un modelo de Aprendizaje Supervisado en base a Clasificación. Se utilizó el algoritmo de Arbol de decisión, para luego, mediante Validación Cruzada, evaluar el modelo a través de la implementación de otros. De esta forma, se consideraron los que tuvieron mejor Correctitud, los cuales fueron Logistic Regression, LDA, AdaBoost y Gradient Boostin. Finalmente, a través del Ensamble de Modelos, los modelos mencionados fueron combinados para obtener uno final, el cual aportará más estabilidad y tendrá un mejor poder predictivo.

4. Resultados

Dado al análisis de modelos hecho en el código y explicado previamente en la sección anterior, se tienen resultados numéricos para cada indicador dependiendo del modelo utilizado.

4.1. Validación cruzada

Validación cruzada:

[0.82994262, 0.84080239, 0.79497439, 0.79492104, 0.78462441, 0.8943662 , 0.76520487, 0.86598378, 0.82522407, 0.7836478]

A continuación, se muestran los resultados por modelo:

4.2. Logistic Regression

- Correctitud: 0.89
- Presunción: 0.81
- Sensibilidad: 0.53
- F1 Score: 0.64
- Tasa real: 0.75

4.3. LDA

- Correctitud: 0.89
- Presunción: 0.79
- Sensibilidad: 0.55
- F1 Score: 0.65
- Tasa real: 0.76

4.4. AdaBoost

- Correctitud: 0.88
- Presunción: 0.70
- Sensibilidad: 0.59
- F1 Score: 0.64
- Tasa real: 0.77

4.5. Gradient Boosting

- Correctitud: 0.89
- Presunción: 0.74
- Sensibilidad: 0.58
- F1 Score: 0.65
- Tasa real: 0.77

4.6. Modelo Ensamble

- Correctitud: 0.89
- Presunción: 0.74
- Sensibilidad: 0.58
- F1 Score: 0.65
- Tasa real: 0.77

Además, como equipo, se decidió escoger el modelo que entrega LDA debido al score que presenta frente a los otros modelos realizados. Teniendo así, un mayor desempeño en general.

5. Conclusiones

En este proyecto se aplicó la herramienta *Machine Learning* a un conjunto de datos que contenían información sobre condiciones climáticas de una ciudad al sur de Australia llamada Wagga Wagga, la cual es conocida por ser un centro agrícola importante, en consecuencia la predicción de lluvias en la zona es fundamental para el crecimiento económico de la industria agricultura en la región.

Una vez aplicados los modelos y comparándolos entre sí, se debe notar que el score general del modelo LDA es superior a los demás modelos presentados en la predicción de lluvia. Debido a esto, es claro que el modelo principal es el modelo LDA.

Finalmente, esto solucionaría nuestra problemática planteada, debido a que usamos un modelo donde se tiene un porcentaje de correctitud bastante bueno, por lo que podría ayudar a muchos trabajadores del sector agrícola a planificar sus labores sabiendo que días es muy probable que llueva logrando así maximizar su producción y/o generar productos de mayor calidad haciendo que sus negocios puedan prosperar.