



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



NOMBRE: WALTER RAUL PEREZ MACHINENA

MATRICULA: 1603647

GRUPO: 003

NOMBRE ACTIVIDAD: MODELO DE CLASIFICACION

NUMERO DE ACTIVIDAD: 3

MATERIA: APRENDIZAJE AUTOMATICO

DOCENTE: JOSE ANASTACIO HERNANDEZ SALDAÑA



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

Contenido

Introducción.....	3
Hallazgos.....	4
Grafica	5
Conclusiones.....	6
Referencias	6



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



Introducción

En esta actividad se consideró una base de datos sobre eficiencia energética en los 12 diferentes formas de construcción de edificios, que considera 8 variables descriptivas, donde una de ellas es categórica y dos variables objetivo, siendo una para la calefacción y otra para el enfriamiento.

En este caso nuestra variable objetivo es una discreta que es la X6, por referirse a la orientación del edificio.

Se busca obtener un modelo de clasificación utilizando modelos de regresión lineal como (knn, decisión tree, svm, logistic regression) para identificar cual es el mejor modelo utilizando también validación cruzada (cross-validation) entre varias modelos.

Tabla de variables

Nombre de la variable	Role	Tipo	Descripción
X1	Característica	Continuo	Compacidad relativa
X2	Característica	Continuo	Área de superficie
X3	Característica	Continuo	Área de la pared
X4	Característica	Continuo	Área del techo
X5	Característica	Continuo	Altura total
X6	Característica	Entero	Orientación
X7	Característica	Continuo	Área de acristalamiento
X8	Característica	Entero	Distribución del área de acristalamiento
Y1	Objetivo	Continuo	Carga de calefacción
Y2	Objetivo	Continuo	Carga de enfriamiento



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



Hallazgos

Modelo	Cross validation accuracy
KNN	0.04 (+/- 0.04)
Arbol de Decision	0.01 (+/- 0.03)
SVN	0.18 (+/- 0.07)
Logistic Regression	0.22 (+/- 0.07)

Considerando la información obtenida el mejor modelo es: Logistic Regression.

Además de que el Accuracy es de: 0.1363

Este accuracy significa que el modelo no está prediciendo correctamente la mayoría de las veces.

Se realiza matriz de confusión.

5	24	5	8
2	11	3	10
5	24	2	13
7	30	2	3

Aquí se muestra como las predicciones se distribuyen en las clases reales.

Y se muestra un reporte de clasificación.

Classification Report:				
	precision	recall	f1-score	support
2	0.26	0.12	0.16	42
3	0.12	0.42	0.19	26
4	0.17	0.05	0.07	44
5	0.09	0.07	0.08	42
accuracy			0.14	154
macro avg	0.16	0.16	0.13	154
weighted avg	0.16	0.14	0.12	154

En este reporte podemos visualizar cual clase tiene más precisión, siendo ella la numero 2, de la misma manera se visualiza el Recall, el cual señala que las



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

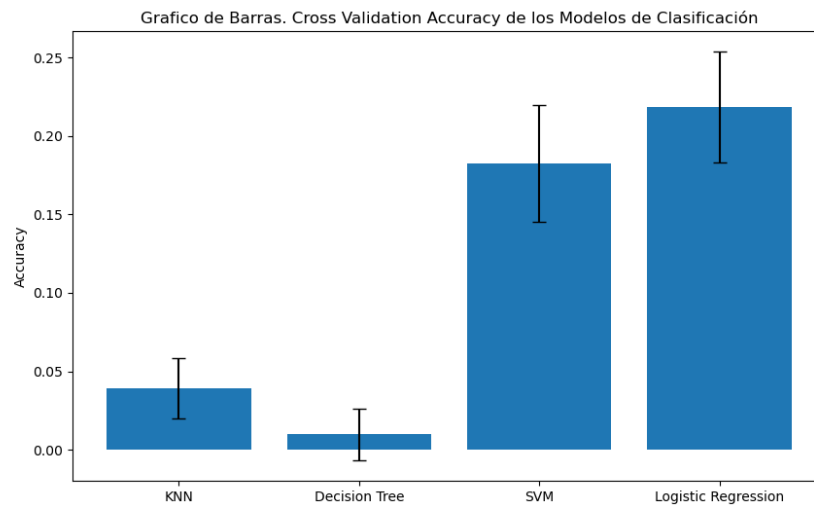
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



instancias de la clase 3 fueron correctamente identificadas, con el F1 podemos identificar que se tiene un rendimiento muy bajo que no sobre pasa el 0.19.

En la última columna podemos ver el número de ocurrencias.

Grafica



En esta grafica podemos visualizar que el modelo de Logistic Regression, es el mejor modelo en este caso



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

Conclusiones

Con esta actividad logramos determinar el mejor modelo aplicable para nuestros datos es **Logistic Regression**, si bien tiene un porcentaje pobre, en cuestión de los modelos utilizados es el mejor de ellos. Se hizo uso de todos los modelos mencionados en la rúbrica con la finalidad de identificar que modelo sería el mejor para este caso de estudio. Se hizo uso de la validación cruzada entre varios modelos según el criterio utilizado.

Referencias

Guido, A. C. (2016). *Introduction to Learning with Python*. California: O'REILLY.

Scikit-Learn. (2024, 07 21). *Scikit-Learn*. Retrieved from <https://scikit-learn.org/stable/>

Xifara, A. T. (2024, 07 21). *archive.ics.uci*. Retrieved from <https://archive.ics.uci.edu/dataset/242/energy+efficiency>

Material de Clase.