



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

UNIVERSIDAD AUTONOMA DE NUEVO LEON
FACULTAD DE CIENCIAS FISICO MATEMATICO



FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

Inteligencia Artificial

Act 14: Programando K-Nearest-Neighbor en Python

Docente: Luis Ángel Gutiérrez Rodríguez

Alumno: Jhoana Esmeralda Escobar Barron. 1950748.

Gpo:031

1 Introducción

K-Nearest Neighbors (K-NN) es un algoritmo de aprendizaje supervisado utilizado para clasificación y regresión. Su importancia radica en su simplicidad y efectividad, ya que clasifica un punto basándose en la mayoría de sus vecinos más cercanos en un espacio de características.

2 Metodología

Para realizar el análisis con K-NN, se siguieron los siguientes pasos:

1. Carga del conjunto de datos desde un archivo CSV.
2. Exploración y visualización de los datos.
3. Normalización de los datos mediante `MinMaxScaler`.
4. División del conjunto de datos en entrenamiento y prueba.
5. Aplicación del modelo K-NN con diferentes valores de k .
6. Evaluación del rendimiento mediante matrices de confusión y reportes de clasificación.

A continuación, se presenta un fragmento del código implementado:

```
X = dataframe[['wordcount', 'sentimentValue']].values
y = dataframe['Star Rating'].values
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0)
scaler = MinMaxScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

knn = KNeighborsClassifier(n_neighbors=7)
knn.fit(X_train, y_train)
print("Accuracy on training set:", knn.score(X_train, y_train))
print("Accuracy on test set:", knn.score(X_test, y_test))
```

3 Resultados

Los resultados obtenidos muestran la efectividad del modelo K-NN en la clasificación de datos. Se obtuvieron las siguientes precisiones:

- Precision en el conjunto de entrenamiento: 90
- Precision en el conjunto de prueba: 85

Además, la matriz de confusión y el reporte de clasificación muestran la distribución de las predicciones y el rendimiento del modelo.

4 Conclusion

Se observo que el modelo K-NN puede proporcionar buenos resultados para la clasificacion de datos cuando se escoge un valor de k adecuado. La normalizacion de los datos y la seleccion de caracteristicas relevantes juegan un papel clave en la efectividad del algoritmo. En futuras iteraciones, se podrian probar otras tecnicas de optimizacion para mejorar la precision.